TRAFFIC THEORY APPROACH FOR HETEROGENEOUS MOBILE COMMUNICATION NETWORKS

Stephan Goebbels

RWTH Aachen University, Faculty 6, Chair of Communication Networks, Aachen, Germany
sgs@comnets.rwth-aachen.de

***Abstract* – This paper presents a new traffic theory based approach for the analysis of heterogeneous wireless communication networks. While traffic theory is widely spread for the analysis of wired networks like the Internet the adaptation of these methods for wireless and especially heterogeneous networks is not yet accomplished. New upcoming mobile communication networks will be ruled by continuously switching radio access technology and varying transmission rates. Also newly developed technologies for the improvement of wireless communication like Smart Caching are not yet investigated by means of traffic theory.**

Therefore the new method is also used to investigate the applicability and the capabilities of Smart Caching in future integrated networks.

Keywords – Traffic Theory, Queuing Theory, Smart Caching.

1. Introduction

Future wireless communication networks will be dominated by the integration of different Radio Access Networks (RAN). To fully illuminate urban areas with broadband coverage and to provide always the best service (Always Best Connected - ABC) it is necessary to integrate the available wireless networks and constantly check which network meets the service requirements of the current online session best. Potentially the terminal has to handover to a better suitable network [1]. The coverage areas of the different wireless networks will overdraw urban areas like a patchwork, as depicted in Figure 1. The continuous changes of the RAN will come along with heterogeneity in e.g. available bandwidth and expected packet delay.

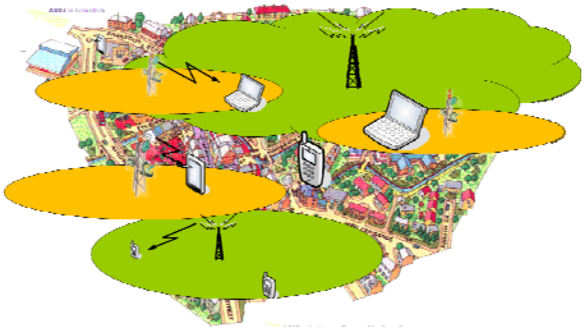

Figure 1 : Heterogeneous Network Structure

While for fixed and cabled networks the modeling of links and nodes could be handled by simple queuing systems it has to be adapted to the special characteristics of future wireless networks.

Also new invented technologies for the specific enhancement of wireless communication like Smart Caching [2] can be investigated with this new approach. Smart Caching bases on the provisioning of potential user data as close as possible to the end terminal in order to minimize delivery latencies and to optimize the utilization of the wireless link. The analytical results will be inspected on their relevance and impact for the applicability of Smart Caching.

2. Motivation

Traffic theory is capable by abstraction to model real communication networks and their behavior quite accurately. Moreover it implies stochastic reliability which is, depending on the implementation, hard to achieve by simulative investigations. The masking of irrelevant side effects due to the abstraction of the system to its core components allows getting a more general impression of the system behavior and its performance.

Also the easy adaptation of the model to changing properties allows a rapid investigation of similar systems with reasonable effort. For other methods like simulations it might mean a complete redesign. Of course, for a final assessment of system performance a modeling of computer networks and its stochastic evaluation is usually not capable. But especially at the initial design of new communication system it represents a fast and effective method to check the impact of the new method. Also new network features can firstly be tested against their applicability before they are implemented in prototypes.

Modeling of communication systems by queuing models is widely spread for communication networks [3][4]. The extension of this method for the investigation of heterogeneous wireless networks and in particular of the Smart Caching approach is therefore the next reasonable step.

3. Analytical communication model

The stochastic modeling of communication networks is a widely spread approach to achieve information about their performance. Each link in a network can be seen as a combination of a queue and a server. The queue resembles the buffer in network routers and access nodes while the serving unit presents the transportation capacity of the link to the next hop of the communication chain.

In Figure 2 this abstraction is depicted. Packets arrive at a network node with a certain average data rate λ. Within the network node they have to wait until they get forwarded. The delay is sourced to a small extent by the processing of e.g. routing algorithms but it is dominated by the limited bandwidth of the next link. In packet switched networks the packets of all currently served connections line up in the queue and wait for being forwarded. Although for channel switched networks each connection has a dedicated channel the arrival rate may exceed the current available transmission rate so that also buffering is necessary.

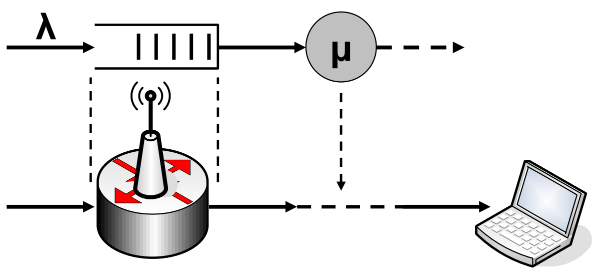

Figure 2 : Queuing Model for Network Nodes

As streams in the Internet overlap each other at each node the final arrival of the packets can be sufficiently modeled by a Poisson process [5]. Therefore in our abstraction it can be assumed that the packets arrive in the queue with negative exponentially distributed inter-arrival time and a mean arrival rate λ equal to the average data rate of the data stream.

The service time of the different packets can also be assumed to be negative exponentially distributed with reasonable accuracy so that the serving process obeys a Markov Process with an average service rate μ. For a scenario with just one RAN the system could be sufficiently modeled by an M/M/1 queue.

For Smart Caching in wireless networks the buffering of packets gets an even more significant meaning. In this approach the packets are gathered at a central point which is the shortest connection between all involved wireless network access nodes (compare Figure 3). Here all packets of the data stream are buffered before they are forwarded to the RAN currently serving the terminal. It has the advantage that in case of a handover a session recovery takes place very fast as not the server somewhere in the backbone has to be contacted but instead the closely located Smart Cache is responsible. Moreover it is possible to fill up the cache in periods of only partial connectivity in order to fully utilize the link capacity if higher transmission rates are provided. So in times of UMTS access the backbone traffic is not slowed down but simply all packets which cannot be delivered to the terminal due to the overloaded situation are buffered in the Smart Cache. Contrary to that periods of WiFi coverage can be used to carry the normal data stream and in addition to that also the cached data. Hence, as long as packets are available in the cache the stream can use up all available bandwidth in the wireless network which increases simultaneously the network utilization. This also leads to the fact that the input of the queue is not affected by the current data rate of the wireless link as this part of the end-to-end connection is separated from the rest. So the assumption of a Poisson arrival process is proven feasible for our system.

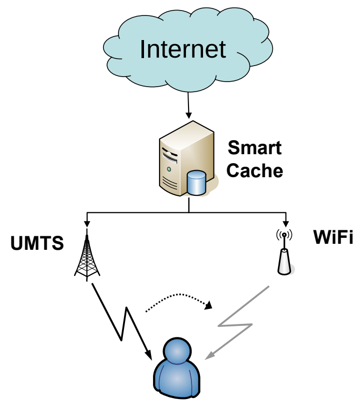

Figure 3 : Smart Caching in Heterogeneous Networks

Contrary to a single RAN system where the serving process could be assumed to be negative exponentially distributed in a multi RAN system the average transmission rate varies when changing the radio technology. However, this can be included in the queuing system by introducing a phase model for the service process as shown in Figure 4. Each path of the model represents one RAN and the according transmission process. Within each path the service time can still be modeled by the negative exponential distribution with average service rate μ_x. The resulting distribution is called hyperexponential with two paths (H2). Which of the paths is traversed in the model depends on the probabilities p_x. The sum of all of them obviously has to be equal to 1.

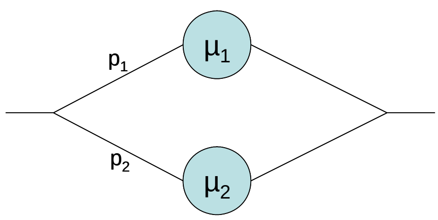

Figure 4 : Hyperexponential phase model (H2)

In the following a combined system out of a WiFi (μ_1; p_1) and a UMTS (μ_2; p_2) network is considered. Each of them represents one path in the phase model. The probability which path is chosen depends on the ratio of the utilization of each system. For ABC, it can be assumed that the WiFi network is used if the terminal is within its coverage area. And furthermore the complete scenario is fully covered by the UMTS network. Therefore the proportion of the size of the coverage areas (C_1 and C_2) has a direct impact on p_1 and p_2.

However, C is not proportional to p itself but only to the fraction of time the corresponding path of the hyperexponential phase model is used. The path probability p reflects the fraction of packets which use one specific path. Therefore the path probability has to be transformed into an occupation rate per path. This value, which is p·µ then has to be proportional to the coverage rate C. Together with the fact that the sum of all path probabilities is equal to one the following expression can be derived:

$$p1 = \frac{C1/\mu 2}{C1/\mu 2 + C2/\mu 1} \qquad (1)$$
$$p2 = 1 - p1$$

The service rate composes out of the average transmission capacity of the wireless link and the assumed packet size.

$$\mu 1 = \frac{WiFi\_Rate}{Packet\_Size} \qquad (2)$$

$$\mu 2 = \frac{UMTS\_Rate}{Packet\_Size} \qquad (3)$$

Therefore finally the model consists out of an M/H2/1 queue with the given parameters (μ_x; p_x).

Although the mean waiting time and other significant parameters like average queue length are easy to derive and hence widely used in scientific analysis it is sometimes not sufficient for the analysis. More information is provided by the full probability distributions. They are usually only computable for specific parameter sets and not given in a closed representation.

However, in the further analysis also the percentiles of the waiting time distribution are considered. Percentiles (Percentage values) of the waiting time distribution represent how much percent of all packets have a waiting time less or equal than a given value. This is especially important for QoS contemplations where mean values are often not satisfactory.

In general M/G/1 queuing systems the Laplace Stieltjes Transform (LST) of the waiting time distribution is given by [6]

$$W^*(s) = \frac{(1-\rho)s}{s - \lambda + \lambda \cdot B^*(s)} \qquad (4)$$
$$B^*(s) = \sum_{i=1}^{2} p_i \frac{\mu_i}{\mu_i + s}$$

where ρ is the overall utilization of the queuing system and B*(s) the service time distribution of a hyperexponential phase model with two phases.

After an inverse LST transformation it results in the waiting time distribution in the time domain of [7]

$$P(W \le t) = 1 - \rho(qe^{-n_1 \cdot t} + (1-q)e^{-n_2 \cdot t}). \qquad (5)$$

The parameters q, n_1, and n_2 can be calculated from the above mentioned scenario parameters μ_x, p_x, and λ. With this general formula the percentile could be computed and used for the following analysis.

4. Application Scenarios

For the evaluation of an integrated WiFi-UMTS system the aforementioned traffic model is used. The cooperation between a cellular network with effective Quality of Service and a broadband wireless LAN will provide the advantages of both systems. Reliable data traffic with guaranteed QoS criteria for basic and control communication in combination with a wideband network for the transfer of larger amounts of data like video streams or ftp traffic.

The considered user application is video streaming. Already nowadays a lot of traffic in the Internet is sourced by such type of applications. And furthermore upcoming services like Video on Demand (VoD) or later on TV over IP (IPTV) will even push the rate of video traffic in the Internet. This will obviously also spread towards wireless end devices if the deployment, the applicability and the performance of broadband networks further growths.

In the example scenario the video stream is buffered in a Smart Cache at the connection point of the UMTS and the WiFi network. Therefore the arrival process of packets, with an average size of 1280 Bytes, is not influenced by the rest of the system and if they cannot be forwarded over the wireless link the packets are stored until enough capacity is available to simultaneously carry the incoming traffic as well as to empty the buffer.

If it is assumed that Smart Cache and the different access nodes are connected with a high capacity link so that the influence of the forward process between cache and access node has no substantial influence on the overall results and both entities can be modeled by one queue.

The capacity of the wireless link does not only vary between the different systems but also within each network. Depending on e.g. interference situation or network load the performance per user or channel might dramatically decrease.

Thus, the capacity of both systems represented by the service rates is varied in three classes: A high, a medium and a low performance class. The chosen values are taken from currently accomplishable performance levels for WLANs (like 802.11a/g or WiMAX) and operating UMTS systems.

As assumed in the later analysis the UMTS and WiFi coverage is equally distributed on both system so that C_1=C_2=50%. With these values also the cumulated data rate of an integrated system can be calculated. All data rates are summarized in Table 1.

In a last scenario a system with only WiFi coverage is contemplated. Between the areas of coverage no radio link is provided. The distance between two successive WiFi cells is assumed to be covered on average within 10 seconds.

High Performance:	
WiFi Data Rate	54 Mbit/s
UMTS Data Rate	384 kbit/s
Cumulated Data Rate	27.2 Mbit/s
Medium Performance:	
WiFi Data Rate	10 Mbit/s
UMTS Data Rate	64 kbit/s
Cumulated Data Rate	5 Mbit/s
Low Performance:	
WiFi Data Rate	1 Mbit/s
UMTS Data Rate	16 kbit/s
Cumulated Data Rate	0.5 Mbit/s

Table 1 : Scenario Parameters

This can be reached by adapting the service rate of the second path of the phase model in such a way that the service of one single packet takes exactly the time of the idle period. Simultaneously the path probability of the second path will be so much diminished that the probability of two successive packets routed through this path is close to zero. This results in a scenario of broadband wireless access (WiFi capacity is set to high performance) interrupted by idle gaps without radio coverage.

As in the prior scenario also a WiFi network coverage of 50% is chosen it is assumed that the terminal faces coverage half of the time while in the remaining no service is available.

5. Evaluation Results

Within the analysis it has shown up that the packet size has some major influence on the expected packet delay and consequently also on the overall throughput if specific QoS criteria have to be fulfilled.

With the given scenario parameters the mean waiting time W´ of an M/H2/1 queue can be simplified to

$$W' = \frac{(P1\mu 2 + P2\mu 1)\rho}{\mu 1\mu 2(1-\rho)}. \qquad (6)$$

In Figure 5 this waiting time is shown depending on the average packet size and parameterized with different overall system utilization. The network data rates were assumed to be equal to the high performance values of Table 1.

This fact is also stressed by the dependency of the curves on the overall utilization. The more the network capacity is used up the more dramatic is the increase in the average waiting time. Although even for a 90% use up of the network capacity and the maximum packet size of 2000 Byte the values are still below 200ms and therefore in reasonable boundaries the tendency is clearly observable.

To evaluate the achievable performance of an integrated WiFi-UMTS system with special regards to QoS criteria the 99[th] percentile of the waiting time depending on the system utilization is depicted in Figure 6. The network capacity is also varied according to the values of Table 1 for low, medium, and high performance. As a reference also the isolated WiFi scenario is investigated where the network capacity is set to 54 Mbit/s.

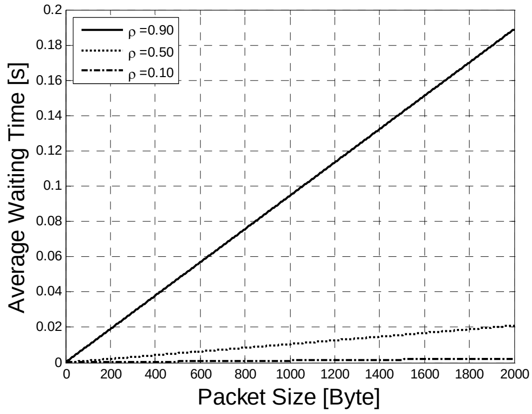

Figure 5 : W´ depending on Packet Size

This fact is also stressed by the dependency of the curves on the overall utilization. The more the network capacity is used up the more dramatic is the increase in the average waiting time. Although even for a 90% use up of the network capacity and the maximum packet size of 2000 Byte the values are still below 200ms and therefore in reasonable boundaries the tendency is clearly observable.

To evaluate the achievable performance of an integrated WiFi-UMTS system with special regards to QoS criteria the 99[th] percentile of the waiting time depending on the system utilization is depicted in Figure 6. The network capacity is also varied according to the values of Table 1 for low, medium, and high performance. As a reference also the isolated WiFi scenario is investigated where the network capacity is set to 54 Mbit/s.

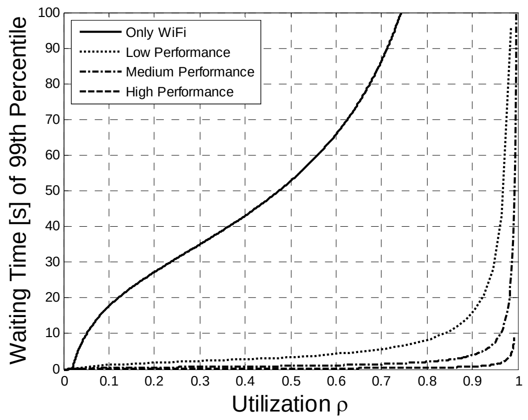

Figure 6 : 99th Percentile of W

All three curves for the integrated approach show the same behavior. After slowly increase in the region below 80% utilization the waiting dramatically escalates the closer the system comes to its capacity limit.

Due to the fact that the curves are depicted depending on the system utilization the network throughput cannot be easily compared. The utilization has to be associated with the maximum cumulated throughput of the system in order to get the actual system throughput. Therefore points of curves with the same x-value do only share the same utilization but not the same overall performance.

However, it can be noticed that, although the same utilization for all network setups is used the waiting time increases with lower network performance. The reason for it is reflected in Formula 6. The mean waiting time is not only influenced by the system utilization but also from the service rates. The lower the achievable throughput of the network is the higher the expected waiting time.

The results show that even with a Smart Cache at the connection point of the two networks the resulting waiting time for 80% system utilization is within reasonable boundaries for delay resilient applications like e.g. video streaming services.

By using the Smart Cache the maximum traffic within the backbone has an average value of λ while in a non-integrated system the load varies between the two data rates of the different wireless communication networks. Therefore the link to the core network can be smaller dimensioned, especially as the targeted applications are usually already the tasks with the highest bandwidth requirements. This fact is even more interesting in new network designs based on relays and multi-hop links where the smoothing of traffic reduces the required maximum link capacities.

The graph for the reference scenario with only 50% WiFi coverage gives also information about the applicability of Smart Caching in areas of discontinuous coverage. Up to 20% of the overall capacity of the network can be used if a delay of 30s is accepted. For streaming services this implies that enough data in the end device has to be buffered to bridge exactly such a period.

In Figure 7 the waiting time distributions for the reference scenario are shown. As a parameter the offered traffic load (λ·Packet_Size) was varied.

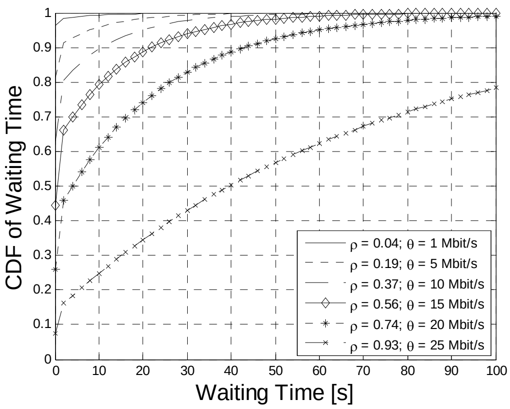

Figure 7 : CDF of Waiting Time in only-WiFi Scenario

The graphs show that the overall system, with an average achievable throughput of 27 Mbit/s (50% network coverage with 54 Mbit/s), can be operated in areas of 10-15Mbit/s offered traffic without suffering delays which cannot be absorbed be buffering techniques in the end device. Such an operation mode requires that the Smart Cache has stored enough data that in case of network coverage the wireless link can be fully utilized. If no Smart Cache would be available the backbone link would limit the performance if the end-to-end bandwidth would be smaller than 54Mbit/s which is a reasonable assumption.

6. Summary and Conclusion

Within this paper a traffic theory based approach for wireless heterogeneous networks was presented. A detailed derivation of the system model was given and the investigated application scenarios were motivated.

The results show that heterogeneous systems can be used to increase the overall performance of communication networks. Especially the applicability and capability of Smart Caching is proven. By using a cache at the connection point of the different wireless networks it is guaranteed that the utilization of the systems is maximized and that the backbone traffic gets much smoother. Without Smart Caching improvements the overall throughput might be diminished.

For the future it remains to improve the system model. Especially the serving process might be refined by using more than 2 phases. However, this will require new analysis methods as a calculation of the probability distribution in a closed form is no longer achievable.

It is also targeted to apply the introduced analysis to other wireless network scenarios and to investigate new network features.

REFERENCES

[1] N. Niebert et al. *"Ambient Networks: An Architecture for Communication Networks beyond 3G"*. IEEE Wireless Communications, Apr. 2004.

[2] S. Goebbels, and R. Probokoesoemo. *"Intelligent Caching Strategy for Mobile Communication Networks"*. Proceedings of European Wireless 2005, Nicosia, Cyprus.

[3] F.P. Kelly, S. Zachary, and I. Ziendins. *"Stochastic Networks – Theory and Applications"*. Clarendon Press, Oxford 1996.

[4] G. Latouche and V. Ramaswami. *"Introduction to Matrix Analytic Methods in Stochastic Modeling"*. American Statistical Association, 1999.

[5] Jin Cao, William S. Cleveland, Dong Lin, and Don X. Sun. *"Internet Traffic Tends Toward Poisson and Independent as the Load Increases"*. Lecture Notes in Statistics, Springer, New York, 2003.

[6] L. Kleinrock. *"Queueing Systems"*. Wiley, 1975.

[7] J. Shortle et al. *"An Algorithm to Compute the Waiting Time Distribution for the M/G/1 Queue"*. Informs Journal on Computing, Vol. 16, No. 2, Spring 2004.