Balancing High-Load Scenarios with Next Cell Predictions and Mobility Pattern Recognition

Stefan Michaelis
Artificial Intelligence Group
Faculty of Computer Science
Dortmund University of Technology
Dortmund, Germany
Email: stefan.michaelis@tu-dortmund.de

***Abstract*—Knowing where a mobile user will be next can deliver a tremendous increase in network performance under high load, as this knowledge enables pro-active load balancing. To derive this information, sequences of traversed cells are fed into pattern detection algorithms. After the training phase the learned model predicts each user's next cell. Even for complex scenarios, the prediction accuracy can exceed 90%. Predictions are used to rearrange mobile connections in a simulated high-load scenario centered around an event at a soccer stadium. To prevent call drops for mobile users targeting the stadium, apropriate resources in the predicted next cell are reserved. The results exceed 20% in improvements for throughput and call drop rates, enabling the network to bear a much higher load before stalling.**

***Keywords*-Handoff Optimization; Mobility Prediction; Load Balancing**

I. The problem of users changing cells

Seamless handoff from basestation to basestation is essential for preserving mobility in cellular networks. Here we provide an additional indicator for handoff, which complements existing decision algorithms and can be used to manage overall mobile network load. The major advantage of this approach is the early availability of the handoff indicator, being in the range of several seconds compared to short-term measurements of signal strength and quality.

The idea is, that moving users are bound to the geographical topology, i.e., street and rail networks, and therefore are forced to partial deterministic behavior. Each movement provides a trail of traversed cells, which deliver a coverage fingerprint for the mobile network. Using knowledge discovery or data mining algorithms to learn the historical sequences of cells lead to a prediction of the most likely next cell each time a user enters a new cell.

This document consists of two main parts: In Section II, the overall achievable next-cell prediction accuracy is calculated for a sample geographical topology. The scenario demonstrates, how the artificial intelligence algorithm performs for varying road and railway networks, depending on the available input data for training the algorithm. The way the mobile data is handled, the privacy of the user's is respected and it is unnecessary to trace complete profiles on a per-user basis.

In Section III, the same methods are applied to a high load scenario: Mobile users moving to and from a soccer stadium.

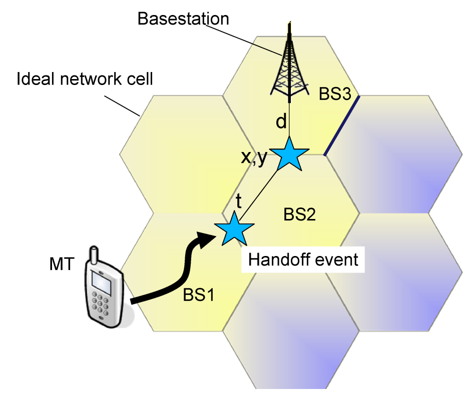

Fig. 1. Available features during user movements

The knowledge of the predicted next cell users are moving into is used to balance the load in the mobile network and enhance the user's quality of experience. The scenario is built using network coverage measurements, the underlying road and railway network, typical network traffic, and finally, the numbers of visitors for each means of transport. The results show the potential for gaining benefits by actively rearranging connections with knowledge about expected handoffs.

II. Space-domain prediction of next cells

Space-domain prediction of cells estimates *where* a mobile user will move next and identify the basestation candidate for the new association. In this section, the complete process from building mobility traces up to prediction of expected next cells with pattern detection algorithms is examined.

A. Mobility trace generation for pattern recognition

Every mobile cellular network needs a component, which is informed about the current location of the user based on the associated cell. Where this information is available depends on the type of network, e.g., mobile switching centers and location registers for mobile networks or remote authentication servers for wireless local area networks (WLAN). The known location is typically rather coarse and one reason why smartphones deploy GPS receivers for location based services.

 ISBN: 978-1-61208-164-9

Here we rely on the most common denominator independent of the specific network type, the basestation identifier. In addition, the duration t each user is connected to each basestation can be easily derived from the sequence of association/deassociation events. Further parameters may include position or distance to the basestation (see Figure 1), but are proved not to be mandatory for a good next cell prediction accuracy. An example sequence generated by a mobile user consisting of information tuples *BS-ID, Residence Time* may look like *BS1, 20s, BS2, 35s, BS4, 32s, …* The length of these sequences is limited by the overall call duration, as mobile nodes in many networks can only be tracked during active connections or otherwise in large location areas bundling several cells. Very long sequences of traversed cells may occur for example when driving on vacation and the kids playing mobile online games during the trip on the backseat of the car. As very long sequences in most cases do not provide more information, all generated sequences are limited in length and split into shorter sequences. The optimal upper bound of length is also part of the analysis and results for sequence lengths between 3 and 6 cells are compared.

The set of all generated traces are used as training data for the pattern detection algorithm. The goal is to correctly predict the last basestation in each sequence.

B. Related research in cellular predictions

Predicting the next cell for moving mobile users has been in focus of mobile positioning research for several years. Macroscopic mobility prediction as discussed here sets the focus on the cellular level, which is useful in network load balancing.

In [3], a fundamental approach has been described for macromobility predictions: A variant of the ZIP-compression algorithm called LeZi is used to build a tree per user from the cell sequences. This algorithm delivers a good prediction accuracy for complete sequences, i.e., without missing values or changes in the cell sizes due to radio effects, and different variants are still popular today (see [7]) due to its simplicity and low consumption of computing resources.

The work on algorithms for mobility prediction can be classified into several categories as defined in [4]: Domain-specific, user dependent and usage of time.

In our previous work in [1], we demonstrated that the selection of the specific algorithm used for predicting the position is of secondary importance. While of course some algorithms may deliver higher accuracy compared to others, in most cases the question whether the mobility sequences contain learnable patterns of movements at all is more critical. Typically, general purpose data mining algorithms as the Support Vector Machine used in this publication are able to extract a minimum of patterns in the data if existent. Therefore we only investigate *domain-independent algorithms* without the need for mobile network specific parameters and keep the pattern detection algorithm replaceable

The feasibility of predictions per user profile has been demonstrated in several recent publications as [11], [12], in [6] for WLAN or in [8] with a prediction accuracy up to 93%. Nevertheless, learning individual movement patterns comes at the price of impacting privacy. The training data used for the predictions in these scenarios has every user identification removed, resulting in pattern detection *independent of specific user behavior*. Of course, approaches like this can only work in case the geographical topology restricts the users in their movement (e.g., on highways or in trains), so that meaningful patterns are generated.

Beside the spatial prediction, predicting the time of the handoff to the next cell is also necessary to reserve resources promptly. This timing can be integrated into the prediction model itself as demonstrated in [5]. This approach is reasonable if different points of time lead to different user behavior (e.g., weekend/weekdays, morning/evening). For short term changes, e.g., during traffic jam, incorporating time into the model increases the complexity of the training process. Our previous work in [2] presented an approach to deal with short term behavioral changes. For the work presented here we are *independent of absolute timing* and simply use cell residence time as a learning attribute.

C. Dynamic user-agent based mobility models

The feasibility of next-cell predictions strongly correlates with constraints moving users have to face due to the geographical topology, network coverage and most important the degree of determinism in the movements itself.

For this work several mobility models are combined to include different behavior. Essentially, these models are *Path Follower*, *Gravity* and *Random Walk* models. The path follower model can closely resemble commuting behavior: Following a preset path, staying at the target area for a certain amount of time and following a similar path back to the origin. This mobility model presents the highest level of determinism in the traces, introducing uncertainty only in variance of speed or residence times at target areas.

The gravity model assigns for different areas a so-called gravity value. This parameter sets a level of attractiveness to the areas, defining the probability for selecting this area as the target for movements.

Finally, the random walk model provides no determinism, but is still useful to generate a certain amount of *background noise* for the pattern recognition algorithm. Nevertheless, the random movement of course is still constrained by the road network, leading partly to the same traces as the other mobility model, e.g., on highways without a chance to leave at will. Random mobility is valuable to generate traces for areas, where the road density is high compared to the diameter of network cells, for example for GSM cells covering dense urban areas.

All mobile users are modelled as *Agents* without a fixed mobility model. This enables user traveling by car and switching to walking at the destination.

Figure 2 presents an exemplary scenario combining the road network, mobile network coverage and mobile user agents for simulation. The focus is put on situations where

 ISBN: 978-1-61208-164-9

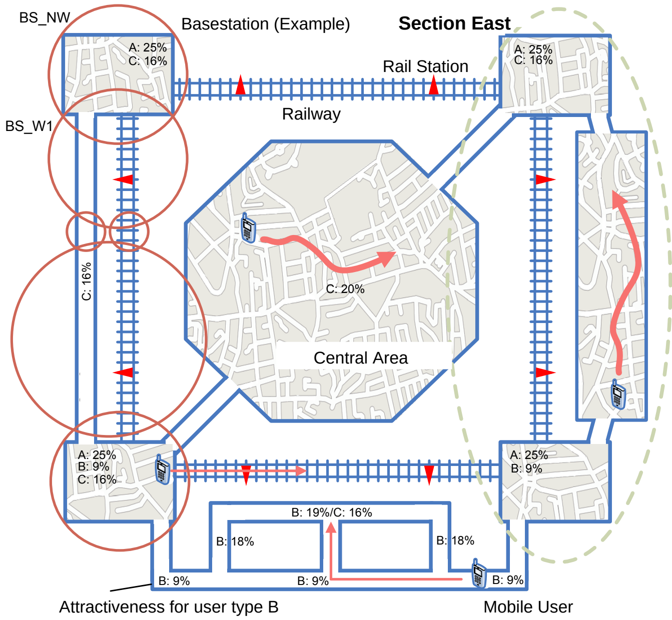

Fig. 2. Simulation scenario with different mobility models

mobile networks may easily receive high load: Highways and especially railways. Here a large amount of potential mobile application users switch cells nearly simultaneously and stay in a cell only for a short amount of time. In contrast, the rail network disables freedom of movement, forcing the users to certain sequences of cell transitions. The next section delivers results for the predictability of these sequences.

The geographical topology consists of several different areas: In the center one large area like an urban center, enabling random movements. This area is adjacent to the four areas to each side, introducing noisy patterns to the more regular streets and rails in these outer areas. Each of the four areas provides a different combination of possible user mobility: Rails only in the north, rails parallel to a simple street network in the south, to an area to the east and parallel to a highway in the west. Each line of rails also incorporates stations, where the simulated train stops for several seconds.

All mentioned mobility models are integrated into the scenario. For the gravity model the attractiveness distributions of user types $A - C$ is given. The topology is covered by overlapping cells, most of them are not show in Figure 2 for sake of simplicity. The western area as an exception shows some cells, as this part of the scenario is especially difficult for pattern recognition. Highway and railway users generate identical sequences of traversed cells except for two tiny cells individual to each path. A correct prediction for these cells is only possible in case the pattern detection algorithm can distinguish users on the parallel tracks.

D. Predicting next cells with pattern detection

The generated sequences are used to train pattern detection algorithms and predict the next cell (the target class) for new sequences. As classification of examples is a well-known task for pattern detection, several algorithms are available for this classification task. The more expressive the algorithm is, the better it can be adapted to complex traces, but the more difficult it is to find the optimal set of parameters for the algorithms. In parallel, the input data has to be selected carefully: What is the optimal maximum length for mobility sequences? Which features beside the basestation identifier enhance the pattern detection process?

For the results presented here the well-known *Support Vector Machine (SVM)* machine learner has been used for the prediction process, see [9]. SVMs try to separate the data samples by optimal hyperplanes and new examples are classified depending on which side of the hyperplane they are positioned. The hyperplane's location is defined by the so-called support vectors, which consist of a selected subset of all provided examples. The plane is considered optimal, if it minimizes the number of samples on the wrong side and maximizes the distance to the support vectors.

As a simple plane can not always capture the nature of data distributions, kernel functions allow to transform the input data into a modified space. A popular kernel is for example the polynomial kernel with the degree as a parameter. Selection of kernel and parameters like degree has to be done consistently for every example set.

To extend the SVM's ability to predict more than two possible classes (due to only comparing the side of the hyperplane of the example in question), the problem of multiple classes, as necessary for predicting the next cell, can be covered by pairwise predictions between each class.

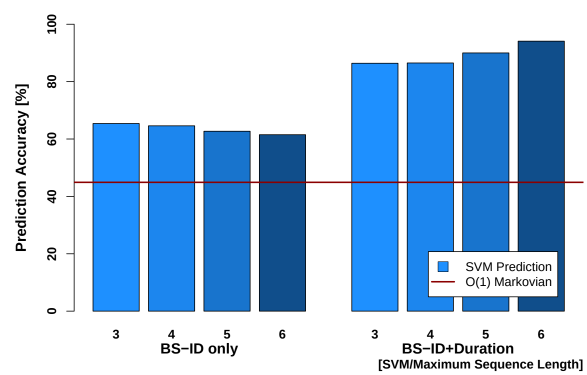

Fig. 3. Cell prediction accuracy for different features

Figure 3 presents the prediction accuracy as relation of correct to all predictions. Four different maximum sequence lengths 3-6 have been evaluated as well as using only the basestation identifier (on the left side) and identifier in combination with cell residence time (right side of the figure). The SVM can handle a combination of nominal data (BS-Id) and continuous data (time) without changing the algorithm. All results are generated using a ten-fold stratified cross-validation, delivering a 90% confidence interval in the range of ± 1.65.

The reference prediction (horizontal line) is calculated using a Markovian O(1)-predictor. This simple classifier uses only

 ISBN: 978-1-61208-164-9

the currently associated basestation as input and predicts the next which occurred most frequently in the training examples. The O(1)-predictor therefore delivers an estimation of the learning complexity, with more neighboring basestations and a uniform transition probability resulting in a lower accuracy.

The results in Figure 3 show for the id-only case an accuracy of around 65%, which is 20% higher compared to the O(1)-predictor, but still not sufficient for reliable enhancements of handoff and network management. For longer sequences of up to six cells the accuracy even slightly decreases. Effects like this appear in cases, where the added data masks the valuable bits of information provided by the rest of the features (here the higher information value of the latest basestation compared to previous ones.

A great boost in prediction accuracy can be seen for the second evaluation, BS-Id with residence times. Users traveling by car provide different patterns compared to users traveling by rail. Using the duration in each cell these users become separable, increasing the prediction accuracy up to 94%. Here the predictions even benefit from longer sequences, as the likelihood of identifying a user's means of travel increases with more durations available.

Predicted cell						
SW	W3	W1	NW	W2b	W2a	Real
0.81	0.00	0.00	0.00	0.00	0.00	SW
0.02	**0.79**	0.01	0.00	0.00	0.00	W3
0.00	0.00	**0.93**	0.01	0.02	0.01	W1
0.00	0.04	0.02	**0.94**	0.00	0.00	NW
0.01	0.00	0.00	0.01	**0.56**	0.56	W2b
0.01	0.00	0.00	0.01	0.42	**0.43**	W2a

TABLE I
CONFUSION MATRIX, WEST SIDE OF SCENARIO

Predicted cell						
SW	W3	W1	NW	W2b	W2a	Real
0.99	0.01	0.00	0.00	0.00	0.00	SW
0.00	**0.99**	0.01	0.00	0.00	0.00	W3
0.00	0.00	**0.99**	0.00	0.00	0.00	W1
0.00	0.00	0.00	**1.00**	0.00	0.00	NW
0.00	0.00	0.00	0.00	**1.00**	0.00	W2b
0.00	0.00	0.00	0.00	0.00	**1.00**	W2a

TABLE II
CONFUSION MATRIX, WEST SIDE, INCLUDING RESIDENCE TIME

Tables I and II enable a detailed comparison of this effect per basestation for the western part of the scenario. Cells with Ids *W2a* and *W2b* are the small cells distinct for highway and railway. Table I presents an overall good accuracy for most cells with the exception of these two cells (56% and 43%). Including the duration needed to cross the cells into the training data increases the accuracy for all cells and enables perfect predictions of W2a and W2b. The duration enables to distinguish users without any further information like GPS positions, knowing in advance which of the cells will be next.

III. BALANCING HIGH-LOAD SCENARIOS

This section applies the next cell predictions of the former section to balance network load in the mobile network itself. The early knowledge about users entering a new cell delivers a convenient time frame for reservation of resources.

A. *Scenario description: Soccer stadium*

The scenario used for evaluating the effect of predictions is based on the same principles as the scenario presented before and incorporates a real geographical topology, network coverage measurements and user movement profiles.

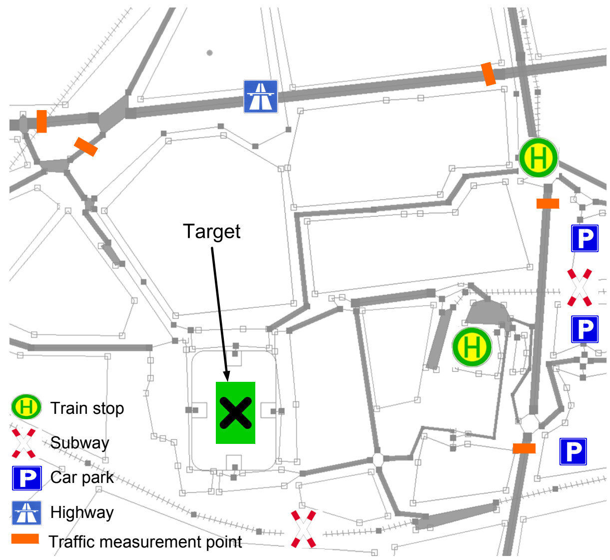

Fig. 4. Topology for mobility simulation of the stadium scenario

Figure 4 illustrates the scenario: The central point of interest is the soccer stadium in Dortmund, a German city with more than 500,000 residents. During an event at the stadium more than 60,000 people are arriving and leaving the stadium; 20,000 people unrelated to the event are expected to move in the region. Data provided by the local Department for Traffic, the regional transport and the stadium operator enables detailed modelling of the movement behaviors. The floating car data is measured using sensors in the streets and have been provided for several days, with and without events at the stadium to calculate the difference in paths and car density. The distribution of visitors arriving by train, car and foot determine the parameterization of the simulated agents, which are again able to switch mobility models. Visitors arriving by car change to a walk model after arriving at the parking sites etc.

For later evaluation two main paths have been selected: At the northern top the urban high way *B1* crosses the scenario from east to west. Secondly, a railway track from north-west to south-east provides one main access route to the stadium.

Together with user movements, the traversed basestations need to be captured. To gain a realistic view of the coverage, the basestations in range have been measured. Figure 5 displays results of measurements by car and foot. Each measurement has been associated with GPS positions and

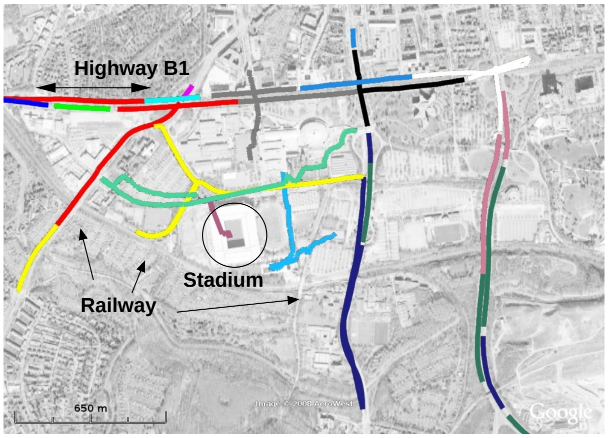

Fig. 5. Coverage measurements for stadium scenario

shows the associated primary UMTS-HSPA basestation. At most positions an active set of 4 was available, showing a high overlap of adjacent cells. This is a necessary precondition to enable rearrangement of connections into neighbor cells.

Interestingly, the measurements highlight a classical handoff parameter, the handover margin. The position of handover is shifted due to this margin depending on the direction of travel, as can be seen for the B1 at the north of the figure.

B. Mobile network management

This section concentrates on applying the next-cell predictions for different dynamic network management techniques like reserving radio resources for expected users or rearranging existing connections to maximize data throughput.

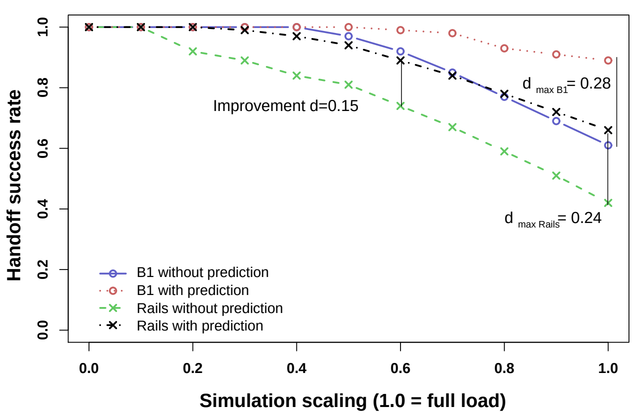

Fig. 6. Handoff success rate based on scenario load

Figure 6 presents the enhancements for handoff success rate. According to data provided by the German mobile network operators, the aggregated data traffic exceeds 1,000 Erlang voice calls equivalent during the event in and in direct neighborhood to the stadium. In case a user with an active connection gets into a cell without free resources, the connection has to be dropped. When a user successfully enters a new cell, the next cell is predicted and the resources for this user are blocked in the expected cell.

To examine the effects for different load situations, the simulation has been scaled for different percentages up to the full load simulating all 80,000 mobile users. Please be aware, that a scaling factor of 0 includes still one user for each mobility model.

As to expect, for a small load scale, the handoff is equal to or nearly 100%, as no cell is completely filled with connections. The success rate starts to decline with increasing load. Figure 6 displays the success rate for two paths, B1 and railway, and for two modes: With and without using the predictions. The success rate declines faster for users arriving by train, as these users travel faster and in higher numbers, increasing the probability for arriving at resource depleted cells.

Reserving resources can not completely avoid this effect, but significantly improve the success rate. The decline is slowed and the improvement of handoff success can go up by 28% for the fully loaded scenario.

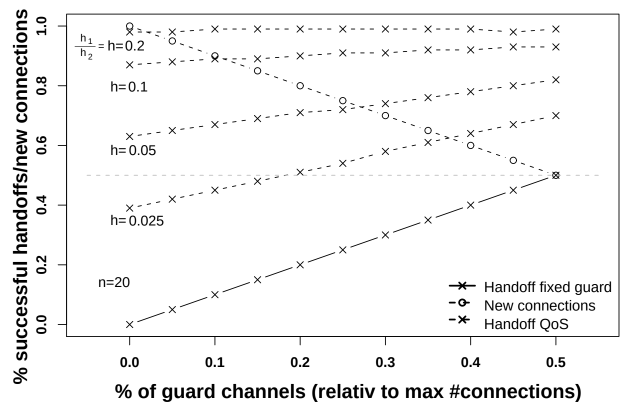

Fig. 7. Handoff success rate based on call holding time

Whether a reservation based on the prediction can be successfully executed, depends on the holding time of connections. Mobile operators report an ever increasing utilization, starting with one minute mean call duration for voice calls and two minutes for video calls in pre-iPhone times. Figure 7 presents the handoff-success rate based on mean holding times and a pre-reserved guard channel. Most network operators prioritize active over new connections and set a fixed amount of cell capacity for users arriving in the cell, decreasing the amount of capacity for new connections accordingly.

Users targeting a fully loaded cell get a reservation in case for a prediction (Handoff QoS), when an existing connection is closed or the guard channels are not completely used. The success rate therefore depends on ratio of the mean cell transient time h_1 of the moving users and the mean holding time h_2 of the resident users in the target cell, $h = \frac{h_1}{h_2}$.

Figure 7 presents the results for a fixed h_1 and varying h_2 of $n = 20$ users resident to the cell. For large ratios $h \geq 0.2$ nearly all handoffs can be handled perfectly without the need for fixed guard channels. For increased holding times and smaller ratios $h = 0.025$, the probability of ending connections in the target cell drops and the success rate is below 40%.

 ISBN: 978-1-61208-164-9

This can be compensated with the classical guard channels. Nevertheless, using the predictions for channel reservation, a smaller fixed guard channel is needed to achieve the same rate of successful handoffs.

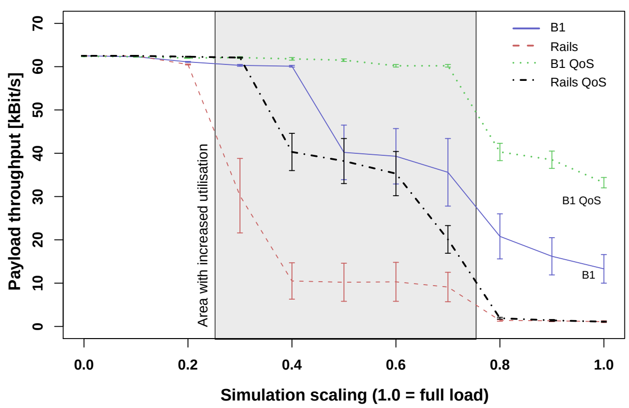

Fig. 8. Network throughput with rearrangement of connections

Lastly, the results presented here further enhance network management by rearranging active connections to neighboring cells for predicted incoming connections to the cell. This handles situations, where throughput can be maximized by distributing the traffic more evenly. Figure 8 illustrates the effect again for two sample users with a reference bitrate of 64 kBit/s TCP traffic. Moving into high-load cells, especially with mixed traffic types and the need to compete with UDP, degrades the mean throughput to 0 kBit/s for the highest load in the scenario.

As for the handoff success rate, the throughput starts to decline with increased load. The error bars for each point present the median absolute deviation, increasing with higher variance and bias in the data transmission. Again, a higher amount of Quality of Service can be guaranteed by using the predictions and moving existing connections from predicted next cells. Nevertheless, in the end with full load in the scenario, the TCP connection looses against other traffic source like UDP. This leaves an area between the two extremes of underused and exhausted network, where the rearrangement balances network load and improves QoS for all mobile users.

IV. The benefit of users changing cells

Instead of compensating for the effects of moving users on the network routing, we actively use knowledge about previously visited cells to predict the next location and balance traffic load accordingly.The results demonstrate, that even with simple features like identifiers and cell residence times the geographical constraints mobile users face can be detected.

Predicting user's next cell with a high accuracy of more than 90% provides mobile network operators with a powerful tool to rearrange traffic. This enhances quality of service for the users as well as saving costs for operators due to more efficient utilization of infrastructure. The approach is non-intrusive and intended to co-exist with mandatory network management for handoff, call admission control and routing. User's privacy is preserved as no individual patterns need to be learned. The only time where the user's id can be associated with the sequence of basestations is when preparing for predicting the next cell. In the future, further methods to make users anonymous like proposed in [10] may enable to provide the data to external location based service providers without breaching privacy.

The final step, before the methods proposed here are considered ready for production use, relates to the selection of subsets of cells for model training. As an example, for the region of Dortmund for the combined network types from GSM to 3G, including sectorization, more than 500 cell ids can be measured. For the whole country this will result in an amount of cells too large for most data mining algorithms. Future research concentrates on distributed data mining for automatically generated clusters of cells.

Acknowledgments

Part of the work in this paper is supported by Deutsche Forschungsgemeinschaft (DFG) within the Collaborative Research Center SFB 876 ”Providing Information by Resource-Constrained Analysis”, A1: http://sfb876.tu-dortmund.de

References

[1] Ms. ichaelis, A. Lewandowski, K. Daniel, and C. Wietfeld: *Macromobility Prognosis for high-priority Resource Reservation in Wireless Networks*, 5th IEEE International Symposium on Wireless Communication Systems (ISWCS), Reykjavik 2008

[2] S. Michaelis, A. Lewandowski, K. Daniel, F.Z. Yousaf, and C. Wietfeld: *A comprehensive mobility management solution for handling peak load in cellular network scenarios*, 6th ACM International Symposium on Mobility Management and Wireless Access (MobiWac), Vancouver 2008

[3] A. Bhattacharya, and S.K. Das: *LeZi-Update: An Information-Theoretic Approach to Track Mobile Users in PCS Networks*, Proceedings of ACM/IEEE International Conference on Mobile Computing and Networking, MobiCom '99, Seattle 1999

[4] C. Cheng, R. Jain, and E.v.d. Berg: *Location Prediction for Mobile Wireless Systems*, in: Furht, B. (Editor): *Wireless Internet Handbook*, pp. 245-264, CRC Press, Boca Raton 2003

[5] J.-M. François, G. Leduc, and S. Martin: *Learning movement patterns in mobile networks: a generic method*, Proceedings of European Wireless, Barcelona 2004

[6] L. Song, D. Kotz, and R. Jain: *Evaluating Next-Cell Predictors with Extensive Wi-Fi Mobility Data*, Transactions on Mobile Computing, Vol. 5, No. 12, pp. 1633-1650, IEEE 2006

[7] A. Rodriguez-Carrion, C. Garcia-Rubio, and C. Campo: *Performance Evaluation of LZ-Based Location Prediction Algorithms in Cellular Networks*, in IEEE Communications Letters, Volume 14 Issue:8, pp. 707-709, 2010

[8] C. Song, Z. Qu, N. Blumm, and A.-L. Barab'asi: *Limits of Predictability in Human Mobility*, Science Vol. 327 no. 5968 pp. 1018-1021, 2010

[9] J. Platt: *Fast Training of Support Vector Machines using Sequential Minimal Optimization*, in Advances in Kernel Methods - Support Vector Learning, B. Schoelkopf, C. Burges, A. Smola (Editoren), MIT Press, Cambridge 1998

[10] A. Monreal, R. Trasarti, C. Renso, D. Predreschi, and V. Bogorny: *Preserving privacy in semantic-rich trajectories of human mobility*, Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Security and Privacy in GIS and LBS, San Jose 2010

[11] P. Salvador, and A. Nogueira: *Markov Modulated Bi-variate Gaussian Processes for Mobility Modeling and Location Prediction*, in Proceedings of NETWORKING 2011, LNCS, Volume 6640/2011, pp. 227-240, Springer Berlin 2011

[12] N. Zhao, W. Huang, G. Song, and K. Xie: *Discrete Trajectory Prediction on Mobile Data*, in Proceedings of Web Technologies and Applications 2001, LNCS, Volume 6612/2011, pp. 77-88, Springer Berlin 2011

 ISBN: 978-1-61208-164-9