Balancing Sum Rate and TCP Throughput in OFDMA based Wireless Networks

Toktam Mahmoodi, Vasilis Friderikos, Oliver Holland, Hamid Aghvami
Centre for Telecommunications Research (CTR)
King’s College London

Outline

- Motivations.
  - Global mobile data traffic forecast, and applications using TCP.
  - Fairer distribution among TCP flows.
- Orthogonal Frequency Division Multiple Access (OFDMA) in new generations of wireless.
- Resource allocation scheme in OFDMA.
- TCP-aware resource allocation problem.
  - TCP steady state throughput as the capacity of the end-to-end path.
  - Cross-Layer Interactions.
- Performance observations.
  - Fairness among end-to-end flows.
- Conclusions.
Motivations

- TCP is the predominantly used transport layer protocol to achieve reliable end-to-end data transfer in IP based networks.
  - According to the statistics, TCP is responsible for 90% of the Internet traffic.

- Various applications are using TCP as their transport layer. Examples include:
  - Elastic applications such as HTTP, FTP.
  - Video Streaming, e.g., Real Media and Windows Media.
  - Signalling for VoIP applications, e.g., Skype.

- There might be a large diversity among competing end-to-end TCP flows.
  - Various end-to-end RTTs.
  - Different versions of TCP at the end-hosts.

- This is especially important if there is a wireless link in the end-to-end connection.
  - For example, resource allocation algorithms such as channel/power allocation, FEC coding rate selection, and the ARQ mechanism at the wireless base station can be optimised utilising the information about the end-host TCP connection.
Motivations

- **Fairness among TCP flows**

  - The presented fairness Index is Jain’s Index that is proposed in 1984 to measure fairness in computer networks.
  - \( x_i \) is the actual throughput normalised by the optimal achievable throughput.
  - \( FI \) is in range \([0,1]\) and \( FI = 1 \) is the complete fair allocation.

\[
FI = \frac{\left( \sum_{i=1}^{n} x_i \right)^2}{n \cdot \sum_{i=1}^{n} x_i^2}.
\]

**Equal data rate allocation (BW = 30Mbps)**

- Opt. Thr. 15Mbps
- Opt. Thr. 25Mbps
- Opt. Thr. 5Mbps

\( FI = 0.68 \)

**Proportional data rate allocation (BW = 30Mbps)**

- Opt. Thr. 15Mbps
- Opt. Thr. 25Mbps
- Opt. Thr. 5Mbps

\( FI \sim 1 \)
Resource Allocation problems in OFDMA

- Various objectives for efficient resource allocation have been studied in the literature.
  - Rate Maximisation problem subject to power or BER constraints.
  - Power Minimisation problem subject to the minimum data rate constraint, which is of more interest for applications requiring fixed data rate.
  - Either Rate Maximisation or Power Minimisation subject to the requested QoS or Fairness.
TCP-aware Resource Allocation Scheme

- TCP-aware resource allocation problem is defined as a balance between maximising sum rate and the achieved end-to-end (TCP) throughput.

- Multi objective optimisation problem attempts to minimise the gap between the allocated data rate to each flow and the theoretically achievable throughput by its TCP.

- Given $R_i$ the allocated rate to the $i^{th}$ user, and $D_i$ the difference between $R_i$ and the achievable TCP send rate (steady-state TCP rate) by TCP of flow $i$, our objective is given by:

$$\text{Maximise } \sum_i R_i - \mu \cdot \sum_i D_i$$
TCP-aware Resource Allocation Scheme

- We use the well-studied approach to combine the multiple objectives into a single objective whose solution is **Pareto optimal**.
  - The optimal solution is not unique and it depends on the value of $\mu$ that balances the two objectives.
  - In this problem increasing $\mu$ can move the allocation balance towards TCP throughput, while decreasing $\mu$ move the balance towards data rate maximisation.

- The proposed resource allocation scheme is constrained by the available power and available subcarriers (wireless resources).
The steady state throughput of an end-to-end TCP flow \((B)\) is a function of the packet loss probability \((PER)\), the end-to-end \(RTT\), and the actual version of TCP congestion control [1].

For Example, TCP Reno is described by the following expression,

\[
B_i = MSS \cdot \frac{\frac{1-e_i}{e_i} + E[W_i]}{RTT_i \left(\frac{1}{2} \cdot E[W_i] + 2\right)},
\]

\[
E[W_i] = -\frac{1}{3} + \sqrt{\frac{8(1-e_i)}{3e_i} + \left(\frac{1}{3}\right)^2}.
\]

\(e_i\) : PER of flow \(i\).

\(MSS\): TCP Maximum Segment Size.

---

Cross Layer Interaction

- The cross-layer information should pass from TCP at the end-host to the link-layer (e.g. MAC) at the wireless base station.
Cross Layer Interaction

- RTT and PER are required to calculate the theoretical TCP throughput.
  - Various methods are presented in the literature to estimate the RTT passively at any middle point of the end-to-end path [2] such as wireless base station.
  - RTT can be calculated using the Timestamp option of the TCP header.
  - Assuming that wireless link is the bottleneck, PER can be calculated given the wireless BER.

Example

- Using information from the end-host TCP layer results in a fairer allocation of the wireless resources.

- Theoretical throughput of the TCP connection is considered to be the capacity of the end-to-end path.
- A two-user scenario is studied.
- Wireless resources are allocated proportionally to the Theoretical TCP Throughput of each flow.
- Jain’s Fairness Index is calculated based on the achieved throughputs.
Solving approaches

This joint power-subcarrier allocation represents a mixed integer non-linear programming problem which pose a high computational complexity.

For real-time implementation and to allow larger instances of the problem to be solved we present a greedy allocation which provides suboptimal but feasible solutions.

Maximise \[ \sum_i R_i - \mu \cdot \sum_i D_i \]

\[ R_i = a_{ij} w_j \log \left( 1 + \frac{p_{ij} G_{ij}}{\sigma^2 c} \right) \]

\[ D_i = |\alpha \cdot B_i - R_i| \]

s.t. \[ \sum_i \sum_j a_{ij} p_{ij} \leq P_t \]

\[ \sum_i a_{ij} \leq 1, \forall j \]

\[ p_{ij} \geq 0, \forall i, j \]

\[ a_{ij} \in \{0,1\}, \forall i, j. \]

\( a_{ij} \): Subcarrier allocation index.

\( p_{ij}, G_{ij} \): Allocated power, and channel gain of user \( i \) over subcarrier \( j \).

\( w_j \): Bandwidth of subcarrier \( j \).

\( \sigma^2 \): Thermal noise power.

\( c \): Constant \( c \) is a function of BER (fixed BER is assumed over all subcarriers, that can be achieved by M-QAM.)
Real-Time Solution

- We decouple the problem to two separate problems.
  - The subcarrier allocation problem is solved with a heuristic method assuming equal power allocation.
  - Afterwards, power is distributed optimally over the allocated set of subcarriers using the waterfilling approach.

- The principle of the subcarrier allocation algorithm is,
  - for each user to allocate the subcarrier with the highest channel gain available.
  - Afterwards, at each iteration, the user with the lowest value of \( R_i - \mu D_i \) chooses a subcarrier.
  - Finally \( \Omega_i \) is the set of assigned subcarriers to user \( i \).
Subcarrier Allocation

Algorithm 1 Subcarrier Allocation Algorithm

a) Initialization
   1) Set $R_i = 0$ and $\Omega_i = \phi$ for $i = 1$ to $n$ and $C = \{1, 2, ..., m\}$.
   2) Sort the users’ index in the descending order of $B_i$.

b) for $i = 1$ to $n$
   1) Find the subcarrier $k$ satisfying $|G_{ik}| > |G_{ij}|$ for all $j \in C$.
   2) Let $\Omega_i = \Omega_i \cup \{k\}$ and $C = C - \{k\}$.
   3) Update $R_i$

c) while $C \neq \phi$
   1) Find user $l$ satisfying $R_l - \mu D_l < R_i - \mu D_i$ for all $i \in \{1, ..., n\}$.
   2) For user $l$, find the subcarrier $k$ satisfying $|G_{ik}| > |G_{ij}|$ for all $j \in C$.
   3) Let $\Omega_l = \Omega_l \cup \{k\}$ and $C = C - \{k\}$.

By the end of the allocation, $\Omega_i$ is the set of allocated subcarriers to user $i$. 
Simulation Parameters

- Simulation parameters are as follows,
  - Number of OFDMA subcarriers: 52
  - Bandwidth: 5 Mhz
  - Available Power at the BS: 40 dBm
  - Target BER: 1E-4
  - Average SNR: 20 dB
  - Thermal noise power: -107 dBm
  - TCP Maximum Segment Size: 1460B (Ethernet size).
  - End-to-End RTT: Uniformly distributed random (10ms, 200ms)
  - Wireless link is highly utilised.
- Presented results are based on 150 Monte Carlo simulations (RTT and SNR are randomly generated in each run).
- The benchmark scheme is the same resource allocation when $\mu = 0$, i.e. purely rate maximisation.
Performance Investigations

- The achieved fairness among users, and the utilised wireless capacity in the two-user, and ten-user scenarios are studied.
  - Enhancements of up to 45% in the fairness index can be observed.
  - The degradation in the sum capacity is approximately 5%.
  - Up to 10% more balance towards the average TCP throughput is also achieved.
- Effect of changing the value of $\mu$ on the performance of our scheme.
  - In our investigated scenarios, $\mu=0.3$ show the best performance enhancements.
Results: Data Rate Vs. $\mu$

- Two-User scenario: aggregated data rate on the wireless link is plotted on the blue bar and the difference between the allocated rate and the average TCP throughput on the red bar.

By increasing the value of multiplier $\mu$ from 0 to 2:

- The achieved sum rate on wireless link is decreased.
- At the same time the achieved data rate gets closer to its optimal value from the TCP perspective.
Results: Data Rate Vs. $\mu$

- Ten-User scenario: aggregated data rate on the wireless link is plotted on the blue bar and the difference between the allocated rate and the average TCP throughput on the red bar.

Across the range of values for multiplier $\mu$:

- The utilized capacity of wireless network is decreased by 14%.
- The overall achieved data rate is 20% closer to the average end-to-end capacity, which is defined by theoretical TCP throughput.
Results: Fairness Vs. $\mu$

- Increasing $\mu$ from 0 to 0.3, the average fairness index is increased by 30% and 45% in the two-user and ten-user scenarios, respectively.
Conclusions

- TCP-aware resource allocation scheme in the context of OFDMA wireless is investigated.
  - This allocation accomplish balance between maximising sum rate and the achieved end-to-end (TCP) throughput.
  - Fairness among end-to-end flows increase significantly.
  - The aggregated throughput is minimally affected.
- This research is motivated by:
  - The use of TCP for various existing applications.
  - The large increase in the mobile/wireless data traffic in the Internet.