INFORMATIK

Selfish Traffic Allocation for Server Farms

Artur Czumaj Piotr Krysta
Berthold Vöcking

MPI–I–2003–1–011 April 2003

FORSCHUNGSBERICHT RESEARCH REPORT

MAX-PLANCK-INSTITUT
FÜR
INFORMATIK

Stuhlsatzenhausweg 85 66123 Saarbrücken Germany

Authors' Addresses

Artur Czumaj
Department of Computer Science
New Jersey Institute of Technology
University Heights
Newark, NJ 07102, USA
czumaj@cis.njit.edu
Partially supported by NSF grant CCR-0105701 and SBR grant No. 421090.

Piotr Krysta
Max-Planck-Institut für Informatik
Stuhlsatzenhausweg 85
66123 Saarbrücken, Germany
krysta@mpi-sb.mpg.de
Partially supported by DFG grant Vo889/1-1, and IST program of the EU under contract IST-1999-14186 (ALCOM-FT).

Berthold Vöcking
Dept. of Computer Science
Universität Dortmund
Baroper Str. 301
44221 Dortmund, Germany
voecking@ls2.cs.uni-dortmund.de
Partially supported by DFG grant Vo889/1-1.

Abstract

We study the price of selfish routing in non-cooperative networks like the Internet. In particular, we investigate the price of selfish routing using the coordination ratio and other (e.g., bicriteria) measures in the recently introduced game theoretic network model of Koutsoupias and Papadimitriou. We generalize this model towards general, monotone families of cost functions and cost functions from queueing theory. A summary of our main results for general, monotone cost functions is:

- We give an exact characterization of those cost functions that have a bounded/unbounded coordination ratio. For example, the coordination ratio for cost functions describing the expected delay in queueing systems is unbounded.
- In addition, we show that an unbounded coordination ratio implies an extremely high performance degradation under bicriteria measures as well. In fact, the price of selfish routing can be as high as a bandwidth degradation by a factor that is linear in the network size.
- We separate the game theoretic (integral) allocation model from the (fractional) flow model by demonstrating that even a very small, in fact negligible, amount of integrality can lead to a dramatic performance degradation.
- We unify recent results on selfish routing under different objectives by showing that an unbounded coordination ratio under the min-max objective implies an unbounded coordination ratio under the average cost objective and vice versa.

Our special focus lies on cost functions describing the behavior of Web servers that can open only a limited number of TCP connections. In particular, we compare the performance of queueing systems that serve all incoming requests with servers that reject requests in case of overload. We come to the conclusion that queueing systems without rejection cannot give any reasonable guarantee on the expected delay of requests under selfish routing even when the injected load is far away from the capacity of the system. In contrast, Web server farms that are allowed to reject requests can guarantee a high quality of service for every individual request stream even under relatively high injection rates.

NOTE: A preliminary version of this paper appeared in the Proc. of the 34th ACM Symposium on Theory of Computing (**STOC**), Montreal, Canada, May 19–21, 2002.

Keywords

Traffic Allocation, Selfish Routing, Game Theory

1 Introduction

The *coordination ratio* is a game theoretic measure introduced by Koutsoupias and Papadimitriou in [14] to reflect the price of selfish routing in *non-cooperative network systems* (like the *Internet*). In large-scale communication networks it is typically impossible to regulate network traffic to achieve an optimal performance of the system. In such networks, a common strategy of individual users is to follow the most rational approach, that is, to behave selfishly to minimize its cost (or maximize profits). This motivates investigations in game theoretical flavor of the relationship between systems in which each user is aware of the situation facing all other users and trying to optimize its own strategy (that is, being in a *Nash equilibrium*) and optimal strategies in such systems.

It is well known that Nash equilibria do not always optimize the overall performance. Therefore, in order to understand the phenomenon of non-cooperative systems Koutsoupias and Papadimitriou [14] initiated investigations of the behavior of the worst-case *coordination ratio*, which is the ratio between the worst possible Nash equilibrium and the social (i.e., overall) optimum. In other words, this analysis seeks the price of uncoordinated individual utility-maximizing decisions ("the price of anarchy"). Koutsoupias and Papadimitriou [14, 19] proposed that model to study routing problems in which a set of several agents want to send traffic from the sources to destinations along parallel links with linear cost functions.

In this paper, we aim to generalize these kind of routing problems towards more realistic cost functions. For this purpose, we focus on a particular example in which parallel links are the natural model. In particular, we investigate the example of a Web server farm. Suppose some companies maintain servers all over the world and offer content providers to store data for them. Such servers could store, e.g., large pictures and other embedded files as this kind of data makes up most of the load. The request streams that would normally go to the servers of the content provider must now be redirected to these new Web servers. Clearly, this defines a load balancing problem in which streams must be mapped to the servers such that a high quality of service can be guaranteed for every stream. Important aspects that have to be taken into account is that different streams might have different characteristics, e.g., caused by different file length. For practical studies that investigate the reasons and impacts of this variable traffic see, e.g., [2, 4, 5, 22, 23].

In todays server farms the mapping of data streams to servers is typically done by a centralized or distributed algorithm that is under control of the provider of the server farm. We can imagine, however, that such a service can be offered in a completely different way without global control. For example, each stream of requests is managed by a selfish agent (e.g., the content provider) which decides to which server the stream is directed. In this case, every agent would aim to minimize its cost. These cost could be the expected latency experienced by the individual requests in the stream or the fraction of requests that are rejected as servers are overloaded.

In this paper, we make a first attempt to investigate coordination ratios under more realistic cost functions. Our general approach is that we investigate families of cost functions that are defined as general as possible. In most of our studies, we assume that data streams can be described by a single weight (depending on the rate of requests and the expected length of the requested files). In this case, the load on a server is defined to be the sum of the weights of the streams mapped to the server. However, the cost occurring at a server can depend in a non-trivial way on this load, e.g., we consider general, continuous and nondecreasing cost functions.

In addition, we analyze selfish routing under more complex cost functions that do not only depend on the sum of the injected rates but also on distributions of session length that might be different for different streams. We want to point out that our approach is not perfect in the sense that it still assumes Poisson arrivals of requests in every individual stream. However, we are able to derive the first rigorous analysis for selfish routing taking into account the heterogeneous nature of Internet traffic in form of general session length distributions.

Outline. Section 2 contains the description of a part of the model that is not related to game theory, whereas Section 3 introduces concepts related to game theory. Previous and related work is outlined in Section 4. Section 5 presents our results concerning general monotone cost functions, and Section 6 contains our results about queueing theoretical cost functions. Finally, Section 7 has some concluding remarks. We keep in this paper the policy that in order to follow the discussion of our new results and their consequences the reader basically does not need to go to the proofs of these results presented in Sections 5 and 6. Thus, reading the content of Section 5 and 6, skipping the proofs therein, gives a precise description of our results. A quite technical proof of Theorem 12 is shifted to Appendix A.

2 Description of the Routing Problem

The routing problem described above can be formally defined as a scheduling problem with m servers and n data streams. The set of servers and streams is denoted by $[m]$ and $[n]$, respectively. The data streams shall be mapped to the servers such that a cost function (describing typically either waiting times or deny of service) is minimized. The particular cost functions that we consider will depend on the traffic and server model that will be defined next in this section. In general, we will compare the solution (i.e., a mapping of streams to servers) achieved by selfish agents with the optimal solution to a min-max optimization problem based on coordination ratios that will be defined in the following, too.

2.1 Traffic model

Traffic streams are assumed to be sequences of requests for service. These sequences are assumed to be of stochastic nature. For simplicity, we assume that requests are issued by independent customers and hence are Poisson. Let r_i denote the *injection rate* of data stream i. The *lengths* of the sessions of stream i are determined by a probability distribution $\mathcal{D}_i$, i.e., the *service time* or *session length distribution* of stream i. We define the *weight* of stream i to be $\lambda_i = r_i \cdot \mathbf{E}[\text{session length with respect to } \mathcal{D}_i]$.

A *server farm* is a set of m servers, all using the same policy to serve requests. Different servers, however, may have different *bandwidths*. Let b_j denote the bandwidth of server j. Without loss of generality, we assume $b_1 \leq b_2 \leq \cdots \leq b_m$. Unless otherwise stated, we will assume that streams are *unsplittable*, that is, every stream must be mapped to exactly one server.

2.2 Server models

We distinguish two general server farm models, one consisting of m parallel servers and another of m sequential servers. A *parallel server* has multiple service channels. On each *service channel* it can serve a session independently from other channels. The number of channels on server j is b_j, its bandwidth, and each channel serves requests with fixed service rate one, that is, the time a channel needs to serve a session is equal to the length of the session. (Recall that session lengths in stream i are determined by a probability distribution $\mathcal{D}_i$.) For example, the number of channels may correspond to the number of TCP connections that can be opened simultaneously on a web server. In contrast, a *sequential server* has only one service channel. This channel works at service rate b_j. Thus, the time a channel needs to serve a session is equal to the length of the session divided by b_j. We will consider both farms of parallel servers and farms of sequential servers.

Another important aspect that leads to different cost functions is how *blocked requests* are handled, i.e., requests that can not be served immediately because all channels are occupied. Again, we distinguish two extreme models, namely, queueing and rejection models. In the *queueing model*, every server holds an FCFS queue in which it inserts blocked requests. When a service channel finishes a session and the queue is non-empty it immediately starts serving the next request in the queue. (All channels on the same server share the same queue.) In this model, the objective is to minimize the maximum expected delay. In the *rejection model*, blocked requests are rejected and disappear from the system. A natural objective in this model is to minimize the number of rejected requests.

In queueing theoretical notation, a server with k channels that delays blocked requests corresponds to a so-called M/$\mathcal{D}$/k/∞ or short M/$\mathcal{D}$/k queue, where $\mathcal{D}$ corresponds to the service time (session length) distribution of the injected request stream. When blocked customers are rejected then the corresponding queue is denoted by M/$\mathcal{D}$/k/k.

2.3 Cost functions and families of cost functions

The cost occurring at the servers under some fixed *allocation* (i.e., a mapping of streams to servers) is defined by *families of cost functions* $\mathcal{F}_B = \{f_b | b \in B\}$, where B denotes the domain of possible bandwidth values and and f_b describes the cost function for servers with bandwidth $b \in B$. Typically, we will assume $B = \mathbb{R}_{>0}$ or $B = \mathbb{N}_{>0}$ but additionally we will study finite domains of bandwidth. For example, a collection of identical servers with some specified bandwidth b can be described by a family of cost functions $\mathcal{F}_B = \{f_b\}$.

A cost function f is called a *simple* if it depends only on the injected load, that is, if the cost of a server is a function of the sum of the weights mapped to the server (but does not depend on other characteristics like, e.g., the session length distribution). A simple cost function is called *monotone* if it is positive, continuous, and nondecreasing. For an ordered set B, a family of simple cost functions $\mathcal{F}_B$ is called *monotone* if (i) f_b is monotone for every $b \in B$ and (ii) the cost functions are nonincreasing with b, i.e., $f_b(\lambda) \geq f_{b'}(\lambda)$, for every $\lambda \geq 0$ and $0 < b \leq b'$. A typical example for a monotone family of cost functions that we will consider is derived from the formula for the expected system time (delay) on an M/M/1 server with injection rate λ and service rate b, namely

$$\frac{1}{b - \min\{b, \sum_{i=1}^{n} \lambda\}}$$

Already Koutsoupias and Papadimitriou in their seminal work [14] ask for the price of selfish routing under cost functions of this form. We want to point out, however, that this is only one particular example for cost functions that we will investigate.

3 Preliminaries in Game Theory

We give a brief introduction containing all required definitions. We distinguish between the *integral* allocation model and the *fractional* flow model.

3.1 Integral allocation model

The *integral allocation model* distinguish between mixed and pure strategies. The set of *pure strategies* for agent $i \in [n]$ is $[m]$, that is, a pure strategy maps every stream to exactly one server. The mapping is described by a matrix $(x_i^j)_{i\in[n],j\in[m]}$, where x_i^j is an indicator variable with $x_i^j = 1$ if stream i chooses strategy j and 0 otherwise. A *mixed strategy* is defined to be a probability distribution over pure strategies. Let p_i^j denote the probability that agent i maps its stream to server j. We define the *load* of server j under an allocation $(x_1^j, \dots, x_n^j)$ by $w_j = \sum_{i=1}^{n} \lambda_i \cdot x_i^j$ and the *cost* of server j by

$$C_j = f_{b_j}(x_1^j, \dots, x_n^j) .$$

Observe that w_j and C_j are random variables with respect to the probabilities p_i^j. Let ℓ_j denote the *expected cost* on server j, that is, $\ell_j = \mathbf{E}[C_j]$. For a stream i, the expected cost of stream i on server j is defined by $c_i^j = \mathbf{E}[C_j \mid x_i^j = 1]$.

In this paper, we focus our attention on the problem of minimization of the maximum cost of the servers. In this model, we define the *social cost* C by

$$C = \mathbf{E}[\max_{j\in[m]} C_j]$$

and the *optimal cost (social optimum) opt* to be the minimum cost of a pure strategy:

$$opt = \min_{(x_i^j)_{i\in[n],j\in[m]}} \max_{j\in[m]} f_{b_j}(x_1^j, \dots, x_n^j) ,$$

where the minimum is taken over all $(x_i^j)_{i\in[n],j\in[m]}$ such that $x_i^j \in \{0, 1\}$ for $i \in [n], j \in [m]$ and $\sum_{j=1}^{m} x_i^j = 1$ for $i \in [n]$.

In order to formalize the notion of selfish behavior, we introduce Nash equilibria of server farms. We say the probabilities $(p_i^j)_{i\in[n],j\in[m]}$ define a (system in a) *Nash equilibrium* if and only if $p_i^j > 0$ implies $c_i^j \leq c_i^q$, for every $i \in [n]$ and $j, q \in [m]$. Thus, a Nash equilibrium is characterized by the property that there is no incentive for any task to change its strategy. Then, in such a model, the *coordination ratio*, for a fixed set of servers and streams, is defined by $\max \frac{C}{opt}$, where the maximum is over all Nash equilibria. In other words, the coordination ratio specifies by which factor the cost can increase due to selfish behavior. For a fixed family of cost functions $\mathcal{F}_B$, the *worst-case coordination ratio* R determines the maximum coordination ratio over all possible server farms, typically described as an asymptotic function in m.

In our study, we will identify several instances of cost functions in which R is unbounded. In this case, we will investigate the *bicriteria coordination ratio* $\overline{R}$. Let opt_Γ denote the value of an optimal solution assuming that all injection rates r_i are increased by a factor of Γ. Then the *bicriteria coordination ratio* $\overline{R}$ is defined to be the smallest Γ satisfying $C \leq opt_\Gamma$ over all Nash equilibria. In other words, we ask by which factor the injected load must be decreased so that the worst-case cost in Nash equilibrium cannot exceed the optimal cost for the original rates.

3.2 Fractional flow model

The motivation behind this model is to assume that every stream consists of infinitely many units each carrying an infinitesimal (and thus negligible) amount of flow. Each such a unit behaves in a selfish way. Intuitively, we expect each such unit to be allocated (selfishly) to a server promising minimum cost, taking into account the behavior of other units of flow. Assuming infinitesimal small units of flow, we come to the *fractional* variant of the integral allocation model (this model was also considered in [8, 24, 25, 26, 27]).

To define the model formally, in the fractional model the variables x_i^j that describe the mapping of streams to servers can take arbitrary *real* values from $[0,1]$ subject to the constraint $\sum_{j\in[m]} x_i^j = 1$, for every $i \in [n]$. The model does not distinguish between mixed and pure strategies, and in particular, the load on server j is simply defined by $w_j = \sum_{i\in[n]} x_i^j$. There are several equivalent ways to define a Nash equilibrium in this model. We use the characterization of Wardrop [29], see also [24]. A fractional allocation is in *Nash equilibrium* if $x_i^j > 0$ implies $C_j \leq C_q$, for every $i \in [n]$ and $j, q \in [m]$. The *coordination ratio* is defined analogously to the integral allocation model, and the *worst-case coordination ratio* is denoted by R^*.

3.3 Integral allocations with negligible weights

The fractional flow model is a simplification that aims to model the situation in which each stream carries only a negligible fraction of the total load. We will investigate in details the relationship between fractional flow and integral allocations of streams with tiny weights. For this purpose we define the notion of "ϵ-small streams". Assuming a server farm with identical servers and homogeneous traffic, we come to the following definition. Stream i is called *ϵ-small* if

$$\lambda_i \leq \frac{\epsilon}{m} \sum_j \lambda_j ,$$

that is, the stream has at most an $\frac{\epsilon}{m}$ fraction of the overall weight.

In case of heterogenous servers or heterogeneous traffic we need a slightly more technical definition. Let us fix a server farm and a set of streams with positive weights. Let *opt** denote the minimum maximum cost over all fractional flow allocations. For a stream $i \in [n]$, define the *scaled stream* i to be a stream with rate $r'_i = r_i/\epsilon$ and session length distribution $\mathcal{D}'_i = \mathcal{D}_i$. (Observe that this implies that the weight of the scaled stream is $\lambda'_i = \lambda_i/\epsilon$.) Then stream i is called ϵ-*small* if, for every $j \in [m]$, the cost of server j is at most *opt**, assuming this server gets assigned only the scaled stream i.

We define R_ϵ to be the worst-case coordination ratio under the restriction that all streams are ϵ-small.

3.4 Min-max versus average-cost objective functions

Besides to the min-max objective introduced above, we shall also briefly consider another objective function that has been investigated by Roughgarden and Tardos [24]. The *average-cost objective function* aims at minimizing the expected weighted average cost over all streams (see also, e.g., [6, 28], for related results). Technically, the cost under this objective function is defined by

$$C_{ave} \quad = \quad \frac{1}{\lambda} \sum_{j \in [m]} \mathbf{E}[w_j \cdot C_j] \quad ,$$

and the social optimum is defined by

$$opt_{ave} = \min \left(\frac{1}{\lambda} \sum_{j \in [m]} C_j \cdot f_{b_j}(x_1^j, \ldots, x_n^j) \right) \quad ,$$

where $\lambda = \sum_{i \in [n]} \lambda_i$ is the total injected weight and the minimum is taken over the same vectors as in the definition of the social optimum under the min-max objective.

We shall consider various coordination ratios for the average-cost objective function similarly as for the min-max objective function. In particular, the average-cost coordination ratios are defined in the same way as for the min-max model considered before, the only difference is that now one compares average social cost with average social optimum. The worst-case coordination ratios for integral, fractional, and ϵ-small integral allocation under this objective function are denoted by Σ, Σ^*, and Σ_ϵ, respectively.

4 Previous Research

The model investigated in this paper has been introduced recently by Koutsoupias and Papadimitriou [14]. In that paper, only the integral allocation model are studied and only for linear cost functions (that is, functions of the form $f_b(x_1, \ldots, x_n) = \sum_{i=1}^{n} \lambda_i x_i/b$). The motivation behind studying these cost functions was to investigate the price of selfish routing under simple cost functions. Koutsoupias and Papadimitriou [14] give some basic results for the worst-case coordination ratio in this model (e.g., tight bounds for two links). These results have been later extended in [3, 13, 16], and in particular, tight bounds in this model are established by Czumaj and Vöcking in [3]. No results for non-linear cost functions are known in this model.

Roughgarden and Tardos [24] (see also, e.g., [1, 6, 28]) considered a related general network model, where the streams may be required to be routed in a network from a source to a destination, and cost of allocating the stream to any edge in the routing path is taken into account. They focus mainly on the fractional flow model making the assumption that each stream consists of infinitely many units each of which behaves in a selfish way (see the fractional model described in Section 3.1). The network parameter (to be minimized) is the *total latency* experienced by the system. They showed that when the latency functions of the edges are linear, then the average latency in a Nash equilibrium is at most $\frac{4}{3}$ times the average latency in an optimal routing. For arbitrary nondecreasing and continuous latency functions, they show the existence of Nash equilibria whose total latency may be arbitrarily bigger than the average latency in an optimal routing. In particular, they give a bicriteria result that the total latency in Nash equilibrium is at most twice the total latency in an optimal routing of double the amount of flow. A similar model has been also investigated recently by Friedman [8], who studied how the amount of flow in the system influence the coordination ratios under the average-cost objective. Friedman studies also the queueing model M/M/1 in this context.

Furthermore, Roughgarden and Tardos [24] consider also integral allocations. They give an example of a network in which a bicriteria coordination ratio is unbounded. (This is in contrast to the fractional flow where the same ratio is bounded above by two.) In addition, they prove a sufficient condition under which the bicriteria coordination ratio is bounded. The sufficient condition is that streams are so small that adding any stream to any server increases the total latency at most by a factor α, for any fixed $\alpha > 0$. They remark that this condition is restricted to pure allocations and its application requires families of cost functions with $f_b(0) > \gamma$, for any $b > 0$ and fixed $\gamma > 0$.

5 Results for Monotone Cost Functions

5.1 Fractional allocation

Our first theorem shows that all monotone cost functions behave very nicely under fractional allocation.

Theorem 1 *For every server farm whose servers are described by monotone cost functions, the worst-case coordination ratio R^* in the fractional flow model is* 1.

This theorem follows almost directly from the definition of Nash equilibria specifying that all servers with positive flow at Nash equilibrium must have same cost, which under monotone cost functions implies that all Nash equilibria have the same social cost $C = opt$ (see, e.g., [24, Lemma 2.5]). As a consequence of this theorem, let us observe that it immediately separates (in the fractional flow model) the two different ratios investigated by Koutsoupias and Papadimitriou

[14] (the coordination ratio R^*) and by Roughgarden and Tardos [24] (the average-cost coordination ratio Σ^*). Both ratios compare the cost of a Nash equilibria with the optimal cost but use different definitions of cost. In sharp contrast to Theorem 1, Roughgarden and Tardos [14] prove that there exist instances of fractional flow for monotone families of cost functions in which the average-cost coordination ratio is unbounded [24]. This shows that minimum total latency and minimum maximum costs can differ arbitrarily under general cost functions. In the context we consider here, the server farms, the maximum cost seems to be the natural choice as it guarantees fairness and efficiency for every customer (agent) simultaneously.

5.2 Integral allocation

Now let us turn to the integral allocation model. We say a coordination ratio R over a family of cost functions $\mathcal{F}_B$ is *bounded* if for every m and every server farm with m servers with cost functions from $\mathcal{F}_B$ there exists $\Gamma > 0$ such that for every set of streams the value of the worst-case Nash equilibrium is at most $\Gamma \cdot opt$. (Observe that Γ might depend on m.) Otherwise, the coordination ratio is *unbounded*. Our first result is a complete characterization of monotone families of class functions for which the cost functions are bounded (vs. unbounded).

Theorem 2 *The coordination ratio R over a monotone family $\mathcal{F}_B$ of cost functions is bounded if and only if*

$$\exists \alpha \geq 1 \quad \forall b \in B \quad \forall \lambda > 0 \quad f_b(2\,\lambda) \leq \alpha \cdot f_b(\lambda) \ .$$

The proof of this theorem is shifted to Section 5.4. Notice that this characterization of bounded vs. unbounded coordination ratios can be applied also to server farms with identical servers. (Recall that such farms are described by families of cost functions consisting only of a single cost function.)

Observe that for every family of monotone cost functions we can identify a minimum $\alpha \in \mathcal{R}_{\geq 1} \cup \{\infty\}$ that fulfills the conditions specified in the theorem. A natural question is, how does the coordination ratio depend on α? In fact, our analysis (see the proof of Theorem 2 and Lemma 5) shows that the coordination ratio is at most $m^{O(\log \alpha)}$. Furthermore, observe that α is constant if the family of cost functions is assumed to be fixed. Thus, we can conclude that for every fixed family of cost functions the coordination ratio is either unbounded or it is polynomially bounded in the number of servers, m.

Let us illustrate the power of the above theorem by investigating some examples. First, we consider families over *polynomial cost functions*, i.e., functions of the form $\sum_{r=0}^{k} a_i \cdot \lambda^i$, for a fixed $k \geq 0$. For these families we can pick $\alpha = a_k \cdot 2^k$ to conclude that here the coordination ratio is bounded. In contrast, there is no such for *exponential cost functions*, i.e., cost functions for which an additive increase in the load leads to a multiplicative increase in the cost. The same is true for *linear threshold functions* of the form $\max\{0, \lambda - 1\}$.

Corollary 3 *The coordination ratio for server farms with polynomial cost functions is bounded, whereas the coordination ratio for server farms of (identical) servers with exponential or linear threshold cost functions is unbounded.*

In the next sections we shall discuss several other, practically motivated examples of families of cost functions with unbounded coordination ratios based on well-known formulas from queueing theory that describe, for example, the expected delay or the expected fraction of rejections.

5.3 Integral allocations with negligible weights

A motivation for considering fractional flow instead of integral allocations is that these two models are often assumed to be "essentially equivalent," see, e.g., Remark 2.3 in [24]. We want to investigate the relationship between these two models closer. Recall that R_ϵ is the coordination ratio for Nash equilibria over ϵ-small streams. Then we can show the following result whose proof is shifted to Section 5.4.

Theorem 4 *Given any monotone family of cost functions $\mathcal{F}_B$. For every $\epsilon > 0$, the coordination ratio R_ϵ over ϵ-small streams is bounded if and only if the coordination ratio R is bounded.*

Interestingly, this theorem *separates fractional flow from integral allocations with negligible weights*. On one hand, there are several practical examples for families of cost functions (e.g., M/M/1 waiting time or Erlang loss formulas) for which our results imply that for arbitrary small ϵ the coordination ratios R_ϵ (and hence R) are unbounded. On the other hand, Theorem 1 implies that R^*, the worst-case coordination ratio in the fractional model, is 1. Moreover, the instances proving the characterization of unbounded coordination ratios use only pure strategies. Thus, even *pure allocations with negligible weights are different from fractional flow*.

Finally, we want to point out that unbounded coordination ratios are not only a special phenomenon of cost functions having a pole or an unbounded first derivative. Later in the paper, we will see a practical example of a family of cost functions (based on the Erlang loss formula) that have an unbounded coordination ratio although these functions as well as their first derivatives are bounded above by one.

5.4 Proof of Theorems 2 and 4

We prove Theorems 2 and 4 in a single proof using two lemmas. The first lemma proves the sufficient conditions for a bounded coordination ratio, while the second lemma proves the necessary conditions.

Lemma 5 *Suppose we are given a server farm with m servers having cost functions from a fixed, monotone family $\mathcal{F}_B$ satisfying*

$$\exists \alpha \geq 1 \quad \forall b \in B \quad \forall \lambda > 0 \quad f_b(2\,\lambda) \;\leq\; \alpha f_b(\lambda) \;.$$

Then, for every set of streams the worst-case cost over all Nash equilibria is upper bounded by $opt \cdot m^{O(1)}$.

Proof. Fix an arbitrary allocation in Nash equilibrium. Let C denote the cost of this allocation. We will show that $C \leq m \cdot \alpha^{\lceil \log m \rceil} \cdot opt = m^{O(1)} \cdot opt$. First, we observe that

$$f_b(s\,\lambda) \leq \alpha^{\lceil \log s \rceil} \cdot f_b(\lambda) \;\;,$$

for every $s \geq 1$. Let $X := \sum_{i\in[n]} \lambda_i$ denote the total injected load. Let x denote the load on server 1 (with the biggest bandwidth) under an *optimal fractional allocation*, i.e., an allocation with minimum maximum cost over all servers assuming that streams can be split arbitrarily. Without loss of generality, we assume that server 1 has the maximum load over all servers, and hence $X \leq m\,x$. (Recall that since server 1 is the server with the biggest bandwidth and each function $f_b \in \mathcal{F}_B$ is nondecreasing in λ and nonincreasing in b, there exists an optimal fractional allocation with maximum load for server 1.) In this way,

$$f_{b_1}(X) \;\leq\; f_{b_1}(m\,x) \;\leq\; \alpha^{\lceil \log m \rceil} \cdot f_{b_1}(x) \;\leq\; \alpha^{\lceil \log m \rceil} \cdot opt \;\;.$$

Let M denote the set of servers $j \in [m]$ with $\sum_{i\in[n]} p_i^j > 0$. Pick any server j in M. Let $i \in [n]$ denote a stream with $p_i^j > 0$. Then, the Nash equilibrium property guarantees that $c_i^j \leq c_i^1$. Hence, for every $j \in M$,

$$\mathbf{E}\left[C_j\right] \;\leq\; \mathbf{E}\left[C_j | x_i^j = 1\right] \;=\; c_i^j \;\leq\; c_i^1 \;\leq\; f_{b_1}(X) \;\leq\; \alpha^{\lceil \log m \rceil} \cdot opt \;\;.$$

Furthermore, for $j \in [m] \setminus M$, we have $\mathbf{E}\left[C_j\right] = f_{b_j}(0) \leq opt \leq \alpha^{\lceil \log m \rceil} opt$. As a consequence,

$$C \;=\; \mathbf{E}\left[\max_{j\in[m]} C_j\right] \;\leq\; m \cdot \max_{j\in[m]} \mathbf{E}\left[C_j\right] \;\leq\; m \cdot \alpha^{\lceil \log m \rceil} \cdot opt \;=\; m^{O(1)} \cdot opt \;\;,$$

which completes the proof of Lemma 5. □

Now, we prove a sufficient condition for an unbounded coordination ratio. Observe that the negation of the property gives the sufficient condition for a bounded coordination ratio as specified in Theorem 2.

Lemma 6 *Let $\epsilon > 0$ be chosen arbitrarily. Suppose we are given a server farm with only two identical servers, each with the same monotone cost function f satisfying*

$$\forall \alpha \geq 1 \quad \exists \lambda > 0 \quad f(2\,\lambda) > \alpha \cdot f(\lambda) \;\;.$$

Then, for every $\Gamma \geq 1$, there exists a pure Nash equilibrium over ϵ-small streams with cost $C > \Gamma \cdot opt$.

Proof. First, let us show that the above property of the function f implies that

$$\forall \alpha \geq 1 \quad \beta > 0 \quad \exists \lambda > 0 \quad f((1+\beta)\cdot\lambda) > \alpha \cdot f(\lambda) \;\;. \tag{1}$$

Indeed, if we consider the negations of the two statements then we observe that $\exists \alpha \geq 1, \beta > 0 \quad \forall \lambda > 0 \quad f((1+\beta)\,\lambda) \leq \alpha\, f(\lambda)$ implies that $\exists \alpha' \geq 1 \quad \forall \lambda > 0 \quad f(2\,\lambda) \leq \alpha' f(\lambda)$, e.g., we can set $\alpha' = \alpha^{1/\lceil \log(1+\beta) \rceil}$.

Now, consider a server farm with two identical servers, each with the same monotone cost function f satisfying condition (1). For the purpose of contradiction, assume that there exists Γ such that $C \le \Gamma \cdot opt$, for every Nash equilibrium over streams of maximum weight ϵ. Therefore, by (1), there exists $\lambda > 0$ such that $f((1+\epsilon/10)\,\lambda) > \Gamma \cdot f(\lambda)$. Using this assumption, we will now define a Nash equilibrium over streams of maximum weight $\epsilon > 0$ with cost $C > \Gamma \cdot opt$.

First, let us consider streams of identical weight $w \in [\lambda\epsilon/2, \lambda\epsilon]$ and assign them to the servers so that the cost on each server is exactly $f(\lambda)$. Let τ be the number of streams per server in this allocation. Now, let us slightly change the instance by taking two streams, one from each server, and "break" each of them into two smaller streams, one of weight $\frac{3}{5}\,w$, the other of weight $\frac{2}{5}\,w$. It is easy to see that the optimal allocation for this instance has cost $opt = f(\lambda)$.

Let us consider a different allocations of the streams to the servers. We assign $\tau - 1$ streams of weight w and two streams of weight $\frac{3}{5}\,w$ to the first server and the remaining streams to the second server. In this way, the first server has cost $f(\lambda + \frac{1}{5}\,w)$ whereas the second server has cost $f(\lambda - \frac{1}{5}\,w)$. This allocation defines a Nash equilibrium, because the streams have minimum weight $\frac{2}{5}\,w$ and therefore there is no incentive for any of them to change its strategy. The cost of this Nash equilibrium, however, is

$$C \;=\; f \cdot \left(\lambda + \frac{w}{5}\right) \;\ge\; f \cdot \left(\lambda + \frac{\lambda \cdot \epsilon}{10}\right) \;>\; \Gamma \cdot f(\lambda) \;=\; \Gamma \cdot opt \;\;.$$

Clearly this contradicts our initial assumption that $C \le \Gamma \cdot opt$ for any Nash equilibrium over streams of maximum weight ϵ. This completes the proof of Lemma 6. □

Now, Theorems 2 and 4 follow immediately from Lemmas 5 and 6. □

5.5 Bicriteria Coordination Ratios

It is not surprising that selfish routing can lead to a dramatic cost increase when the cost function has a pole. In principle, bicriteria measures can be much more informative as they filter out the extreme behavior of cost functions and look directly at the quality of the load balancing obtained by selfish routing. The following theorem, however, shows that an unbounded coordination ratio R implies a very poor worst-case behavior under bicriteria measures as well.

The next theorem shows that given a monotone family of cost functions, the bicriteria coordination ratios $\bar{R}$ as well as $\bar{R}_\epsilon$ (for every $\epsilon > 0$) are at least m.

Theorem 7 *Consider a server farm with m servers, with cost functions drawn from a monotone family $\mathcal{F}_B$. If the coordination ratio R is unbounded, then the bicriteria coordination ratio $\overline{R}$ has value at least m, even over ϵ-small streams.*

Proof. Let n denote, as usual, the number of streams, and let $\alpha = m^{n-1}$. Assume, that n is such that $\frac{m}{n} \le \epsilon$. Since ratio R is unbounded, Theorem 2 implies that there exist $f_b \in \mathcal{F}_B$ and $\lambda > 0$, such that $f_b(2\lambda) > \alpha f_b(\lambda)$.

Assume that we have m identical servers, each with bandwidth $b_j = b$ ($j \in [m]$), and a set of n identical data streams, each having weight of $\lambda_i = \frac{\lambda m}{\Gamma n}$ ($i \in [n]$), where $\Gamma > 0$ will be specified later. Define the probabilities p^i_j as $p^i_j = \frac{1}{m}$, for each $i, j \in [m]$. These probabilities define a Nash equilibrium, since all the expected costs c^i_j have the same value.

Let us fix a server j. The probability that all the data streams are assigned to server j is $\Pi_i p^i_j = m^{-n}$. In this case, the cost on server j is $f_b\left(\frac{\lambda m}{\Gamma}\right)$, since we have n streams, each of weight $\frac{\lambda m}{\Gamma n}$, and so their total weight is $\frac{\lambda m}{\Gamma n} n = \frac{\lambda m}{\Gamma}$.

Therefore, with probability m^{-n}, the cost on a particular server j is at least $f_b\left(\frac{\lambda m}{\Gamma}\right)$, and thus also $\max_j C_j \geq f_b\left(\frac{\lambda m}{\Gamma}\right)$, over all $j \in [m]$. Additionally, these events corresponding to different servers are pairwise disjoint. Using these observations, we can estimate $C = \mathbf{E}\left[\max_j C_j\right]$ as

$$\mathbf{E}\left[\max_j C_j\right] \geq m \cdot f_b\left(\frac{\lambda m}{\Gamma}\right) \cdot m^{-n} = \frac{f_b\left(\frac{\lambda m}{\Gamma}\right)}{m^{n-1}} \quad .$$

We want to show that if $C \leq opt_\Gamma$, then $\Gamma \geq m$. Assume towards a contradiction that there is Γ such that $C \leq opt_\Gamma$ and $\Gamma < m$. Then, we obtain

$$\frac{f_b\left(\frac{\lambda m}{\Gamma}\right)}{m^{n-1}} \leq opt_\Gamma \leq f_b\left(\frac{\lambda m}{\Gamma n} \cdot \frac{n}{m} \cdot \Gamma\right) = f_b(\lambda) \quad ,$$

where the last inequality follows by observing that the value of opt_Γ is at most the value of a solution in which we assign $\frac{n}{m}$ (we can assume that $\frac{n}{m}$ is a positive integer) data streams to each server, after blowing each data stream up by Γ. Then, we obtain

$$f_b\left(\frac{\lambda m}{\Gamma}\right) \leq m^{n-1} \cdot f_b(\lambda). \tag{2}$$

By our assumption $\forall \alpha \geq 1\ \exists \lambda > 0\ f_b(2\lambda) > \alpha f_b(\lambda)$. This, by the argument from the proof of Lemma 6, implies that $\forall \alpha \geq 1, \beta > 0\ \ \exists \lambda > 0\ \ f_b((1+\beta)\lambda) > \alpha f_b(\lambda)$. Now, we plug $\alpha = m^{n-1}$, $1 + \beta = \frac{m}{\Gamma}$, to obtain a contradiction with inequality (2).

Since all the servers are identical, a data stream of weight λ_i is ϵ-small if $\lambda_i \leq \epsilon \frac{\sum_j \lambda_j}{m}$. Since in our case $\lambda_i = \frac{\lambda m}{\Gamma n}$ for all $i \in [n]$, the last condition is equivalent to $\frac{\lambda m}{\Gamma n} \leq \epsilon \frac{\lambda}{\Gamma}$, which gives $\frac{m}{n} \leq \epsilon$. The last inequality is true by the choice of n, and so our data streams are ϵ-small. □

The example proving the bad ratio from Theorem 7 is a server farm of identical servers. For this case one can easily show that a coordination ratio of m is the worst possible as a Nash equilibrium cannot be worse than mapping all streams to the same server and this, of course, is not worse than an optimal allocation with all rates blown up by a factor of m.

Even worse, our proof of this negative result gives a construction of a Nash equilibrium in which also all streams are identical and the total injected load $\sum_{i \in [n]} \lambda_i$ is less than the bandwidth of a single server. Thus, the coordination ratio can be very poor even in extremely lightly loaded cases.

5.6 Average-cost objective functions

As we have discussed it before, Roughgarden and Tardos [24] consider an objective function that is different from the one of Koutsoupias and Papadimitriou [14]. Their objective function aims to minimize average cost. Now, we investigate the relationship between the different objective functions under integral allocation. Recall that the coordination ratios for the average-cost objective function are denoted by Σ, Σ^*, and Σ_ϵ.

Theorem 8 *Given any monotone family of cost functions. Σ is bounded if and only if R is bounded. Furthermore, for every $\epsilon > 0$, Σ_ϵ is bounded if and only if R is bounded as well.*

Proof. Fix a monotone family $\mathcal{F}_B = \{f_b | b \in B\}$ of cost functions. First let us show that the ratio Σ for this family is bounded if the coordination ratio R is bounded. Clearly, if R is bounded then we obtain from Theorem 2

$$\exists \alpha \geq 1 \quad \forall b \in B \quad \forall \lambda > 0 \quad f_b(2\lambda) \leq \alpha f_b(\lambda) \ .$$

We will show that Σ is bounded provided this property is given. Fix any Nash equilibrium with probabilities p_i^j, $i \in [n]$, $j \in [m]$. We have to give an upper bound on the ratio between the expected total latency given by the probabilities p_i^j, on one hand, and the optimal total latency, on the other hand. Let $w = w_1, \ldots w_m$ denote a vector of random variables with w_j describing the injected load of server j. Let w^* denote a corresponding load vector of an optimal allocation that minimizes the total latency. We have to show that there exists $\Gamma \geq 1$ such that

$$\mathbf{E}\left[\sum_{j \in [m]} w_j f_{b_j}(w_j)\right] \leq \Gamma \sum_{j \in [m]} w_j^* f_{b_j}(w_j^*) \ .$$

Define $X = \sum_{j \in [m]} w_j = \sum_{j \in [m]} w_j^*$. Then $f_{b_1}(X)$ corresponds to the cost that is obtained by assigning all the load to the fastest server. Using the same arguments as in the proof of Lemma 5, we obtain $\mathbf{E}\left[f_{b_j}(w_j)\right] \leq f_{b_1}(X)$, for every $j \in [m]$, such that

$$\begin{aligned}
\mathbf{E}\left[\sum_{j \in [m]} w_j f_{b_j}(w_j)\right] &\leq \sum_{j \in [m]} X \, f_{b_1}(X) \\
&\leq m^2 \, \alpha^{\lceil 1 + \log m \rceil} \frac{X}{m} \, f_{b_1}\!\left(\frac{X}{m}\right) \\
&\leq m^2 \, \alpha^{\lceil 1 + \log m \rceil} \sum_{j \in [m]} w_j^* f_{b_j}(w_j^*) \ .
\end{aligned}$$

The second inequality follows from our assumptions on the family of cost functions. The third inequality needs some more explanation. Observe that there exists $j \in [m]$ with $w_j^* \geq X/m$. Applying monotonicity of our cost function family, we obtain $\frac{X}{m} f_{b_1}(\frac{X}{m}) \leq w_j^* f_{b_j}(w_j^*)$, which yields the inequality. As a consequence, $\Sigma \leq m^2 \, \alpha^{\lceil 1 + \log m \rceil}$, that is, Σ is bounded.

It remains to show that an unbounded coordination ratio R implies an unbounded ratio Σ. If R is unbounded then the family of cost functions satisfies

$$\forall \alpha \geq 1 \quad \exists b \in B \quad \exists \lambda > 0 \quad f_b(2\lambda) > \alpha f_b(\lambda) \ .$$

In Lemma 6 we describe an instance with two identical servers and a set of ϵ-small streams such that both servers have identical load in an optimal allocation but there is a Nash equilibrium in which one of the servers receives more load than the other server. Using the above property we show that if one server has load at least $(1 + \epsilon/10)$ times average load then the cost on this server can deviate by an arbitrary large factor from the optimal cost, which then proves that R is unbounded. A straightforward adaption of this argument, shows that also Σ is unbounded if the above condition is fulfilled. This completes the proof of Theorem 8. □

We want to point out that the property proven in Theorem 8 is non-trivial. In general, the behavior of the two objective functions can be quite different. For example, for the fractional case, Roughgarden and Tardos give examples with unbounded coordination ratio Σ^* whereas Theorem 1 shows that $R^* = 1$.

6 Results for Queueing Theoretical Cost Functions

Koutsoupias and Papadimitriou [14] asked to investigate the coordination ratio of realistic cost functions of the form $\frac{1}{b-\min\{\lambda,b\}}$. This particular cost function describes the expected waiting time (delay) under an M/M/1 queueing process with injection rate λ and service rate b. Our characterization of bounded and unbounded coordination ratios given in Theorem 2 immediately implies that the integral coordination ratios for this family of functions are unbounded. This resolves an open problem from [14]. In the following, we want to consider the behavior of cost functions motivated by queueing theory in more details. In particular, we give in Section 6.3 (under "The local exchange lemma") a simple greedy exact algorithm for solving a related fractional scheduling problem with P-K cost functions under heterogeneous traffic.

6.1 Queueing systems without rejection

Several other studies are concerned with similar cost functions motivated by queueing systems, see, e.g., [11, 12, 15, 18, 19, 20] In order to avoid discussions about what is exactly the right queueing model (e.g., M/M/1, M/D/1, M/G/1, M/G/c, ...) and what is exactly the right cost model (e.g., expected waiting time or expected system time), let us introduce the unifying concept of "monotone queueing functions." A family $\mathcal{F}_B$ of monotone cost function is called a *monotone queueing function* if it satisfies $\lim_{\lambda \to b^-} f_b(\lambda) = \infty$. (Observe that the monotonicity implies $f_b(\lambda) = \infty$, for every $\lambda \geq b$). This assumption is motivated by the fact that expected waiting as well as expected system time in every queueing process without rejection goes to infinity when the injection rate approaches the service rate. An immediate consequence of the pole of these functions is that $\alpha = \infty$. Thus, we obtain the following corollary to Theorems 1, 2 and 7.

Corollary 9 *For every family $\mathcal{F}_{\mathbb{R}_{>0}}$ of monotone queueing functions $R^* = 1$, $R = \infty$, and $\bar{R} \geq m$.*

Recall that the instances proving the unbounded coordination ratio R are constructed from pure strategies only. However, the bad instances for the bicriteria ratio $\bar{R}$ that we have seen until now use mixed strategies. This motivates us to investigate whether the randomness introduced due to the choice of mixed strategies is the only source of bad bicriteria ratios. The following theorem demonstrates that bicriteria ratios can also be poor in case of pure strategies only.

Theorem 10 *Let $\mathcal{F}_{\mathbb{R}_{>0}}$ be any family of monotone queueing functions. Suppose there exists $b > \sqrt{m}$ such that $f_b\left(\left(1-\frac{b}{m}\right)\cdot b\right) < f_1\left(\frac{b}{m}\right)$. Then the bicriteria worst-case coordination ratio over pure strategies is at least $\frac{m}{2\,b}$.*

The following lemma directly implies Theorem 10.

Lemma 11 *Let $\mathcal{F}_{\mathbb{R}_{>0}}$ be a family of monotone queueing functions. Assume, we are given m servers with bandwidths $1 = b_1 = b_2 = \ldots = b_m \leq b_{m+1} \leq b_{m+2} = b$, such that $f_{b_j} \in \mathcal{F}_{\mathbb{R}_{>0}}$, for each $j = 1, 2, \ldots, m+2$. Assume moreover, that there exists an $\varepsilon > 0$, such that functions f_1 and f_b fulfill*

$$f_b\left(\left(1-\frac{1}{2\Gamma}\right)b+\varepsilon\right) \leq f_1\left(\frac{1}{2\Gamma}\right), \tag{3}$$

where $\Gamma = \frac{m}{2b}$ and $\frac{b}{2\Gamma} - \varepsilon/2 > 1$. Then, in this system, $C > opt_\Gamma$, where C is the maximum value over all Nash equilibria assuming only pure strategies. Thus, this gives a lower bound on the bicriteria coordination ratio $\overline{R}$ for pure strategies, $\overline{R} \geq \frac{m}{2b}$.

Proof. We define the following instance of the problem. Let the servers $1, 2, \ldots, m$ be called slow. We also have a fast server with $b_{m+2} = b$, and an additional server with bandwidth $b_{m+1} = \frac{b}{2\Gamma}$. Assume, that each slow server holds one small data stream with weight $\frac{1}{2\Gamma}$, and let the fast server have one large data stream with weight $\frac{b}{2\Gamma} - \varepsilon/2$. The additional server does not hold any data stream.

In the solution we have just defined, each slow server has cost (after blowing the streams up by Γ) $f_1\left(\frac{1}{2}\right)$, and the fast server has cost $f_b\left(\frac{b}{2} - \frac{\varepsilon\Gamma}{2}\right)$. Let

$$cost(\varepsilon) = \max\left(f_{1/\Gamma}\left(\frac{1}{2\Gamma}\right), f_{b/\Gamma}\left(\frac{b}{2\Gamma} - \varepsilon/2\right)\right),$$

and observe that by the definition of monotone queueing functions, the value of $cost(\varepsilon)$ is finite even if $\varepsilon \to 0$, and obviously $opt_\Gamma \leq cost(\varepsilon)$.

Now, we will define a Nash equilibrium for our instance. We put as many small data streams on the fast server as possible, to fill this server almost to its bandwidth, that is b. More precisely, we assign to the fast server a total amount of $\left(1 - \frac{1}{2\Gamma}\right) b + \varepsilon$ of small streams. Observe, that we have enough small streams to achieve this, since their total size $= \frac{m}{2\Gamma} > \left(1 - \frac{1}{2\Gamma}\right) \frac{m}{2\Gamma} = \left(1 - \frac{1}{2\Gamma}\right) b$ (note, that if ε is not small enough to achieve this, we can use a smaller value of $\varepsilon > 0$ having still assumption (3) and $\frac{b}{2\Gamma} - \varepsilon/2 > 1$ of the theorem guaranteed by the monotonicity of function f_b). The large stream cannot now be assigned to the fast server (it would exceeded its bandwidth), and by $\frac{b}{2\Gamma} - \varepsilon/2 > 1$ the large stream cannot also be put on any slow server. Therefore, the large data stream must be assigned to the additional server $m+1$.

We argue, that we have just defined a Nash equilibrium. First, none of the small streams would go from the fast server to a slow server, since by (3) we have $f_b\left(\left(1 - \frac{1}{2\Gamma}\right) b + \varepsilon\right) \leq f_1\left(\frac{1}{2\Gamma}\right)$. Also, none of the small streams would go from the fast server to the additional server, since the remaining space on the additional server is $\varepsilon/2$, and it can be made arbitrarily small (in particular smaller than the size of a small stream, $\frac{1}{2\Gamma}$). Finally, the large stream cannot go from the additional server to any slow server (since $\frac{b}{2\Gamma} - \varepsilon/2 > 1$), or to the fast server (since it would exceed the capacity on the fast server). (Notice, that it is possible that a small remaining data stream, if any, can go from a slow server to the fast server. But this is no harm for our construction.)

The cost on the additional server in this Nash equilibrium is $f_{b/(2\Gamma)}\left(\frac{b}{2\Gamma} - \varepsilon/2\right)$, and so by the properties of monotone queueing functions, $\lim_{\varepsilon\to 0} f_{b/(2\Gamma)}\left(\frac{b}{2\Gamma} - \varepsilon/2\right) = \infty$. This implies that there exists an $\varepsilon > 0$, such that C, i.e. the value of the Nash equilibrium, is arbitrarily large, and in particular larger than $cost(0)$. By monotonicity, we have that $cost(0) \geq cost(\varepsilon)$. Finally, we obtain that $opt_\Gamma \leq cost(\varepsilon) \leq cost(0) < C$. $\square$

Theorem 10 needs some explanation. For example, if we assume the cost function to be *M/M/1 waiting time* then the worst-case coordination ratio over pure strategies is $\Omega(n^{1/3})$. On the other hand, for *M/M/1 system time* the coordination ratio is $\Omega(n^{1/2})$. In both cases, the total injected load needed to construct an example that leads to the specified coordination ratios is very small. In Appendix we investigate this closer and show that the $\Omega(n^{1/3})$ bound for M/M/1 waiting time is actually tight. More precisely, we can prove the following result, whose quite technical proof is shifted to Appendix A.

Theorem 12 *Let us consider the integral allocation model and cost function on a server being the M/M/1 waiting time. The bicriteria coordination ratio with pure strategies, $\overline{R}$, in this model has value at most $O(m^{1/3}\log(m))$, where m is the number of servers.*

Thus summarizing, even for pure strategies and under a small total injection rate the slowdown due to the lack of coordination can be significant.

6.2 An alternative bicriteria measure

For the family of monotone queueing functions there is another interesting bicriteria measure. It is a very natural question to ask by how much one has to decrease the bandwidths of the servers such that an optimal assignment under the decreased bandwidths is at least as expensive as a Nash equilibrium for the original system. Let $\overline{R}_{\mathrm{bw}}$ denote the corresponding worst-case bicriteria ratio.

It turns out that for most functions from Queueing Theory, the effect of changing the bandwidths is larger than the effect of changing the injection rate. In fact, most of these functions show *super-linear scaling*, i.e., $f_b(\lambda) \leq \frac{1}{\alpha} \cdot f_{b/\alpha}(\lambda/\alpha)$, for every $\lambda \in [0, b)$ and $\alpha \geq 1$. Applying this property, we can determine the bicriteria bandwidth ratio exactly.

Theorem 13 *For every family $\mathcal{F}_B$ of monotone queueing functions with super-linear scaling, $\overline{R}_{\rm bw} = m$, where m is the number of servers.*

Let OPT_Γ denote the (value of the) optimum solution in a system where bandwidths of all servers are slowed down by a factor of Γ. We first prove the following lemma.

Lemma 14 *Fix an arbitrary monotone queueing function $f = f_b$. Consider a server farm with m identical servers with cost function f. Then for every $\epsilon > 0$ and $\Gamma < m$ there exists a Nash equilibrium over ϵ-small streams such that $C > OPT_\Gamma$.*

Proof. Assume that we have m identical servers, each with bandwidth $b_j = b$ ($j \in [m]$), and a set of n identical data streams, each having weight of $\lambda_i = \frac{b-\delta}{n}$ ($i \in [n]$), where n is such that $\frac{b}{n} \le \epsilon$, and $\delta \in [0, b)$ will be specified later. Define the probabilities p^i_j as follows, $p^i_j = \frac{1}{m}$, for each $i, j \in [m]$. These probabilities define a Nash equilibrium, since all the expected costs c^i_j have the same value.

Let us fix a server j. The probability that all the data streams are assigned to server j is $\Pi_i p^i_j = m^{-n}$. In this case, the cost on server j is $f_b(b-\delta)$, since we have n streams, each of weight $\frac{b-\delta}{n}$, and so their total weight is $\frac{b-\delta}{n} n = b - \delta$.

Therefore, with probability m^{-n}, the cost on a particular server j is at least $f_b(b-\delta)$, and thus also $\max_j C_j$ is at least $f_b(b-\delta)$, for $j \in [m]$. Additionally, these events corresponding to different servers are pairwise disjoint. Using these observations, we can estimate $C = \mathbf{E}\left[\max_j C_j\right]$ as

$$\mathbf{E}\left[\max_j C_j\right] \ge m \cdot f_b(b-\delta) \cdot m^{-n} = \frac{f_b(b-\delta)}{m^{n-1}} \quad .$$

We want to show that if $C \le OPT_\Gamma$, then $\Gamma \ge m$. Assume towards a contradiction that $C \le OPT_\Gamma$ and $\Gamma < m$. Then, we obtain

$$\frac{f_b(b-\delta)}{m^{n-1}} \le OPT_\Gamma \le f_{b/\Gamma}\left(\frac{b-\delta}{n} \cdot \frac{n}{m}\right) = f_{b/\Gamma}\left(\frac{b-\delta}{m}\right) \quad ,$$

where the last inequality follows by observing that the value of OPT_Γ is at most the value of a solution in which we assign $\frac{n}{m}$ (we can assume that $\frac{n}{m}$ is a positive integer) data streams to each server, after slowing the server down by Γ. Then, by the last inequality and by continuity of functions f_b and $f_{b/\Gamma}$, we have

$$\lim_{\delta \to 0} \frac{f_b(b-\delta)}{m^{n-1}} \le \lim_{\delta \to 0} f_{b/\Gamma}\left(\frac{b-\delta}{m}\right) = f_{b/\Gamma}\left(\frac{b}{m}\right) .$$

By our assumption the RHS is finite, but the LHS is infinity by the properties of function f_b. Therefore, there exists a small enough $\delta > 0$, such that $\frac{f_b(b-\delta)}{m^{n-1}} > f_{b/\Gamma}\left(\frac{b}{m}\right) \ge f_{b/\Gamma}\left(\frac{b-\delta}{m}\right)$ (by monotonicity). Contradiction. $\square$

Proof. (of Theorem 13) Lemma 14 implies $\overline{R} \ge m$. Therefore, we have to show that $C \le OPT_m$. Assume, that $b_j, j \in [m]$ are the bandwidths of the servers. Let $\Lambda = \sum_i \lambda_i$, and observe, that we can assume $\Lambda \le \frac{1}{m}\sum_j b_j$. Otherwise, there is a server, say j, in OPT_m solution with load strictly greater than mb_j, so $OPT_m = \infty$ and we are done. By this assumption, we see that $\sum_i \lambda_i \le \max_j b_j$. Therefore, we can assign all the data streams (deterministically) to the fastest server, say 1, and let $B = b_1 = \max_j b_j$.

Now, using linearity of expectation, we have the following

$$C = \mathbf{E}\left[\max_j C_j\right] \le \mathbf{E}\left[\sum_j C_j\right] = \sum_j \mathbf{E}\left[C_j\right] \le m \cdot \max_j \mathbf{E}\left[C_j\right].$$

We now argue, that $\max_j \mathbf{E}\left[C_j\right] \le f_B(\Lambda)$. Assume towards a contradiction, that there is a server, say j_0, such that $\mathbf{E}\left[C_{j_0}\right] > f_B(\Lambda)$. Then, obviously, there is a data stream, say i_0, with $p_{i_0}^{j_0} > 0$. Also, $c_{i_0}^{j_0} = \mathbf{E}\left[C_{j_0} | x_{i_0}^{j_0} = 1\right] \ge \mathbf{E}\left[C_{j_0}\right]$, which follows from the fact that the random variable $\left(C_{j_0} | x_{i_0}^{j_0} = 1\right)$ can not assume smaller values than the random variable C_{j_0}. Similarly, $f_B(\Lambda) \ge c_{i_0}^1 = \mathbf{E}\left[C_1 | x_{i_0}^1 = 1\right]$, since the value $f_B(\Lambda)$ corresponds to the situation where all probabilities $p_i^1 = 1$ for all i, and $p_i^j = 1$ for all i and $j \ne 1$ (recall, 1 is the fastest server). Thus, $p_{i_0}^{j_0} > 0$ and $c_{i_0}^{j_0} > c_{i_0}^1$, which is a contradiction with the Nash equilibrium property. Therefore, we have

$$C \le m \cdot \max_j \mathbf{E}\left[C_j\right] \le m \cdot f_B(\Lambda).$$

By the super-linear scaling we obtain

$$C \le m \cdot f_B(\Lambda) \le f_{B/m}\left(\frac{\Lambda}{m}\right).$$

Let us now consider $OPT_m^{\mathtt{frac}}$, which is the cost of an optimum solution for the fractional flow model in the system, where the bandwidths of all servers are reduced by a factor of m. We claim, that $f_{B/m}\left(\frac{\Lambda}{m}\right) \le OPT_m^{\mathtt{frac}}$. To show this, we first argue that in the optimal fractional solution, the fastest server has largest load. Let us fix two servers, with bandwidths b_1, b_2 such that $b_1 < b_2$, and let x_1, x_2 be the loads on this servers in the optimal fractional solution. It is easy to observe, that $f_{b_1}(x_1) = f_{b_2}(x_2)$, and by the super-linear scaling, we obtain $f_{b_2}\left(\frac{b_2}{b_1}x_1\right) \le \frac{b_1}{b_2} f_{b_1}(x_1) < f_{b_2}(x_2)$, which by the monotonicity property implies that $x_1 \le x_2$.

Now, by a simple averaging argument, there is a server in the optimal fractional solution with load at least $\frac{\Lambda}{m}$. Using the fact we have just shown, we see that the fastest server (with bandwidth B) must have load at least $\frac{\Lambda}{m}$. This shows $f_{B/m}\left(\frac{\Lambda}{m}\right) \le OPT_m^{\mathtt{frac}}$. We can finish the proof by observing the following

$$C \le f_{B/m}\left(\frac{\Lambda}{m}\right) \le OPT_m^{\mathtt{frac}} \le OPT_m,$$

where the last inequality is obvious. □

We want to emphasize that Theorem 13 gives tight results, e.g., for expected waiting time or system time in the queueing systems M/M/1, M/D/1, M/G/1, M/M/c or M/G/1. This follows by the fact that one can show that the respective cost functions of these queueing systems obey the super-linear scaling.

6.3 Queueing systems under heterogeneous traffic

Until now we implicitly assume homogeneous traffic, i.e., all streams have the same (general) session length distribution. However, several practical studies show that Internet traffic is far away from being homogenous [2, 5, 22]. Assuming heterogeneous traffic, one has to take into account different session lengths distributions.

Let us investigate the consequences of heterogeneous traffic using a particular instance of a cost function. The *Pollaczek-Khinchin* (P-K) formula (see, e.g., [10]) describes expected waiting time in M/G/1 queues. Observe that this is a very general class of queueing systems that allows arbitrary service time distributions. We transform the P-K formula into a family of cost functions describing the expected time on servers with bandwidth b using only two parameters that specify the injected mix of streams into server, namely the expectation λ, and variance V of the injected load. Suppose every stream i is characterized by two weights λ_i and V_i (corresponding to the expected load of stream i and the variance of this load, respectively). Then the P-K family of cost functions can be defined as follows:

$$f_b(x_1, \dots, x_n) \;=\; \frac{\sum_{i=1}^{n} V_i\, x_i}{b\,(b - \sum_{i=1}^{n} \lambda_i\, x_i)} \;\; .$$

The remarkable fact here is that both parameters, the expected load and the variance, can be aggregated independently in a simple linear fashion. That is, the expected load injected into the server is $\lambda = \sum_{i=1}^{n} \lambda_i\, x_i$ and the variance of this load is $V = \sum_{i=1}^{n} V_i\, x_i$. For more details on the derivation of this family of cost functions, see Appendix B.

Observe that if we assume $\lambda_i = V_i$ then we are back in the homogeneous model with identical session length distribution and we obtain a monotone queueing function with only one parameter, λ. Consequently $R = \infty$, $R_\epsilon = \infty$, and $\overline{R} \geq m$ for the P-K cost function family. In the fractional flow model, however, we will come to different results. Recall that $R^* = 1$ under homogeneous traffic.

Theorem 15 *The fractional coordination ratio R^* for the P-K cost function family is unbounded. If the ratio between the fastest and slowest server is bounded by S then the worst-case ratio is $R^* = S$.*

We need some technical tools before we present the proof of Theorem 15. Recall, that S denotes the ratio of the largest server bandwidth to the smallest one.

The local exchange lemma. We introduce here some tools. Consider the following problem. We are given m servers, each with two non-negative parameters p_j and q_j ($j \in [m]$). We are also given n data streams, each with two non-negative parameters v_i and e_i ($i \in [n]$). We want to solve (fractionally) the following linear program, which we call LP.

$$\begin{array}{lll} min & \sum_{j=1}^{m} \sum_{i=1}^{n} (v_i p_j + e_i q_j) x_{ij} & \qquad (4) \\ s.t. & \sum_{i=1}^{n} (v_i p_j + e_i q_j) x_{ij} \leq 1 \quad \forall j \in [m] & \qquad (5) \\ & \sum_{j=1}^{m} x_{ij} = 1 \qquad\qquad \forall i \in [n] & \\ & x_{ij} \geq 0 \qquad\qquad\qquad \forall i, j. & \end{array}$$

The value of variable x_{ij} represents a fraction of data stream i assigned to server j. We will show first a very simple combinatorial algorithm to solve this LP. For simplicity, we assume that $e_i > 0$, for all $i \in [n]$. (We note here, that the results can be extended to the case where $e_i = 0$ for some i's.)

Assume, that the servers are sorted such that $\frac{p_1}{q_1} \geq \frac{p_2}{q_2} \geq \ldots \geq \frac{p_m}{q_m}$. We also sort the data streams such that $\frac{v_1}{e_1} \leq \frac{v_2}{e_2} \leq \ldots \leq \frac{v_n}{e_n}$. Let $i = 1, j = 1$. Now, we perform the following greedy assignment: if setting $x_{ij} = 1$ does not violate constraint (5) then set $x_{ij} = 1$ and set $i \leftarrow i + 1$, otherwise set x_{ij} to the greatest value such that constraint (5) is still valid and set $j \leftarrow j + 1$.

Lemma 16 *The described algorithm solves the LP optimally.*

Proof. Assume towards a contradiction that there are two streams, say i, i', such that $\frac{v_i}{e_i} > \frac{v_{i'}}{e_{i'}}$, and a fraction of i is assigned to server j, and a fraction of i' is assigned to server j', with $j < j'$. Assume, that $\frac{p_j}{q_j} > \frac{p_{j'}}{q_{j'}}$. Let the fraction of i which is assigned to server j be $y_i = x_{ij}$. Then, the space that this fraction occupies on server j, or in other words the contribution of this fraction to the objective function of LP, is equal to $c_j = y_i(v_i p_j + e_i q_j)$. Consequently, $y_{i'} = \frac{c_j}{v_{i'} p_j + e_{i'} q_j}$ is the fraction of data stream i' that can be reassigned from server j' to server j to fill the space c_j (note, that if $x_{i',j'} < y_{i'}$, then we can just use a smaller, but still positive, fraction y_i of stream i). Then, $c_{j'} = y_{i'}(v_{i'} p_{j'} + e_{i'} q_{j'})$ is the space left on server j', and $\overline{c_{j'}} = y_i(v_i p_{j'} + e_i q_{j'})$ is the space that the fraction y_i of stream i would occupy if reassigned to server j'. Thus, we reassign fraction y_i of stream i from server j to j', and we reassign fraction $y_{i'}$ of stream i' from server j' to server j. We claim that $c_{j'} > \overline{c_{j'}}$, which would mean that such a reassignment would reduce the value of the objective, and thus would give a contradiction. $c_{j'} > \overline{c_{j'}}$ is equivalent to

$$\frac{v_{i'} p_{j'} + e_{i'} q_{j'}}{v_{i'} p_j + e_{i'} q_j} > \frac{v_i p_{j'} + e_i q_{j'}}{v_i p_j + e_i q_j},$$

which rewrites into

$$v_i v_{i'} p_j p_{j'} + e_i v_{i'} p_{j'} q_j + e_{i'} v_i p_j q_{j'} + e_i e_{i'} q_j q_{j'} > v_i v_{i'} p_j p_{j'} + e_i v_{i'} p_j q_{j'} + e_{i'} v_i p_{j'} q_j + e_i e_{i'} q_j q_{j'},$$

and we further obtain

$$e_i v_{i'} p_{j'} q_j + e_{i'} v_i p_j q_{j'} > e_i v_{i'} p_j q_{j'} + e_{i'} v_i p_{j'} q_j.$$

The last inequality can be written as $(p_j q_{j'} - p_{j'} q_j)(e_{i'} v_i - e_i v_{i'}) > 0$, which by our assumptions that $\frac{v_i}{e_i} > \frac{v_{i'}}{e_{i'}}$ and $\frac{p_j}{q_j} > \frac{p_{j'}}{q_{j'}}$ is obviously true. $\square$

We now show how to use the tools we have just presented. Consider the heterogeneous model under the P-K cost function. Then, the optimum solution in the fractional model can be characterized by the following program.

$$\begin{array}{lll} min & z & \\ s.t. & \frac{\sum_{i=1}^{n} V_i x_i^j}{b_j (b_j - \sum_{i=1}^{n} \lambda_i x_i^j)} \leq z & \forall j \in [m] \\ & \sum_{j=1}^{m} x_{ij} = 1 & \forall i \in [n] \\ & x_{ij} \geq 0 & \forall i, j. \end{array}$$

This program can be equivalently rewritten into the following program, called P_1.

$$\begin{array}{lll} min & z & \\ s.t. & \sum_{i=1}^{n}\left(\frac{V_i}{zb_j^2}+\frac{\lambda_i}{b_j}\right)x_i^j \le 1 & \forall j \in [m] \\ & \sum_{j=1}^{m} x_{ij} = 1 & \forall i \in [n] \\ & x_{ij} \ge 0 & \forall i, j. \end{array}$$

Observe, that if z is fixed, then P_1 is a linear program that is very similar to linear program LP. This similarity can be captured by defining $v_i = V_i$ and $e_i = \lambda_i$, for $i \in [n]$, and $p_j = 1/(zb_j^2)$, $q_j = 1/b_j$, for $j \in [m]$. We can easily observe that, for any fixed value of z, we can test feasibility of the program P_1 using the above greedy assignment algorithm. Finally, it can be shown that we can use binary search for z to solve program P_1. The order of the servers is given by decreasing values of $\frac{p_j}{q_j} = \frac{1}{zb_j}$, which means by increasing values of bandwidths b_j. The order of data streams is given by increasing values of $\frac{v_i}{e_i} = \frac{V_i}{\lambda_i}$. To be able to run the greedy assignment algorithm we need to know the optimum value of z. However, we will show that already the knowledge of the order of servers and streams is useful.

It can easily be noticed that in a fractional Nash equilibrium, the costs on each server must be the same. Using this observation, we can characterize the worst-case Nash fractional equilibrium by the following program.

$$\begin{array}{lll} max & z & \\ s.t. & \frac{\sum_{i=1}^{n} V_i x_i^j}{b_j(b_j - \sum_{i=1}^{n}\lambda_i x_i^j)} = z & \forall j \in [m] \\ & \sum_{j=1}^{m} x_{ij} = 1 & \forall i \in [n] \\ & x_{ij} \ge 0 & \forall i, j. \end{array}$$

A similar line of arguments shows that in the worst-case Nash equilibrium, if we keep the order of the servers by increasing values of bandwidths b_j, then the order of data streams is given by decreasing values of $\frac{v_i}{e_i} = \frac{V_i}{\lambda_i}$. Again, we will use this fact in the proofs below.

Proof. (of Theorem 15) *An upper bound.* Assume, that the bandwidths of the servers are such that $b_1 \ge b_2 \ge \ldots \ge b_m$. We are given a set of data streams with weights and variances λ_i, V_i, $i \in [n]$. We first consider a relaxation of the problem, by replacing each data stream λ_i, V_i, by two new data streams $0, V_i$ and $\lambda_i, 0$. We also delete from the new problem instance all data streams of the form $0, 0$.

The value C of the worst-case Nash equilibrium for the relaxed problem is obviously not smaller than C for the original problem. Also, the value *opt* of the optimum solution for the relaxed problem is not greater than for the original problem. Therefore, an upper bound on the coordination ratio R^* for the relaxed problem implies the same upper bound for the original coordination ratio. From now on we work with the relaxed problem.

Define $\Lambda = \sum_{i\in[n]} \lambda_i$, $V = \sum_{i\in[n]} V_i$, $\tilde{S} = \sum_{j\in[m]} b_j$, and $G = \tilde{S} - \Lambda$. We will show that *opt* $\ge \frac{V}{b_1 G}$ and $C \le \frac{V}{b_m G}$, which implies $R^* \le b_1/b_m = S$.

We use now the order of data streams in the *opt* solution. (Note, that we can extend the local exchange lemma for cases where $\lambda_i = 0$ for some data streams.) This order implies, that there exists a j' such that *opt* maps all streams of the form $\lambda_i, 0$ to a subset of servers $\{j', j'+1, \ldots, m\}$. Assume, that each server $j \geq j'$ is filled with $\lambda(j)$ amount of data streams weights, and therefore server $j \geq j'$ is filled up to a "gap" of

$$g_j = b_j - \lambda(j) \ .$$

Furthermore, the *opt* solution maps all streams of the form $0, V_i$ to the servers $\{1, 2, \ldots, j''\}$, for some j''. Let $V(j)$ denote the "amount of variance" received by server $j \in \{1, \ldots, m\}$.

We first observe, that $j'' = m$. Assume this is false, i.e. there is a server $j_1 \in \{j''+1, \ldots, m\}$ with $V(j_1) = 0$. Then, we can just move a fraction of some type $0, V_i$ stream from a server $j_2 \in \{1, 2, \ldots, j''\}$ to server j_1. The cost on server j_1 was zero before, and after this operation increases marginally. But then, we can also reassign type $0, V_i$ streams more evenly on the servers $\{1, 2, \ldots, j''\}$ using the free space on server j_2. This decreases the cost of the solution, and thus gives a contradiction with the *opt* solution.

The cost on all servers $\{1, 2, \ldots, m\}$ must be identical, that is,

$$\frac{V(1)}{(b_1)^2} = \frac{V(2)}{(b_2)^2} = \cdots = \frac{V(j'-1)}{(b_{j'-1})^2} = \frac{V(j')}{b_{j'} g_{j'}} = \frac{V(j'+1)}{b_{j'+1} g_{j'+1}} = \cdots = \frac{V(m)}{b_m g_m} \ .$$

Consequently,

$$V(m) = \frac{V b_m g_m}{\sum_{j=1}^{j'-1} (b_j)^2 + \sum_{j=j'}^{m} b_j g_j} \ .$$

Therefore, the waiting time at server m and, hence, the value of *opt* is

$$\frac{V(m)}{b_m g_m} = \frac{V}{\sum_{j=1}^{j'-1} (b_j)^2 + \sum_{j=j'}^{m} b_j g_j} \geq \frac{V}{b_1 \left(\sum_{j=1}^{j'-1} b_j + \sum_{j=j'}^{m} g_j \right)} = \frac{V}{b_1 G} \ .$$

Analogously, we can use the reverse order of the streams in the worst-case Nash equilibrium, and we can argue that the variances must be split on all the servers (otherwise, the Nash property would be violated). This, with an appropriate definition of a "gap" g_j and j' shows that the value of Nash is

$$\frac{V}{\sum_{j=1}^{j'} b_j g_j + \sum_{j=j'+1}^{m} (b_j)^2} \leq \frac{V}{b_m \left(\sum_{j=1}^{j'} g_j + \sum_{j=j'+1}^{m} b_j \right)} = \frac{V}{b_m G} \ .$$

This completes the proof of the upper bound.

A lower bound. We show here a lower bound in Theorem 15, that is $R^* \geq S$. Assume, we are given two servers with $b_1 = 1$ and $b_2 = S$, and two data streams with $\lambda_1 = 0, V_1 = 1$ and $\lambda_2 = S, V_2 = 0$. We can use now the local exchange lemma to show that $R^* \geq S$, but we will show an analytical argument here.

Let $\lambda(j)$ be the amount of stream 2 assigned to server j, $j = 1, 2$, and let $V(j)$ be the amount of stream 1 assigned to server j, $j = 1, 2$. We have that $\lambda(1) + \lambda(2) = S$ and $V(1) + V(2) = 1$. We know that in *opt* and in the Nash solutions, the costs are the same on each server. This let us write that

$$\frac{V(1)}{1-\lambda(1)} = \frac{V(2)}{S(S-\lambda(2))} \quad .$$

Now, we have that $\lambda(2) = S - \lambda(1)$ and $V(2) = 1 - V(1)$. After plugging these values in the condition above, we obtain $V(1) = \frac{1-\lambda(1)}{(S-1)\lambda(1)+1}$. Consider now function

$$f(\lambda(1)) = \frac{V(1)}{1-\lambda(1)} = \frac{\frac{1-\lambda(1)}{(S-1)\lambda(1)+1}}{1-\lambda(1)}.$$

It can be shown that f has minimum when $\lambda(1) \to 1$, equal to $1/S$, and so $opt \leq 1/S$ (we use de l'Hospital Rule to prove this fact). We can also easily show that f has maximum when $\lambda(1) = 0$, which gives that $C \geq 1$. Therefore, we obtain finally that $R^* = C/opt \geq S$, which completes the proof of Theorem 15. □

In light of Theorem 15 we conclude that the optimality of fractional flow in Nash equilibrium is a special property of homogeneous traffic on parallel links.

6.4 Servers with parallel channels and rejection

Until now we assumed that all requests are served, regardless of how long they have to wait for service. In practice, however, Web servers can and have to reject requests when they are overloaded. For simplicity, let us assume that a server rejects requests whenever all service channels are occupied and then these requests disappear from the system. In this case, the fraction of rejected requests is completely independent of the service time distribution. In other words, there is no difference between homogeneous and heterogeneous traffic under this service model. The fraction of rejected requests is described by the *Erlang loss formula* (see, e.g., [8, 9]), giving the following family of cost functions.

$$f_b(x_1, \ldots, x_n) \;=\; \frac{\lambda^b/b!}{\sum_{k=0}^{b} \lambda^k/k!}$$

describing the fraction of rejected requests, where $\lambda = \sum_{i=1}^{n} \lambda_i\, x_i$. To simplify notation, let us simply consider f_b as a function of λ. Then it is easy to check that the family of Erlang loss functions is monotone. Thus, Theorem 1 immediately implies the following on the fractional coordination ratio.

Corollary 17 *For the family $\mathcal{F}_{\mathbb{N}_{>0}}$ of the Erlang loss cost functions, $R^* = 1$.*

On the first glance, the family of Erlang loss functions makes an innocent impression. Indeed, these functions are continuous, convex, monotonically increasing in λ and $f_b'(\lambda) \leq 1/b$, for every $\lambda \geq 0$. Nevertheless, the following corollary shows that the integral coordination ratio R (and hence also $\lim_{\epsilon\to 0} R_\epsilon$) for the Erlang loss family is unbounded.

Corollary 18 *For the family $\mathcal{F}_{\mathbb{N}_{>0}}$ of Erlang loss cost functions, $R = \infty$.*

Intuitively, this phenomenon can be explained as follows. Let us consider the family of functions $\mathcal{F}_{\mathbb{N}_{>0}}$ with $F_b(x) = f_b(b\,x)$, i.e., the Erlang loss functions in terms of relative load. We observe, for every $x \geq 0$,

$$\lim_{b\to\infty} F_b(x) = \max\left\{0, \frac{x-1}{x}\right\} \quad .$$

Thus, the limit of these functions behaves in a very extreme way around $x = 1$, where the cost increases by an unbounded factor. This behavior implies $\alpha = \infty$ for the family of Erlang loss cost functions, and so the corollary above is an immediate consequence of Theorem 2.

In contrast to the monotone queueing functions from Section 6.1, the poor bounds for coordination ratios for Erlang loss cost functions are not caused by unbounded cost in Nash equilibrium but they are a consequence of the tiny, close to zero, optimal cost. Our next result gives the absolute cost in Nash equilibrium in selfish routing.

Theorem 19 *Consider a server farm of m servers with bandwidths $b_1 \geq \cdots \geq b_m$ and cost functions from the Erlang loss families. Let $\sum_{i\in[n]} \lambda_i \leq \sum_{j\in[m]} \frac{b_j}{6\,e}$ and let $\max_{i\in[n]} \lambda_i \leq \frac{b_m}{3\,\alpha\,\log m}$, for any $\alpha > 0$ with $\alpha\,\log m \geq 2$. Then, any Nash equilibrium has social cost at most $m^{-\alpha+1} + m \cdot 2^{-b_m/4}$.*

Proof. Suppose $\sum_{i\in[n]} \lambda_i \leq \sum_{j\in[m]} b_j/6\,e$. Suppose the maximum weight over all streams is $W = b_m/(3\,\alpha\,\log m)$, for some given $\alpha > 0$ with $\alpha\,\log m \geq 2$. We have to show that for every Nash equilibrium $C \leq m^{-\alpha+1} + m \cdot 2^{-b_m/4}$.

For $j \in [m]$, let X_j be the random variable describing the injected load on server j. A simple averaging argument shows the existence of a server $q \in [m]$ with

$$\mathbf{E}[X_q] \;=\; \sum_{i\in[n]} p_i^q\,\lambda_i \;\leq\; \frac{b_q}{6\,e} \quad .$$

This server has expected cost $\mathbf{E}[C_q] = \mathbf{E}\left[f_{b_q}(X_q)\right]$. Next, consider an arbitrary server j. Suppose there exists $i \in [n]$ with $p_i^j > 0$. Then the Nash equilibrium property gives $c_i^j \leq c_i^q$ so that

$$\mathbf{E}[C_j] \;\leq\; \mathbf{E}[C_j | x_i^j = 1] \;=\; c_i^j \;\leq\; c_i^q \;=\; \mathbf{E}[C_q | x_i^q = 1] \;\leq\; \mathbf{E}\left[f_{b_q}(X_q + W)\right] \quad .$$

If $p_i^j = 0$, for every $i \in [n]$, then $\mathbf{E}[C_j] = 0 \leq \mathbf{E}\left[f_{b_q}(X_q + W)\right]$ as well. Furthermore, observe that $C_j \leq 1$ by the definition of the cost function. Hence, we can bound the expected cost of any server j by

$$\mathbf{E}[C_j] \;\;\leq\;\; \mathbf{Pr}\left[X_q \leq \frac{b_q}{3}\right] \cdot f_{b_q}\left(\frac{b_q}{3} + W\right) + \mathbf{Pr}\left[X_q > \frac{b_q}{3}\right] \cdot 1 \;\leq\; f_{b_q}\left(\frac{b_q}{3} + W\right) + \mathbf{Pr}\left[X_q > \frac{b_q}{3}\right] \quad .$$

Let us estimate these two terms as follows. Observe that $W \leq b_m/6 \leq b_q/6$. Consequently,

$$f_{b_q}\left(\frac{b_q}{3}+W\right) \;\leq\; f_{b_q}\left(\frac{b_q}{2}\right) \;\leq\; \frac{\left(\frac{b_q}{2}\right)^{b_q}/b_q!}{\sum_{k=0}^{b_q}\left(\frac{b_q}{2}\right)^{k}/k!} \;\leq\; \frac{\left(\frac{b_q}{2}\right)^{\lfloor b_q/2\rfloor}\left\lceil\frac{b_q}{2}\right\rceil!}{b_q!} \;\leq\; 2^{-b_q/4} \;\leq\; 2^{-b_m/4} \;.$$

Furthermore, using $\mathbf{E}[X_q] \leq b_q/(6e)$ and the upper bound for the maximum weight over all streams of W, we can apply the standard Hoeffding bound (see, e.g., [3]) to obtain

$$\mathbf{Pr}\left[X_q > \frac{b_q}{3}\right] \;\leq\; 2^{-b_q/3W} \;\leq\; m^{-\alpha}$$

because $b_q \geq b_m$ and $W = b_m/(3\,\alpha\,\log m)$. Hence, we can conclude

$$C \;=\; \mathbf{E}\left[\max_{j\in[m]} C_j\right] \;\leq\; m\cdot\max_{j\in[m]}\left(\mathbf{E}[C_j]\right) \;\leq\; m^{-\alpha+1} + m\,2^{-b_m/4} \;.$$

This completes the proof of Theorem 19. □

In other words, Theorem 19 says that, if the total injected load is at most a constant fraction of the total bandwidth and every stream has not too large weight, that is, streams are $O(\frac{1}{\log m})$-small, then the fraction of rejected requests is at most $m^{-\alpha+1} + m\,2^{-b_m/4}$, where α denotes an arbitrary constant. Under the same conditions, an optimal allocation would reject a fraction of $2^{-\Theta(b_1)}$ packets. Taking into account that typical Web servers can open several hundred TCP connections simultaneously, so that b_m can be assumed to be quite large, we conclude that the cost of selfish routing is very small in absolute terms even though the coordination ratio comparing these cost with the optimal cost is unbounded.

7 Conclusions

Our results have important algorithmic consequences in that the choice of the queueing discipline should take into account the possible performance degradation due to selfish and uncoordinated behavior of network users.

We have shown that the coordination ratio for closed queueing systems is unbounded. The same is true for server farms that reject requests in case of overload. However, there is a fundamental difference between these two kinds of queueing policies. Because of the infinity pole the delay under selfish routing in closed queueing systems is in general unbounded. In fact, we have explicitely shown that selfish routing in closed queueing systems can lead to an arbitrary large delay even when the total injected load can potentially be served by a single server. In contrast, the fraction of rejected requests under selfish routing can be bounded above by a function that is exponentially small in the number of TCP connections that can be opened simultaneously.

We conclude that server farms that serve all requests, regardless how long requests have to wait, cannot give any reasonable guarantee on the quality of service when selfish agents manage the traffic. However, if requests are allowed to be rejected then it is possible to guarantee a high quality of service for every individual request stream. Thus, the typical practice of rejecting requests in case of overload is a necessary condition to ensure efficient service under game theoretic measures.

References

[1] M. Beckmann, C. B. McGuire, and C. B. Winston. *Studies in the Economimices of Transportation*. Yale University Press, 1956.

[2] M. E. Crovella and A. Bestavros. Self-similarity in world wide web traffic - evidence and possible causes. In *Proceedings of ACM Sigmetrics'96*, pages 160–169, 1996.

[3] A. Czumaj and B. Vöcking. Tight bounds for worst-case equilibria. In *Proceedings of the 13th Annual ACM-SIAM Symposium on Discrete Algorithms*, 2002.

[4] Fred Douglis, Anja Feldmann, Balachander Krishnamurthy, and Jeffrey Mogul. Rate of change and other metrics: a live study of the World Wide Web. In *Proceedings of the Symposium on Internetworking Systems and Technologies*. USENIX, December 1997.

[5] Anja Feldmann, Anna C. Gilbert, Polly Huang, and Walter Willinger. A study of the role of variablity and the impact of control. In *Proc. of the ACM/SIGCOMM'99*, 1999.

[6] M. Frank. The Braess paradox. *Mathematical Programming Study*, 20:283–302, 1981.

[7] P. Franken, D. Konig, U. Arndt, and V. Schmidt. *Queues and Point Processes*. Akademie-Verlag, Berlin; Wiley, Chichester, 1982.

[8] E. Friedman. A generic analysis of selfish routing. Manuscript, 2001.

[9] D. Gross and C. M. Harris. *Queueing Theory. Third Edition* John Wiley & Sons, New York, NY, 1998.

[10] L. Kleinrock. *Queueing Systems. Volume I: Theory.* John Wiley & Sons, New York, NY, 1975.

[11] Y. A. Korilis, A. A. Lazar, and A. Orda. Capacity allocation under noncooperative routing. In *IEEE Transactions on Automatic Control*, 42(3):309–325, 1997.

[12] Y. A. Korilis, A. A. Lazar, and A. Orda. Avoiding the Braess paradox in noncooperative networks. In *Journal of Applied Probability*, 36(1):211–212, 1999.

[13] E. Koutsoupias, M. Mavronicolas, and P. Spirakis. Personal communication.

[14] E. Koutsoupias and C. H. Papadimitriou. Worst-case equilibria. In *Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science*, pages 404–413, 1999.

[15] A. A. Lazar, A. Orda, and D. E. Penderakis. Virtual path bandwidth allocation in multiuser networks. In *IEEE/ACM Transactions on Networking*, 5:861–871, 1997.

[16] M. Mavronicolas and P. Spirakis. The price of selfish routing. In *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing*, pages 510–519, 2001.

[17] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, New York, NY, 1995.

[18] A. Orda, R. Rom, and N. Shimkin. Competitive routing in multi-user communication networks. In *IEEE/ACM Transactions on Networking*, 1:510–521, 1993.

[19] C. H. Papadimitriou. Algorithms, games, and the Internet. In *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing*, pages 749–753, 2001.

[20] C. H. Papadimitriou. Game theory and mathematical economics: A theoretical computer scientist's introduction (Tutorial). In *Proceedings of the 42th IEEE Symposium on Foundations of Computer Science*, pages 4–8, 2001.

[21] C. H. Papadimitriou and M. Yannakakis. On complexity as bounded rationality. In *Proceedings of the 26th Annual ACM Symposium on Theory of Computing*, pages 726–733, 1994.

[22] K. Park, G. Kim, and M.E. Crovella. On the relationship between file sizes, transport protocols, and self-similar network traffic. In *Proc. IEEE International Conference on Network Protocols*, pages 171–180, 1996.

[23] V. Paxson and S. Floyd. Wide area traffic: The failure of Poisson modeling. *IEEE/ACM Transactions on Networking*, 3:226–244, 1995.

[24] T. Roughgarden and É. Tardos. How bad is selfish routing? In *Proceedings of the 41st IEEE Symposium on Foundations of Computer Science*, pages 93–102, 2000.

[25] T. Roughgarden. Stackelberg scheduling strategies. In *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing*, pages 104–113, 2001.

[26] T. Roughgarden. Designing networks for selfish users is hard. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science*, pages 472–481, 2001.

[27] T. Roughgarden. How unfair is optimal routing? In *Proceedings of the 13th Annual ACM-SIAM Symposium on Discrete Algorithms*, 2002.

[28] R. Steinberg and W. I. Zangwill. The prevalence of Braess' paradox. *Transportation Science*, 17(3):301–318, 1983.

[29] J. G. Wardrop. Some theoretical aspects of road traffic research. In *Proceedings of the Institution of Civil Engineers*, Part II, volume 1, pages 325–362, 1952.

A Upper Bounds for Pure Strategies: Proof of Theorem 12

Let us first recall the formulation of Theorem 12. We present below its proof which is quite technical.

Theorem *Let us consider the integral allocation model and cost function on a server being the M/M/1 waiting time. The bicriteria coordination ratio with pure strategies, $\overline{R}$, in this model has value at most $O(m^{1/3}\log(m))$, where m is the number of servers.*

Proof. Recall, that the cost function on a server with bandwidth b (called here the service rate) being the M/M/1 waiting time is given by the formula $f_b(\lambda) = \frac{\lambda}{b(b-\lambda)}$.

Assume that $b_1 \leq \ldots \leq b_m$, and let n be the number of data streams. Observe, that if the maximum cost on the servers is z for a solution given by variables x_i^j, then we have

$$\frac{\sum_{i=1}^n \lambda_i x_i^j}{b_j(b_j - \sum_{i=1}^n \lambda_i x_i^j)} \leq z, \quad \forall j \in [m].$$

Let us rescale the service rates of all servers so that the maximum service rate is $S > 0$. This means, that we replace each original service rate b_j by $(S/b_m)b_j$. By the observation made above, we have for any server j:

$$\frac{\sum_{i=1}^n \lambda_i x_i^j}{b_j(b_j - \sum_{i=1}^n \lambda_i x_i^j)} \leq z, \quad \text{equivalent to} \quad \frac{\sum_{i=1}^n S'\lambda_i x_i^j}{S'b_j(S'b_j - \sum_{i=1}^n S'\lambda_i x_i^j)} \leq z/S',$$

where $S' = S/b_m$. Thus, rescaling results in changing the objective value from z into z/S', and in replacing each λ_i by $S'\lambda_i$. We will work with such a rescaled instance of the problem, and in particular λ_i, b_j will denote the new rescaled values of the data stream weights and service rates. We will use x_{ij} to denote x_i^j in what follows.

Let us fix the value of parameter $\Gamma = \beta\sqrt{S}\log(S+1)$, where a constant $\beta > 0$ will be specified later. Recall, that *opt*$_\Gamma$ means the optimum solution in a system, where we blow the weights λ_i up by Γ. Now, fix any *opt*$_\Gamma$ solution, i.e. the value of this solution is *opt*$_\Gamma$. (If there is no feasible solution for the problem where we blow the λ_i's up, then we are done.) Define α to be a parameter such that *opt*$_\Gamma = \frac{\alpha}{\sqrt{S}}$.

We classify now the servers as follows. All servers with service rate $b_j < 1$ belong to group (set) G_0, and all servers with service rate b_j s.t. $2^{k-1} \leq b_j < 2^k$ define group G_k, for $k \in [K]$. Clearly, $K = O(\log(S))$. All the data streams assigned by the *opt*$_\Gamma$ solution to servers in group G_k define a group of data streams T_k, $k \in [K]$. We can easily see that for each data stream $i \in T_k$, $\lambda_i < \frac{2^k}{\Gamma}$.

Let us now fix any Nash equilibrium, say $x^{\mathtt{N}} \in \{0,1\}^{n\times m}$. Let $k \in \{0\} \cup [K]$, and define $T^j = \{i : x_{ij}^{\mathtt{N}} = 1\}$, i.e. the set of all data streams assigned by the Nash solution to server j, and $T(k) = T_0 \cup T_1 \cup \ldots \cup T_k$. Then $S(k) = \{j : T^j \cap T(k) \neq \emptyset\}$ is the set of all servers having at least one data stream from set $T(k)$. We will denote by $w_j^{\mathtt{N}}$ the cost on server j in the Nash solution, i.e. $w_j^{\mathtt{N}} = f_{b_j}(\sum_{i\in T^j}\lambda_i)$. We also abbreviate $f_{b_j}(\cdot)$ to $f_j(\cdot)$. A direct application of the Lagrange Mean Value Theorem gives the following.

Lemma 20 *Let $b_j > 0$, and $x, y \in (0, b_j)$ with $x < y$. Then there exists a number $\rho \in (x, y)$ such that $f_j(y) = f_j(x) + f_j'(\rho) \cdot (y - x)$.*

Claim 21 *Let $f(x) = \frac{x}{b(b-x)}$, and $x_0 \in (0, b)$ be such that $f(x_0) \le \frac{\beta k}{\Gamma}$, where $k \le \log(S+1)$ and $2^{k-1} \le b$. Then $f'(\rho) \le \frac{\mu}{b}$, for any $\rho \in [x_0, x_0 + \frac{2^k}{\Gamma}]$, where $\mu = (\sqrt{3}+1)^2 \approx 7.464$.*

Proof. By a simple expression manipulation, we can see that $f(x_0) \le \frac{\beta k}{\Gamma}$ implies

$$x_0 \le \frac{\beta k b^2}{\beta k b + \Gamma}.$$

Observe, that $f'(x) = \frac{1}{(b-x)^2}$. Thus, we have to show that for any $\rho \in [x_0, x_0 + \frac{2^k}{\Gamma}]$, we have $\frac{1}{(b-\rho)^2} \le \frac{\mu}{b}$. This is equivalent to showing that $\rho \le b - \sqrt{\frac{b}{\mu}}$.

Since the derivative $f'(\cdot)$ is non-decreasing, it is enough to prove that $x_0 + \frac{2^k}{\Gamma} \le b - \sqrt{\frac{b}{\mu}}$. And since $x_0 \le \frac{\beta k b^2}{\beta k b + \Gamma}$, it suffices to prove that

$$\frac{\beta k b^2}{\beta k b + \Gamma} + \frac{2^k}{\Gamma} \le b - \sqrt{\frac{b}{\mu}}.$$

Multiplying both sides of this inequality by $\Gamma(\beta k b + \Gamma)$, we get

$$\beta k b^2 \Gamma + \beta k b 2^k + 2^k \Gamma \le \beta k b^2 \Gamma + b\Gamma^2 - \beta k b \Gamma \sqrt{\frac{b}{\mu}} - \Gamma^2 \sqrt{\frac{b}{\mu}}.$$

This inequality is equivalent to

$$\beta k b 2^k + 2^k \Gamma + \beta k b \Gamma \sqrt{\frac{b}{\mu}} + \Gamma^2 \sqrt{\frac{b}{\mu}} \le b\Gamma^2,$$

which by applying $\Gamma = \beta\sqrt{S}\log(S+1)$, rewrites into

$$\beta k b 2^k + \beta 2^k \sqrt{S}\log(S+1) + \beta^2 k b \sqrt{S}\sqrt{\frac{b}{\mu}}\log(S+1) + \beta^2 S \sqrt{\frac{b}{\mu}}\log^2(S+1)$$

$$\le \beta^2 b S \log^2(S+1).$$

Now, we factor out β from both sides, and use the facts that $k \le \log(S+1)$ and $2^k \le 2b$ to upper-bound the left-hand-side of the last inequality to obtain the following

$$\log(S+1)b \cdot 2b + 2b\sqrt{S}\log(S+1) + \beta\log(S+1)b\sqrt{S}\sqrt{\frac{b}{\mu}}\log(S+1) +$$

$$+\beta S\sqrt{\frac{b}{\mu}}\log^2(S+1) \leq \beta bS\log^2(S+1).$$

So if we are able to prove the last inequality, then the claim holds as well. Now, we use $\log(S+1) \leq \log^2(S+1)$ and $\sqrt{b} \leq b$ (recall, that $b \geq 1$) to further upper-bound the left-hand-side of the last inequality to obtain

$$\log^2(S+1)b \cdot 2b + 2b\sqrt{S}\log^2(S+1) + \beta\log^2(S+1)b\sqrt{S}\sqrt{\frac{b}{\mu}}+$$

$$+\beta S\frac{b}{\sqrt{\mu}}\log^2(S+1) \leq \beta bS\log^2(S+1).$$

We factor out b and $\log(S+1)$ from both sides, and we have

$$2b + 2\sqrt{S} + \beta\sqrt{S}\sqrt{\frac{b}{\mu}} + \beta S\frac{1}{\sqrt{\mu}} \leq \beta S.$$

We upper-bound the left-hand-side by applying $b \leq S$, which gives

$$2S + 2\sqrt{S} + \frac{2}{\sqrt{\mu}}\beta S \leq \beta S.$$

Now applying $\sqrt{S} \leq S$ to upper-bound the LHS, and factoring S out implies

$$4 + \frac{2}{\sqrt{\mu}}\beta \leq \beta,$$

which is equivalent to $\beta \geq \frac{4\sqrt{\mu}}{\sqrt{\mu}-2} = \frac{4\sqrt{3}+4}{\sqrt{3}-1} \approx 14.9283$. This is true, since we will choose β such that $\beta \geq 15$. □

We distinguish two cases in the proof, depending on the value of α.

The case of $\alpha \geq 2$. We show the following main fact in this case.

Lemma 22 *There is an absolute fixed constant $\beta \geq 15$, such that for any $k \in \{0\} \cup [K]$, and for any server $j \in S(k)$, we have $w_j^{\mathtt{N}} \leq \frac{\beta(k+1)}{\Gamma}$.*

Proof. The proof is by induction on k.

Induction Base. We show the statement for $k = 0$. Assume towards a contradiction that there is a server $j_1 \in S(0)$ with $w_{j_1}^{\mathtt{N}} > \frac{\beta}{\Gamma}$. There must be at least one data stream, say $i_1 \in T_0$, present on server j_1. By the definition of group T_0, $\lambda_{i_1} \leq \frac{b_0}{\Gamma}$, where $b_0 = 1$ is the maximum service rate of any server in group G_0.

One of the ideas for the rest of the proof is to show that all data streams in group T_0 can be "assigned" to the servers in $G_1 \cup \cdots \cup G_K$, without occupying too much space, such that we could assign also the rest of the streams. In this way we can basically delete all the small servers in the group G_0. The details follow.

Let $\beta' = \beta - 9 \geq 1$. We first define a "bin" for each server $j \in G_1 \cup \cdots \cup G_K$, by just saying that such a bin has size (capacity) c_j such that:

$$f_j(c_j) = \frac{c_j}{b_j(b_j - c_j)} = \frac{\beta'}{\Gamma}.$$

Notice, that $\frac{\beta'}{\Gamma} \leq opt_\Gamma$. Then, a simple calculation shows that

$$c_j = \frac{\beta' b_j^2}{\beta' b_j + \Gamma}. \tag{6}$$

Claim 23 *There is a fixed constant $c > 1$, such that $\beta' > 2c$, and:* **(i)** $\sum_{i \in T_1 \cup \cdots \cup T_K} \lambda_i < \frac{1}{2c} \sum_{j \in G_1 \cup \cdots \cup G_K} c_j$, **(ii)** $\sum_{i \in T_0} \lambda_i < \frac{1}{2c} \sum_{j \in G_1 \cup \cdots \cup G_K} c_j$. *The claim holds under the following assumptions: $c = 2, \beta = 15$, $S = 4m^{2/3}$.*

Proof. *Part* **(i)**. Consider the set, say T, of all streams, placed by opt_Γ on server $j \in G_1 \cup \cdots \cup G_K$. Obviously, $\sum_{i \in T} \lambda_i \leq \frac{b_j}{\Gamma}$. We will show that $\frac{b_j}{\Gamma} < \frac{c_j}{2c}$. By (6), the last inequality is equivalent to

$$\Gamma > \frac{2c\beta' b_j}{\beta' b_j - 2c}. \tag{7}$$

Notice, that $\frac{2c\beta' b_j}{\beta' b_j - 2c} = \frac{2c\beta'}{\beta' - (2c)/b_j}$. Therefore, the right-hand-side in (7) is maximized when b_j is minimum, which is $b_j = 1$. Recall, that $\Gamma = \beta\sqrt{S}\log(S+1)$. So the minimum value of Γ is when $S = 1$, and then $\Gamma = \beta$. Thus, (7) holds if the following holds

$$\beta > \frac{2c\beta'}{\beta' - 2c}.$$

Recall, that also $\beta' = \beta - 9$. It is easy to check that if $c = 2$, then the last inequality holds provided that $\beta \geq 15$. Summing over all the streams in $T_1 \cup \cdots \cup T_K$ we obtain the claim.
Part **(ii)**. We know that for each server $j \in G_0$ its speed b_j is at most 1. By the definition of opt_Γ we know that the sum of the sizes of all data streams assigned by opt_Γ solution to server j is at most $\frac{b_j}{\Gamma}$. Thus, $\sum_{i \in T_0} \lambda_i \leq \frac{|G_0|}{\Gamma} \leq \frac{m}{\Gamma}$. Recall, that m is the number of all servers. The inequality claimed in Part **(ii)** would definitely hold if we can show that $\frac{m}{\Gamma} < \frac{1}{2c} c_S$, where c_S is the bin size for the fastest server with speed S. By (6), the last inequality is equivalent to

$$\frac{m}{\Gamma} < \frac{\beta' S^2}{2c(\beta' S + \Gamma)},$$

which, using $\Gamma = \beta\sqrt{S}\log(S+1)$, can be rewritten as

$$2c\beta'\sqrt{S}m + 2c\beta\log(S+1)m < \beta S^2 \log(S+1),$$

and, after plugging $\beta' = \beta - 9$ and $c = 2$ (from the proof of Part **(i)**), as

$$4(\beta - 9)\sqrt{S}m + 4\beta \log(S+1)m < \beta S^2 \log(S+1).$$

It is now easy to observe that S must be asymptotically at least $\Omega(m^{2/3})$. In fact it suffices to take $S = 4m^{2/3}$, and $\beta = 15$. This proves the second part. $\square$

From now on we assume, following Claim 23, that $c = 2, \beta = 15$, and $S = 4m^{2/3}$. By Claim 23, $\sum_{i \in T_0 \cup \cdots \cup T_K} \lambda_i < \frac{1}{c} \sum_{j \in G_1 \cup \cdots \cup G_K} c_j$, with $c = 2$. Therefore, by a simple counting argument, there is a server, say $j_2 \in G_1 \cup \cdots \cup G_K$, in the Nash solution, with load at most $\frac{1}{2} c_{j_2}$. This implies that $w^{\mathtt{N}}_{j_2} < \frac{\beta'}{\Gamma}$. Let us now assign stream i_1 to server j_2. Then, the new waiting time on server j_2 is at most:

$$f_{j_2}\left(\frac{1}{2}c_{j_2} + \lambda_{i_1}\right) \le f_{j_2}\left(\frac{1}{2}c_{j_2} + \frac{1}{\Gamma}\right).$$

Using Lemma 20 with $x = \frac{1}{2}c_{j_2}$ and $y = \frac{1}{2}c_{j_2} + \frac{1}{\Gamma}$, we have that there is a $\rho \in (x, y)$ such that:

$$f_{j_2}\left(\frac{1}{2}c_{j_2} + \lambda_{i_1}\right) = f_{j_2}\left(\frac{1}{2}c_{j_2}\right) + \frac{\frac{1}{\Gamma}}{(b_{j_2} - \frac{1}{2}c_{j_2} - \frac{1}{\Gamma})^2}.$$

By the fact that function $f_{j_2}(\cdot)$ is increasing, we have:

$$f_{j_2}\left(\frac{1}{2}c_{j_2} + \lambda_{i_1}\right) < f_{j_2}(c_{j_2}) + \frac{\frac{1}{\Gamma}}{(b_{j_2} - \frac{1}{2}c_{j_2} - \frac{1}{\Gamma})^2}.$$

By (6), we easily see that $c_j < b_j$, for each server j. This allows us to write:

$$f_{j_2}\left(\frac{1}{2}c_{j_2} + \lambda_{i_1}\right) < f_{j_2}(c_{j_2}) + \frac{\frac{1}{\Gamma}}{(b_{j_2} - \frac{1}{2}b_{j_2} - \frac{1}{\Gamma})^2} = f_{j_2}(c_{j_2}) + \frac{\frac{1}{\Gamma}}{(\frac{1}{2}b_{j_2} - \frac{1}{\Gamma})^2} \le$$

$$f_{j_2}(c_{j_2}) + \frac{\frac{1}{\Gamma}}{(\frac{1}{2} - \frac{1}{\Gamma})^2}.$$

The last inequality follows from the fact that $b_{j_2} \ge 1$, since server was in a group G_p for some $p \ge 1$.

By the definition of c_j, $w^{\mathtt{N}}_{j_2} = f_{j_2}(c_{j_2}) \le \frac{\beta'}{\Gamma}$. This gives:

$$f_{j_2}\left(\frac{1}{2}c_{j_2} + \lambda_{i_1}\right) < \frac{\beta'}{\Gamma} + \frac{\frac{1}{\Gamma}}{(\frac{1}{2} - \frac{1}{\Gamma})^2}. \tag{8}$$

We now show that $\frac{1}{(\frac{1}{2} - \frac{1}{\Gamma})^2} < 9$. This inequality is equivalent to $(\frac{2\Gamma}{\Gamma - 2})^2 < 9$, which holds iff $\frac{2\Gamma}{\Gamma - 2} < 3$, or equivalently $\Gamma > 6$. This is of course true, since $\Gamma = \beta\sqrt{S}\log(S+1) \ge \beta \ge 15$.

Having $\frac{1}{(\frac{1}{2} - \frac{1}{\Gamma})^2} < 9$, and knowing that $\beta' = \beta - 9$, we obtain from (8) that:

$$f_{j_2}\left(\frac{1}{2}c_{j_2} + \lambda_{i_1}\right) < \frac{\beta}{\Gamma}.$$

At the beginning of the induction base proof we have assumed that $w^{\mathtt{N}}_{j_1} > \frac{\beta}{\Gamma}$. We have just proved that moving stream i_1 from server j_1 to j_2 gives waiting time of value strictly less than $\frac{\beta}{\Gamma}$. This contradicts the definition of the Nash equilibrium.
Induction Step. We assume that the claim holds for $k-1$ ($k \geq 1$), and we will show that it also holds for k. Assume towards a contradiction that there is a server $j_1 \in S(k)$ with $w^{\mathtt{N}}_{j_1} > \frac{\beta(k+1)}{\Gamma}$. By the definition of $S(k)$, server j_1 has a data stream, say i_1, s.t. $i_1 \in T(k) = T_0 \cup T_1 \cup \ldots \cup T_k$.

By the definition of sets T_p's and of the opt_Γ solution, the total sum of the weights λ_i's of all the streams from $T' = T_k \cup \cdots \cup T_K$ assigned (in the Nash solution) to servers in $G_k \cup \cdots \cup G_K$ is at most $\frac{1}{\Gamma}\sum_{j'' \in G_k \cup \cdots \cup G_K} b_{j''}$. This, by an averaging argument implies that there is a server, say $j_2 \in G_k \cup \cdots \cup G_K$, with load w.r.t. the streams in T' at most $\frac{1}{\Gamma} b_{j_2}$. Formally, if $T^{j_2} = \{i : x^{\mathtt{N}}_{ij_2} = 1\}$ is the set of all streams assigned to server j_2, then $\sum_{i \in T^{j_2} \cap T'} \lambda_i \leq \frac{1}{\Gamma} b_{j_2}$.

If there is a stream from set $T(k-1)$ assigned to server j_2, then by the induction assumption for $k-1$, $w^{\mathtt{N}}_{j_2} \leq \frac{\beta k}{\Gamma}$, and we are done. If only streams from set T' are present on server j_2, then its load is $\sum_{i \in T^{j_2}} \lambda_i = \sum_{i \in T^{j_2} \cap T'} \lambda_i \leq \frac{1}{\Gamma} b_{j_2}$, and thus

$$w^{\mathtt{N}}_{j_2} \leq f_{j_2}(\frac{1}{\Gamma} b_{j_2}) = \frac{\frac{1}{\Gamma} b_{j_2}}{b_{j_2}(b_{j_2} - \frac{1}{\Gamma} b_{j_2})}.$$

Therefore, we have

$$w^{\mathtt{N}}_{j_2} \leq \max\left(\frac{\beta k}{\Gamma}, \frac{\frac{1}{\Gamma} b_{j_2}}{b_{j_2}(b_{j_2} - \frac{1}{\Gamma} b_{j_2})}\right). \tag{9}$$

We now upper-bound the second expression in the $\max$ formula above:

$$\frac{b_{j_2}/\Gamma}{b_{j_2}(b_{j_2} - b_{j_2}/\Gamma)} = \frac{1}{b_{j_2}(\Gamma - 1)} \leq \frac{1}{2^{k-1}(\Gamma - 1)} \leq \frac{1}{\Gamma - 1} < \frac{\beta}{\Gamma} \leq \frac{\beta k}{\Gamma}.$$

The estimates above are based respectively on the following facts: $b_{j_2} \geq 2^{k-1}$ (since $j_2 \in G_k \cup \cdots \cup G_K$), $k \geq 1$, $\Gamma = \beta\sqrt{S}\log(S+1) \geq \beta = 15 > 2$. Thus, by (9) we have that $w^{\mathtt{N}}_{j_2} \leq \frac{\beta k}{\Gamma}$.

Let us now put stream i_1 on server j_2 (obviously, $j_1 \neq j_2$.). Since $i_1 \in T(k)$, $\lambda_{i_1} < \frac{2^k}{\Gamma}$. The waiting time of server j_2 after putting stream i_1 on it is at most (let x_{j_2} be the load on server j_2 before putting stream i_1):

$$w'_{j_2} = f_{j_2}(x_{j_2} + \lambda_{i_1}) \leq f_{j_2}(x_{j_2} + \frac{2^k}{\Gamma})$$

(note, that $f_{j_2}(\cdot)$ is an increasing function). By Lemma 20, there is $\rho \in (x_{j_2}, x_{j_2} + \frac{2^k}{\Gamma})$ so that

$$w'_{j_2} \leq f_{j_2}(x_{j_2}) + \frac{2^k}{\Gamma} \cdot f'_{j_2}(\rho)$$

Observe, that $f_{j_2}(x_{j_2}) = w^{\mathtt{N}}_{j_2} \leq \frac{\beta k}{\Gamma}$. Now, we invoke Claim 21 with $f = f_{j_2}$, $b = b_{j_2}$, $x_0 = x_{j_2}$, since $f(x_0) \leq \frac{\beta k}{\Gamma}$ and $b_{j_2} \geq 2^{k-1}$ (by the fact that $j_2 \in G_k \cup \cdots \cup G_K$), we have by Claim 21 that $f'_{j_2}(\rho) \leq \frac{\mu}{b_{j_2}}$. Putting all the pieces together implies

$$w'_{j_2} \leq f_{j_2}(x_{j_2}) + \frac{2^k}{\Gamma} \cdot f'_{j_2}(\rho) \leq w^{\mathtt{N}}_{j_2} + \frac{2^k}{\Gamma} \cdot \frac{\mu}{b_{j_2}} \leq \frac{\beta k}{\Gamma} + \frac{2^k}{\Gamma} \cdot \frac{\mu}{b_{j_2}}.$$

Using $b_{j_2} \geq 2^{k-1}$, we obtain from the last inequality

$$w'_{j_2} \leq \frac{\beta k}{\Gamma} + \frac{2^k}{\Gamma} \cdot \frac{\mu}{2^{k-1}} = \frac{\beta k + 2\mu}{\Gamma} < \frac{\beta(k+1)}{\Gamma}.$$

The last inequality follows by observing that $2\mu < \beta$. Thus $w'_{j_2} < \frac{\beta(k+1)}{\Gamma}$ after putting a stream from server j_1 to server j_2. But the waiting time on server j_1 is at least $w^{\mathtt{N}}_{j_1} > \frac{\beta(k+1)}{\Gamma}$, which is a contradiction with the Nash property. This finishes the proof of the induction step. □

By Lemma 22, $w^{\mathtt{N}}_j \leq \frac{\beta(K+1)}{\Gamma}$ for each server $j \in S(K)$. If there is a server, say j, that has no streams assigned in the Nash solution, then of course $w^{\mathtt{N}}_j = 0$. Observe, that $S(K)$ is the set of all servers that have at least one stream assigned. Therefore, we see that $w^{\mathtt{N}}_j \leq \frac{\beta(K+1)}{\Gamma}$ for each server j. This means that $C \leq \frac{\beta(K+1)}{\Gamma}$, and by $\Gamma = \beta\sqrt{S}\log(S+1)$, we have $C \leq \frac{K+1}{\sqrt{S}\log(S+1)}$. Recall, that $2^K \leq S$, and so $K \leq \log(S)$. Therefore, we have $C \leq \frac{\log(S)+1}{\sqrt{S}\log(S+1)} < \frac{2}{\sqrt{S}} \leq \frac{\alpha}{\sqrt{S}} = opt_\Gamma$. Since $S = 4m^{2/3}$, we finally get $C \leq opt_\Gamma$, where $\Gamma = O(m^{1/3}\log(m))$, which proves Theorem 12 when $\alpha \geq 2$.

The case of $\alpha < 2$. The proof in this case has to take into account the fact that α can now be arbitrarily small, and this may cause problems, since we want to upper-bound C by $opt_\Gamma = \frac{\alpha}{\sqrt{S}}$. Similarly to the case of $\alpha \geq 2$, we show the following fact in this case. The proof of this fact has similar main steps to those in the proof of Lemma 22, but the details are quite different.

Lemma 24 *There is an absolute fixed constant $\beta \geq 15$, such that for any $k \in \{0\} \cup [K]$, and for any server $j \in S(k)$, we have $w^{\mathtt{N}}_j \leq \frac{\alpha\beta(k+1)}{\Gamma}$.*

Proof. The proof is by induction on k.
Induction Base. We show first the statement for $k = 0$. Assume towards a contradiction that there is a server $j_1 \in S(0)$ with $w^{\mathtt{N}}_{j_1} > \frac{\alpha\beta}{\Gamma}$. There must be at least one data stream, say $i_1 \in T_0$, present on server j_1. By the definition of group T_0, stream i_1 is placed by opt_Γ on a server from group G_0 with speed b_0, say. Thus $\frac{\Gamma\lambda_{i_1}}{b_0(b_0-\Gamma\lambda_{i_1})} \leq opt_\Gamma = \frac{\alpha}{\sqrt{S}}$, which gives

$$\frac{\Gamma\lambda_{i_1}}{b_0(b_0 - \Gamma\lambda_{i_1})} \leq \frac{\alpha}{\sqrt{S}} \quad \text{is equivalent to} \quad \lambda_{i_1} \leq \frac{\alpha b_0^2}{\alpha b_0 \Gamma + \sqrt{S}\Gamma}. \tag{10}$$

Since $b_0 \leq 1$, we obtain that $\lambda_{i_1} \leq \frac{\alpha}{\alpha\Gamma + \sqrt{S}\Gamma}$

Let $\beta' = \beta - 9 \geq 1$. We define a "bin" for each server $j \in G_1 \cup \cdots \cup G_K$, by saying that such a bin has capacity c_j such that:

$$f_j(c_j) = \frac{c_j}{b_j(b_j - c_j)} = \frac{\alpha\beta'}{\Gamma}.$$

A simple calculation shows that

$$c_j = \frac{\alpha\beta' b_j^2}{\alpha\beta' b_j + \Gamma}. \tag{11}$$

Claim 25 *There is a fixed constant* $c > 1$*, such that* $\beta' > 2c$*, and:* **(i)** $\sum_{i \in T_1 \cup \cdots \cup T_K} \lambda_i < \frac{1}{2c} \sum_{j \in G_1 \cup \cdots \cup G_K} c_j$*,* **(ii)** $\sum_{i \in T_0} \lambda_i < \frac{1}{2c} \sum_{j \in G_1 \cup \cdots \cup G_K} c_j$*. The claim holds under the assumptions:* $c = 2, \beta = 15$*, and* $S = 4m^{2/3}$*.*

Proof. *Part* **(i)**. Consider the set, say T, of all streams, placed by opt_Γ on server $j \in G_1 \cup \cdots \cup G_K$. Let $\Lambda_0 = \sum_{i \in T} \lambda_i$, then $\frac{\Gamma\Lambda_0}{b_j(b_j - \Gamma\Lambda_0)} \le \frac{\alpha}{\sqrt{S}}$. This, by (10), implies $\Lambda_0 \le \frac{\alpha b_j^2}{\alpha b_j \Gamma + \sqrt{S}\Gamma}$.

We will show that $\frac{\alpha b_j^2}{\alpha b_j \Gamma + \sqrt{S}\Gamma} < \frac{c_j}{2c}$. By (11), the last inequality is equivalent to

$$\frac{1}{\alpha b_j \Gamma + \sqrt{S}\Gamma} < \frac{\beta'}{2c(\alpha\beta' b_j + \Gamma)}. \tag{12}$$

This inequality is equivalent to

$$2c\alpha\beta' b_j + 2c\Gamma < \alpha\beta' b_j \Gamma + \beta'\sqrt{S}\Gamma.$$

Since $\beta' = \beta - 9 \ge 6$, $c = 2$ and $S \ge 1$, it is easy to see that $2c\Gamma < \beta'\sqrt{S}\Gamma$. Thus, to prove the last inequality, it suffices to show that $2c\alpha\beta' b_j < \alpha\beta' b_j \Gamma$, which rewrites to $2c < \Gamma$. Since $b \ge 1$, $\Gamma = \beta\sqrt{S}\log(S+1) \ge \beta \ge 15$ and $c = 2$, the inequality $2c < \Gamma$ surely holds. This proves (12). Summing over all the streams in $T_1 \cup \cdots \cup T_K$ we obtain part **(i)** of the claim.

Part **(ii)**. Assume, that Λ_j is the sum the sizes of all streams assigned by opt_Γ solution to server j. By (10), we can estimate Λ_j as $\Lambda_j \le \frac{\alpha b_j^2}{\alpha b_j \Gamma + \sqrt{S}\Gamma}$. Since $j \in G_0$, $b_j \le 1$, and so $\Lambda_j \le \frac{\alpha}{\alpha\Gamma + \sqrt{S}\Gamma}$.

Thus, $\sum_{i \in T_0} \lambda_i \le |G_0| \frac{\alpha}{\alpha\Gamma + \sqrt{S}\Gamma} \le m \frac{\alpha}{\alpha\Gamma + \sqrt{S}\Gamma}$. Recall, that m is the number of all servers. The inequality claimed in Part **(ii)** would definitely hold if $m \frac{\alpha}{\alpha\Gamma + \sqrt{S}\Gamma} < \frac{1}{2c} c_S$, where c_S is the bin size for the fastest server with speed S. By (11), the last inequality is equivalent to

$$\frac{m}{\alpha\Gamma + \sqrt{S}\Gamma} < \frac{\beta' S^2}{2c(\alpha\beta' S + \Gamma)},$$

which, using $\Gamma = \beta\sqrt{S}\log(S+1)$, can be rewritten as

$$2c\alpha\beta' Sm + 2c\beta\sqrt{S}\log(S+1)m < \alpha\beta\beta'\sqrt{S}S^2\log(S+1) + \beta\beta' S^3\log(S+1),$$

We prove this inequality by comparing the LHS and RHS term-wise. Comparing first terms gives

$$2c\alpha\beta' Sm < \alpha\beta\beta'\sqrt{S}S^2\log(S+1),$$

which rewrites into

$$2cm < \beta S^{3/2}\log(S+1).$$

The last inequality holds if $S = 4m^{2/3}$ and $c = 2$, since $\beta \geq 15$. Now, we compare the second terms of the LHS and RHS, which leads to

$$2c\beta\sqrt{S}\log(S+1)m < \beta\beta' S^3 \log(S+1),$$

and it is equivalent to

$$2cm < \beta' S^{5/2}.$$

The last inequality obviously holds when $S = 4m^{2/3}$, $c = 2$, and $\beta' = \beta - 9 \geq 6$. This proves the second part of the claim. $\square$

From now on we assume, following Claim 25, that $c = 2, \beta = 15$, and $S = 4m^{2/3}$. By Claim 25, $\sum_{i\in T_0\cup\cdots\cup T_K} \lambda_i < \frac{1}{c}\sum_{j\in G_1\cup\cdots\cup G_K} c_j$, with $c = 2$. Thus, by a counting argument, there is a server, say $j_2 \in G_1 \cup\cdots\cup G_K$, in the Nash solution, with load at most $\frac{1}{2}c_{j_2}$. This gives $w^{\mathtt{N}}_{j_2} < \frac{\alpha\beta'}{\Gamma}$.

We assign stream i_1 to server j_2. The new waiting time on server j_2 is at most:

$$f_{j_2}\left(\frac{1}{2}c_{j_2} + \lambda_{i_1}\right) \leq f_{j_2}\left(\frac{1}{2}c_{j_2} + \frac{\alpha}{\alpha\Gamma + \sqrt{S}\Gamma}\right).$$

Using Lemma 20 with $x = \frac{1}{2}c_{j_2}$ and $y = \frac{1}{2}c_{j_2} + \frac{\alpha}{\alpha\Gamma+\sqrt{S}\Gamma}$, we have that there is a $\rho \in (x, y)$ such that:

$$f_{j_2}\left(\frac{1}{2}c_{j_2} + \lambda_{i_1}\right) = f_{j_2}\left(\frac{1}{2}c_{j_2}\right) + \frac{\alpha}{\alpha\Gamma + \sqrt{S}\Gamma} \cdot \frac{1}{(b_{j_2} - \frac{1}{2}c_{j_2} - \frac{\alpha}{\alpha\Gamma+\sqrt{S}\Gamma})^2}.$$

Function $f_{j_2}(\cdot)$ is non-decreasing, therefore:

$$f_{j_2}\left(\frac{1}{2}c_{j_2} + \lambda_{i_1}\right) \leq f_{j_2}(c_{j_2}) + \frac{\alpha}{\alpha\Gamma + \sqrt{S}\Gamma} \cdot \frac{1}{(b_{j_2} - \frac{1}{2}c_{j_2} - \frac{\alpha}{\alpha\Gamma+\sqrt{S}\Gamma})^2}.$$

By (11), we see that $c_j < b_j$, for each server j. This allows to write:

$$f_{j_2}\left(\frac{1}{2}c_{j_2} + \lambda_{i_1}\right) < f_{j_2}(c_{j_2}) + \frac{\alpha}{\alpha\Gamma + \sqrt{S}\Gamma} \cdot \frac{1}{(b_{j_2} - \frac{1}{2}b_{j_2} - \frac{\alpha}{\alpha\Gamma+\sqrt{S}\Gamma})^2} =$$

$$= f_{j_2}(c_{j_2}) + \frac{\alpha}{\alpha\Gamma + \sqrt{S}\Gamma} \cdot \frac{1}{(\frac{1}{2}b_{j_2} - \frac{\alpha}{\alpha\Gamma+\sqrt{S}\Gamma})^2} \leq f_{j_2}(c_{j_2}) + \frac{\alpha}{\alpha\Gamma + \sqrt{S}\Gamma} \cdot \frac{1}{(\frac{1}{2} - \frac{\alpha}{\alpha\Gamma+\sqrt{S}\Gamma})^2}.$$

The last inequality follows from $b_{j_2} \geq 1$, since server j_2 was from a group G_p for some $p \geq 1$.

By the definition of c_j, $w^{\mathtt{N}}_{j_2} = f_{j_2}(c_{j_2}) \leq \frac{\alpha\beta'}{\Gamma}$, which gives:

$$f_{j_2}\left(\frac{1}{2}c_{j_2} + \lambda_{i_1}\right) < \frac{\alpha\beta'}{\Gamma} + \frac{\alpha}{\alpha\Gamma + \sqrt{S}\Gamma} \cdot \frac{1}{(\frac{1}{2} - \frac{\alpha}{\alpha\Gamma+\sqrt{S}\Gamma})^2}. \tag{13}$$

We now show that $\frac{\alpha}{\alpha\Gamma+\sqrt{S}\Gamma} \cdot \frac{1}{(\frac{1}{2}-\frac{\alpha}{\alpha\Gamma+\sqrt{S}\Gamma})^2} < \frac{9\alpha}{\Gamma}$. This inequality is equivalent to $\frac{1}{\alpha+\sqrt{S}} \cdot \frac{1}{(\frac{1}{2}-\frac{\alpha}{\alpha\Gamma+\sqrt{S}\Gamma})^2} < 9$, which would hold if $\frac{1}{(\frac{1}{2}-\frac{\alpha}{\alpha\Gamma+\sqrt{S}\Gamma})^2} < 9$, or equivalently $\frac{1}{3} < \frac{1}{2} - \frac{\alpha}{\alpha\Gamma+\sqrt{S}\Gamma}$. The last inequality rewrites into $6\alpha < \alpha\Gamma + \sqrt{S}\Gamma$, which is true, since $\Gamma = \beta\sqrt{S}\log(S+1) \geq \beta \geq 15$.

Applying $\frac{\alpha}{\alpha\Gamma+\sqrt{S}\Gamma} \cdot \frac{1}{(\frac{1}{2}-\frac{\alpha}{\alpha\Gamma+\sqrt{S}\Gamma})^2} < \frac{9\alpha}{\Gamma}$, and $\beta' = \beta - 9$, we obtain from (13):

$$f_{j_2}\left(\frac{1}{2}c_{j_2} + \lambda_{i_1}\right) < \frac{\alpha\beta}{\Gamma}.$$

At the beginning of the induction base proof we have assumed that $w^{\mathtt{N}}_{j_1} > \frac{\beta}{\Gamma}$. We have just showed that moving data stream i_1 from server j_1 to server j_2 gives waiting time on that server of value strictly less than $\frac{\alpha\beta}{\Gamma}$. This contradicts the definition of the Nash equilibrium, and thus proves the induction base.

Induction Step. We assume that the claim holds for $k-1$ ($k \geq 1$), and we will show that it also holds for k. Assume towards a contradiction that there is a server $j_1 \in S(k)$ with $w^{\mathtt{N}}_{j_1} > \frac{\alpha\beta(k+1)}{\Gamma}$. By the definition of $S(k)$, server j_1 has a stream, say i_1, s.t. $i_1 \in T(k) = T_0 \cup T_1 \cup \ldots \cup T_k$.

Let us fix any server $j \in G' = G_k \cup \cdots \cup G_K$ and let $j \in G_p$ for some $p \in \{k, \ldots, K\}$. By the definition of T_p and the opt_Γ solution, the total sum, say x_j, of the sizes of all the streams (they all are from T_p) assigned by opt_Γ to server j fulfills $\frac{\Gamma x_j}{b_j(b_j - \Gamma x_j)} \leq \frac{\alpha}{\sqrt{S}}$. By (10), this implies $x_j \leq \frac{\alpha b_j^2}{\alpha b_j \Gamma + \sqrt{S}\Gamma}$.

This, by a simple averaging argument implies that there is a server, say $j_2 \in G_k \cup \cdots \cup G_K$, with load w.r.t. the streams in $T' = T_k \cup \cdots \cup T_K$ at most $\frac{\alpha b_{j_2}^2}{\alpha b_{j_2}\Gamma + \sqrt{S}\Gamma}$. (Otherwise, the sum over the sizes of all streams in T' would be strictly greater than $\sum_{j \in G'} \frac{\alpha b_j^2}{\alpha b_j \Gamma + \sqrt{S}\Gamma}$, which by the argument above is false.) Formally, if $T^{j_2} = \{i : x^{\mathtt{N}}_{i,j_2} = 1\}$ is the set of all streams assigned to server j_2, then $\sum_{i \in (T^{j_2}) \cap T'} \lambda_i \leq \frac{\alpha b_{j_2}^2}{\alpha b_{j_2}\Gamma + \sqrt{S}\Gamma}$.

If there is a stream from set $T(k-1)$ assigned to server j_2, then by the induction assumption for $k-1$, the waiting time $w^{\mathtt{N}}_{j_2}$ on server j_2 is at most $\frac{\alpha\beta k}{\Gamma}$, and we are done. If only streams from set T' are present on server j_2, then its load is at most $\sum_{i \in T^{j_2}} \lambda_i = \sum_{i \in T^{j_2} \cap T'} \lambda_i \leq \frac{\alpha b_{j_2}^2}{\alpha b_{j_2}\Gamma + \sqrt{S}\Gamma}$, and thus

$$w^{\mathtt{N}}_{j_2} \leq f_{j_2}\left(\frac{\alpha b_{j_2}^2}{\alpha b_{j_2}\Gamma + \sqrt{S}\Gamma}\right) = \frac{\frac{\alpha b_{j_2}^2}{\alpha b_{j_2}\Gamma + \sqrt{S}\Gamma}}{b_{j_2}(b_{j_2} - \frac{\alpha b_{j_2}^2}{\alpha b_{j_2}\Gamma + \sqrt{S}\Gamma})}.$$

Therefore, we have

$$w^{\mathtt{N}}_{j_2} \leq \frac{\alpha b_{j_2}^2}{\alpha b_{j_2}\Gamma + \sqrt{S}\Gamma} \cdot \frac{1}{b_{j_2}(b_{j_2} - \frac{\alpha b_{j_2}^2}{\alpha b_{j_2}\Gamma + \sqrt{S}\Gamma})} = \frac{\alpha}{\alpha b_{j_2}\Gamma + \sqrt{S}\Gamma} \cdot \frac{1}{(1 - \frac{\alpha b_{j_2}}{\alpha b_{j_2}\Gamma + \sqrt{S}\Gamma})} =$$

$$\frac{\alpha}{\alpha b_{j_2}\Gamma + \sqrt{S}\Gamma - \alpha b_{j_2}} = \frac{\alpha}{\alpha b_{j_2}(\Gamma - 1) + \sqrt{S}\Gamma} \leq \frac{\alpha}{\sqrt{S}\Gamma} \leq \frac{\alpha\beta k\sqrt{S}}{\sqrt{S}\Gamma} = \frac{\alpha\beta k}{\Gamma}.$$

The last two inequalities in the estimate above follow by observing that $\alpha b_{j_2}(\Gamma - 1) \geq 0$, $\beta \geq 15$, $k \geq 1$ and $\sqrt{S} \geq 1$. Thus, we finally have that $w^{\mathtt{N}}_{j_2} \leq \frac{\alpha\beta k}{\Gamma}$.

Let us now put stream i_1 on server j_2 (obviously, $j_1 \neq j_2$.). Since $i_1 \in T(k)$, stream i_1 was assigned by opt_Γ to a server, say j, such that $j \in G_0 \cup \cdots \cup G_k$. Similarly to the previous arguments, we have that $\frac{\Gamma \lambda_{i_1}}{b_j(b_j - \Gamma \lambda_{i_1})} \leq \frac{\alpha}{\sqrt{S}}$, and by (10), $\lambda_{i_1} \leq \frac{\alpha b_j^2}{\alpha b_j \Gamma + \sqrt{S}\Gamma}$.

Let x_{j_2} be the load on server j_2 before putting stream i_1 on it, then by the considerations above, we have

$$x_{j_2} \leq \frac{\alpha b_{j_2}^2}{\alpha b_{j_2} \Gamma + \sqrt{S}\Gamma}. \tag{14}$$

The waiting time of server j_2 after putting stream i_1 on it is at most:

$$w'_{j_2} = f_{j_2}(x_{j_2} + \lambda_{i_1}) \leq f_{j_2}\left(x_{j_2} + \frac{\alpha b_j^2}{\alpha b_j \Gamma + \sqrt{S}\Gamma}\right)$$

(note, that $f_{j_2}(\cdot)$ is a non-decreasing function). By Lemma 20, there is $\rho \in (x_{j_2}, x_{j_2} + \frac{\alpha b_j^2}{\alpha b_j \Gamma + \sqrt{S}\Gamma})$ so that

$$w'_{j_2} \leq f_{j_2}(x_{j_2}) + \frac{\alpha b_j^2}{\alpha b_j \Gamma + \sqrt{S}\Gamma} \cdot f'_{j_2}(\rho). \tag{15}$$

Observe, that

$$f_{j_2}(x_{j_2}) = w^{\mathtt{N}}_{j_2} \leq \frac{\alpha \beta k}{\Gamma}. \tag{16}$$

We will now prove that

$$\frac{\alpha b_j^2}{\alpha b_j \Gamma + \sqrt{S}\Gamma} \cdot f'_{j_2}(\rho) \leq \frac{\alpha \beta}{\Gamma}. \tag{17}$$

Since the derivative $f'_{j_2}(\cdot)$ is non-decreasing (recall, that $f_{j_2}(\cdot)$ is a convex function), it suffices to prove the following

$$\frac{\alpha b_j^2}{\alpha b_j \Gamma + \sqrt{S}\Gamma} \cdot f'_{j_2}\left(x_{j_2} + \frac{\alpha b_j^2}{\alpha b_j \Gamma + \sqrt{S}\Gamma}\right) \leq \frac{\alpha \beta}{\Gamma},$$

and using (14), it is enough to prove that

$$\frac{\alpha b_j^2}{\alpha b_j \Gamma + \sqrt{S}\Gamma} \cdot f'_{j_2}\left(\frac{\alpha b_{j_2}^2}{\alpha b_{j_2} \Gamma + \sqrt{S}\Gamma} + \frac{\alpha b_j^2}{\alpha b_j \Gamma + \sqrt{S}\Gamma}\right) \leq \frac{\alpha \beta}{\Gamma}.$$

We now factor-out α, Γ, and divide both sides by β to get

$$\frac{b_j^2}{\beta(\alpha b_j + \sqrt{S})} \cdot f'_{j_2}\left(\frac{\alpha b_{j_2}^2}{\alpha b_{j_2} \Gamma + \sqrt{S}\Gamma} + \frac{\alpha b_j^2}{\alpha b_j \Gamma + \sqrt{S}\Gamma}\right) \leq 1,$$

which can also be rewritten into

$$\frac{b_j^2}{\beta(\alpha b_j + \sqrt{S})} \le \left[f'_{j_2}\left(\frac{\alpha b_{j_2}^2}{\alpha b_{j_2}\Gamma + \sqrt{S}\Gamma} + \frac{\alpha b_j^2}{\alpha b_j \Gamma + \sqrt{S}\Gamma} \right) \right]^{-1} .$$

Since $f_{j_2}(y) = \frac{1}{(b_{j_2}-y)^2}$, the inequality above is equivalent to

$$\sqrt{\frac{b_j^2}{\beta(\alpha b_j + \sqrt{S})}} \le b_{j_2} - \left(\frac{\alpha b_{j_2}^2}{\alpha b_{j_2}\Gamma + \sqrt{S}\Gamma} + \frac{\alpha b_j^2}{\alpha b_j \Gamma + \sqrt{S}\Gamma} \right),$$

or to

$$\frac{b_j}{\sqrt{\beta(\alpha b_j + \sqrt{S})}} + \frac{\alpha b_{j_2}^2}{\alpha b_{j_2}\Gamma + \sqrt{S}\Gamma} + \frac{\alpha b_j^2}{\alpha b_j \Gamma + \sqrt{S}\Gamma} \le b_{j_2}.$$

Since $j \in G_0 \cup \cdots \cup G_k$, $b_j < 2^k$, and also $b_{j_2} \ge 2^{k-1}$, we have $b_j < 2b_{j_2}$. Observe, that if we increase b_j in the first and third term in the LHS of the last inequality, the value of LHS increases. Therefore, it is enough to prove that

$$\frac{2b_{j_2}}{\sqrt{\beta(2\alpha b_{j_2} + \sqrt{S})}} + \frac{\alpha b_{j_2}^2}{\alpha b_{j_2}\Gamma + \sqrt{S}\Gamma} + \frac{4\alpha b_{j_2}^2}{2\alpha b_{j_2}\Gamma + \sqrt{S}\Gamma} \le b_{j_2}.$$

We now factor b_{j_2} out from both sides to obtain

$$\frac{2}{\sqrt{\beta(2\alpha b_{j_2} + \sqrt{S})}} + \frac{\alpha b_{j_2}}{\alpha b_{j_2}\Gamma + \sqrt{S}\Gamma} + \frac{4\alpha b_{j_2}}{2\alpha b_{j_2}\Gamma + \sqrt{S}\Gamma} \le 1.$$

Replacing $2\alpha b_{j_2}$ by 0 in the first term, and replacing α by 2 (recall, that $\alpha < 2$) in the second and third terms again increases the LHS, and we have

$$\frac{2}{\sqrt{\beta\sqrt{S}}} + \frac{2b_{j_2}}{2b_{j_2}\Gamma + \sqrt{S}\Gamma} + \frac{8b_{j_2}}{4b_{j_2}\Gamma + \sqrt{S}\Gamma} \le 1.$$

We increase further the LHS by replacing b_{j_2} with S (recall, that $b_{j_2} \le S$) in the second and the third terms, which gives

$$\frac{2}{\sqrt{\beta\sqrt{S}}} + \frac{2S}{2S\Gamma + \sqrt{S}\Gamma} + \frac{8S}{4S\Gamma + \sqrt{S}\Gamma} \le 1.$$

Finally, since $S \ge 1$, we have $\Gamma = \beta\sqrt{S}\log(S+1) \ge \beta\sqrt{S}$, we replace Γ to get

$$\frac{2}{\sqrt{\beta\sqrt{S}}} + \frac{2S}{\beta\sqrt{S}(2S + \sqrt{S})} + \frac{8S}{\beta\sqrt{S}(4S + \sqrt{S})} \le 1,$$

which is equivalent to

$$\frac{2}{\sqrt{\beta\sqrt{S}}}+\frac{2}{\beta(2\sqrt{S}+1)}+\frac{8}{\beta(4\sqrt{S}+1)}\leq 1,$$

and by using $S \geq 1$, it finally suffices to show that

$$\frac{2}{\sqrt{\beta}}+\frac{2}{3\beta}+\frac{8}{5\beta}\leq 1.$$

This final inequality can easily be checked to hold when $\beta \geq 15$. This proves (17).

Putting (15), (16) and (17) together implies

$$w'_{j_2} \leq \frac{\beta(k+1)}{\Gamma}.$$

Thus $w'_{j_2} \leq \frac{\beta(k+1)}{\Gamma}$ after putting a stream from server j_1 to server j_2. But the waiting time on server j_1 is at least $w^{\mathtt{N}}_{j_1} > \frac{\beta(k+1)}{\Gamma}$, which is a contradiction with the Nash property. This finishes the proof of the induction step. □

Now, by Lemma 24, $w^{\mathtt{N}}_j \leq \frac{\alpha\beta(K+1)}{\Gamma}$ for each server $j \in S(K)$. If there is a server, say j, that has no streams assigned in the Nash solution, then $w^{\mathtt{N}}_j = 0$. Observe, that $S(K)$ is the set of all servers that have at least one stream assigned. Thus, we see that $w^{\mathtt{N}}_j \leq \frac{\alpha\beta(K+1)}{\Gamma}$ for each server j. This means that $C \leq \frac{\alpha\beta(K+1)}{\Gamma}$, and by $\Gamma = \beta\sqrt{S}\log(S+1)$ (in fact we can use a slightly greater Γ, namely $\Gamma = 2\beta\sqrt{S}\log(S+1)$, since this can only increase opt_Γ), we have $C \leq \frac{\alpha(K+1)}{2\sqrt{S}\log(S+1)}$. Recall, that $2^K \leq S$, and so $K \leq \log(S)$. Therefore, we have $C \leq \frac{\alpha(\log(S)+1)}{2\sqrt{S}\log(S+1)} < \frac{2\alpha}{2\sqrt{S}} = \frac{\alpha}{\sqrt{S}} = opt_\Gamma$. Since $S = 4m^{2/3}$, we finally get $C \leq opt_\Gamma$, where $\Gamma = O(m^{1/3}\log(m))$, which proves Theorem 12 when $\alpha < 2$. This completes the proof of Theorem 12. □

B Derivation of the P-K cost function family

Consider a single M/G/1 queue. Let r denote the injection rate, i.e., the expected number of customers per time unit (= expected number of requests per time unit). Let b denote the service rate (= number of processed data units per time unit.) Fix the service time (= session length) distribution. Let S denote a random variable describing the service time for a fixed customer (= requested data units per request = session length). Then the well-known Pollaczek-Khinchin equation states

$$W=\frac{\lambda(\mathbf{E}[S]^2+\mathbf{Var}\,[S])}{2b(b-r\mathbf{E}[S])}\ ,$$

where W is the expected waiting time in this queue (= server) in steady state. In the following, we rewrite this formula using a different notation that will enable us to aggregate streams in a simple, linear fashion.

Let X denote a random variable describing the number of injected units of load in any fixed time interval of length one. We define

$$\begin{aligned}\lambda &= \mathbf{E}[X] &= r\mathbf{E}[S]\\ V &= \mathbf{Var}[X] &= r(\mathbf{E}[S]^2 + \mathbf{Var}[S])\ .\end{aligned}$$

In words, λ denotes the expectation of the injected load (= requested data units) per time unit and V the variance of this load. Using this non-standard notation, we obtain

$$W = \frac{\mathbf{Var}[X]}{2b(b - \mathbf{E}[X])}\ .$$

Next we describe how streams can be aggregated under this cost function. For every stream $i \in [n]$,

- let λ_i denote the expected load injected per time unit, and
- let V_i denote the variance of the same random variable.

Furthermore let x_i denote the fraction of stream i that is directed to the considered server. For example, imagine that every single request is directed to the considered server with probability x_i. As requests are Poisson, the stream obtained by aggregating all streams in this way has expected load $\lambda = \sum_{i\in[n]} \lambda_i$ and variance $V = \sum_{i\in[n]} V_i$. Thus, the expected waiting time on the expected server is

$$W = \frac{\sum_{i\in[n]} \lambda_i}{2b(b - \sum_{i\in[n]} V_i)}\ .$$

For simplicity in notation, we drop the factor $\frac{1}{2}$ and define that the cost of the considered server under the allocation $x_1, \ldots, x_n$ is $2W$.

I N F O R M A T I K

Below you find a list of the most recent technical reports of the Max-Planck-Institut für Informatik. They are available by anonymous ftp from `ftp.mpi-sb.mpg.de` under the directory `pub/papers/reports`. Most of the reports are also accessible via WWW using the URL `http://www.mpi-sb.mpg.de`. If you have any questions concerning ftp or WWW access, please contact `reports@mpi-sb.mpg.de`. Paper copies (which are not necessarily free of charge) can be ordered either by regular mail or by e-mail at the address below.

Max-Planck-Institut für Informatik
Library
attn. Anja Becker
Stuhlsatzenhausweg 85
66123 Saarbrücken
GERMANY
e-mail: `library@mpi-sb.mpg.de`

MPI-I-2003-NWG2-002	F. Eisenbrand	Fast integer programming in fixed dimension
MPI-I-2003-NWG2-001	L.S. Chandran, C.R. Subramanian	Girth and Treewidth
MPI-I-2003-4-002	C. Theobalt, M. Li, M. Magnor, H. Seidel	A Flexible and Versatile Studio for Synchronized Multi-view Video Recording
MPI-I-2003-4-001	M. Tarini, H.P.A. Lensch, M. Goesele, H. Seidel	3D Acquisition of Mirroring Objects
MPI-I-2003-4–003	I. Ivrissimtzis, W. Jeong, H. Seidel	Neural Meshes: Statistical Learning Methods in Surface Reconstruction
MPI-I-2003-2-002	M. Jaeger	A Representation Theorem and Applications to Measure Selection and Noninformative Priors
MPI-I-2003-2-001	P. Maier	Compositional Circular Assume-Guarantee Rules Cannot Be Sound And Complete
MPI-I-2003-1-011	P. Krysta, A. Czumaj, B. Voecking	Selfish Traffic Allocation for Server Farms
MPI-I-2003-1-010	H. Tamaki	A linear time heuristic for the branch-decomposition of planar graphs
MPI-I-2003-1-009	B. Csaba	On the Bollobás – Eldridge conjecture for bipartite graphs
MPI-I-2003-1-008	P. Sanders	Soon to be published
MPI-I-2003-1-007	H. Tamaki	Alternating cycles contribution: a strategy of tour-merging for the traveling salesman problem
MPI-I-2003-1-006	H. Tamaki, M. Dietzfelbinger	On the probability of Rendezvous in Graph
MPI-I-2003-1-005	M. Dietzfelbinger, P. Woelfel	Almost Random Graphs with Simple Hash Functions
MPI-I-2003-1-004	E. Althaus, T. Polzin, S.V. Daneshmand	Improving Linear Programming Approaches for the Steiner Tree Problem
MPI-I-2003-1-003	R. Beier, B. Vcking	Random Knapsack in Expected Polynomial Time
MPI-I-2003-1-002	P. Krysta, P. Sanders, B. Vcking	Scheduling and Traffic Allocation for Tasks with Bounded Splittability
MPI-I-2003-1-001	P. Sanders, R. Dementiev	Asynchronous Parallel Disk Sorting
MPI-I-2002-4-002	F. Drago, W. Martens, K. Myszkowski, H. Seidel	Perceptual Evaluation of Tone Mapping Operators with Regard to Similarity and Preference
MPI-I-2002-4-001	M. Goesele, J. Kautz, J. Lang, H.P.A. Lensch, H. Seidel	Tutorial Notes ACM SM 02 A Framework for the Acquisition, Processing and Interactive Display of High Quality 3D Models
MPI-I-2002-2-008	W. Charatonik, J. Talbot	Atomic Set Constraints with Projection
MPI-I-2002-2-007	W. Charatonik, H. Ganzinger	Symposium on the Effectiveness of Logic in Computer Science in Honour of Moshe Vardi

MPI-I-2002-1-008	P. Sanders, J.L. Trff	The Factor Algorithm for All-to-all Communication on Clusters of SMP Nodes
MPI-I-2002-1-005	M. Hoefer	Performance of heuristic and approximation algorithms for the uncapacitated facility location problem
MPI-I-2002-1-004	S. Hert, T. Polzin, L. Kettner, G. Schfer	Exp Lab A Tool Set for Computational Experiments
MPI-I-2002-1-003	I. Katriel, P. Sanders, J.L. Trff	A Practical Minimum Scanning Tree Algorithm Using the Cycle Property
MPI-I-2002-1-002	F. Grandoni	Incrementally maintaining the number of l-cliques
MPI-I-2002-1-001	T. Polzin, S. Vahdati	Using (sub)graphs of small width for solving the Steiner problem
MPI-I-2001-4-005	H.P.A. Lensch, M. Goesele, H. Seidel	A Framework for the Acquisition, Processing and Interactive Display of High Quality 3D Models
MPI-I-2001-4-004	S.W. Choi, H. Seidel	Linear One-sided Stability of MAT for Weakly Injective Domain
MPI-I-2001-4-003	K. Daubert, W. Heidrich, J. Kautz, J. Dischler, H. Seidel	Efficient Light Transport Using Precomputed Visibility
MPI-I-2001-4-002	H.P.A. Lensch, J. Kautz, M. Goesele, H. Seidel	A Framework for the Acquisition, Processing, Transmission, and Interactive Display of High Quality 3D Models on the Web
MPI-I-2001-4-001	H.P.A. Lensch, J. Kautz, M. Goesele, W. Heidrich, H. Seidel	Image-Based Reconstruction of Spatially Varying Materials
MPI-I-2001-2-006	H. Nivelle, S. Schulz	Proceeding of the Second International Workshop of the Implementation of Logics
MPI-I-2001-2-005	V. Sofronie-Stokkermans	Resolution-based decision procedures for the universal theory of some classes of distributive lattices with operators
MPI-I-2001-2-004	H. de Nivelle	Translation of Resolution Proofs into Higher Order Natural Deduction using Type Theory
MPI-I-2001-2-003	S. Vorobyov	Experiments with Iterative Improvement Algorithms on Completely Unimodel Hypercubes
MPI-I-2001-2-002	P. Maier	A Set-Theoretic Framework for Assume-Guarantee Reasoning
MPI-I-2001-2-001	U. Waldmann	Superposition and Chaining for Totally Ordered Divisible Abelian Groups
MPI-I-2001-1-007	T. Polzin, S. Vahdati	Extending Reduction Techniques for the Steiner Tree Problem: A Combination of Alternative-and Bound-Based Approaches
MPI-I-2001-1-006	T. Polzin, S. Vahdati	Partitioning Techniques for the Steiner Problem
MPI-I-2001-1-005	T. Polzin, S. Vahdati	On Steiner Trees and Minimum Spanning Trees in Hypergraphs
MPI-I-2001-1-004	S. Hert, M. Hoffmann, L. Kettner, S. Pion, M. Seel	An Adaptable and Extensible Geometry Kernel
MPI-I-2001-1-003	M. Seel	Implementation of Planar Nef Polyhedra
MPI-I-2001-1-002	U. Meyer	Directed Single-Source Shortest-Paths in Linear Average-Case Time
MPI-I-2001-1-001	P. Krysta	Approximating Minimum Size 1,2-Connected Networks
MPI-I-2000-4-003	S.W. Choi, H. Seidel	Hyperbolic Hausdorff Distance for Medial Axis Transform
MPI-I-2000-4-002	L.P. Kobbelt, S. Bischoff, K. Khler, R. Schneider, M. Botsch, C. Rssl, J. Vorsatz	Geometric Modeling Based on Polygonal Meshes
MPI-I-2000-4-001	J. Kautz, W. Heidrich, K. Daubert	Bump Map Shadows for OpenGL Rendering
MPI-I-2000-2-001	F. Eisenbrand	Short Vectors of Planar Lattices Via Continued Fractions
MPI-I-2000-1-005	M. Seel, K. Mehlhorn	Infimaximal Frames: A Technique for Making Lines Look Like Segments
MPI-I-2000-1-004	K. Mehlhorn, S. Schirra	Generalized and improved constructive separation bound for real algebraic expressions
MPI-I-2000-1-003	P. Fatourou	Low-Contention Depth-First Scheduling of Parallel Computations with Synchronization Variables
MPI-I-2000-1-002	R. Beier, J. Sibeyn	A Powerful Heuristic for Telephone Gossiping
MPI-I-2000-1-001	E. Althaus, O. Kohlbacher, H. Lenhof, P. Mller	A branch and cut algorithm for the optimal solution of the side-chain placement problem