TCP-aware Resource Allocation in Wireless Networks via Utility Maximisation

Toktam Mahmoodi, Vasilis Friderikos, Hamid Aghvami
Centre for Telecommunications Research, King's College London
{toktam.mahmoodi, vasilis.friderikos, hamid.aghvami}@kcl.ac.uk

***Abstract*—Motivated by the fact that a plethora of different applications, ranging from streaming video to file transferring, are using TCP we propose a set of utility functions that capture these application and optimise the transmission of TCP over wireless via a utility maximisation framework. We show that defining such a set of utility functions in the design of the wireless resource allocation problem can significantly benefit the overall efficiency metrics such as the end-to-end throughput and fairness among end-to-end flows . We quantify the achieved gains via numerical investigations by detailing a resource allocation problem based on the proposed framework.**

I. Introduction

As the Internet grows both in terms of users and available applications, one of the challenging issues is providing fair and efficient allocation of the available network resources. The Transmission Control Protocol (TCP), is the default transport layer protocol used in the Internet to provide reliable end-to-end communications and is responsible for more than 90% of all Internet traffic. However, TCP exhibits a number of shortages when the underlying medium detracts from the reliable, wired medium which it was originally designed to serve [1]. This is more crucial knowing that wireless data traffic is expected to increase drastically over the next few years [2].

Thereby, a significant amount of research has evolved over the past few years, aimed at improving the performance of TCP over wireless networks [3]. Two schools of thought have emerged: the first investigates changes to the transport-layer protocol, while the second explores the potential to enhance the characteristics of lower layers to improve the end-to-end performance of TCP. For compatibility reasons with the existing TCP implementations over Internet, this research focuses on the latter approach, aiming to improve the performance of TCP flows by utilising information from the lower layers [4][5].

To this end, we present a utility maximisation problem which brings the constraints of the end-to-end TCP flow and the wireless link of the corresponding flow together with the requirements of the utilised application. The proposed utility function can define the resource allocation problem of various wireless access techniques in the application and TCP-aware manner. In this regard, as an example of the proposed utility function we present a novel resource allocation scheme in Orthogonal Frequency Division Multiple Access (OFDMA) based wireless networks, thus investigate the achieved performance.

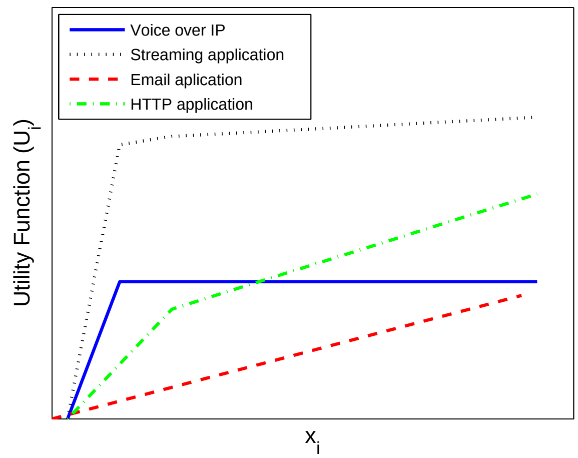

Figure 1. Utility functions for various Internet traffic types.

The remainder of this paper is organised as follows. In section II, the proposed utility functions are detailed. Section III provides a thorough description of a TCP-aware resource allocation problem as an application of the proposed utility functions. We present numerical investigations in section IV to highlight the main concepts. Finally, this paper concludes in section V together with some avenues of future research.

II. Problem Definition and Utility Functions

We investigate the problem of achieving the system optimal rates in the sense of maximising the aggregate utility, using cross-layer information from application layer, TCP, and the wireless link. Figure 1 gives an insight into the design methodology of our proposed utility functions for various applications.

Clearly, applications such as Voice over IP (VoIP) and video streaming require to ramp up to a specific data rate quickly. On the other hand, the utility of elastic applications such as Email and HTTP can grow in a more smooth (concave) manner with the allocated rate [6]. This effect can be observed in Figure 1. It is worthwhile noting that, Real Media and Windows Media, the two dominant streaming media applications, both are based on TCP streaming.

In data centric networks, a common objective of various resource allocation problems is to maximise the overall throughput. Moreover, user requirements such as fairness and QoS provisioning can also be implicitly or explicitly considered in such a setting. On the other hand, utilising side information such as the minimum data rate requirements, or the optimal achievable data rate for each user can provide significant benefits on the design of more efficient resource allocation strategies.

We present a utility maximisation approach to resource allocation problem, which incorporates not only the behaviour of the end-to-end TCP flow but also the application that uses TCP. In general for an elastic traffic source (as is the case of FTP traffic), the utility function should be a continuously differentiable, increasing and strictly concave function [7].

Given T, the set of active TCP flows, and Λ the capacity of wireless system, we can formulate the above described problem as follows,

$$\text{Maximise } \sum_{i \in T} U_i(x_i),$$

subject to:

$$X \in \Lambda, \tag{1}$$

$$X_{min} \leq X \leq X_{max}. \tag{2}$$

In the above equations, $[X]_i = x_i$; the minimum and maximum achievable values of X are defined by X_{max} and X_{min} in Equation (2), which can represent numerous per user or per TCP flow constraints. Finally, Λ can determine various capacity constraints in the wireless system such as total available power, and total channel capacity. The utility function U_i is defined based on the plot in Figure 1, that represent the shape of these functions for different applications.

The minimum required data rate can be defined based on the congestion window size that the TCP congestion control state transits from slow start to the congestion avoidance. Moreover, the performance of TCP has been analysed in [8], where closed-form expressions are derived based on the probability of a packet in error, and the end-to-end RTT. These formulas describe the average achievable data rate by the end-to-end TCP flow, and can be used in the allocation of resources in the wireless link.

Assuming OFDMA as the access method, we present an example of a resource allocation problem based on the proposed utility function. In addition, the effect of the proposed framework on the end-to-end performance is investigated.

III. TCP-Aware Resource Allocation Algorithm for OFDMA

Next generation wireless technologies, such as Long Term Evolution (LTE) [9] and IEEE 802.16e [10], specify OFDMA as their access method. The multiple access in OFDMA is achieved by assigning subsets of subcarriers to individual users. It is well known that dynamic allocation of subcarriers can significantly improve the overall performance of OFDMA systems. Thus, the implied problem of the joint subcarrier assignment and resource (bit and power) allocation for the OFDMA has been a prominent area of research over the past few years.

In the context of OFDMA, the problem of resource allocation is to determine the users' subcarrier allocation matrix $[A]_{ij} = a_{ij}$, and power distribution matrix $[P]_{ij} = p_{ij}$, while $i \in T$ and $j \in S$ in which S is the set of subcarriers. In these two matrices, a_{ij} is a binary value which represents the allocation of subcarrier j to flow i, and p_{ij} is the amount of transmission power assigned to flow i in the subcarrier j. In this respect, objectives such as maximising the overall data rate subject to power or Bit Error Rate (BER) constraints is studied in a number of previous research works [11]. On the other hand, a resource allocation problem can also address the minimisation of the overall power consumption with a minimum allocated rate constraint per user [12]. The later concept is of more interest for applications requiring a fixed data rate. None of the above approaches, however, considers issues regarding fair allocation among users.

Alternative formulations do, however, consider fairness, either by prioritisation using the weighted sum rate method [13], or by introducing proportional rate constraints [14]. Another approach is presented in [15], in which fairness is considered by maximising the lowest achieved data rate among the user set. The research presented in [16] addresses proportional fairness in OFDMA resource allocations, based on the Nash bargaining solution. Although in some of previous works the issue of fairness and QoS with respect to the allocated data rate over the wireless link is investigated, the aspects of fairness in end-to-end data transmissions are not sufficiently addressed.

A thorough overview of cross-layer design for resource allocation algorithms in the third generation wireless networks is given in [17], where TCP over CDMA is also addressed. TCP-aware resource allocation algorithms over a CDMA network are also studied in [18], in which the objective is to maximise the overall throughput. The proposed algorithm in the above paper uses information from the TCP state machine (i.e., slow start or congestion avoidance) to appropriately allocate the data rate at the wireless link. A joint congestion control and power allocation scheme for CDMA based wireless networks is proposed in [19], based on a generalised network utility maximisation framework. In the context of IEEE 802.16 wireless networks, reference [20] proposes a TCP-aware allocation algorithm which estimates the bandwidth demand based on the long-term available data rate, and allocates resources accordingly. In OFDMA based wireless networks, the research work presented in [21] propose a modifications to the power/subcarrier allocations in order to provide the proportional rate among users with regard to their end-to-end capacity. Presented results in [21] show significant improvement gains for TCP flows that compete over a wireless link.

In this paper, a power/subcarrier allocation problem is presented based on the proposed utility maximisation framework. In the resource allocation scheme, which is detailed in the sequel, we attempt to achieve an optimal design in the sense that it satisfies the lower layer requirements, and also provides fairer allocation with respect to TCP, thus enhance the performance over the end-to-end paths. Therefore, constraints similar to the ones used in [21] are utilised here.

In our defined problem, capacity Λ represents the total available power and the total number of subcarriers at the base station. Moreover, per flow data rates are constrained to be proportional to their optimal TCP throughput. Considering the above, the addressed resource allocation problem can be formally expressed as follows (problem - P1),

$$(P1): \text{ Maximise } \sum_{i \in T} U_i(R_i),$$

subject to:

$$P, A \in \Lambda, \tag{3}$$

$$\frac{R_1}{B_1} = \frac{R_i}{B_i}. \qquad \forall i \in T \tag{4}$$

In the above formulation, R_i expresses the data rate of the i^{th} flow over the wireless link. Moreover, B_i is the steady state throughput that end-to-end TCP flow i can achieve [8].

We define the utility function as a linear function for each application, as can be seen in Figure 1. Given the slopes ϕ_i of the linear functions (assuming that $\phi_1 > \phi_2 > ...\phi_6$) the utility functions shown in Figure 1 can be written as follows.

1) Media streaming:

$$U_i(R_i) = \begin{cases} \phi_1 \cdot R_i & R_i \leq R_{min-Media} \\ \phi_6 \cdot R_i & R_i > R_{min-Media} \end{cases} \tag{5}$$

2) VoIP:

$$U_i(R_i) = \begin{cases} \phi_2 \cdot R_i & R_i \leq R_{min-VoIP} \\ \phi_2 \cdot R_{min-VoIP} & elsewhere \end{cases} \tag{6}$$

3) HTTP:

$$U_i(R_i) = \begin{cases} \phi_3 \cdot R_i & R_i \leq R_{min-HTTP} \\ \phi_5 \cdot R_i & R_i > R_{min-HTTP} \end{cases} \tag{7}$$

4) Email:

$$U_i(R_i) = \phi_4 \cdot R_i \tag{8}$$

The above linear definition of the utility functions is an abstract model for the four applications that are discussed in this paper. Although these utilities are not accurately model the requirement of these applications, they can capture the overall behaviour of them. As mentioned before, VoIP or video streaming applications require to ramp up to a specific data rate quickly, thus larger slopes of ϕ_1 and ϕ_2 are assigned to their utility definitions. Similar argument to the one in section II is valid for assigning the smaller slopes of ϕ_3 and ϕ_4 to the HTTP and Email utilities. Using the above utility functions in problem (P1), the end-to-end performance of the proposed scheme is investigated.

TABLE I
SIMULATION PARAMETERS [9].

Bandwidth	5MHz
Target BER	10^{-4}
Channel model	ITU Pedestrian
Shadowing standard deviation	8dB
Total power at the base station	43dBm
Total power at the mobile user	23dBm

IV. PERFORMANCE INVESTIGATION

We simulate an OFDMA system with 52 subcarriers (equal to the number of OFDM subcarriers in IEEE 802.11a). The rest of simulation parameters are similar to the ones used in [9], and are summarised in Table I. The available bandwidth is 5MHz, the maximum available power at the base station and at the mobile terminals is 43dBm and 23dBm respectively. The thermal noise power, σ^2, is -107dBm (Johnson-Nyquist noise over 5MHz bandwidth), and the target BER is 10^{-4}. The wireless channel is modelled with the ITU pedestrian model ($PL = 40 \log_{10} d + 30 \log_{10} f + 49$, where d express users' distance from the base station and f is the operating frequency). Also a frequency selective slow fading is assumed with standard deviation of 8dB.

The MSS of the TCP flows is set to the standard maximum transfer unit of an Ethernet network which is 1460 bytes. We further assume that the end-to-end RTT for any of the TCP flows is a uniformly distributed random value in the range $[10ms, 200ms]$. We perform 150 Monte Carlo simulations; thus each simulation round runs with a new set of random RTT values per TCP flow. The matrix of utilised slopes for the various utility functions is given by $\Phi = [2, 1, 0.4, 0.1, 0.09, 0.05]$. The threshold values for $R_{min-Media}$, $R_{min-HTTP}$, and $R_{min-VoIP}$ are given 3Mbps, 2Mbps, and 1.5Mbps consecutively.

Our benchmark is problem (P1) but subject to constraint (3) only, which corresponds to the pure rate maximisation problem. The comparison parameter between the two resource allocation schemes is the well-known Jain's fairness index [22] as well as the aggregated data rate over the wireless link.

We simulate a scenario with four flows each of a different traffic type, while the average SNR over the wireless channel is 20dB. Figure 2 shows the results of this scenario. It can be seen that the minimum, average, and the maximum achievable fairness among competing flows are increased by 45%, 30%, and 10% respectively.

In the next simulation scenario the wireless channel average SNR varies from 5dB to 20dB and the aim is to investigate the effect of these variations on the achieved fairness index for both resource allocation schemes. Observe from Figure 3 that as the average SNR increases the achievable values of the fairness index are increasing. It is worth noting that over the whole range of the wireless channel SNR values, the fairness

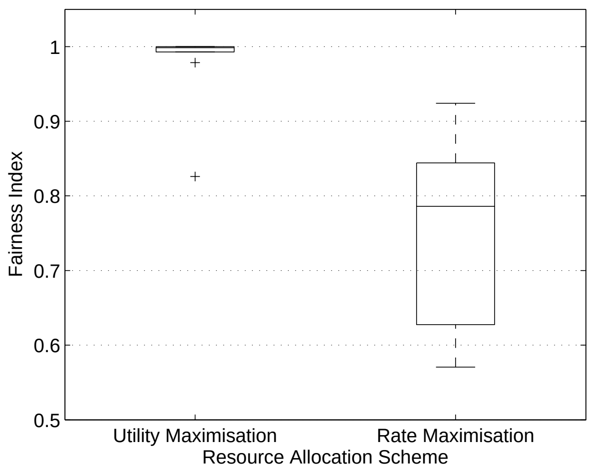

Figure 2. Fairness index among four competing TCP flows for the wireless channel with the 20dB average SNR.

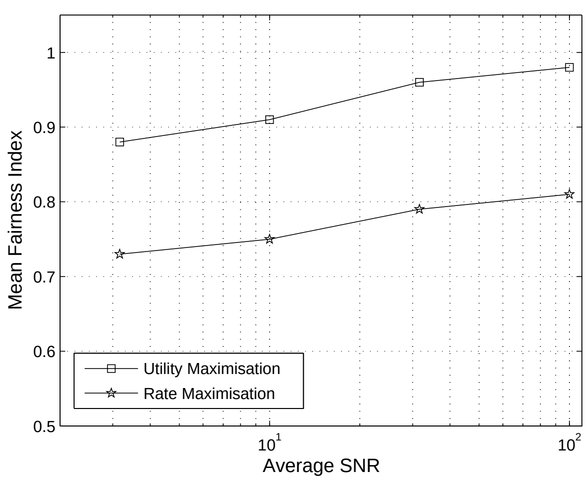

Figure 3. Fairness index among four competing TCP flows versus the average SNR of the wireless channel.

index is increased on average by 20%.

Results of this scenario in terms of the sum data rate are presented in Figure 4. These results show that along with the improvement in fairness index presented in Figure 2 and 3, the aggregated achievable data rate is penalised by the average of 13%. Clearly, we expect this degradation in the overall throughput in comparison to the pure rate maximisation scheme.

V. Conclusions

In this paper, a utility maximisation framework combined with TCP-aware resource allocation is proposed. An example of the proposed utility functions is detailed for OFDMA based wireless networks. Numerical investigations for different applications and range of wireless channel conditions are presented . The numerical investigations reveal that up to 45% improvement in the level of fairness among competing TCP

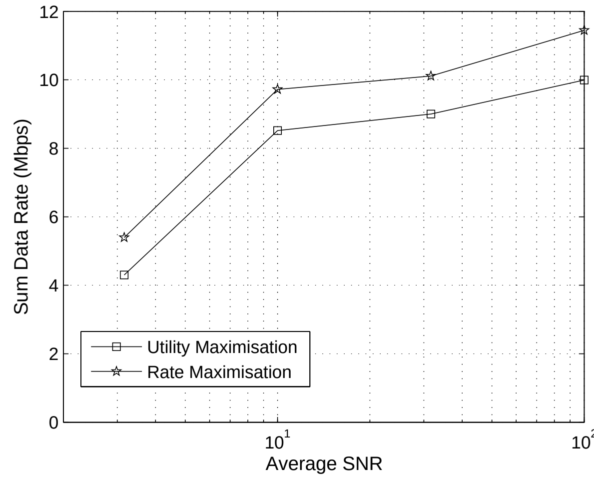

Figure 4. Aggregated data rate (Mbps) versus the average SNR of the wireless channel.

flows can be achieved, comparing with a baseline scheme in OFDMA based wireless networks that maximises the sum rate.

The linear expressions of the utility functions in this work entail a low computational complexity but at the same time may not give a precise description of each application's behaviour. Thus, an area of interesting future work would be to devise more accurate utility functions for the various applications. In addition, the definition of utility function can be per class or category of applications instead of per application, thus variants of utility function can be decreased and extra complexity will be avoided. Clearly, there is a trade-off between maximizing individual utilities and achieving in an efficient manner (in terms of power for example) the capacity at the air-interface. Such trade-offs which are dependent on the shape of the individual utility functions are another interesting areas of future research.

VI. Acknowledgment

The work reported in this paper has formed part of the Delivery Efficiency Core Research Programme of the Virtual Centre of Excellence in Mobile & Personal Communications, Mobile VCE, www.mobilevce.com. This research has been funded by EPSRC and by the Industrial Companies who are Members of Mobile VCE. Fully detailed technical reports on this research are available to Industrial Members of Mobile VCE.

References

[1] V. Tsaoussidis and I. Matta, "Open Issues on TCP for Mobile Computing," *Wireless Commun. and Mobile Comput.*, vol. 2, pp. 3–20, Feb. 2002.

[2] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update," *White Paper*, Jan. 2009.

[3] A. A. Hanbali and E. Altman, "A survey of TCP over Ad Hoc Networks," *IEEE Commun. Surveys & Tutorials*, vol. 7, no. 3, pp. 22–36, 2005.

[4] T. Mahmoodi, V. Friderikos, O. Holland, and A. H. Aghvami, "Cross-layer Optimization to Maximize Fairness among TCP Flows of different TCP Flavors," *Proc. IEEE GLOBECOM'08*, Nov. 2008.

[5] T. Mahmoodi, O. Holland, V. Friderikos, and H. Aghvami, "Cross-Layer Optimization of the Link-Layer based on the Detected TCP Flavor," *WWRF 20th*, Ottawa, Canada, May 2008.

[6] R. J. La and V. Anantharam, "Utility-Based Rate Control in the Internet for Elastic Traffic," *IEEE/ACM Trans. Net.*, vol. 10, pp. 272–286, Apr. 2002.

[7] S. Shenker, "Fundamental Design Issues for the Future Internet, IEEE Journal on Selected Areas in Communication," *IEEE J Sel. Areas Commun.*, vol. 13, pp. 1176–1188, Sep. 1995.

[8] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose, "Modeling TCP Throughput: A Simple Model and its Empirical Validation," *Proc. ACM SIGCOMM '98*, pp. 303–314, 1998.

[9] 3GPP Std. TR 25.814, *3rd Generation Partnership Project, Technical Specification Group Radio Access Network: Physical Layer Aspects for Evloved Universal Radio Access (UTRA)(Release 7)*, Sep. 2006.

[10] IEEE Std. 802.16e-2005, *Air Interface for Fixed and Mobile Broadband Wireless Access Systems*, Feb. 2006.

[11] I. C. Wong and B. L. Evans, "Optimal Downlink OFDMA Resource Allocation with Linear Complexity to Maximize Ergodic Rates," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 962–971, Feb. 2008.

[12] Z. Han, Z. Ji, and K. J. R. Liu, "Power Minimization for Multi-Cell OFDM Networks Using Distributed Non-cooperative Game Approach," *Proc. IEEE GLOBECOM'04*, vol. 6, pp. 3742–3747, Nov. 2004.

[13] L. Hoo, B. Halder, J. Tellado, and J. M. Cioffi, "Multiuser Transmit Optimization for Multicarrier Broadcast Channels: Asympotic FDMA capacity region and algorithms," *IEEE Trans. Commun.*, vol. 52, pp. 922–930, June 2004.

[14] Z. Shen, J. G. Andrews, and B. L. Evans, "Adaptive Resource Allocation in Multiuser OFDM Systems With Proportional Rate Constraints," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 2726–2737, Nov. 2005.

[15] W. Rhee and J. M. L. Cioffi, "Increase in Capacity of Multiuser OFDM System using Dynamic Subchannel Allocation," *Proc. IEEE VTC Spring*, pp. 1085–1089, May 2000.

[16] Z. Han, Z. Ji, and K. J. R. Liu, "Fair Multiuser Channel Allocation for OFDMA Networks Using Nash Bargaining Solutions and Coalitions," *IEEE Trans. Commun.*, vol. 35, pp. 1366–1375, Aug. 2005.

[17] H. Jiang, W. Zhuang, and X. Shen, "Cross-Layer Design for Resource Allocation in 3G Wireless Networks and Beyond," *IEEE Commun.*, vol. 43, pp. 120–126, Dec. 2005.

[18] M. Ghaderi, A. Sridharan, H. Zang, D. Towsley, and R. Cruz, "TCP-Aware Channel Allocation in CDMA Networks," *IEEE Trans. Mobile Comput.*, vol. 8, pp. 14–28, Jan. 2009.

[19] M. Chiang, "Balancing Transport and Physical Layers in Wireless Multihop Networks: Jointly Optimal Congestion Control and Power Control," *IEEE J Sel. Areas Commun.*, vol. 23, pp. 104–116, Jan. 2005.

[20] S. Kim and I. Yeom, "TCP-aware Uplink Scheduling for IEEE 802.16," *IEEE Commun. Letters*, vol. 11, pp. 146–148, Feb. 2007.

[21] T. Mahmoodi, V. Friderikos, O. Holland, and A. H. Aghvami, "TCP-aware Resource Allocation Problem in OFDMA based Wireless Networks," *Proc. IWCLD*, Jun. 2009.

[22] R. Jain, D. Chiu, and W. Hawe, "A Quantitative Measure Of Fairness And Discrimination For Resource Allocation In Shared Computer Systems," DEC Research Report TR-301, Sep. 1984.