Stability vs. Optimality in Selfish Ring Routing*

Bo Chen[a] Xujin Chen[b,†] Jie Hu[c] Xiaodong Hu[b]

[a]Warwick Business School, University of Warwick, Coventry CV4 7AL, United Kingdom

[b]Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing 100190, China

[c]State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

Abstract

We study asymmetric atomic selfish routing in ring networks, which has diverse practical applications in network design and analysis. We are concerned with minimizing the maximum latency of source-destination node-pairs over links with linear latencies. We obtain the first constant upper bound on the price of anarchy and significantly improve the existing upper bounds on the price of stability. Moreover, we show that there exists an optimal solution that is a good approximate Nash equilibrium. Finally, we present better performance analysis and fast implementation of pseudo-polynomial algorithms for computing approximate Nash equilibria.

Keywords Selfish routing Price of anarchy Price of stability, Minimum maximum linear latency

1 Introduction

Recent trends in the design and analysis of network routing take into account rational and selfish behaviors of network users. *Selfish routing* [25] models network routing from a game-theoretic perspective, in which network users are viewed as independent *players* participating in a non-cooperative game. Each player, with his own pair of source and destination in the network, aims to establish a communication path (between his source and destination) along which he experiences latency as low as possible, given the link congestion caused by all the players. In the absence of a central authority who can impose and maintain globally efficient routing strategies on the network traffic [20], network designers are often interested in a (pure) Nash equilibrium that is as close to the system optimum as possible, where the *Nash equilibrium* is a "stable state" among the players, from which no player has the incentive to deviate unilaterally. The notion of *price of anarchy* (PoA) (resp. *price of stability* (PoS)) was introduced in [18] (resp. [2]) to capture the gap between the worst (resp. best) possible Nash equilibrium and the globally optimal solution. They respectively quantify the maximum and minimum penalties in network performance required to ensure a stable outcome.

The PoA and PoS of selfish routing depend on, among others, the network topologies, the number of players, the latency functions on network links, as well as the system and individual objectives. In this paper, we are concerned with selfish routing in ring networks with multiple players and linear load-dependent latencies, whose PoA and PoS are evaluated against the social objective of minimizing the maximum latency. We denote such a *selfish ring routing* model as the SRR for short.

Motivations and related works The SRR model under consideration falls within the general framework of *network congestion games*, which are guaranteed to admit at least one Nash equilibrium [23]. In contrast to the symmetric setting of one single strategy set for all the players (see e.g. [10, 13, 15, 18]), the congestion game of the SRR is asymmetric (equivalently, it is a multi-commodity game) and models more realistic and

*Supported in part by the NSF of China under Grant No. 10531070, 10771209, 10721101, and Chinese Academy of Sciences under Grant No. kjcx-yw-s7.

†Corresponding author: xchen@amss.ac.cn.

difficult scenarios where multiple players may have different locations in the network and thus different sets of strategies to choose from (see e.g. [10, 16]). Compared with an essentially complete understanding of the PoA and the PoS with respect to the utilitarian objective – the average latency – for linear congestion games (see e.g. [4, 10, 17]), our current understanding with respect to the egalitarian objective – the maximum latency – for multi-commodity linear congestion games is relatively limited, and this will be enhanced by studying SRR. As splitting the traffic usually causes the problem of packet reassembly at the receiver and thus should be generally avoided (see e.g. [4]), the SRR model is *atomic* and *unsplittable* [4, 7] in the sense that the unit traffic demand from a source to a destination must be satisfied by choosing a single path between the source and the destination.

The vast majority of the work on bounding the PoA and PoS in congestion games has been focused on the criterion of the average latency of all players. Christodoulou and Koutsouplas [10] and Awerbuch et al. [4] independently proved that the PoA of the atomic congestion game (symmetric or asymmetric) with linear latency is 2.5. The PoA grows to 2.618 for weighted demands [4]. The non-atomic congestion games was considered by Roughgarden and Tardos [25] where they showed that for linear latencies the PoA is 4/3. They also extended this result to polynomial latencies. Roughgarden [24] proved that, as far as average latency is concerned with non-atomic players, it is actually the class of allowable latency function and not the specific topology of a network that determines the PoA.

When the system performance is measured by the maximum latency (whose minimization is thus desirable), the PoA of atomic congestion games [23] with linear latency is 2.5 in single-commodity network, but it explodes to $\Theta(\sqrt{m})$ in m-commodity networks [10]. In the model of restricted parallel links [17], where players must be routed on n parallel links under the restriction that the link for each player be chosen from a certain set of allowed links for the player, the PoA lies in $[n-1, n)$. For non-atomic weighted selfish routing with linear latency, recent work by Correa et al. [12] proved the existence of an optimal flow in single-commodity network that is “fair”. Recently, Lin et al. [22] showed that the PoA of selfish routing in single-commodity n-node networks with arbitrary continuous and non-decreasing latency functions is $n-1$. For multi-commodity games, the authors exhibited an infinite family of two-commodity networks, related to the Fibonacci numbers, in which both the PoA grows exponentially with the network size. The construction demonstrates that numerous known selfish routing results for single-commodity networks have no analogues in networks with two or more commodities.

Our motivation of studying selfish routing on the ring topology is threefold. Firstly, the PoS (hence the PoA) of selfish routing with respect to minimizing the maximum latency in general networks can be unbounded even if all latency functions are linear, which can be demonstrated in the following example of an undirected network illustrated in Figure 1. An example of directed network has been provided in [9]. Our example is similar in spirit to the examples of [29] and [12], which show that the PoA with respect to maximum latency objective for nonatomic linear multicommodity games can be unbounded.

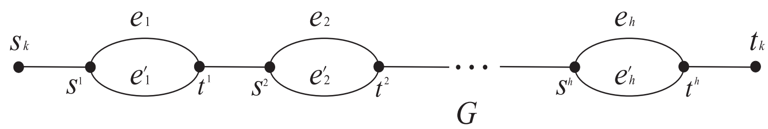


Fig. 1. Unbounded PoS in undirected networks.

Example In the selfish routing on the undirected graph G in Fig. 1, there are in total $k = h^3 + 1$ ($h \geq 2$) players $1, 2, \dots, k$. Each player i ($1 \leq i \leq k$) sends one unit of flow along a path P_i between node s_i and node t_i, where $s_{(j-1)h^2+1} = s_{(j-1)h^2+2} = \cdots = s_{jh^2} = s^j$ and $t_{(j-1)h^2+1} = t_{(j-1)h^2+2} = \cdots = t_{jh^2} = t^j$ for $j = 1, 2, \dots, h$, meaning that players $1, 2, \dots, h^3$ are evenly partitioned into h groups, and all h^2 players in the jth group ($1 \leq j \leq h$) have (s^j, t^j) as their source-destination pair. Let e be a link of G, and x be the number of players who use e in sending their flows. The latency on e is hx if $e = e_j$ for some $1 \leq j \leq h$, and x if $e = e'_j$ for some $1 \leq j \leq h$, and 0 otherwise. Player i experiences a latency equal to the sum of latencies on edges of path P_i, $1 \leq i \leq k$. It is easy to see that the maximum latency among all players is minimized when all players experience an identical latency of h^2 in such a way that all P_i, $i = 1, 2, \dots, k-1$, avoid using $e_1, e_2, \dots, e_h$ except P_k, which uses all of these links. On the other hand, at any Nash equilibrium of the selfish routing, every e_j (resp. e'_j), $1 \le j \le h$, must be contained by at least $h-1$ (resp. h^2-h+1) paths in $P_1, P_2, \ldots, P_k$; otherwise some player i could experience a latency at least h^2-h+3 on e'_j (resp. a latency at least h^2+h on e_j), and would strictly lower his own latency by using e_j in stead of e'_j (resp. e'_j in stead of e_j) in his path. Thus player k always experiences a latency greater than $h(h^2-h)$ in every Nash equilibrium. It follows that the PoS of the selfish routing on G is greater than $h-1$, which turns to infinity as $h \to \infty$. In light of this negative example, practical (undirected) network design has to pay much attention to selecting suitable topologies so that small PoS, as well as small PoA, can be guaranteed.

Secondly, rings have been a fundamental topology frequently encountered in communication networks, and attract considerable attention and efforts from the research community [3, 5, 6, 8, 26, 28], especially in design of approximation algorithms for combinatorial optimization problems. Our study of selfish routing on the ring topology attempts not only to provide a good starting point for evaluating the PoA and PoS in asymmetric network congestion games, but also to enhance the diversity of network topologies amenable to the minimax criterion.

Thirdly, even in a ring, the problem of routing in response to communication requests is not trivial. It has not been known until the present work whether the SRR admits a constant PoA. Upper bounds of 6.83 and 4.57 on the PoS respectively with linear latency and homogenous linear latency have been established in [9]. The authors have also proved the existence of an optimal solution which approximates a Nash equilibrium by a factor of 54. Improving these bounds or showing their tightness is very desirable for better quantifying the PoS and the instability of efficient solutions, which in turn will provide improved guidelines for achieving a good balance between stability and efficiency in the SRR network design.

Main contributions With new ideas and techniques in addition to more elaborate analysis, we contribute to the study of the SRR and of atomic selfish routing in multi-commodity networks [4, 21, 22] by proving four groups of main results: (1) The PoA has a constant upper bound of 16; (2) The PoS is at most 3.9, which reduces to 3.5 for homogenous latency; (3) Any optimal solution is a 9-approximate Nash equilibrium (see Definition 2.2); (4) A polynomial-time combinatorial algorithm and pseudo-polynomial-time convergence combined compute a $(1, 11.7)$-approximate Nash equilibrium (see Definition 2.2). In addition to establishment of new results, our improvements upon previous results in [9] rely on a novel bi-parametric analysis into the impacts of individual players strategies on the ring latency. In summary, our work provides a strong justification on more attractive features of the ring topology compared with general networks [12], apart from simplicity and fault-tolerance of rings in real-world applications.

Paper organization The SRR model is formally defined and some basic properties are presented in Section 2. After establishing a constant bound on the PoA in Section 3, we present evaluation of the PoS with improved bounds in Section 4. Then we prove in Section 5 the existence of $(9, 1)$-approximate Nash equilibria. In Section 6 we provide algorithms for finding good approximate Nash equilibria in pseudo-polynomial time. Finally, we conclude the paper in Section 7 with computational study of the PoS in the SRR of 2 players and 3 players, respectively, which shows that the corresponding PoSs are 1.25 and 1.2565, respectively.

2 The selfish ring routing model

This section introduces the problem formulation, as well as concepts and notation to be used in the paper. The basic properties of Nash equilibria established will play an important role in our theoretical proofs and algorithm design.

2.1 The model

Our selfish ring routing (SRR) model is specified by a triple $(R, l, (s_i, t_i)_{i=1}^k)$, usually called an *SRR instance*. As illustrated in Fig. 2, the underlying network is a ring $R = (V, E)$, an undirected cycle, with node set $V = \{v_1, v_2, \ldots, v_n\}$ of n nodes and link set $E = \{e_i = v_i v_{i+1} : i = 1, 2, \ldots, n\}$ of n links, where $v_{n+1} = v_1$.

By writing $P \subseteq R$, we mean that P is a subgraph of R (possibly R itself) with node set $V(P)$ and link set $E(P)$. Each link $e \in E$ is associated with a load-dependent linear *latency* (*function*) $l_e(x) = a_e x + b_e$, where a_e, b_e are nonnegative constants, and x is an integer variable indicating the load on e.

$$\text{Without loss of generality, all } a_e \text{ and } b_e,\ e \in E, \text{ are assumed to be integers.} \tag{2.1}$$

The latency l is said to be *homogeneous* if $b_e = 0$ for all $e \in E$. There are k source-destination node-pairs (s_i, t_i), $i = 1, 2, \ldots, k$, corresponding to k players $1, 2, \ldots, k$. Each player i $(1 \le i \le k)$ has a communication request for routing one unit of flow from his source $s_i \in V$ to his destination $t_i \in V$, and his strategy set consists of two internally disjoint paths P_i and $\bar{P}_i$ in ring R with ends s_i and t_i satisfying

$$V(P_i) \cap V(\bar{P}_i) = \{s_i, t_i\} \text{ and } P_i \cup \bar{P}_i = R,\ i = 1, 2, \ldots, k. \tag{2.2}$$

We set $\bar{\bar{P}}_i := P_i$ for $i = 1, 2, \ldots, k$. Different players may have the same source-destination pair, and vertices $s_i, t_i, i = 1, 2, \ldots, k$ are not necessarily distinct. On the other hand, $k \ge 2$ and $s_i \ne t_i$, $i = 1, 2, \ldots, k$, are assumed to avoid triviality.

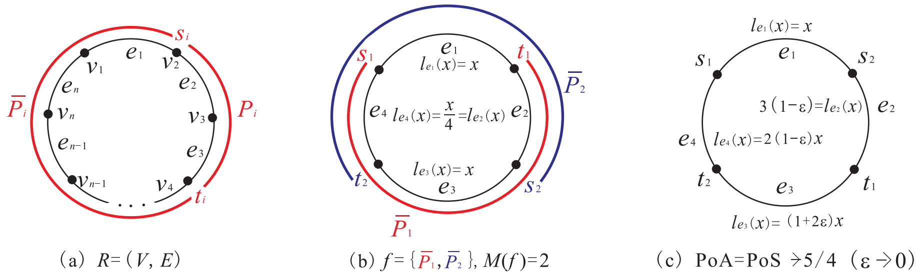


Fig. 2. The SRR instances.

A (*feasible*) *routing* f for the SRR instance is a 0-1 function f on multiset $\mathcal{P} := \cup_{i=1}^{k}\{P_i, \bar{P}_i\}$ such that $f_{P_i} + f_{\bar{P}_i} = 1$ for every $i = 1, 2, \ldots, k$. In view of the correspondence between f and player strategies adopted for the SRR instance, we abuse the notation slightly by writing $f = \{Q_1, Q_2, \ldots, Q_k\}$ with the understanding that, for each $i = 1, 2, \ldots, k$, the one unit of flow requested by player i is routed along path $Q_i \in \{P_i, \bar{P}_i\}$, and correspondingly $f(Q_i) = 1 > 0 = f(\bar{Q}_i)$. Also we write $Q_i \in f$ for $i = 1, 2, \ldots, k$ and call Q_i the f*-route* of player i. Each link $e \in E$ bears a *load* with respect to f defined as the integer

$$f_e := \sum_{P \in \mathcal{P}: e \in E(P)} f(P) = |\{Q_i : e \in E(Q_i), i = 1, 2, \ldots, k\}|$$

equal the number of paths in $\{Q_1, Q_2, \ldots, Q_k\}$ each of which goes through e. Every $P \subseteq R$ is associated with a nonnegative integer $l_P(f) := \sum_{e \in E(P)} l_e(f_e) = \sum_{e \in E(P)} (a_e f_e + b_e)$, which indicates roughly the total latencies of links on P experienced in f. (The wording "indicates roughly" changes to "equals" when every link of P is used by some player in the routing f.) Naturally, the maximum latencies experienced by individuals and the system are

$$M_i(f) := l_{Q_i}(f) \text{ for } i = 1, 2, \ldots, k, \text{ and } M(f) := \max_{i=1}^{k} M_i(f), \tag{2.3}$$

where $M_i(f)$ is the *(maximum) latency of player* i with respect to f (the "maximum" can be dropped in view that the routing is unsplittable; $M_i(f)$ is also referred to as f*-latency of player* i or *of path* Q_i), and $M(f)$ is the *maximum latency of the routing* f. A routing f^* is *optimal* if $M(f^*)$ is minimum among all routings for the SRR instance.

2.2 Approximate Nash equilibria

A Nash equilibrium is characterized by the property that no player has the incentive to change his strategy unilaterally. A routing $f = \{Q_1, Q_2, \ldots, Q_k\}$ is a *Nash equilibrium* or simply a *Nash routing* if

$$l_{Q_i}(f) \le \sum_{e\in E(\bar{Q}_i)} l_e(f_e+1) \text{ for all } i = 1,2,\ldots,k. \tag{2.4}$$

As a network congestion game [14], the SRR possesses at least one Nash routing whose existence can be proved by using *potential function* Φ, defined as follows:

$$\Phi(f) = \sum_{e\in E}\sum_{x=1}^{f_e} l_e(x). \tag{2.5}$$

The domain of the potential function is the set of routings for the SRR instance. For link e, we call $\sum_{x=1}^{f_e} l_e(x)$ the *potential of f on e*. For routing $f = \{Q_1, Q_2, \ldots, Q_k\}$, reversing the summations, the potential of f becomes

$$\Phi(f) = \sum_{i=1}^{k}\sum_{e\in E(Q_i)} l_e\big(|\{Q_h : h \le i, e \in E(Q_h)\}|\big),$$

from which one can easily derive the following well-known result [23, 19].

Lemma 2.1 *Let routing $\tilde{f}$ result from routing f due to a single player i changing his adopted strategy (route). Then the following hold:*

(i) $\Phi(f) - \Phi(\tilde{f}) = M_i(f) - M_i(\tilde{f})$.

(ii) f is a Nash routing if and only if $\Phi(f)$ is a local minimum of Φ. □

Definition 2.2 *Let f^* be an optimal routing and $\alpha, \beta \ge 1$ be two real numbers. A routing $f = \{Q_1, Q_2, \ldots, Q_k\}$ is called an α-*approximate Nash routing *if*

$$l_{Q_i}(f) \le \alpha \sum_{e\in E(\bar{Q}_i)} l_e(f_e+1), \textit{ for all } i = 1,2,\ldots,k.$$

*If additionally $M(f) \le \beta M(f^*)$, then f is called an (α,β)-*approximate Nash routing.

When $\alpha = 1$, routing f is a Nash equilibrium, and thus also referred to as a $(1,\beta)$-*Nash routing*. In the SRR instance $(R, l, (s_i,t_i)_{i=1}^k)$, the *price of stability* (PoS) is defined as the minimum β for which $(1,\beta)$-Nash routing exists; and the *price of anarchy* (PoA) is defined as the minimum β for which every Nash routing is a $(1,\beta)$-Nash routing. The notions of the PoS and PoA extend to the *SRR problem* of all SRR instances, whose *PoS* (resp. *PoA*) is set to be the supremum of PoS (resp. PoA) over all SRR instances.

As an example, for the SRR instance depicted in Fig. 2(c), where $0 < \varepsilon < 1/2$, enumeration of all four feasible routings shows that its unique optimal routing $f^* = \{s_1s_2t_1, s_2t_1t_2\}$ has maximum latency $M(f^*) = 4-\varepsilon$, and is a $(\frac{4-\varepsilon}{4-2\varepsilon}, 1)$-approximate Nash routing, while its unique Nash routing $f = \{s_1s_2t_1, s_2s_1t_2\}$ has maximum latency $M(f) = 5-3\varepsilon$. Hence the PoA and PoS of this instance both tend to 5/4 as ε approaches 0. In addition, the example suggests a small improvement on the lower bound of the PoS $\ge 8/7$ for 2-player SRR stated in Theorem 2 of [9].

Remark 2.3 *The price of stability is at least* $5/4$ *for the SRR problem with* $k = 2$ *players.*

2.3 Basic properties

We investigate Nash routings for an arbitrary SRR instance $I = (R, l, (s_i, t_i)_{i=1}^k)$. For any $P \subseteq R$ and any routing f for I, we often consider

$$l_P(f) := \sum_{e \in E(P)} l_e(f_e) = \sum_{e \in E(P)} (a_e f_e + b_e)$$

as the sum of

$$l_P^a(f) := \sum_{e \in E(P)} a_e f_e \text{ and } l_P^b(f) := \sum_{e \in E(P)} b_e.$$

Define the *a-norm*, *b-norm* and *norm* of P to be

$$||P||_a := \sum_{e \in E(P)} a_e, \quad ||P||_b := \sum_{e \in E(P)} b_e, \text{ and } ||P|| := ||P||_a + ||P||_b, \text{ respectively.}$$

It is worth noting that the equation $l_P^b(f) = ||P||_b$ always holds, though in contrast such an equality does not generally hold between $l_P^a(f)$ and $||P||_a$. So for any routing f we particularly have

$$l_P(f) = l_P^a(f) + l_P^b(f) = l_P^a(f) + ||P||_b. \tag{2.6}$$

When P ($\subseteq R$) is a path, complementary to it is the other path $\bar{P} \subseteq R$ whose link-disjoint union with P forms R. In particular, we will make explicit or implicit use of the following equations in our discussion:

$$||P||_a + ||\bar{P}||_a = ||R||_a, \ ||P||_b + ||\bar{P}||_b = ||R||_b, \text{ and } ||P|| + ||\bar{P}|| = ||R||. \tag{2.7}$$

For any subgraphs P and Q of the ring R, by $P \cup Q$ (resp. $P \cap Q$) we mean the subgraph of R with node set $V(P) \cup V(Q)$ (resp. $V(P) \cap V(Q)$) and link set $E(P) \cup E(Q)$ (resp. $E(P) \cap E(Q)$).

Throughout the paper, we denote by f^N an arbitrary Nash routing for SRR instance $I = (R, l, (s_i, t_i)_{i=1}^k)$, and by $f^\triangledown$ an arbitrary routing for I other than f^N. Permuting the players' names if necessary, we may assume that

$$f^\triangledown = \{Q_1, \ldots, Q_j, Q_{j+1}, \ldots, Q_k\} \text{ and } f^N = \{\bar{Q}_1, \ldots, \bar{Q}_j, Q_{j+1}, \ldots, Q_k\}$$

only differ at the first j locations, and this index j is as small as possible. Note from $f^\triangledown \neq f^N$ that $j \geq 1$. If $\bar{Q}_p = Q_q$ for some p, q with $1 \leq p \neq q \leq j$, then without loss of generality $\{p, q\} = \{j-1, j\}$; it follows that $Q_{j-1} = \bar{Q}_j \in f^N$, $Q_j = \bar{Q}_{j-1} \in f^N$, and we can express f^N as $f^N = \{\bar{Q}_1, \ldots, \bar{Q}_{j-2}, Q_{j-1}, \ldots, Q_k\}$. This contradicts the minimality of j, and gives

$$\{\bar{Q}_1, \ldots, \bar{Q}_j\} \cap \{Q_1, \ldots, Q_j\} = \emptyset. \tag{2.8}$$

In view of the arbitrary choices of $f^\triangledown$ and f^N, to accomplish our task of bounding PoA and PoS of SRR, it suffices to investigate the ratio $M(f^N)/M(f^\triangledown)$. In particular, we wish to obtain upper bounds of $M(f^N)$ as accurate as possible. The following upper bounds have been proved in [9] using equilibrium property of f^N, see (3.6) and (3.7) over there.

$$M(f^N) \leq \frac{l_R(f^N) + ||R||_a}{2} = \frac{l_R^a(f^N) + ||R||_a + ||R||_b}{2} = \frac{l_R^a(f^N) + ||R||}{2}. \tag{2.9}$$

$$M(f^N) \leq l_{Q_i}(f^N) + \frac{||Q_i||_a}{2} + \frac{||R||_a}{2}, \text{ for } i = 1, 2, \ldots, j. \tag{2.10}$$

Since $||R||_a$ and $||R||_b$ are constants, these inequalities suggest us to upper bound

- *ring latency* $l_R(f^N) = \sum_{Q \in f^N} ||Q||$ or $l_R^a(f^N) = \sum_{Q \in f^N} ||Q||_a$, and

- the minimum $\min_{i=1}^{j} l_{Q_i}(f^N) + \frac{1}{2}||Q_i||_a$.

We next define parameter γ to be the highest ratio between a-norms of f^N-route and $f^\triangledown$-route of the same player who follows different routes in f^N and $f^\triangledown$. Without loss of generality suppose the maximum is attached by player 1,

$$\gamma := \max_{i=1}^{j} \frac{||\bar{Q}_i||_a}{||Q_i||_a} = \frac{||\bar{Q}_1||_a}{||Q_1||_a}. \tag{2.11}$$

First, the parameter γ relates ring latency of f^N to that of $f^\triangledown$ by

$$l_R^a(f^N) \le \max\{\gamma, 1\}\, l_R^a(f^\triangledown), \tag{2.12}$$

which together with (2.10) bounds $M(f^N)/M(f^\triangledown)$ from above as follows:

$$\begin{aligned}
\beta := \frac{M(f^N)}{M(f^\triangledown)} \le \frac{l_R^a(f^N) + ||R||}{2M(f^\triangledown)} &\le \frac{\max\{\gamma, 1\} l_R^a(f^\triangledown) + ||R||}{2M(f^\triangledown)} \\
&\le \frac{\max\{\gamma, 1\} l_R(f^\triangledown) - l_R^b(f^\triangledown) + ||R||}{2M(f^\triangledown)} \\
&= \max\{\gamma, 1\} \frac{l_R(f^\triangledown)}{2M(f^\triangledown)} + \frac{||R||_a}{2M(f^\triangledown)}
\end{aligned}$$

In addition to γ, the ratio between $l_R(f^\triangledown)$ and $M(f^\triangledown)$ plays an important role in bounding β. Roughly speaking, to prove smaller $M(f^N)$, we need smaller $l_R(f^\triangledown)$. Let ρ denote a positive integer satisfying

$$l_R(f^\triangledown)/M(f^\triangledown) \le 2\rho. \tag{2.13}$$

This way we obtain the following upper bound on $\beta = M(f^N)/M(f^\triangledown)$ in terms of γ, ρ and $||R||_a/M(f^\triangledown)$.

Lemma 2.4 $\beta = M(f^N)/M(f^\triangledown) \le \rho \max\{\gamma, 1\} + ||R||_a/(2M(f^\triangledown))$. □

Second, in estimating $\min_{i=1}^{j} l_{Q_i}(f^N) + \frac{1}{2}||Q_i||_a$, we intuitively think that shorter $||Q_i||_a$ may more or less imply smaller $l_{Q_i}(f^N)$. So we investigate

$$\min_{i=1}^{j} ||Q_i||_a = ||Q_1||_a = \frac{||R||_a}{\gamma + 1}, \tag{2.14}$$

and $\min_{i=1}^{j} l_{Q_i}(f^N) + \frac{1}{2}||Q_i||_a \le l_{Q_1}(f^N) + \frac{1}{2}||Q_1||_a$, where the second equation of (2.14) follows from $||Q_1||_a + ||\bar{Q}_1||_a = ||R||_a$ in (2.7) and $||\bar{Q}_1||_a = \gamma ||Q_1||_a$ in (2.11). The next lemma is devoted to showing that $l_{Q_1}(f^N)$ is not too large.

Lemma 2.5 *If $\beta > \rho$, then $(\beta\gamma - \beta - 2\rho)\, l_{Q_1}(f^N) \le 2\rho(\beta\gamma - \rho) M(f^\triangledown) + (\beta + \rho)||Q_1||_a + \rho||R||_a - (\beta - \rho)||R||_b$.*

Despite the complexity of the expression, a non-rigorous explanation for the desirable upper bound of $l_{Q_1}(f^N)$ can be derived from the equilibrium properties

$$l_{\bar{Q}_i}(f^N) \le l_{Q_i}(f^N) + ||Q_i||_a \text{ for } i = 1, 2, \ldots, j, \tag{2.15}$$

which rephrase the first j inequalities of (2.4) for the Nash routing $f^N = \{\bar{Q}_1, \ldots, \bar{Q}_j, Q_{j+1}, \ldots, Q_k\}$. Indeed, suppose $l_{Q_1}(f^N)$ is much larger than $l_R(f^\triangledown)$. Then γ is very large by (2.12). Since $l_{Q_1}(f^N)$ is much larger than $l_R(f^\triangledown) \ge l_{Q_1}(f^\triangledown)$, the minimality of $||Q_1||_a$ in (2.14) and the difference between $f^\triangledown$ and f^N impliy that a large portion of $l_{Q_1}(f^N)$ would be experienced by a large number of players in $\{1, \ldots, j\}$ as parts of their f^N-latencies. If the f^N-routes of these players contribute a lot to $l_{\bar{Q}_1}(f^N)$, i.e., intersect $\bar{Q}_1$ with long (in terms of a-norm) paths, then $l_{\bar{Q}_1}(f^N) \le l_{Q_1}(f^N) + ||Q_1||_a$ stated in (2.15) is violated; else the very large γ together with (2.11) says that the a-norm of $\bar{Q}_1$ is much larger than that of Q_1, and most of these players have their $f^{\triangledown}$-routes intersect $\bar{Q}_1$ with very long (in terms of a-norm) paths, therefore experiencing very high $f^{\triangledown}$-latencies which are higher than $M(f^{\triangledown})$ and yield a contradiction.

Proof of Lemma 2.5. The main idea is establishing a lower bound of $l_{\bar{Q}_1}(f^N)$ in terms of $l_{Q_1}(f^N)$ and other parameters and quantities, see (2.18). Since the lower bound is equal to or smaller than the upper bound $l_{Q_1}(f^N) + ||Q_1||_a$ of $l_{\bar{Q}_1}(f^N)$ in (2.15), the inequality presenting this fact is then manipulated algebraically to give the expression in the conclusion of the lemma, see the last two groups of inequalities of the proof.

The derivation of (2.18) relies on the following inequalities, saying that the f^N-latency of player 1 is at least the total contributions to $\bar{Q}_1$ made by the f^N-routes of the first j players.

$$l^a_{\bar{Q}_1}(f^N) \geq \sum_{i=1}^{j} ||\bar{Q}_i \cap \bar{Q}_1||_a \geq \sum_{i=1}^{j} \left(||\bar{Q}_i||_a - ||Q_1||_a\right) = \left(\sum_{i=1}^{j} ||\bar{Q}_i||_a\right) - j \cdot ||Q_1||_a. \tag{2.16}$$

So our *first* step is lower bounding $\sum_{i=1}^{j} ||\bar{Q}_i||_a$. To this end, we notice that this sum equals $l^a_R(f^N) - \sum_{i=j+1}^{k} ||Q_i||_a$, and then consider $l^a_R(f^N) \geq 2M(f^N) - ||R|| = 2\beta M(f^{\triangledown}) - ||R||$ which is implied by (2.9). Thus $l^a_R(f^N) \geq \frac{\beta}{\rho}\, l_R(f^{\triangledown}) - ||R||$ by the definition of ρ in (2.13), and the inequality can be expressed using (2.6) as

$$\sum_{i=1}^{j} ||\bar{Q}_i||_a + \sum_{i=j+1}^{k} ||Q_i||_a \geq \frac{\beta}{\rho} \sum_{i=1}^{j} ||Q_i||_a + \frac{\beta}{\rho} \sum_{i=j+1}^{k} ||Q_i||_a + \frac{\beta}{\rho}||R||_b - ||R||_a - ||R||_b.$$

As our concern is $\sum_{i=1}^{j} ||\bar{Q}_i||_a$, we substitute $||R||_a - ||\bar{Q}_i||_a$ for $||Q_i||_a$, $i = 1, 2, \ldots, j$ (see (2.7)) in the above inequality, and obtain

$$\sum_{i=1}^{j} ||\bar{Q}_i||_a \geq \frac{\beta}{\rho}\left(j \cdot ||R||_a - \sum_{i=1}^{j} ||\bar{Q}_i||_a\right) + \left(\frac{\beta}{\rho} - 1\right) \sum_{i=j+1}^{k} ||Q_i||_a + \left(\frac{\beta}{\rho} - 1\right) ||R||_b - ||R||_a.$$

Rearranging terms yields

$$\left(\frac{\beta}{\rho} + 1\right) \sum_{i=1}^{j} ||\bar{Q}_i||_a \geq \left(\frac{\beta}{\rho} j - 1\right) ||R||_a + \left(\frac{\beta}{\rho} - 1\right) \sum_{i=j+1}^{k} ||Q_i||_a + \left(\frac{\beta}{\rho} - 1\right) ||R||_b.$$

Since $\beta/\rho > 1$ by the hypothesis, we can ignore the nonnegative middle term on the right-hand side of the inequality. Then dividing both sides by positive number $\beta/\rho + 1$, we derive the following lower bound of $\sum_{i=1}^{j} ||\bar{Q}_i||_a$:

$$\sum_{i=1}^{j} ||\bar{Q}_i||_a \geq \frac{\beta j - \rho}{\beta + \rho}||R||_a + \frac{\beta - \rho}{\beta + \rho}||R||_b.$$

This lower bound along with (2.16) gives

$$l^a_{\bar{Q}_1}(f^N) \geq \frac{\beta j - \rho}{\beta + \rho}||R||_a - j \cdot ||Q_1||_a + \frac{\beta - \rho}{\beta + \rho}||R||_b.$$

In turn, using $||R||_a = (\gamma + 1)||Q_1||_a$ in (2.14), we obtain

$$\begin{aligned} l^a_{\bar{Q}_1}(f^N) &\geq j\left(\frac{\beta(\gamma + 1)}{\beta + \rho} - 1\right)||Q_1||_a - \frac{\rho}{\beta + \rho}||R||_a + \frac{\beta - \rho}{\beta + \rho}||R||_b \\ &\geq \frac{\beta\gamma - \rho}{\beta + \rho} \sum_{i=1}^{j} ||\bar{Q}_i \cap Q_1||_a + \frac{(\beta - \rho)||R||_b - \rho||R||_a}{\beta + \rho}. \end{aligned} \tag{2.17}$$

We now proceed to the *second* step: lower bounding $\sum_{i=1}^{j} ||\bar{Q}_i \cap Q_1||_a$, which equals the total contributions of f^N-routes $\bar{Q}_1, \bar{Q}_2, \ldots, \bar{Q}_j$ to the value of $l^a_{Q_1}(f^N)$. Clearly, the sum of the contributions equals $l^a_{Q_1}(f^N) - \sum_{i=j+1}^{k} ||\bar{Q}_1 \cap Q_i||_a$, which is at least

$$l^a_{Q_1}(f^N) - \sum_{i=j+1}^{k} ||Q_i||_a \geq l^a_{Q_1}(f^N) - l^a_R(f^\nabla),$$

and thus at least $l^a_{Q_1}(f^N) - l_R(f^\nabla) + ||R||_b$ by (2.6). It follows from $l_R(f^\nabla) \leq 2\rho M(f^\nabla)$ in (2.13) that

$$\sum_{i=1}^{j} ||\bar{Q}_i \cap Q_1||_a \geq l^a_{Q_1}(f^N) - 2\rho M(f^\nabla) + ||R||_b,$$

which in combination of (2.17) implies the following lower bound of $l^a_{\bar{Q}_1}(f^N)$ in terms of $l^a_{Q_1}(f^N)$ and other parameters and quantities:

$$l^a_{\bar{Q}_1}(f^N) \geq \frac{\beta\gamma - \rho}{\beta + \rho}\Big(l^a_{Q_1}(f^N) - 2\rho M(f^\nabla) + ||R||_b\Big) + \frac{(\beta - \rho)||R||_b - \rho||R||_a}{\beta + \rho}. \tag{2.18}$$

In the *final* step, we make use of equilibrium property (2.15) for player 1 to eliminate term $l^a_{\bar{Q}_1}(f^N)$ in our formulas by algebraic operations, and obtain the inequality for $l_{Q_1}(f^N)$ as desired. Applying (2.15) and (2.6), we have

$$l_{Q_1}(f^N) + ||Q_1||_a \geq l_{\bar{Q}_1}(f^N) = l^a_{\bar{Q}_1}(f^N) + ||\bar{Q}_1||_b.$$

Combining the above inequality with (2.18) and using $||R||_b = ||Q_1||_b + ||\bar{Q}_1||_b \geq ||Q_1||_b$, we deduce that

$$\begin{aligned}
l_{Q_1}(f^N) + ||Q_1||_a &\geq \frac{\beta\gamma - \rho}{\beta + \rho}\Big(l^a_{Q_1}(f) - 2\rho M(f^\nabla) + ||R||_b\Big) + \frac{(\beta - \rho)||R||_b - \rho||R||_a}{\beta + \rho} + ||\bar{Q}_1||_b \\
&\geq \frac{\beta\gamma - \rho}{\beta + \rho}\Big(l^a_{Q_1}(f^N) - 2\rho M(f^\nabla) + ||Q_1||_b\Big) + \frac{\beta\,(||Q_1||_b + ||\bar{Q}_1||_b) - \rho||R||_a - \rho||R||_b}{\beta + \rho} \\
&\geq \frac{\beta\gamma - \rho}{\beta + \rho}\Big(l_{Q_1}(f^N) - 2\rho M(f^\nabla)\Big) + \frac{\beta\,||R||_b - \rho||R||_a - \rho||R||_b}{\beta + \rho} \\
&= \frac{\beta\gamma - \rho}{\beta + \rho}\Big(l_{Q_1}(f^N) - 2\rho M(f^\nabla)\Big) + \frac{(\beta - \rho)||R||_b - \rho||R||_a}{\beta + \rho}
\end{aligned}$$

Thus we obtain

$$(\beta + \rho)\big(l_{Q_1}(f^N) + ||Q_1||_a\big) \geq (\beta\gamma - \rho)\Big(l_{Q_1}(f^N) - 2\rho M(f^\nabla)\Big) + (\beta - \rho)||R||_b - \rho||R||_a,$$

which is equivalent to the inequality in the conclusion. The lemma is then proved. □

Lemmas 2.4 and 2.5 constitute technical preparation for proving upper bounds on $M(f^N)/M(f^\nabla)$ as stated in the next two lemmas.

Lemma 2.6 *If $l_R(f^\nabla) \leq 8M(f^\nabla)$ and $||R||_a \leq 3.5M(f^\nabla)$, then $\beta = M(f^N)/M(f^\nabla) \leq 16$.*

Proof. Recalling (2.13), we can take $\rho = 4$. Assume to the contrary $\beta > 16$. We deduce from Lemma 2.4 that parameter γ is at least 3.56:

$$\gamma = \max\{\gamma, 1\} \geq \frac{\beta}{\rho} - \frac{||R||_a}{2\rho M(f^\nabla)} > \frac{16}{4} - \frac{3.5}{8} = 3.5625, \tag{2.19}$$

saying that player 1 uses a much longer (in terms of a-norm) route in f^N than in f^∇: $||\bar{Q}_1||_a \geq 3.56||Q_1||_a$, and presenting more challenges to assure $l_{\bar{Q}_1}(f^N) \leq l_{Q_1}(f^N) + ||Q_1||_a$ in (2.15). Large γ would imply contradiction to the stability of f^N as explained before the proof of Lemma 2.5. In particular, as $\beta\gamma-\beta-8>0$ by $\beta>16$ and $\gamma\ge 3.56$ in (2.19), Lemma 2.5 asserts that $l_{Q_1}(f^N)$ is bounded above by

$$\frac{8(\beta\gamma-4)M(f^\triangledown)+(\beta+4)||Q_1||_a+4||R||_a-(\beta-4)||R||_b}{\beta\gamma-\beta-8}\le\frac{8(\beta\gamma-4)M(f^\triangledown)+(\beta+4)||Q_1||_a+4||R||_a}{\beta\gamma-\beta-8}.$$

With (2.10), the upper bound of $l_{Q_1}(f^N)$ implies that $M(f^N)$ is not very large:

$$\begin{aligned}M(f^N) &\le l_{Q_1}(f^N)+\frac{||Q_1||_a}{2}+\frac{||R||_a}{2}\\ &\le \frac{8(\beta\gamma-4)}{\beta\gamma-\beta-8}M(f^\triangledown)+\left(\frac{\beta+4}{\beta\gamma-\beta-8}+\frac{1}{2}\right)||Q_1||_a+\left(\frac{4}{\beta\gamma-\beta-8}+\frac{1}{2}\right)||R||_a.\end{aligned}$$

To bound the ratio $M(f^\triangledown)/M(f^N)$, we use the relation $||Q_1||_a=\frac{||R||_a}{\gamma+1}\le 3.5\frac{M(f^\triangledown)}{\gamma+1}$ available from (2.14) and the hypothesis of the lemma. It follows that

$$\begin{aligned}M(f^N) &\le \frac{8(\beta\gamma-4)}{\beta\gamma-\beta-8}M(f^\triangledown)+\frac{\beta(\gamma+1)}{2(\beta\gamma-\beta-8)}\cdot\frac{3.5M(f^\triangledown)}{\gamma+1}+\frac{\beta(\gamma-1)}{2(\beta\gamma-\beta-8)}\cdot 3.5M(f^\triangledown)\\ &= \frac{19.5\beta\gamma-64}{2(\beta\gamma-\beta-8)}M(f^\triangledown).\end{aligned}$$

As $\gamma>0$ by (2.19), the derivative of $\frac{19.5\beta\gamma-64}{2(\beta\gamma-\beta-8)}$ with respect to β is negative for all $\beta>0$. So, using $\beta>16$, we obtain

$$16<\beta=\frac{M(f^N)}{M(f^\triangledown)}\le\frac{19.5\beta\gamma-64}{2(\beta\gamma-\beta-8)}\le\frac{19.5(16\gamma)-64}{2(16\gamma-16-8)}=\frac{312\gamma-64}{32\gamma-48}.$$

Now $\frac{312\gamma-64}{32\gamma-48}>16$ implies $\gamma<3.52$, a contradiction to (2.19), proving the lemma. □

Along a similar line to that of Lemma 2.6, we have the following result whose proof can be found in Appendix.

Lemma 2.7 *If $l_R(f^\triangledown)<2M(f^\triangledown)$ and $||R||<2M(f^\triangledown)$, then $\beta=M(f^N)/M(f^\triangledown)$ is at most 3.9, and at most 3.5 when $||R||_b=0$.* □

3 A constant upper bound on the price of anarchy

The first constant upper bound 16 on the PoA of all SRR instances is established in this section.

Theorem 3.1 *The price of anarchy of the SRR problem is at most 16.*

To prove the theorem, we compare an arbitrary Nash routing f^N with an optimal routing f^* on ring R, and show that $l_R(f^N)\le 16M(f^*)$ whenever $f^N\ne f^*$ and the conditions in Lemma 2.6 are not satisfied, which implies $M(f^N)/M(f^*)\le 16$ as $M(f^N)\le l_R(f^N)$. Our proof is based on contradiction assumption $l_R(f^N)>16M(f^*)$, under which we get $l_Q(f^N)>7.5M(f^*)$ for all paths Q with $Q\in f^*$ and $\bar{Q}\in f^N$. Furthermore, we can find at most three f^N-routes whose union contains all paths $\bar{Q}$ with $\bar{Q}\in f^N$ and $Q\in f^*$ (see Cases 1 – 3 in the proof of Theorem 3.1). These paths either help us to upper bound $l_R(f^*)$ and $||R||_a$ by $8M(f^*)$ and $3.5M(f^*)$, respectively, or have their intersection T satisfy $l_T(f^N)\le 4M(f^*)<6M(f^*)<l_T(f^*)$. In the former case, Lemma 2.6 gives $M(f^N)/M(f^*)\le 16$; in the latter case, it can be deduced that some path Q with $Q\in f^*$ and $\bar{Q}\in f^N$ is contained in T, yielding $l_Q(f^N)\le l_T(f^N)\le 4M(f^*)$, a contradiction. The detailed argument goes as follows.

Proof of Theorem 3.1. Consider an arbitrary Nash routing f^N for an SRR instance $I=(R,l,(s_i,t_i)_{i=1}^k)$. Clearly I admits an optimal routing f^* that is *irredundant* in the sense that any two f^*-routes P,Q whose union $P\cup Q=R$ covers R are link-disjoint.

Set $\beta := M(f^N)/M(f^*)$. It suffices to show $\beta \le 16$. To this end, we may assume $f^* = f^\triangledown = \{Q_1, \ldots, Q_j, Q_{j+1}, \ldots, Q_k\} \ne f^N = \{\bar{Q}_1, \ldots, \bar{Q}_j, Q_{j+1}, \ldots, Q_k\}$ as described in Section 2.3, as otherwise $\beta = 1$ and we are done.

If some $\bar{Q}_g$ and $\bar{Q}_h$ with $1 \le g < h \le j$ are link-disjoint, then $Q_g \cup Q_h = R$, and since $f^\triangledown$ is irredundant, it must be the case that $\bar{Q}_g = Q_h$ and $\bar{Q}_h = Q_g$, a contradiction to (2.8). Hence

$$E(\bar{Q}_g) \cap E(\bar{Q}_h) \ne \emptyset \text{ for all } 1 \le g < h \le j. \tag{3.1}$$

Since R is the link-disjoint union of Q_i and $\bar{Q}_i$ for every $i = 1, 2, \ldots, k$, we may assume

$$l_{Q_i}(f^N) + l_{\bar{Q}_i}(f^N) = l_R(f^N) > 16M(f^\triangledown) \text{ for all } i = 1, 2, \ldots, k, \tag{3.2}$$

as otherwise (2.3) implies $M(f^N) \le l_R(f^N) \le 16M(f^\triangledown)$ giving $\beta \le 16$.

By definition, $||Q_i|| \le l_{Q_i}(f^\triangledown) \le M(f^\triangledown)$ for all $i = 1, 2, \ldots, k$. For the Nash routing f^N, we deduce from (2.15) and (3.2) that

$$l_{Q_i}(f^N) \ge l_{\bar{Q}_i}(f^N) - M(f^\triangledown) \text{ and } l_{Q_i}(f^N) \ge \frac{l_{\bar{Q}_i}(f^N) + l_{Q_i}(f^N) - M(f^\triangledown)}{2} > 7.5M(f^\triangledown) \text{ for } 1 \le i \le j. \tag{3.3}$$

If some Q_g with $1 \le g \le j$ is link-disjoint from $\cup_{i=1}^j \bar{Q}_i$, then $l_{Q_g}(f^N) \le l_{Q_g}(f^\triangledown) \le M(f^\triangledown)$ indicates a contradiction to (3.3). So we have

$$E(Q_g) \cap \left(\cup_{i=1}^j E(\bar{Q}_i)\right) \ne \emptyset \text{ for all } 1 \le g \le j; \text{ in particular } j \ge 2. \tag{3.4}$$

It is not difficult to see from (3.1) and (3.4) that one of the following three cases (illustrated in Fig. 3) must be true:

- Case 1: There exist p, q, and r with $1 \le p < q < r \le j$ such that $\bar{Q}_p \cup \bar{Q}_q \subsetneq R$, $\bar{Q}_q \cup \bar{Q}_r \subsetneq R$, $\bar{Q}_r \cup \bar{Q}_p \subsetneq R$, and $\bar{Q}_p \cup \bar{Q}_q \cup \bar{Q}_r = R$.
- Case 2: There exist p and q with $1 \le p < q \le j$ such that $\bar{Q}_p \cup \bar{Q}_q = R$.
- Case 3: There exist p and q with $1 \le p < q \le j$ such that $\cup_{i=1}^j \bar{Q}_i \subseteq \bar{Q}_p \cup \bar{Q}_q \subsetneq R$.

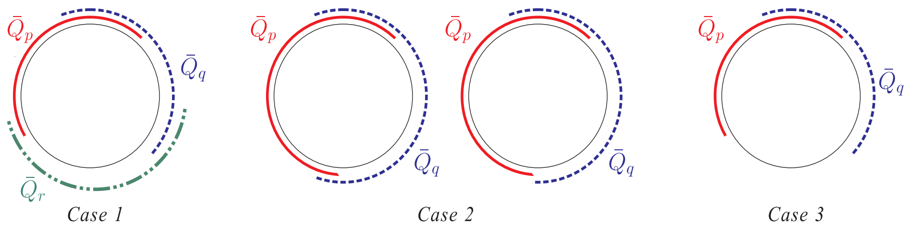


Fig. 3. Possible configurations of f^N when $l_R(f^N) > 16M(f^\triangledown)$.

To see that Cases 1 − 3 cover all possible scenarios, as $j \ge 2$ by (3.4), we may take distinct $p, q \in \{1, 2, \ldots, j\}$ such that $\bar{Q}_p \cup \bar{Q}_q$ contains as many links as possible. Recall from (3.1) that $\bar{Q}_p$ and $\bar{Q}_q$ have some link in common. When $j = 2$, clearly either Case 2 or Case 3 is true. When $j > 2$, let $r \in \{1, 2, \ldots, j\} - \{p, q\}$ be arbitrary. If $\bar{Q}_p \cup \bar{Q}_q \cup \bar{Q}_r = R$, then either Case 1 or Case 2 holds. Else $\bar{Q}_p \cup \bar{Q}_q \cup \bar{Q}_r \subsetneq R$. The maximality of $|E(\bar{Q}_p \cup \bar{Q}_q)|$ enforces that $\bar{Q}_p \cup \bar{Q}_q$ is not properly contained in $\bar{Q}_r \cup \bar{Q}_p$ nor $\bar{Q}_r \cup \bar{Q}_q$. Hence $E(\bar{Q}_r) \cap E(\bar{Q}_p) \ne \emptyset \ne E(\bar{Q}_r) \cap E(\bar{Q}_q)$ in (3.1) implies $\bar{Q}_r \subseteq \bar{Q}_p \cup \bar{Q}_q$ showing Case 3. Next, we proceed to case analysis.

Case 1. It is easy to see that $Q_p \cup Q_q \cup Q_r = R$, which implies

$$||R||_a \le ||R|| \le l_R(f^\triangledown) \le l_{Q_p}(f^\triangledown) + l_{Q_q}(f^\triangledown) + l_{Q_r}(f^\triangledown) \le 3M(f^\triangledown).$$

Hence Lemma 2.6 guarantees $\beta \le 16$ as desired.

Case 2. Notice that $\bar{Q}_p \supseteq Q_q$ and $\bar{Q}_q$ is the link-disjoint union of $\bar{Q}_p \cap \bar{Q}_q$ and Q_p. It follows from (3.3) that $l_{Q_p}(f^N) \geq l_{\bar{Q}_p}(f^N) - M(f^\triangledown) \geq l_{Q_q}(f^N) - M(f^\triangledown) \geq l_{\bar{Q}_q}(f^N) - 2M(f^\triangledown)$, yielding

$$l_{\bar{Q}_p\cap\bar{Q}_q}(f^N) = l_{\bar{Q}_q}(f^N) - l_{Q_p}(f^N) \leq 2M(f^\triangledown). \tag{3.5}$$

Observe that R is the link-disjoint union of Q_p, Q_q and $\bar{Q}_p \cap \bar{Q}_q$, implying

$$||R||_a \leq ||Q_p||_a + ||Q_q||_a + \tfrac{1}{2} l_{\bar{Q}_p\cap\bar{Q}_q}(f^N) \leq 3M(f^\triangledown).$$

When $l_R(f^\triangledown) \leq 8M(f^\triangledown)$, Lemma 2.6 gives $\beta \leq 16$ as desired. When $l_R(f^\triangledown) > 8M(f^\triangledown)$, we have

$$l_{\bar{Q}_p\cap\bar{Q}_q}(f^\triangledown) = l_R(f^\triangledown) - l_{Q_p}(f^\triangledown) - l_{Q_q}(f^\triangledown) > 8M(f^\triangledown) - 2M(f^\triangledown) = 6M(f^\triangledown).$$

Let $S := \{Q \in f^\triangledown : Q \subseteq \bar{Q}_p \cap \bar{Q}_q\}$ denote the set of $f^\triangledown$-routes all contained in $\bar{Q}_p \cap \bar{Q}_q$. Then $\cup_{Q\in S} Q \subseteq \bar{Q}_p \cap \bar{Q}_q$. Note that $E(\cup_{Q\in f^\triangledown} Q) \cap E(\bar{Q}_p \cap \bar{Q}_q) - E(\cup_{Q\in S} Q)$ can be covered by at most four $f^\triangledown$-routes outside S, implying

$$l_{\cup_{Q\in S} Q}(f^\triangledown) + 4M(f^\triangledown) \geq l_{\bar{Q}_p\cap\bar{Q}_q}(f^\triangledown) > 6M(f^\triangledown),$$

so we have

$$l_{\cup_{Q\in S} Q}(f^\triangledown) > 2M(f^\triangledown),$$

which, together with $l_{\bar{Q}_p\cap\bar{Q}_q}(f^N) \leq 2M(f^\triangledown)$ in (3.5), enforces $Q \notin f^N$ for some member $Q \in S$. So $Q = Q_i$ belongs to S for some i with $1 \leq i \leq j$. However, it follows from (3.5) that $l_{Q_i}(f^N) \leq l_{\bar{Q}_p\cap\bar{Q}_q}(f^N) \leq 2M(f^\triangledown)$ contradicting (3.3).

Case 3. In this case $\cup_{i=1}^{j} \bar{Q}_i \subseteq \bar{Q}_p \cup \bar{Q}_q$ implies $l_{Q_p\cap Q_q}(f^N) \leq l_{Q_p\cap Q_q}(f^\triangledown) \leq l_{Q_p}(f^\triangledown) \leq M(f^\triangledown)$. Since Q_q is the link-disjoint union of $Q_p \cap Q_q$ and a subpath of $\bar{Q}_p$, we derive from (3.3) that

$$\begin{aligned} l_{Q_p}(f^N) \geq l_{\bar{Q}_p}(f^N) - M(f^\triangledown) &\geq l_{Q_q}(f^N) - l_{Q_p\cap Q_q}(f^N) - M(f^\triangledown) \\ &\geq l_{Q_q}(f^N) - 2M(f^\triangledown) \geq l_{\bar{Q}_q}(f^N) - 3M(f^\triangledown), \end{aligned}$$

yielding

$$l_{\bar{Q}_p\cap\bar{Q}_q}(f^N) = l_{\bar{Q}_q}(f^N) - \Big(l_{Q_p}(f^N) - l_{Q_p\cap Q_q}(f^N)\Big) \leq 3M(f^\triangledown) + l_{Q_p\cap Q_q}(f^N) \leq 4M(f^\triangledown). \tag{3.6}$$

Notice that

$$\begin{aligned} ||R||_a &\leq l_{Q_p}(f^\triangledown) + l_{Q_q}(f^\triangledown) - l_{Q_p\cap Q_q}(f^\triangledown) + \tfrac{1}{2} l_{\bar{Q}_p\cap\bar{Q}_q}(f^N) \\ &\leq 2M(f^\triangledown) - l_{Q_p\cap Q_q}(f^N) + \tfrac{1}{2}\big(3M(f^\triangledown) + l_{Q_p\cap Q_q}(f^N)\big) \leq 3.5M(f^\triangledown). \end{aligned}$$

Similar to Case 2, when $l_R(f^\triangledown) \leq 8M(f^\triangledown)$, Lemma 2.6 ensures $\beta \leq 16$. When $l_R(f^\triangledown) > 8M(f^\triangledown)$, we have $l_{\bar{Q}_p\cap\bar{Q}_q}(f^\triangledown) > 6M(f^\triangledown)$, and set $S := \{Q \in f^\triangledown : Q \subseteq \bar{Q}_p \cap \bar{Q}_q\}$. Thus $l_{\cup_{Q\in S} Q}(f^\triangledown) + 2M(f^\triangledown) \geq l_{\bar{Q}_p\cap\bar{Q}_q}(f^\triangledown) > 6M(f^\triangledown)$ gives $l_{\cup_{Q\in S} Q}(f^\triangledown) > 4M(f^\triangledown)$, which implies $Q_i \in S$ for some i with $1 \leq i \leq j$ since $l_{\bar{Q}_p\cap\bar{Q}_q}(f^N) \leq 4M(f^\triangledown)$ by (3.6). Now $l_{Q_i}(f^N) \leq l_{\bar{Q}_p\cap\bar{Q}_q}(f^N) \leq 4M(f^\triangledown)$ contradicts (3.3).

We are now able to conclude that $\beta \leq 16$ in all cases, which establishes Theorem 3.1. $\square$

4 Tighter bounds on the prices of stability

This section is devoted to the establishment of the following theorem, which provides tighter upper bounds on PoS for SRR than Theorem 1 of [9] does. These improvements are achieved by more elaborate investigation with a novel bi-parametric analysis into the impacts of individual players strategies on the ring latency (see the proof of Lemma 2.7 in Appendix), instead of mechanically classifying them according to long- or short-path strategies as [9] did.

Theorem 4.1 *The price of stability of the SRR problem is at most* 3.9 *and is at most* 3.5 *if additionally the linear latency functions are homogenous.*

Proof. Let $f^\nabla = \{Q_1, Q_2, \ldots, Q_k\}$ denote an optimal routing for SRR instance $I = (R, l, (s_i, t_i)_{i=1}^k)$. We only need to consider the case where f^∇ is *not* a Nash routing. Therefore, some player $h \in \{1, 2, \ldots, k\}$ can benefit from unilaterally changing his strategy provided strategies of other players remain the same. It follows that the SRR instance admits a routing $f' = \{Q_1, \ldots, Q_{h-1}, \bar{Q}_h, Q_{h+1}, \ldots, Q_k\}$ for which we have

$$0 \le l_{\bar{Q}_h}(f^\nabla) + ||\bar{Q}_h||_a = l_{\bar{Q}_h}(f') < l_{Q_h}(f^\nabla) \le M(f^\nabla), \tag{4.1}$$

$$l_R(f^\nabla) = l_{Q_h}(f^\nabla) + l_{\bar{Q}_h}(f^\nabla) < 2M(f^\nabla) - ||\bar{Q}_h||_a. \tag{4.2}$$

Since $l_{\bar{Q}_h}(f') \ge ||\bar{Q}_h||$ and $l_{Q_h}(f^\nabla) \ge ||Q_h||$, it follows from (2.7) and (4.1) that

$$||R|| = ||Q_h|| + ||\bar{Q}_h|| < 2M(f^\nabla). \tag{4.3}$$

Consider f^N an arbitrary Nash routing for the instance I. Since the conditions of Lemma 2.7 have been satisfied by (4.2) and (4.3), the conclusion of the lemma establishes the theorem. $\square$

Actually, the above proof of Theorem 4.1 provides us with the following stronger results.

Theorem 4.2 *Given any routing f^∇ for an SRR instance, either f^∇ is a Nash routing, or $M(f^N) \le 3.9M(f^\nabla)$ for all Nash routings f^N of the instance, and $M(f^N) \le 3.5M(f^\nabla)$ if all linear latency functions in the instance are homogeneous.*

Corollary 4.3 *Given any SRR instance, either every optimal routing is a Nash routing, or the price of anarchy of the instance is at most* 3.9*, and at most* 3.5 *if all linear latency functions are homogeneous.*

5 Better evaluation of instability

The instance in Fig. 2(c) with $0 < \varepsilon < 0.5$ has the property that its *unique* optimal routing is a $\frac{4-\varepsilon}{4-2\varepsilon}$-approximate Nash routing. This instability ratio approaches $\frac{7}{6}$ as $\varepsilon \to 0.5$. One natural question is: Will the instability ratio grow infinitely when all SRR instances are taken into account? A negative answer has been provided in [9] that every SRR instance possess an optimal routing that approximates a Nash routing within a factor of 54. The gap between $\frac{7}{6}$ and 54 is large, and it is substantially narrowed down by the following theorem.

Theorem 5.1 *The SRR problem admits a* $(9, 1)$*-approximate Nash routing.*

Consider an arbitrary instance $I = (R, l, (s_i, t_i)_{i=1}^k)$. Let $f^* = \{Q_1, Q_2, \ldots, Q_k\}$ be an optimal routing for I such that

the potential $\Phi(f^*)$ is minimum among all optimal routings.

We will show that f^* is a $(9, 1)$-approximate Nash routing for I, which establishes Theorem 5.1. Suppose it were not the case.

- Then the instability of f^* and the minimality of $\Phi(f^*)$ provide two f^*-routes, say Q_1 and Q_2, which overlap at a proper common subpath $Q_1 \cap Q_2 \subsetneq Q_i$, $i = 1, 2$ (see Fig. 4(a)) with relatively high f^*-latency $l_{Q_1 \cap Q_2}(f^*) \ge \frac{8}{9}M(f^*)$, see (C8) below.
- Furthermore, there are two overlapping f^*-routes, say Q_g and Q_h, such that they each share with $Q_1 \cap Q_2$ common f^*-latency $l_{Q_1 \cap Q_2 \cap Q_i}(f^*) > \frac{2}{3}M(f^*)$, $i = g, h$ and $Q_g \cup Q_h$ contains all f^*-routes Q who share with $Q_1 \cap Q_2$ common f^*-latency $l_{Q_1 \cap Q_2 \cap Q}(f^*) > \frac{2}{3}M(f^*)$ (see Fig. 4(c)), where all these f^*-routes Q's have a common path in $Q_1 \cap Q_2$ of f^*-latency higher than $\frac{1}{3}M(f^*)$; in particular $l_{Q_g \cap Q_h}(f^*) > \frac{1}{3}M(f^*)$, see (C13) below.

- Changing Q_g and Q_h to $\bar{Q}_g$ and $\bar{Q}_h$, respectively, we get another optimal routing f (see Fig. 4(d)) of lower potential $\Phi(f) < \Phi(f^*)$, a contradiction. The decrease of potential from $\Phi(f^*)$ to $\Phi(f)$ is realized by saving more potential from $Q_g \cap Q_h$, which bears high f^*-latency (higher than $\frac{1}{3}M(f^*)$, see (C13)), than the addition to $\bar{Q}_g \cap \bar{Q}_h$, which has low f^*-latency (lower than $\frac{1}{9}M(f^*)$, see (C5)).

Recalling (2.5), the marginal change of potential on a link is increasing in the latency of that link. This fact constitutes the basic idea to justify our reasoning. The following rigorous proof elaborates on the contradiction argument in three steps: proving such Q_1 and Q_2 exist, then finding Q_g and Q_h, and finally changing Q_g and Q_h for achieving a smaller potential.

Proof of Theorem 5.1. To simplify description, let us shrink any $e \in E$ with $a_e + b_e = 0$ into a node, which obviously has no effect on our result. The preprocessing reduces us to the setting in which

(C1) $a_e + b_e > 0$ for all $e \in E$.

For any two ordered nodes $u, v \in V$, let $R[u, v]$ denote the clockwise path in R from u to v. In contrast, $R(u, v)$ denotes $V(R[u, v]) \cup E(R[u, v]) - \{u, v\}$. Swapping s_i and t_i if necessary, we assume

(C2) Q_i is the clockwise path in R from s_i to t_i, i.e., $Q_i = R[s_i, t_i]$, for every $i = 1, 2, \ldots, k$.

If $E(Q_i) \cup E(Q_j) = E$ and $E(Q_i) \cap E(Q_j) \neq \emptyset$ for some $1 \leq i < j \leq k$, then the routing f, obtained from f^* by replacing Q_i with $\bar{Q}_i$ and Q_j with $\bar{Q}_j$, is optimal, and it can be deduced from condition (C1) and the definition of Φ in (2.5) that $\Phi(f) < \Phi(f^*)$, a contradiction to the minimality of $\Phi(f^*)$. Hence

(C3) For all $1 \leq i, j \leq k$, either $E(Q_i) \cap E(Q_j) = \emptyset$ or $E(Q_i) \cup E(Q_j) \subsetneq E$ holds.

Step 1: Proving properties of Q_1 *and* Q_2. For each player $i \in \{1, 2, \ldots, k\}$, let routing f^i be obtained from f^* by replacing Q_i with $\bar{Q}_i$. Suppose without loss of generality that

(C4) $M_1(f^*)/M_1(f^1) = \max_{i=1}^{k} M_i(f^*)/M_i(f^i)$. So in f^* player 1 has the highest incentive to change, and no f^*-route can properly contain Q_1 by (C1).

To prove the theorem, we are to show that f^* is a $(9, 1)$-approximate Nash routing for I. To this end, we assume to the contrary that f^* is not. By Definition 2.2, we see from (C4) that player 1 could reduce his latency by a factor larger than 9 via unilateral deviation from f^*, i.e., $M_1(f^*) > 9M_1(f^1)$, which yields

(C5) $l_{\bar{Q}_1}(f^*) \leq l_{\bar{Q}_1}(f^1) = M_1(f^1) < \frac{1}{9}M_1(f^*) \leq \frac{1}{9}M(f^*)$, saying that the latency along $\bar{Q}_1 \cap \bar{Q}_2 \subseteq \bar{Q}_1$ is relatively small.

Combining (C5) and Lemma 2.1, we get $0 < M_1(f^*) - M_1(f^1) = \Phi(f^*) - \Phi(f^1)$. The minimality of $\Phi(f^*)$ enforces that f^1 is not optimal, and thus has maximum latency $M(f^1) > M(f^*)$ higher than f^* does. Since player 1 experiences lower latency in f^1 than in f^*, he does not bear the maximum latency of f^1, i.e., $M_1(f^1) < M_1(f^*) \leq M(f^*) < M(f^1)$. Hence the maximum latency must be experienced by another player $i \in \{2, 3, \ldots, k\}$, say $i = 2$, such that

$$M_2(f^1) = M(f^1) > M(f^*) \geq M_2(f^*).$$

The maximality of $M_2(f^*) = M(f^1)$ together with (C1) asserts that

(C6) No f^1-route can properly contain Q_2.

Note from $l_{\bar{Q}_1}(f^1) = M_1(f^1) < M(f^1) = M_2(f^1) = l_{Q_2}(f^1)$ that f^1-route $\bar{Q}_1$ of player 1 cannot contain f^1-route Q_2 of player 2, which is equivalent to say

$$E(Q_1) \cap E(Q_2) \neq \emptyset, \text{ implying } E - (E(Q_1) \cup E(Q_2)) \neq \emptyset \text{ by (C3)}.$$

Since the latency $M_2(f^1) > M_2(f^*)$ of player 2 is increased after player 1 changes from Q_1 to $\bar{Q}_1$, we have

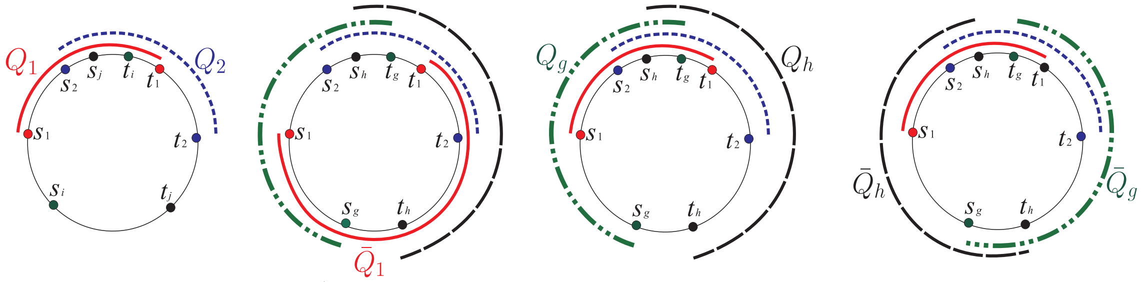
$$E(Q_2) \cap E(\bar{Q}_1) \neq \emptyset, \text{ or equivalently } E(Q_2) - E(Q_1) \neq \emptyset.$$

If $E(Q_1) - E(Q_2) = \emptyset$, then $E(Q_2) \cap E(\bar{Q}_1) \neq \emptyset$ implies $Q_2 \supsetneq Q_1$, in turn condition (C1) implies $\frac{M_2(f^*)}{M_2(f^2)} > \frac{M_1(f^*)}{M_1(f^1)}$, contradicting (C4). Hence

$$E(Q_1) - E(Q_2) \neq \emptyset.$$

These three expressions on nonemptyness along with (C3) implies that Q_2 has one end in $V(Q_1) - \{s_1, t_1\}$ and the other in $V(\bar{Q}_1) - \{s_1, t_1\}$. Symmetry allows us to assume without loss of generality that $s_2 \in V(Q_1) - \{s_1, t_1\}$ and $t_2 \in V(\bar{Q}_1) - \{s_1, t_1\}$. Thereby we arrive at the following configuration.

(C7) Nodes s_1, s_2, t_1 and t_2 are distinct, and located on R in clockwise order (see Fig. 4(a)). Hence $R[t_1, t_2], R[t_2, s_1] \subseteq R[t_1, s_1] = \bar{Q}_1$, and (C5) implies relatively low f^*-latency $l_P(f^*) \leq l_{\bar{Q}_1}(f^*) < \frac{1}{9}M(f^*)$ along paths $P = R[t_1, t_2]$, $R[t_2, s_1]$.



(a) $f^* = \{Q_1, Q_2, \ldots, Q_k\}$ (b) $f^1 = \{\bar{Q}_1, Q_2, \ldots, Q_k\}$ (c) $f^* = \{Q_1, Q_2, \ldots, Q_g, \ldots, Q_h, \ldots\}$ (d) $f = \{Q_1, Q_2, \ldots, \bar{Q}_g, \ldots, \bar{Q}_h, \ldots\}$

Fig. 4. Evaluation of instability.

As seen from (C7) and Fig. 4(a-b), the intersection of Q_1 and Q_2 is the path $R[s_2, t_1] = Q_2 \backslash R(t_1, t_2)$ bearing f^1-latency $l_{R[s_2,t_1]}(f^1) = M_2(f^1) - l_{R[t_1,t_2]}(f^1)$. Since $M_2(f^1) = M(f^1) > M(f^*)$ and $l_{R[t_1,t_2]}(f^1) < \frac{1}{9}M(f^*)$ by (C7), we see that

(C8) $Q_1 \cap Q_2 = R[s_2, t_1]$ bears relative high latency $l_{R[s_2,t_1]}(f^*) \geq l_{R[s_2,t_1]}(f^1) > \frac{8}{9}M(f^*)$, and

(C9) $Q_1 \backslash R(s_2, t_1) = R[s_1, s_2]$ bears relatively low f^*-latency $l_{R[s_1,s_2]}(f^*) = M_1(f^*) - l_{R[s_2,t_1]}(f^*) < M(f^*) - \frac{8}{9}M(f^*) = \frac{1}{9}M(f^*)$.

Despite the low latency of $\bar{Q}_1 \cap \bar{Q}_2$ in (C5) and high latency of $Q_1 \cap Q_2$ in (C8), switching Q_1 and Q_2 to their complements $\bar{Q}_1$ and $\bar{Q}_2$ might increase the maximum latency or the potential of the routing. We make further step to find two routes Q_g and Q_h whose intersection bears high latency against low latency of their complements such that switching them to complements gives rise to an optimal routing with smaller potential.

Step 2: Finding Q_g and Q_h. Since $\bar{Q}_1 \cup \bar{Q}_2$ bears f^*-latency $l_{\bar{Q}_1 \cup \bar{Q}_2}(f^*) \leq \frac{2}{9}M(f^*)$ by (C7) and (C8), the high latencies of Q_g and Q_h must be seen at $Q_1 \cap Q_2$. Due to this observation, we next define sets S and T to contain the candidates for Q_g and Q_h, respectively, as follows (see Fig. 4(a) for illustrations of positions of s_i, t_i, s_j, t_j):

$$\begin{aligned} S &:= \{Q_i : R[s_1, s_2] \subseteq Q_i, \, l_{R[s_2,t_i]}(f^*) > 2M(f^*)/3, 1 \leq i \leq k\}, \\ T &:= \{Q_j : R[t_1, t_2] \subseteq Q_j, \, l_{R[s_j,t_1]}(f^*) > 2M(f^*)/3, 1 \leq j \leq k\}. \end{aligned}$$

It is clear from the locations in (C7) and $l_{R[s_2,t_1]}(f^*) > \frac{8}{9}M(f^*)$ in (C8) that

(C10) $Q_1 \in S - T$ and $Q_2 \in T - S$.

Observe that elements of S (resp. T) are some f^*-routes going through $R[s_1, s_2]$ (resp. $R[t_1, t_2]$) and intersecting $Q_1 \cap Q_2 = R[s_2, t_1]$ with clockwise paths starting from s_2 (resp. anticlockwise paths staring from t_1). By (C4) (resp. (C6)), no route in S (resp. T) can properly contain Q_1 (resp. Q_2). It follows that path $Q_1 \cap Q_2 = R[s_2, s_1]$ contains both nodes t_i and s_j for any $Q_i \in S$ and $Q_j \in T$. In other words,

$$R[s_2, t_i] = Q_i \cap Q_2 \subseteq Q_1 \cap Q_2 \text{ and } R[s_j, t_1] = Q_j \cap Q_1 \subseteq Q_1 \cap Q_2 \text{ for any } Q_i \in S \text{ and } Q_j \in T.$$

In particularly, all f^*-routes in S (resp. T) interset $Q_1 \cap Q_2 = R[s_2, t_1]$ clockwisely (resp. anticlockwisely) with paths of f^*-latency higher than $\frac{2}{3}M(f^*)$.

Consider arbitrary $Q_i \in S$ and $Q_j \in T$. They intersect $Q_1 \cap Q_2 = R[s_2, t_1]$ with paths $R[s_2, t_1]$ and $R[s_j, t_1]$, respectively. As $R[s_2, t_1] \cup R[s_j, t_1] \subseteq Q_1 \cap Q_2 \subseteq Q_1$ does not bear f^*-latency higher than $l_{Q_1}(f^*) \leq M(f^*)$, it follows from $l_{R[s_2,t_i]}(f^*) > 2M(f^*)/3$ and $l_{R[s_j,t_1]}(f^*) > 2M(f^*)/3$ that

(C11) $R[t_i, t_1] = (Q_1 \cap Q_2) - R(s_2, t_i)$ experiences f^*-latency $l_{R[t_i,t_1]}(f^*) < \frac{1}{3}M(f^*)$ for all $Q_i \in S$; and $R[s_2, s_j] = (Q_1 \cap Q_2) - R(s_j, t_1)$ experiences f^*-latency $l_{R[s_2,s_j]}(f^*) < \frac{1}{3}M(f^*)$ for all $Q_j \in T$.

Moreover, from $l_{R[s_2,t_1]}(f^*) + l_{R[s_j,t_1]}(f^*) > 2(2M(f^*)/3)$ we deduce that $R[s_2, t_1] \cap R[s_j, t_1]$ must be a path (on $Q_1 \cap Q_2$) bearing f^*-latency at least $l_{R[s_2,t_1]}(f^*) + l_{R[s_j,t_1]}(f^*) - l_{Q_1}(f^*) > M(f^*)/3$. So $Q_i \cap Q_j \supseteq R[s_2, t_1] \cap R[s_j, t_1] = R[s_j, t_1]$ contains at least one link. It follows from (C3) that $Q_i \cup Q_j \neq R$ and

(C12) $Q_i \cap Q_j = R[s_j, t_i]$ is a subpath of $Q_1 \cap Q_2 = R[s_2, t_1]$ bearing f^*-latency $l_{Q_i \cap Q_j}(f^*) > \frac{1}{3}M(f^*)$ for all $Q_i \in S$ and $Q_j \in T$.

By (C12), the nonemptyness of S and T stated in (C10) allows us to take $Q_g \in S$ and $Q_h \in T$ such that $Q_g \cup Q_h$ covers all routes in $S \cup T$, i.e., $s_i \in R[s_g, s_1]$ for all $Q_i \in S$ and $t_j \in R[t_2, t_h]$ for all $Q_j \in T$ (see Fig. 4(c)). Properties (C11) and (C12) give rise to

(C13) $Q_g \cap Q_h = R[s_j, t_i]$ ($\subseteq R[s_2, t_1]$) bears latency $l_{Q_g \cap Q_h}(f^*) > \frac{1}{3}M(f^*)$.

(C14) $R[t_h, s_g]$ ($\subseteq R[t_2, s_1]$) is link disjoint from all paths in $S \cup T$.

(C15) $l_{R[t_g,t_1]}(f^*) < \frac{1}{3}M(f^*)$, and $l_{R[s_2,s_h]}(f^*) < \frac{1}{3}M(f^*)$.

Step 3: Switching Q_g and Q_h to their complements. Let routing f be obtained from f^* by replacing Q_g with $\bar{Q}_g$ and Q_h with $\bar{Q}_h$ (see Fig. 4(d)). It is clear from (C7) and (C14) that $\bar{Q}_1 = R[t_1, s_1] \supseteq R[t_2, s_1] \supseteq R[t_h, s_g]$. Comparing among routings f, f^* and f^1 (cf. Fig. 4(d-c-b)) leads to

$$\begin{aligned} l_{R[t_1,s_2]}(f) &\leq l_{R[t_1,s_1]}(f^1) + ||R[t_h, s_g]||_a + l_{R[s_1,s_2]}(f^*) \\ &\leq l_{\bar{Q}_1}(f^1) + ||R[t_2, s_1]||_a + l_{R[s_1,s_2]}(f^*) \\ &\leq 2l_{\bar{Q}_1}(f^1) + l_{R[s_1,s_2]}(f^*). \end{aligned}$$

Due to (C7) and (C9), both $l_{\bar{Q}_1}(f^1)$ and $l_{R[s_1,s_2]}(f^*)$ are smaller than $\frac{1}{9}M(f^*)$. Thus

(C16) $\bar{Q}_1 \cup \bar{Q}_2 = R[t_1, s_2]$ bears latency $l_{R[t_1,s_2]}(f) < \frac{1}{3}M(f^*)$.

With (2.5), it is not hard to see from Fig. 4(b-c-d) that

$$\begin{aligned} \Phi(f^*) - \Phi(f) &= 2l_{Q_g \cap Q_h}(f^*) - ||Q_g \cap Q_h||_a - \Big(2l_{R[t_h,s_g]}(f^1) + ||R[t_h, s_g]||_a\Big) \\ &\geq l_{Q_g \cap Q_h}(f^*) - 3l_{R[t_h,s_g]}(f^1). \end{aligned}$$

Then the inequality in (C13) and the inclusion in (C14) yield

$$\Phi(f^*) - \Phi(f) > M(f^*)/3 - 3l_{R[t_2,s_1]}(f^1).$$

Then from $l_{R[t_2,s_1]}(f^*) < \frac{1}{9}M(f^*)$, implied by the inequalities in (C7), we see $\Phi(f^*) - \Phi(f) > 0$. The minimality of $\Phi(f^*)$ asserts that f is not an optimal routing, so $M(f^*) < M(f) = M_q(f)$ for some $q \in \{1, 2, \ldots, k\}$. We distinguish between two cases, depending on whether q belongs to $\{g, h\}$ or not.

Case 1: $q \in \{g, h\}$. Symmetry allows us to assume $q = g$. Then from $R[t_1, s_g] \subseteq R[t_1, s_2]$, as properties (C14) and (C7) guarantee, we derive

$$\begin{aligned} M(f^*) &< M_q(f) = M_g(f) = l_{\bar{Q}_g}(f) = l_{R[t_g,t_1]}(f) + l_{R[t_1,s_g]}(f) \\ &\leq l_{R[t_g,t_1]}(f) + l_{R[t_1,s_2]}(f) = l_{R[t_g,t_1]}(f^*) + l_{R[t_1,s_2]}(f). \end{aligned}$$

By the first inequality in (C15) and the inequality (C16), both $l_{R[t_g,t_1]}(f^*)$ and $l_{R[t_1,s_2]}(f)$ are smaller than $\frac{1}{3}M(f^*)$. Thus the string of inequalities implies $M(f^*) < \frac{2}{3}M(f^*)$, which is absurd.

Case 2: $q \notin \{g, h\}$. Then player q adopts the same strategy Q_q in both f^* and f. Comparing f and f^*, we see $l_e(f_e) \leq l_e(f_e^*)$ for all $e \in E - E(R[t_h, s_g])$. Since $l_{Q_q}(f^*) \leq M(f^*) < M_q(f) = l_{Q_q}(f)$, we must have $E(Q_q) \cap E(R[t_h, s_g]) \neq \emptyset$. Hence (C14) claims $Q_q \notin S \cup T$, implying $l_{Q_q \cap R[s_2,t_1]}(f^*) \leq \frac{2}{3}M(f^*)$ by the definitions of S and T. Notice from the inclusion in (C14) that $E(R[s_2, t_1]) \subseteq E - E(R[t_h, s_g])$. So we have

$$\begin{aligned} M(f^*) &< l_{Q_q}(f) = l_{Q_q \cap R[s_2,t_1]}(f) + l_{Q_q \cap R[t_1,s_2]}(f) \\ &\leq l_{Q_q \cap R[s_2,t_1]}(f^*) + l_{R[t_1,s_2]}(f) \leq 2M(f^*)/3 + l_{R[t_1,s_2]}(f). \end{aligned}$$

Again a contradiction $M(f^*) < M(f^*)$ arises from $l_{R[t_1,s_2]}(f) < \frac{1}{3}M(f^*)$ in (C16).

The contradiction in either case disproves the assumption that f^* is not a $(9, 1)$-approximate Nash routing. The theorem is established. □

6 Fast search for good Nash routings

Given an SRR instance $I = (R, l, (s_i, t_i)_{i=1}^k)$ as an input, there is no loss of generality in assuming $n \leq 2k$, and $W = \max_{e \in E}(a_e + b_e)$ is an integer at least 2 (recall (2.1)). The number of bits in the binary representation of I is $\Omega(k + n \log W)$, which is considered as the input size of the instance. Let opt denote the maximum latency of an optimal routing for I. We first device an $O(nk^3)$ time algorithm to find a routing $\tilde{f}$ for I with $M(\tilde{f}) \leq 3\,\mathsf{opt}$, then from $\tilde{f}$ we reach a Nash routing f in $O(nk^3W)$ time. This convergence time improves upon the one in [9] by a factor of n, and is achieved by exploiting the unique structural property of the ring topology.

6.1 Data structure and subroutines

In our algorithmic implementations, the nodes $v_1, v_2, \ldots, v_n$ of $R = (V, E)$ are ordered in cyclic order. The source s_i and destination t_i, $1 \leq i \leq k$, are input in terms of $v_1, v_2, \ldots, v_n$. Fix the clockwise direction of R to be the one along which $v_1, v_2, \ldots, v_n$ can be encountered in this order. Recalling (2.2), suppose without loss of generality that P_i (resp. $\bar{P}_i$) is a clockwise (resp. counterclockwise) path in R from s_i to t_i, $i = 1, 2, \ldots, k$. We *associate* (record) each path P in the multiset $\mathcal{P} = \cup_{i=1}^k \{P_i, \bar{P}_i\}$ with a unique integer $\pi(P) \in \{1, 2, \ldots, 2k\}$ by putting $\pi(P_i) := 2i - 1, \pi(\bar{P}_i) := 2i$, $i = 1, 2, \ldots, k$. In this way, given $\pi(P)$ with $P \in \mathcal{P}$, we can deduce that

$$\begin{array}{l} P \text{ is a clockwise (resp. counterclockwise) path in } R \\ \text{from } s_{\lfloor \frac{\pi(P)+1}{2} \rfloor} \text{ to } t_{\lfloor \frac{\pi(P)+1}{2} \rfloor}, \text{ when } \pi(P) \text{ is odd (resp. even).} \end{array} \tag{6.1}$$

A routing $f = \{Q_1, Q_2, \ldots, Q_k\}$ for the instance I is recorded by ordered sequence $\pi(Q_1), \pi(Q_2), \ldots, \pi(Q_k)$ of integers in stead of the node-sequence representations of these paths.

We call a path in R with end nodes s and t an *s-t path*. A link in E is often considered a path in R. A path in R is *nontrivial* if it has at least one link. Let $P \subseteq R$ be a nontrivial v_i-v_j path, where $1 \leq i < j \leq n$. Set $\sigma(P)$ be the ordered quadruple (v_i, v_i', v_j', v_j) satisfying $v_i v_i', v_j' v_j \in E(P)$. Note that the v_i-v_j path P with $i < j$ has $\sigma(P)$ either $(v_i, v_{i+1}, v_{j-1}, v_j)$ or $(v_i, v_{i-1}, v_{j+1}, v_j)$, where the additions and subtractions on subscripts are taken modulo n. In the former case, P does not contain the link $v_n v_1 \in E$, $|E(P)| = j - i$, and P is said to *be of type I*. In the latter case, $v_n v_1 \in E(P)$, $|E(P)| = n - j + i$, and P is said to *be of type II*. Hence,

$$\text{Given } \sigma(P) \text{ for any path } P \text{ in } R, \text{ both } |E(P)| \text{ and the type of } P \text{ are determined in } O(1) \text{ time.} \tag{6.2}$$

Moreover, given $\sigma(P)$, the node-sequence representation of P can be produced in $O(n)$ time.

From (6.1) it is easy to see that, given $\pi(P)$ for $P \in \mathcal{P}$, it takes $O(1)$ time to produce $\sigma(P)$. So, by preprocessing, we obtain in $O(k)$ time all $\sigma(P)$, $P \in \mathcal{P}$. Clearly, this $O(k)$ time does not count in the time complexity $O(nk^3)$ and $O(nk^3W)$ to be established in Sections 6.2 and 6.3. Particularly, array Σ has been set up to bind $\pi(P)$ and $\sigma(P)$ together for $P \in \mathcal{P}$ in way of

$$\Sigma[\pi(P)] = \sigma(P),\ P \in \mathcal{P}. \tag{6.3}$$

Given $\pi(P)$ for $P \in \mathcal{P}$, from either (6.1) or (6.3) we see that $||P||_a$ is computable in $O(n)$ time. Thus, in $O(nk)$ time array Θ with

$$\Theta[\pi(P)] = ||P||_a,\ P \in \mathcal{P}, \tag{6.4}$$

has been constructed for providing data needed in future computation. Similarly, the $O(nk)$ time can be ignored.

Lemma 6.1 *Let Q_1 and Q_2 be nontrivial paths in R. Given $\sigma(Q_1)$ and $\sigma(Q_2)$, it takes $O(1)$ time to determine whether $Q_1 \subseteq Q_2$ or not, and to determine whether $E(Q_1) \cap E(Q_2) = \emptyset$ or not.*

Proof. Suppose $\sigma(Q_1) = (v_i, v'_i, v'_j, v_j)$ and $\sigma(Q_2) = (v_p, v'_p, v'_q, v_q)$. When Q_2 is of type I, $Q_1 \subseteq Q_2$ if and only if $p \le i < j \le q$. When Q_1 and Q_2 are of type II, $Q_1 \subseteq Q_2$ if and only if $i \le p < q \le j$. When Q_1 is of type I and Q_2 is of type II, $Q_1 \subseteq Q_2$ if and only if $j \le p$ or $i \ge q$. Hence, by (6.2), a subroutine can be devised for determining whether $Q_1 \subseteq Q_2$ or not in $O(1)$ time.

From $\sigma(Q_2)$ one easily obtains $\sigma(\bar{Q}_2)$ in $O(1)$ time. Note that $E(Q_1) \cap E(Q_2) = \emptyset$ if and only if $Q_1 \subseteq \bar{Q}_2$. The above subroutine runs in $O(1)$ time to determine whether $Q_1 \subseteq \bar{Q}_2$ or not, and hence $E(Q_1) \cap E(Q_2) = \emptyset$ or not. The conclusion follows. $\square$

Lemma 6.2 *Given i with $1 \le i \le k$, $v_p v_q \in E$ with $1 \le p < q \le n$, and $\pi(P)$ with $P \in \mathcal{P}$, it takes $O(1)$ time to determine whether $\{s_i, t_i\} \subseteq V(P)$ or not, and to determine whether $v_p v_q \in E(P)$ or not.*

Proof. Note that $\{s_i, t_i\} \subseteq V(P)$ if and only if $P_i \subseteq P$ or $\bar{P}_i \subseteq P$. Using array Σ in (6.3) and using $\sigma(v_p, v_q) = (v_p, v_q, v_p, v_q)$, Lemma 6.1 implies the result. $\square$

Lemma 6.3 *Given routing f for I represented by $\pi(Q_1), \pi(Q_2), \ldots, \pi(Q_k)$, and $e = v_p v_q \in E$, it takes $O(k)$ time to compute f_e, and takes $O(nk)$ time to compute all $M_i(f)$, $i = 1, 2, \ldots, k$. So $M(f)$ is derivable in $O(nk)$ time.*

Proof. By (6.3) and Lemma 6.2, for every $j \in \{1, 2, \ldots, k\}$, either $e \in Q_j$ or $e \notin E(Q_j)$ is checked in $O(1)$ time. Thus in $O(nk)$ time we get all $f_{e'}$, $e' \in E$, which enables us to compute $M_i(f) = l_{Q_i}(f)$ in $O(n)$ time for every $i \in \{1, 2, \ldots, k\}$. The lemma follows. $\square$

Lemma 6.4 *Given $\pi(Q_1)$ and $\pi(Q_2)$ for $Q_1, Q_2 \in \mathcal{P}$, it takes $O(1)$ time to either verify $E(Q_1) \cap E(Q_2) = \emptyset$ or compute $\sigma(P)$ for every nontrivial maximal subpath P of $Q_1 \cap Q_2$. So $||Q_1 \cap Q_2||_a$ is derivable in $O(n)$ time.*

Proof. Checking with array Σ in (6.3), we get $\sigma(Q_1) = (v_i, v'_i, v'_j, v_j)$ and $\sigma(Q_2) = (v_{\dot{p}}, v'_{\dot{p}}, v'_{\dot{q}}, v_{\dot{q}})$. In view of Lemma 6.1, it remains to consider the case where $E(Q_1) \cap E(Q_2) \neq \emptyset$, $Q_1 \not\subseteq Q_2$ and $Q_2 \not\subseteq Q_1$. So we can denote all of nontrivial maximal subpaths of $Q_1 \cap Q_2$ as X_1, X_g with $g = 1$ or 2. By Lemma 6.2 in $O(1)$ time we can find $\{v_i v'_i, v'_j v_j\} \cap E(Q_2)$ and $\{v_{\dot{p}} v'_{\dot{p}}, v'_{\dot{q}} v'_{\dot{q}}\} \cap E(Q_1)$, where both sets have size g. Clearly in $O(1)$ time we can write $\{\imath, \jmath\} = \{i, j\}$ and $\{p, q\} = \{\dot{p}, \dot{q}\}$ such that $v_\imath v'_\imath \in E(Q_2)$, $v_p v'_p \in E(Q_1)$, $v'_\jmath v_\jmath \in E(Q_2)$ if and only if $g = 2$, and $v'_q v_q \in E(Q_1)$ if and only if $g = 2$. Then $\sigma(X_1)$ turns out to be $(v_\imath, v'_\imath, v'_p, v_p)$ if $\imath < p$ and $(v_p, v'_p, v'_\imath, v_\imath)$ otherwise. In case of $g = 2$, we have $\sigma(X_2) = (v_\jmath, v'_\jmath, v'_q, v_q)$ if $\jmath < q$ and $\sigma(X_2) = (v_q, v'_q, v'_\jmath, v_\jmath)$ otherwise. $\square$

6.2 3-approximation to the optimal routing in $O(nk^3)$ time

For $i = 1, 2, \ldots, k$, let $\{O_i, \bar{O}_i\} = \{P_i, \bar{P}_i\}$ satisfy $||O_i|| \le ||\bar{O}_i||$. The routing $f^\circ := \{O_1, O_2, \ldots, O_k\}$ has the minimum ring latency $l_R(f^\circ)$ among all routings. In order to find a routing of maximum latency at most $3\,\mathsf{opt}$, we use $l_R(f^\circ)$ as a criterion to distinguish between two cases. When $l_R(f^\circ) \le 3\,\mathsf{opt}$, the routing f° has its maximum latency $M(f^\circ) \le l_R(f^\circ)$ not more than thrice the optimal, and therefore is the 3-approximation as desired. When $l_R(f^\circ) > 3\,\mathsf{opt}$, we aim to find an optimal routing $f^* = \{Q_1^*, Q_2^*, \ldots, Q_k^*\}$ by enumerating in polynomial time.

Consider $l_R(f^\circ) > 3\,\mathsf{opt}$. If f° is optimal then we are done by taking $f^* := f^\circ$. So assume $f^* \ne f^\circ$, and $\bar{Q}_g \in f^*$ for some $g \in \{1, 2, \ldots, k\}$. Taking $Q_h^* \in f^*$ such that $||Q_h^*|| = \max_{P \in f^*} ||P|| \ge ||\bar{O}_g|| > ||R||/2$, we see from $||Q_h^*|| + ||\bar{Q}_h^*|| = ||R||$ in (2.7) that $||\bar{Q}_h^*|| < ||R||/2 < ||Q_h^*||$ which says $\bar{Q}_h^* \in f^o$. Therefore

$$Q_h^* = \bar{O}_h \in f^* \text{ and } ||\bar{O}_h|| = \max_{P \in f^*} ||P|| > ||R||/2, \text{ for some } h \in \{1, 2, \ldots, k\}. \tag{6.5}$$

Note that, given h, both $\bar{O}_h$ and O_h are determined in view that R is link-disjoint union of $\bar{O}_h$ and O_h satisfying $||\bar{O}_h|| + ||O_h|| = ||R||$ and $||\bar{O}_h|| > ||R||/2$ (by (6.5)). The path $\bar{O}_h$ partitions $\{1, 2, \ldots, k\}$ into two sets:

$$S := \{i : \{s_i, t_i\} \subseteq V(O_h) \text{ or } V(\bar{O}_h), 1 \le i \le k\} \text{ and } T := \{1, 2, \ldots, k\} - S. \tag{6.6}$$

Since $l_R(f^*) \ge l_R(f^\circ) > 3\,\mathsf{opt}$, we see that

$$O \cup P \cup Q \subsetneq R \text{ for any } O, P, Q \in f^*. \tag{6.7}$$

Hence for any $i \in S$, if $\{s_i, t_i\} \subseteq V(\bar{O}_h)$ then $Q_i^* \subseteq \bar{O}_h = Q_h^*$ by (6.7), else $\{s_i, t_i\} \subseteq V(O_h)$ and $Q_i^* \subseteq O_h = \bar{Q}_h^*$ by the maximality in (6.5). In short,

$$Q_i^* \text{ is uniquely determined by } \bar{O}_h \text{ (in essence by } h) \text{ for any } i \in S. \tag{6.8}$$

Now for any $i \in T$, we observe from (6.6) that Q_i^* uses links from both $E(Q_h^*) = E(\bar{O}_h)$ and $E(\bar{Q}_h^*) = E(O_h)$. Write $\{s_h, t_h\} \cup [\cup_{i \in T}(\{s_i, t_i\} \cap V(O_h))] = \{u_0, u_{|T|+1}\} \cup \{u_i : 1 \le i \le |T|\}$ in way that

$$u_0, u_1, \ldots, u_{|T|}, u_{|T|+1} \text{ are encountered in order in a traverse of } O_h \text{ from } s_h \text{ to } t_h. \tag{6.9}$$

If $E(Q_i^*) \cup E(Q_j^*) \supseteq E(\bar{Q}_h^*)$ for some $i, j \in T$ then $Q_h^* \cup Q_i^* \cup Q_j^* = R$ contradicts (6.7). So $E(Q_i^*) \cup E(Q_j^*) \not\supseteq E(\bar{Q}_h^*)$ for all $i, j \in T$, which assures the existence of a *maximal* subpath Λ of $\bar{Q}_h^* = O_h$ that is nontrivial and link-disjoint from all paths Q_i^*, $i \in T$. By (6.9) we see that

$$\begin{aligned} &\Lambda \text{ is the subpath of } O_h \text{ between } u_j \text{ and } u_{j+1} \text{ for some } 0 \le j \le |T|, \text{ and} \\ &\text{all } Q_i^*,\ i \in T, \text{ are determined by } \Lambda. \end{aligned} \tag{6.10}$$

Thus the combination of (6.8) and (6.10) gives f^*.

To summarize, we make a number of "guesses", and pick the best outcome as an approximation to the optimal routing. Our guesses, held in a set $\mathcal{F}$, include $f^\circ = \{O_1, O_2, \ldots, O_k\}$, and routing f (as a guess of f^*) with respect to every $h \in \{1, 2, \ldots, k\}$ and every possible $\Lambda \subseteq O_h$, in view of (6.5), (6.6), (6.8) and (6.10). In total we have at most $1 + k(k-1) \le k^2$ guesses, each of which is a routing put in $\mathcal{F}$ as specified in the following pseudocode.

Approximate Efficient Routing Algorithm (ApxER_Alg)
Input: An SRR instance $I = (R, l, (s_i, t_i)_{i=1}^k)$ with minimum maximum latency opt.
Output: A routing $\tilde{f}$ for I with $M(\tilde{f}) \le 3\,\mathsf{opt}$.

1. Determine O_i and $\bar{O}_i$ for all $i = 1, 2, \ldots, k$

2. $f^\circ \leftarrow \{O_1, O_2, \ldots, O_k\}$, $\mathcal{F} \leftarrow \{f^\circ\}$
3. **for** $h = 1$ **to** k **do**
4. $\quad S \leftarrow \big\{i : \{s_i, t_i\} \subseteq V(O_h)$ or $V(\bar{O}_h), 1 \leq i \leq k\big\}$, $\quad T \to \{1, 2, \ldots, k\} - S$
5. $\quad$ Let $u_0, u_1, \ldots, u_{|T|}, u_{|T|+1}$ be as defined in (6.9)
6. $\quad$ **for** every $i \in S$ **do**
7. $\quad\quad$ **if** $\{s_i, t_i\} \subseteq V(\bar{O}_h)$ **then** $Q_i^* \leftarrow$ the s_i-t_i path in $\bar{O}_h$
8. $\quad\quad\quad\quad$ **else** $Q_i^* \leftarrow$ the s_i-t_i path in O_h
9. $\quad$ **end-for**
10. $\quad$ **for** $j = 0$ **to** $|T|$ **do**
11. $\quad\quad \Lambda \leftarrow$ the subpath of O_h between u_j and u_{j+1}
12. $\quad\quad Q_i^* \leftarrow$ the s_i-t_i path link-disjoint from Λ, for all $i \in T$
13. $\quad\quad f \leftarrow \{Q_1^*, Q_2^*, \ldots, Q_k^*\}$, $\quad \mathcal{F} \leftarrow \mathcal{F} \cup \{f\}$
14. $\quad$ **end-for**
15. **end-for**
16. Take $\tilde{f} \in \mathcal{F}$ such that $M(\tilde{f})$ is minimum
17. Output $\tilde{f}$

Clearly, Step 1 finishes in $O(kn)$ time. In turn, the construction of f° in Step 2 takes $O(k)$ time. By Lemma 6.2, Step 4 obtains S and T with $|S| \leq k$ and $|T| \leq k$ in $O(k)$ time. It is not hard to see that Step 5 can be accomplished in $O(k \log k)$ time with the help of merge sorting [11]. Subsequently, a single implementation of Steps 7–8 uses $O(1)$ time by Lemma 6.1. In practise, the setting in Step 11 is realized in $O(1)$ time by defining $\sigma(\Lambda)$, as given $\{u_j, u_{j+1}\} \subseteq \{v_1, v_2, \ldots, v_n\}$, the set $\{\sigma(\Lambda), \sigma(\bar{\Lambda})\}$ is derivable in $O(1)$ time, and by Lemma 6.2 the selection of $\sigma(\Lambda)$ from the set takes $O(1)$ time. Consequently, Step 12 finishes in $O(k)$ time by Lemma 6.1 and $|T| \leq k$. Evidently, Step 13 takes $O(k)$ time. Therefore, in $O(k^3)$ time we have all $O(k^2)$ guesses put in $\mathcal{F}$ when the for-loop (Steps 3–15) finishes. Recall from Lemma 6.3 that computing $M(f)$ for an $f \in \mathcal{F}$ takes $O(nk)$ time. It follows that algorithm ApxER_Alg runs in $O(nk^3)$ time and outputs routing $\tilde{f} \in \mathcal{F}$ with $M(\tilde{f})$ minimum. In particular $M(\tilde{f}) \leq M(f^\circ)$ since $f^\circ \in \mathcal{F}$. If $M(f^\circ) \leq 3\,\mathsf{opt}$, then $M(\tilde{f}) \leq 3\,\mathsf{opt}$, else by the above argument some optimal routing f must have been put to $\mathcal{F}$ in Step 13 since all possibilities have been enumerated. In conclusion, we have the following theorem.

Theorem 6.1 *Algorithm* ApxER_Alg *finds in* $O(nk^3)$ *time a routing* $\tilde{f}$ *with maximum latency* $M(\tilde{f}) \leq 3\,\mathsf{opt}$. □

6.3 Convergence to Nash routing in $O(nk^3W)$ time

To obtain a Nash routing for the SRR instance I, we make use of the fact in Lemma 2.1(ii): Starting from $\tilde{f}$, the potential of the current routing is decreased iteratively by changing the strategy of a player who has incentive to deviate, until the potential attains a local minimum. This is accomplished in $O(k^2\mathsf{opt})$ time, and hence in $O(nk^3W)$ time by the following Nash Routing Algorithm (NR_Alg), which provides a more efficient way to identify deviating players, and update the routing data (description) accordingly.

To facilitate our presentation, for any routing f for the instance I and any player i ($1 \le i \le k$), we use Q_i^f to denote the strategy in f of player i, and use f^i to denote the routing obtained from f by only changing the strategy of player i to $\bar{Q}_i^f$. Note that deriving f^i from f takes $O(1)$ time, as $Q_i^f \in \mathcal{P}$ and $Q_i^{f^i}$ $(= \bar{Q}_i^f)$ $\in \mathcal{P}$ are presented by integers $\pi(Q_i^f)$ and $\pi(\bar{Q}_i^f)$, respectively.

Nash Routing Algorithm (NR_Alg)
Input: An SRR instance $I = (R, l, (s_i, t_i)_{i=1}^k)$.
Output: A Nash routing f for I.

1. Apply ApxER_Alg to find a routing $\tilde{f}$ for I with $M(\tilde{f}) \le 3\,\mathsf{opt}$
2. Compute $||P_i \cap P_j||_a$ and $||P_i \cap \bar{P}_j||_a$ for all $1 \le i \ne j \le k$
3. $f \leftarrow \tilde{f}$, Compute f_e for all $e \in E$, and $M_i(f)$ for all $i = 1, 2, \dots, k$
4. $d \leftarrow \sum_{e\in E}[a_e(f_e + 1) + b_e]$, $i \leftarrow 1$
5. **repeat**
6.   **if** $d < 2M_i(f) + ||Q_i^f||_a$
7.   **then** $M_j(f) \leftarrow M_j(f) - ||Q_j^f \cap Q_i^f||_a + ||Q_j^f \cap \bar{Q}_i^f||_a$ for all $j \in \{1, 2, \dots, k\} - \{i\}$
8. $M_i(f) \leftarrow M_i(f^i)$, $f \leftarrow f^i$, Update f_e for all $e \in E$
9.   Go to Step 4
10. $i \leftarrow i + 1$
11. **until** $i = k + 1$
12. Output f

Evidently the above convergence algorithm simply does best response dynamic. Our contribution here is an $O(k)$-time method of doing all updates in response to a single route change.

Theorem 6.2 Nash Routing Algorithm *finds in $O(nk^3W)$ time a $(1, \beta)$-Nash routing with $\beta \le 11.7$, and $\beta \le 10.5$ if the latencies are homogeneous.*

Proof. By Theorem 6.1, in $O(nk^3)$ time, Step 1 finds a routing $\tilde{f}$ such that

$$M(\tilde{f}) \le 3\,\mathsf{opt} \text{ and } \Phi(\tilde{f}) \le kM(\tilde{f}) \le nk^2W. \tag{6.11}$$

The computations in Steps 2 and 3 take $O(nk^2)$ time and $O(nk)$ time, respectively, as guaranteed by Lemmas 6.4 and 6.3. Then NR_Alg spends $O(n)$ time getting value d in Step 4. Observe that

$$M_i(f) + ||Q_i^f||_a + M_i(f^i) = \sum_{e\in E} \big(a_e(f_e + 1) + b_e\big) = d$$

holds for all routings f for I and all players $i = 1, 2, \ldots, k$. The observation shows

$$d < 2M_i(f) + ||Q_i^f||_a \Leftrightarrow M_i(f^i) < M_i(f), \text{ for all } i = 1, 2, \ldots, k.$$

Therefore, it follows from (2.3) and (2.4) that f under investigation is not a Nash routing if and only if NR_Alg finds (by implementing Step 10 a certain number of times) some $i \in \{1, 2, \ldots, k\}$ for which the condition in Step 6 is satisfied, where by (6.4) searching for this i accomplishes in $O(k)$ time. In addition, recalling (2.1) and Lemma 2.1(i), the integrality of $M_i(f)$ and $M_i(f^i)$ implies

$$\Phi(f^i) = \Phi(f) - \big(M_i(f) - M_i(f^i)\big) \leq \Phi(f^i) - 1. \tag{6.12}$$

When f is not a Nash routing, Steps 7 and 8 are implemented in $O(k)$ time and $O(n)$ time, respectively, to reset f as f^i, and update $M_j(f)$ for all $j = 1, 2, \ldots, k$ and f_e for all $e \in E$ correctly, where the $O(n)$ time is enough as f and f^i differ only by the strategy player i adopts. Subsequently, NR_Alg goes back to Step 4. From an implementation of Step 4 till the next, $O(k)$ time elapses as $n \leq 2k$, and $\Phi(f)$ reduces by at least 1 as (6.12) states. Thus, starting from $\tilde{f}$ as Step 3 sets, it takes NR_Alg time $O(nk + k\,\Phi(\tilde{f}))$ to reach a Nash routing f as Step 12 outputs. The correctness of NR_Alg follows directly. By (6.11), the time complexity $O(nk^3 + nk^2 + nk + k\,\Phi(\tilde{f}))$ turns out to be $O(nk^3W)$. The performance ratios β are guaranteed by applying (6.11), and Theorem 4.2 with $f^\triangledown = \tilde{f}$. □

The *pseudo*-polynomial runtime of NR_Alg is complemented in some sense by the PLS-completeness [1] of the problem of finding a Nash equilibrium in an asymmetric congestion game with linear latencies and undirected links. Also useful is the observation that the SRR model does not possess the matroid structure [1] which can guarantee polynomial time convergence to a Nash equilibrium by best response dynamic.

The proof of Theorem 5.1 translates algorithmically to a modification of NR_Alg, which is referred to as ApxNR_Alg and finds a $(9, 3)$-approximate Nash routing in $O(nk^3 + k^2\mathsf{opt})$ time. Similarly, ApxNR_Alg first finds a routing $\tilde{f}$ that satisfies (6.11). Then with initial setting $f := \tilde{f}$, ApxNR_Alg lowers the potential of f iteratively by changing in each iteration strategies of one or two players under the condition that the maximum latency of f does not increase. Finally, at the time the potential cannot be reduced any more, the routing f turns out to be a $(9, 3)$-approximate Nash routing, as otherwise a contradiction in Case 1 or 2 of the proof of Theorem 5.1 would occur with f in place of f^*.

7 Further studies

In this section, Theorem 7.1 exhibits the exact values of PoS for SRR with two or three players. Discussion on future research concludes the paper.

7.1 Empirical study

We undertake some empirical study on the SRR of two and three players, which algorithmically leads us to the following more accurate evaluation of the PoS.

Theorem 7.1 *(i) The price of stability is* 1.25 *for the SRR problem with $k = 2$ players. (ii) The price of stability is approximately* 1.2565 *for the SRR problem with $k = 3$ players, where the absolute error is no more than* 0.0001.

To validate the values, our task here is to come up with an SRR instance $I = (R, l, (s_i, t_i)_{i=1}^k)$ for $k \leq 3$ whose PoS is as large as possible. Clearly, we may assume that the node set of R is $\{s_i, t_i : i = 1, 2, k\}$ (otherwise two links with a common end outside $\{s_i, t_i : i = 1, 2, k\}$ can be merged), and further that the $|\{s_i, t_i : i = 1, 2, k\}| = 2k$ (otherwise insertion of links with constant zero latency function(s) can split identical s_i and s_j or s_i and t_j with $i \neq j$). The links of R are accordingly labeled as $e_1, e_2, \ldots, e_{2k}$ in cyclic order. For $i = 1, 2, \ldots, 2k$, we write the nonnegative numbers a_{e_i} and b_{e_i}, which define the latency function $l_{e_i}(x) = a_{e_i}x + b_{e_i}$ on e_i, as a_i and b_i, respectively. In illustration, let us indicate s_1, t_1 (resp. s_2, t_2) by disks (resp. squares), and indicate s_3, t_3 by solid pentagons when $k = 3$. Fig. 5 exhausts all combinations of positions of source-destination pairs on the ring R (up to renaming players and swapping source and destination of the same player): cases (a') and (b') for 2-player SRR, and cases (a)–(e) for 3-player SRR.

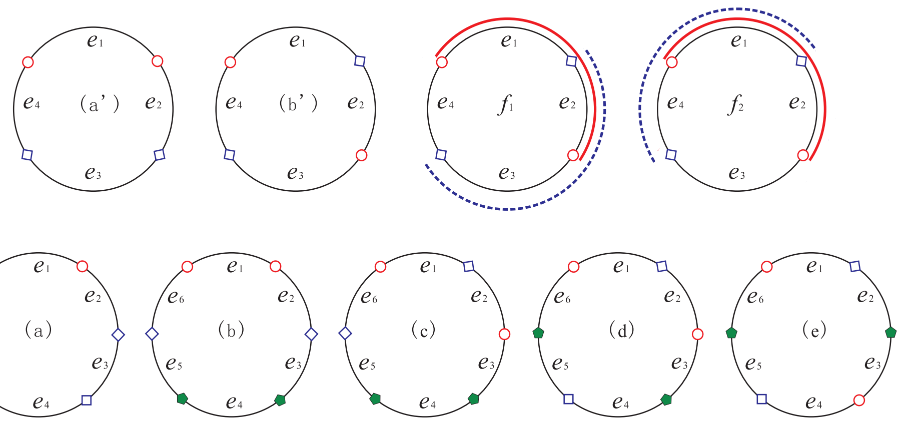


Fig. 5. The SRR of 2 or 3 players.

The 2-player SRR When $k = 2$, it is not hard to see that case (a') gives a PoS of 1 for all nonnegative a_i, b_i, $i = 1, 2$. In dealing with case (b'), we label all $2^k = 4$ routings as f_j, $j = 1, 2, 3, 4$, and suppose without loss of generality that f_1 and f_2 are as depicted in Fig. 5, and that changing the route of player 1 in routings f_1 and f_2 gives routings f_3 and f_4, respectively. The latency of player i in routing f_j can be expressed as a linear function $\chi_{ij} = \chi_{ij}(a_1, b_1, \ldots, a_{2^k}, b_{2^k})$ of variables $a_1, b_1, a_2, b_2, \ldots, a_{2^k}, b_{2^k}$, that is

$$\chi_{11} = a_1 + b_1 + 2a_2 + b_2,\ \chi_{21} = 2a_2 + b_2 + a_3 + b_3, \ldots,\ \chi_{24} = a_1 + b_1 + 2a_4 + b_4.$$

The functions χ_{ij} are then used to describe whether or not a player has an incentive to deviate. For example, let us consider a *sample scenario* when it happens that routings f_1 and f_2 are an optimum routing and the unique Nash routing, respectively; $M(f_1) = \chi_{21}$ and $M(f_2) = \chi_{12}$; and in f_3 player 2 wants to deviate. We are to maximize the PoS $= \chi_{12}/\chi_{21}$ subject to the constraints $\chi_{21} \geq \chi_{11}$ (saying $M(f_1) = \chi_{21}$), and $\chi_{21} > \chi_{22}$, $\chi_{23} > \chi_{24}$, $\chi_{14} > \chi_{12}$ (saying player 2 in f_1, player 2 in f_3, player 1 in f_4 wish to deviate). Recalling (2.1), this amounts to finding the largest constant p such that the system of linear inequalities:

$$(\mathrm{S}) = \left\{ \begin{array}{rcl} \chi_{12} - \mathsf{p} \cdot \chi_{2,1} & \geq & 0, \\ \chi_{21} - \chi_{11} & \geq & 0, \\ \chi_{21} - \chi_{22} & \geq & 1, \\ \chi_{23} - \chi_{24} & \geq & 1, \\ \chi_{14} - \chi_{12} & \geq & 1, \\ a_1, b_1, a_2, b_2, a_3, a_4, a_4, b_4 & \geq & 0 \end{array} \right.$$

has a feasible integer solution $(a_1, b_1, a_2, b_2, a_3, b_3, a_4, b_4)$. The task is accomplished by a binary process of narrowing down the interval $[p_l, p_r)$ that contains this largest value of p. More specifically, the system (S) is always feasible when $\mathsf{p} = p_l$ and infeasible when $\mathsf{p} = p_r$, meaning that the PoS cannot be greater than p_r in the sample scenario. By checking the middle point of the interval, the interval could be replaced with either its left half or its right half. The process terminates when the final interval has a length $p_r - p_l \leq 0.0001$.

Enumerating all scenarios, we similarly construct the corresponding systems of linear inequalities and intervals $[p_l, p_r)$. For all final intervals, we find no p_r greater than 1.2500. It implies that the PoS of the SRR with $k = 2$ players is bounded above by 1.25. In turn, the PoS of exact value 1.25 in Theorem 7.1(i) follows as a corollary of Remark 2.3.

The 3-player SRR When $k = 3$, more complicate enumerations and computations in the same spirit provide the results summarized in Table 1 below, where (a)–(e) refer to the cases in Fig. 5 and $*$ represents any nonnegative number.

	$[p_l, p_r)$	the setting realizing PoS $= p_l$											
		a_1	b_1	a_2	b_2	a_3	b_3	a_4	b_4	a_5	b_5	a_6	b_6
(a)	PoS = 1	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$
(b)	PoS = 1	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$
(c)	[1.2499,1.2500)	19	770351	256748	73	256746	37	10	31	256746	37	256739	84
(d)	[1.2499,1.2500)	19	26	0	33	260258	31	19	780892	260252	71	520558	21
(e)	[1.2564,1.2565)	1152663	21	3227324	32	8	3457818	691582	61	5	4841017	1383109	49

Table 1. The PoS in the SRR of 3 players.

The table gives a universal upper bound 1.2565 on the PoS for all SRR instances with 3 players. From the setting realizing PoS = 1.2564 in case (e), we draw the conclusion (ii) of Theorem 7.1.

7.2 Future research

In addition to the challenges of obtaining the exact PoS in general SRR, the upper bound 16 on the PoA of the SRR (Theorem 3.1) leaves much room for improvement. Also, it remains an interesting problem to explore the possibility of finding efficient (approximate) Nash equilibria for the SRR in polynomial time. Another intriguing direction is suggested by the small PoS in the SRR (Theorem 4.1) and the unbounded PoS in general selfish routing (Fig. 1): Characterizing network topologies of constant PoS deserves further research efforts.

Acknowledgements

This work was supported in part by China 973 Project under Grant No. 2011CB80800, the NSF of China under Grant No. 10531070, 10771209, 10721101, Chinese Academy of Sciences under Grant No. kjcx-yw-s7, and State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University under Contract No. RCS2009ZQ001.

The authors are deeply indebted to the anonymous referees for their invaluable comments and suggestions which have greatly improved the presentation of this paper.

References

[1] H. Ackermann, H. Röglin, B. Vöcking, On the impact of combinatorial structure on congestion games, Journal of the ACM 55 (2008) Article No. 25.

[2] E. Anshelevich, A. Dasgupta, J. Kleinberg, É. Tardos, T. Wexler, T. Roughgarden, The price of stability for network design with fair cost allocation, SIAM Journal on Computing 38 (2008) 1602–1623.

[3] E. Anshelevich, L. Zhang, Path decomposition under a new cost measure with applications to optical network design, ACM Transactions on Algorithms 4 (2008) Article No. 15.

[4] B. Awerbuch, Y. Azar, L. Epstein, The price of routing unsplittable flow, in: Proceedings of the 37th annual ACM Symposium on Theory of Computing, 2005, pp. 57–66.

[5] C. Bentza, M-.C. Costab, L. Létocartc, F. Roupin, Multicuts and integral multiflows in rings, European Journal of Operational Research 196 (2009) 1251–1254.

[6] A. Blum, A. Kalai, J. Kleinberg, Addmission control to minimize rejections, Lecture Notes in Computer Science 2152 (2001) 155–164.

[7] C. Busch, M. Magdon-Ismail, Atomic routing games on maximum congestion, Lecture Notes in Computer Science 4041 (2006) 79–91.

[8] C.T. Cheng, Improved approximation algorithms for the demand routing and slotting problem with unit demands on rings, SIAM Journal on Discrete Mathematics 17 (2004) 384–402.

[9] B. Chen, X. Chen, X. Hu, The price of atomic selfish ring routing, Journal of Comminatorial Optimization 19 (2010) 258–278.

[10] G. Christodoulou, E. Koutsoupias, The price of anarchy of finite congestion games, in: Proceedings of the 37th annual ACM Symposium on Theory of Computing, 2005, pp. 67–73.

[11] T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein, Introduction to Algorithms (2nd ed.), MIT Press and McGraw-Hill, 2001.

[12] J.R. Correa, A.S. Schulz, N.E. Stier-Moses, Fast, fair, and efficient flows in networks, Oprerations Research 55 (2007) 215–225.

[13] E. Even-Dar, A. Kesselman, Y. Mansour, Convergence time to Nash equilibria in load balancing, ACM Transactions on Algorithms 3 (2007) Article No. 32.

[14] A. Fabrikant, C. Papadimitriou, K. Talwar, The complexity of pure Nash equilibria, in: Proceedings of the 36th annual ACM Syposimum on Theory of Computing, 2004, pp. 604–612.

[15] D. Fotakis, Congestion game with linearly independent paths: convergence time and price of anarchy, Lecture Notes in Computer Science 4997 (2008) 33–45.

[16] D. Fotakis, S. Kontogiannisa, P. Spirakis, Selfish unsplittable flows, Theoretical Computer Science 348 (2005) 226–239.

[17] M. Gairing, T. Lücking, M. Mavronicolas, B. Monien, , The price of anarchy for restricted parallel links, Parallel Processing Letters 16 (2006) 117–132.

[18] E. Koutsoupias, C.H. Papadimitriou, Worst-case equilibria, Lecture Notes in Computer Science 1563 (1999) 404–413.

[19] D. Monderer, L. Shapley, Potential games, Games and Economic Behavior 14 (1996) 124–143.

[20] N. Nisan, T. Roughtgarden, É. Tardos, V.V. Vazirani (Eds), Algorithmic Game Theory, Cambridge University Press, Cambridge, UK, 2007.

[21] H. Lin, T. Roughgarden, É. Tardos, A. Walkover, Braess's paradox, Fibonacci Numbers, and exponential inapproximability, Lecture Notes in Computer Science 3580 (2005), 497–512.

[22] H. Lin, T. Roughgarden, É. Tardos, A. Walkover, Stronger bounds on Braess's paradox and the maximum latency of selfish routing, http://theory.stanford.edu/˜tim/papers/mcbp.pdf.

[23] R.W. Rosenthal, A class of games possessing pure-strategy Nash equilibira, International Jouranl of Game Theory 2 (1973) 65–67.

[24] T. Roughgarden, The price of anarchy is independent of the network topology, Jouranl of Computer and System Sciences 67 (2003) 342–364.

[25] T. Roughgarden, É. Tardos, How bad is selfish routing? Journal of the ACM 49 (2002) 236–259.

[26] A. Schrijver, P. Seymour, P. Winkler, The ring loading problem, SIAM Journal on Discrete Mathematics 11 (1998) 1–14.

[27] S. Suri, C.D. Tóth, Y. Zhou, Selfish load balancing and atomic congestion games, in: Proceedings of the 16th annual ACM symposium on Parallelism in algorithms and architectures, 2004, pp. 188-195.

[28] B.F. Wang, Linear time algorithms for the ring loading problem with demand splitting, Journal of Algorithms 54 (2005) 45-57.

[29] D. Weitz, The price of anarchy, Unpublished manuscript, 2001.

Appendix: Proof of Lemma 2.7

We are to show $\beta \le 3.9$ for general linear latencies and $\beta \le 3.5$ for homogeneous latencies. To this end, we assume that

$$\beta = \frac{M(f^N)}{M(f^\triangledown)} > 3.5, \tag{7.1}$$

on which we derive a contradiction in either case. Observe from (2.13) and (7.1) that Lemmas 2.4 2.5 apply with $\rho = 1$, yielding

$$\beta \le \max\{\gamma, 1\} + \frac{||R||}{2M(f^\triangledown)}, \tag{7.2}$$

$$(\beta\gamma - \beta - 2)\, l_{Q_1}(f^N) \le 2(\beta\gamma - 1)M(f^\triangledown) + (\beta+1)||Q_1||_a + ||R||_a - (\beta-1)||R||_b. \tag{7.3}$$

The combination of (7.1) and (7.2) implies

$$\gamma > 2.5. \tag{7.4}$$

It has been proved in the proof of Lemma 1 of [9] that

$$l_R^a(f^N) \le 2\gamma M(f^\triangledown) - (\gamma - 1)||R||_a. \tag{7.5}$$

Homogenous case: In the case of homogeneous linear latency functions, $||\cdot|| = ||\cdot||_a$ holds, subscript and superscript a can be dropped, and everything with subscript or superscript b is 0. If the theorem is not true, then we must have $f = f^\triangledown$ as a non-Nash routing, and a Nash routing f^N as studied above. From (7.5) and $M(f^N) \le (l_R(f^N) + ||R||)/2$ in (2.9), we derive

$$M(f^N) \le \gamma M(f^\triangledown) - \frac{\gamma - 2}{2}||R||. \tag{7.6}$$

Recall from (7.4) that $\gamma > 2.5$. Thus the combination of (7.6) and (7.1) implies

$$\gamma \ge \beta > 3.5. \tag{7.7}$$

So $\beta\gamma - \beta - 2 > 0$ is a positive number. Using it to divide both sides of the inequality in (7.3), we obtain

$$l_{Q_1}(f^N) \le \frac{2(\beta\gamma-1)}{\beta\gamma-\beta-2}M(f^\triangledown) + \frac{\beta+1}{\beta\gamma-\beta-2}||Q_1|| + \frac{1}{\beta\gamma-\beta-2}||R||,$$

which implies

$$\begin{aligned} l_{Q_1}(f^N) \quad &+ \quad \frac{||Q_1||}{2} + \frac{||R||}{2} \\ &\le \quad \frac{2(\beta\gamma-1)}{\beta\gamma-\beta-2}M(f^\triangledown) + \left(\frac{\beta+1}{\beta\gamma-\beta-2} + \frac{1}{2}\right)||Q_1|| + \left(\frac{1}{\beta\gamma-\beta-2} + \frac{1}{2}\right)||R|| \\ &= \quad \frac{2(\beta\gamma-1)}{\beta\gamma-\beta-2}M(f^\triangledown) + \frac{\beta(\gamma+1)}{2(\beta\gamma-\beta-2)}||Q_1|| + \frac{\beta(\gamma-1)}{2(\beta\gamma-\beta-2)}||R||. \end{aligned}$$

Since $M(f^N) \le l_{Q_1}(f^N) + \frac{||Q_1||}{2} + \frac{||R||}{2}$ by (2.10) and $||Q_1|| = \frac{||R||}{\gamma+1}$ by (2.14), we obtain

$$M(f^N) \le \frac{2(\beta\gamma-1)}{\beta\gamma-\beta-2} M(f^\triangledown) + \frac{\beta\gamma}{2(\beta\gamma-\beta-2)} ||R||. \tag{7.8}$$

By (7.7), both $\beta\gamma$ and $(\gamma-2)(\beta\gamma-\beta-2)$ are positive numbers. Observe that the coefficients of $||R||$ in (7.6) and (7.8) are negative and positive, respectively. Let us multiply both sides of (7.6) by $\beta\gamma$, multiply both sides of (7.8) by $(\gamma-2)(\beta\gamma-\beta-2)$, and put the two resulting inequalities together. As a result, we can cancel the terms involving $||R||$, and get

$$\frac{M(f^N)}{M(f^\triangledown)} \le \frac{3\beta\gamma^2 - 4\beta\gamma - 2\gamma + 4}{\beta\gamma^2 - 2\beta\gamma - 2\gamma + 2\beta + 4}, \tag{7.9}$$

which is true since both $\beta\gamma^2 - 2\beta\gamma - 2\gamma + 2\beta + 4$ and $M(f^\triangledown)$ are positive as implied by (7.7) and $M(f^\triangledown) > l_R(f^\triangledown)/2$, respectively. Observe that the right hand side of (7.9) has both numeration and denominator positive. Plugging $M(f^N)/M(f^\triangledown) = \beta$ into (7.9), we have

$$(\gamma^2 - 2\gamma + 2)\beta^2 - (3\gamma^2 - 2\gamma - 4)\beta + 2\gamma - 4 \le 0.$$

Notice from (7.7) that $\gamma^2 - 2\gamma + 2 > 0$, we obtain

$$\beta \le \frac{3\gamma^2 - 2\gamma - 4 + \sqrt{(3\gamma^2 - 2\gamma - 4)^2 - 4(\gamma^2 - 2\gamma + 2)(2\gamma - 4)}}{2(\gamma^2 - 2\gamma + 2)}.$$

Consider the expression on the right hand side of the above inequality as a function $\lambda(\gamma)$ of variable $\gamma \in (3.5, \infty)$ (recalling (7.7)). The unique root of $\lambda'(\gamma) = 0$ in $(3.5, \infty)$ is $\gamma \doteq 4.4562$, at which $\lambda(\gamma)$ attains a local maximum 3.4959. It follows that $\beta < 3.496$, a contradiction to (7.7).

General case: From (2.9) and (7.5), we obtain

$$M(f^N) \le \gamma M(f^\triangledown) - \frac{\gamma-1}{2} ||R||_a + \frac{1}{2} ||R||.$$

Using $||R|| < 2M(f^\triangledown)$, we get an analogue to (7.6):

$$M(f^N) \le (\gamma+1) M(f^\triangledown) - \frac{\gamma-1}{2} ||R||_a. \tag{7.10}$$

Since $\gamma > 1$ by (7.4) and $\beta = M(f^N)/M(f^\triangledown)$ by (7.1), the inequality in (7.10) further enables us to work on the following (from which we will derive a contradiction):

$$\gamma + 1 > \beta > 3.9. \tag{7.11}$$

Hence $\beta\gamma - \beta - 2$ is positive, which allows us to divide both sides of the inequality in (7.3) by $\beta\gamma - \beta - 2$, and obtain

$$l_{Q_1}(f^N) \le \frac{2(\beta\gamma-1)}{\beta\gamma-\beta-2} M(f^\triangledown) + \frac{\beta+1}{\beta\gamma-\beta-2} ||Q_1||_a + \frac{||R||_a - (\beta-1)||R||_b}{\beta\gamma-\beta-2}.$$

It follows from (2.10) that

$$\begin{aligned} M(f^N) &\le l_{Q_1}(f^N) + \frac{||Q_1||_a}{2} + \frac{||R||_a}{2} \\ &\le \frac{2(\beta\gamma-1)}{\beta\gamma-\beta-2} M(f^\triangledown) + \left(\frac{\beta+1}{\beta\gamma-\beta-2} + \frac{1}{2}\right) ||Q_1||_a + \frac{||R||_a - (\beta-1)||R||_b}{\beta\gamma-\beta-2} + \frac{1}{2}||R||_a \\ &= \frac{2(\beta\gamma-1)}{\beta\gamma-\beta-2} M(f^\triangledown) + \frac{\beta(\gamma+1)}{2(\beta\gamma-\beta-2)} ||Q_1||_a + \frac{\beta(\gamma-1)}{2(\beta\gamma-\beta-2)} ||R||_a - \frac{\beta-1}{\beta\gamma-\beta-2} ||R||_b. \end{aligned}$$

Notice from (7.11) that $\frac{\beta-1}{\beta\gamma-\beta-2}||R||_b \geq 0$, which implies

$$M(f^N) \leq \frac{2(\beta\gamma-1)}{\beta\gamma-\beta-2}M(f^\triangledown) + \frac{\beta(\gamma+1)}{2(\beta\gamma-\beta-2)}||Q_1||_a + \frac{\beta(\gamma-1)}{2(\beta\gamma-\beta-2)}||R||_a.$$

Recalling $||Q_1||_a = \frac{||R||_a}{\gamma+1}$ in (2.14), we have

$$M(f^N) \leq \frac{2(\beta\gamma-1)}{\beta\gamma-\beta-2}M(f^\triangledown) + \frac{\beta\gamma}{2(\beta\gamma-\beta-2)}||R||_a. \tag{7.12}$$

Let us multiply both sides of (7.10) by positive number $\beta\gamma$, multiply both sides of (7.12) by positive number $(\gamma-1)(\beta\gamma-\beta-2)$, and then add the resulting inequalities together. The terms involving $||R||_a$ vanish, so we arrive at

$$\beta\gamma M(f^N) + (\gamma-1)(\beta\gamma-\beta-2)M(f^N) \leq \beta\gamma(\gamma+1)M(f^\triangledown) + (\gamma-1)\cdot 2(\beta\gamma-1)M(f^\triangledown).$$

Since $M(f^N) = \beta M(f^\triangledown)$ by (7.1), the above inequality is equivalent to

$$(\beta\gamma^2 - \beta\gamma - 2\gamma + \beta + 2)\cdot \beta M(f^\triangledown) \leq (3\beta\gamma^2 - \beta\gamma - 2\gamma + 2)M(f^\triangledown).$$

Dividing both sides of the inequality by the positive number $M(f^\triangledown)$, we get

$$(\gamma^2-\gamma+1)\beta^2 - (3\gamma^2+\gamma-2)\beta + 2\gamma - 2 \leq 0.$$

By (7.11), $\gamma^2-\gamma+1 > 0$, which enforces

$$\beta \leq \frac{3\gamma^2+\gamma-2+\sqrt{(3\gamma^2+\gamma-2)^2 - 4(\gamma^2-\gamma+1)(2\gamma-2)}}{2(\gamma^2-\gamma+1)} =: \lambda(\gamma).$$

The unique root of $\lambda'(\gamma) = 0$ in interval $(2.9, \infty)$ (recalling (7.11)) is $\gamma \doteq 2.46$, at which $\lambda(\gamma)$ attains a local maximum 3.89. It follows that $\beta < 3.9$, a contradiction to (7.11). Lemma 2.7 is established. $\square$