THE UNIVERSITY OF WARWICK

Original citation:
Chen, Bo, Li, Song-Song and Zhang, Yu-Zhong (2012) Strong stability of Nash equilibria in load balancing games. In: International Symposium on Combinatorial Optimization 2012, Oxford, UK, 17-19 Sep 2012 pp. 1-14.
Permanent WRAP url:
http://wrap.warwick.ac.uk/52397

Strong Stability of Nash Equilibria in Load Balancing Games*

Bo Chen†
University of Warwick, United Kingdom

Song-Song Li‡ Yu-Zhong Zhang‡
Qufu Normal University, China

November 18, 2012

Abstract

We study strong stability of Nash equilibria in the load balancing games of m ($m \geq 2$) identical servers, in which every job chooses one of the m servers and each job wishes to minimize its cost, given by the workload of the server it chooses.

A Nash equilibrium (NE) is a strategy profile that is resilient to unilateral deviations. Finding an NE in such a game is simple. However, an NE assignment is not stable against coordinated deviations of several jobs, while a strong Nash equilibrium (SNE) is. We study how well an NE approximates an SNE.

Given any job assignment in a load balancing game, the improvement ratio (IR) of a deviation of a job is defined as the ratio between the pre- and post-deviation costs. An NE is said to be a ρ-approximate SNE ($\rho \geq 1$) if there is no coalition of jobs such that each job of the coalition will have an IR more than ρ from coordinated deviations of the coalition.

While it is already known that NEs are the same as SNEs in the 2-server load balancing game, we prove that, in the m-server load balancing game for any given $m \geq 3$, any NE is a (5/4)-approximate SNE, which together with the lower bound already established in the literature implies that the approximation bound is tight. This closes the final gap in the literature on the study of approximation of general NEs to SNEs in the load balancing games. To establish our upper bound, we apply with novelty a powerful graph-theoretic tool.

Keywords: load balancing game, Nash equilibrium, strong Nash equilibrium, approximate strong Nash equilibrium

*An earlier version has appeared at the International Symposium on Combinatorial Optimization 2012, September 17–19, 2012, University of Oxford, UK

†Corresponding author: Centre for Discrete Mathematics and Its Applications, Warwick Business School, University of Warwick, Coventry CV4 7AL, UK; `b.chen@warwick.ac.uk`

‡School of Management, Qufu Normal University, Rizhao 276826, China; `{163lisongsong, yuzhongrz}@163.com`

1 Introduction

In game theory, a fundamental notion is *Nash equilibrium* (NE), which is such a state that is *stable* against deviations of any individual game players (agents) in the sense that any such deviation will not bring about benefit to the deviator. Much stronger stability is exhibited by a *strong Nash equilibrium* (SNE), a notion introduced by Aumann [3], at which no coalition of agents exists such that each member of the coalition can benefit from coordinated deviations by the members of the coalition.

Evidentally selfish individual agents stand to benefit from cooperation and hence SNEs are much more preferred to NEs for stability. However, SNEs do not necessarily exist [2] and, even if they do, they are much more difficult to identify and to compute [7, 4]. It is therefore very much desirable to have the advantages of both computational efficiency and strong stability, which motivates our study in this paper. We establish that, for general NE job assignments in the load balancing games, which exist and are easy to compute, their loss of strong stability possessed by SNEs is at most 25%.

In a load balancing game, there are n selfish agents, each representing one of a set $J = \{J_1, \cdots, J_n\}$ of n jobs. In the absence of a coordinating authority, each agent must choose one of m identical servers, $M = \{1, \ldots, m\}$, to assign his job to in order to complete the job as soon as possible. All jobs assigned to the same server will finish at the same time, which is determined by the workload of the server, defined to be the total processing time of the jobs assigned to the server. Let job J_j have a processing time p_j $(1 \leq j \leq n)$ and let S_i denote the set of jobs assigned to server i $(1 \leq i \leq m)$. For convenience, we will use "agent" and "job" interchangeably, and consider job processing time also as their "lengths". The completion time c_j of job $J_j \in S_i$ is the *workload* of its server: $L_i = \sum_{J_j \in S_i} p_j$.

NEs in the load balancing games have been widely studied (see, e.g., [8, 11, 6, 10, 5]) with the main focus of quantifying their loss of global optimality in terms of the price of anarchy, a term coined by Koutsoupias and Papadimitriou [11], as largely summarized in [12]. In this paper, we study NEs in the load balancing games from a different perspective by quantifying their loss of strong stability.

We focus on pure NEs, those corresponding to deterministic job assignments in load balancing games. Finding such an NE is simple and identification of an SNE is strongly NP-hard, while high-quality NEs are easily computed [4]. Given any job assignment in a load balancing game, the *improvement ratio* (IR) of a deviation of a job is defined as the ratio between the pre- and post-deviation costs. An NE is said to be a ρ-approximate SNE ($\rho \geq 1$) (which is called ρ-SE in [1]) if there is no coalition of jobs such that each job of the coalition will have an IR more than ρ from coordinated deviations of the coalition. Clearly, the stability of NE improves with a decreased value of ρ and a 1-approximate SNE is in fact an SNE itself.

For the load balancing game of two servers, one can easily verify that every NE is also an SNE [2]. If there are three or four servers in the game, then it is proved in [7] and [4], respectively, that any NE assignment is a (5/4)-approximate SNE, and the bound is tight. Furthermore, it is a $(2 - 2/(m + 1))$-approximate SNE if the game has m servers for $m \geq 5$ [7].

We establish in this paper that, in the m-server load balancing game ($m \geq 3$), any NE is a (5/4)-approximate SNE, which is tight and hence closes the final gap in the literature on the study of NE approximation of SNE in the load balancing games. To establish our approximation bound, we apply with novelty a powerful graph-theoretic tool.

2 Definitions and Preliminaries

2.1 Graph-theoretic Tool [4]

As a tool of our analysis, we start with the minimal deviation graph introduced by Chen [4]. For convenience we collect into this subsection some basic results on minimal deviation graph from [4]. Given an NE job assignment $S = \{S_i : i \in M\}$ and a coalition Γ of agents (or simply, of jobs), as an NE-based coalitional deviation or simply *coalitional deviation*, we refer to a collective action in which each job of the coalition migrates from its server based on S with a decreased completion time. We introduce deviation graphs to characterize coalitional deviations. In a coalitional deviation, a server i is said to be *participating* or *involved* if its job set changes after the deviation. Given a coalitional deviation $\Delta = \Delta(\Gamma)$ of a coalition Γ, we define the corresponding (directed) deviation graph $G(\Delta) = (V, A)$ as follows:

$$\begin{aligned} V = V(G) &:= \{i : \text{ server } i \text{ is a participating server}\}; \\ A = A(G) &:= \{(u, v) : \text{ a job } J_j \in \Gamma \text{ migrates from } S_u \text{ to } S_v\}. \end{aligned}$$

Given a coalitional deviation Δ, we denote by $L'_i = L_i(\Delta)$ the workload of server i after deviation Δ, and by $\mathrm{IR}(\Delta)$ the minimum of the improvement ratios of all jobs taking part in Δ. For notational convenience, let $v_i = i$ for $i = 1, \cdots, m$. Then we have the following definition and lemmas from [4]:

Lemma 1 *The out-degree $\delta^+(i)$ of any node i of a deviation graph is at least 1, and hence $|S_i| \geq 2$.*

Lemma 2 *If all m servers are involved in a coalitional deviation, then the deviation graph does not contain any node-disjoint directed cycles that span all nodes.*

Definition 1 *Let Γ be a coalition and Δ be a coalitional deviation of Γ. Deviation graph $G = G(\Delta)$ is said to be* minimal *if $IR(\Delta') < IR(\Delta)$ for any coalitional deviation Δ' of Γ' that is a proper subset of Γ.*

Lemma 3 *The in-degree $\delta^-(i)$ of any node i of a minimal deviation graph is at least 1.*

Lemma 4 *A minimal deviation graph is strongly connected.*

2.2 Some Observations

In our study of bounding NE approximation of SNE, we can apparently focus on those coalitional deviations that correspond to minimal deviation graphs. We start with several observations on the coalitional deviation Δ_m of any NE-based coalition Γ involving m ($m \geq 3$) servers, which has corresponding minimal deviation graph $G(\Delta_m)$. For notational simplicity, we omit subscript m for coalitional deviation Δ_m involving all m servers that leads to minimal deviation graph $G(\Delta_m)$. Hence $V(G) = M$.

If two jobs assigned to server $i \in M$ in the NE assignment migrate to server $j \in M$ $(j \neq i)$ together, or do not migrate together, then we can treat them as one single job without loss of generality in our study of minimal deviation graph. With this understanding, if we let a_i $(i \in M)$ denote the number of jobs assigned to server i in the NE assignment, then the following is immediate.

Observation 1 *For any $i \in M$, we have $2 \leq a_i \leq m$. $\delta^+(i) = a_i$ or $\delta^+(i) = a_i - 1$.*

As a result of the above observation, the node set M can be partitioned into two, M' and M'', as follows:

$$\begin{aligned} M' &:= \{i \in M : \ a_i = \delta^+(i)\}, \\ M'' &:= M \backslash M' = \{i \in M : \ a_i = \delta^+(i) + 1\}. \end{aligned}$$

By applying a data scaling if necessary, we assume that $\min_{i \in M} L_i = 1$ without loss of generality.

Observation 2 *For any $i \in M$, we have $L_i \leq a_i/(a_i - 1)$.*

Proof. Suppose to the contrary that $L_i > a_i/(a_i - 1)$, which implies that $a_i > L_i/(L_i - 1)$.

Let x_i denote the length of the shortest job assigned to server i in the NE assignment. We have $L_i \geq a_i x_i$, which leads to $L_i > L_i x_i/(L_i - 1)$, that is, $L_i > x_i + 1$, which implies that the shortest job assigned to server i in the NE assignment can have the benefit of reducing its job completion time by unilaterally migrating to the server of which the workload is 1, contradicting the NE property. □

The following observation states that, if all jobs on a server participate in the migration, then none of the servers they migrate to will have *all* its jobs migrate out.

Observation 3 *If $(i, j) \in A$ and $i \in M'$, then $j \in M''$.*

Proof. Suppose to the contrary that $a_j \neq \delta^+(j) + 1$. According to Observation 1, we have $a_j = \delta^+(j)$, which implies that all the jobs assigned to server i and server j in the NE assignment belong to coalition Γ.

Since $(i, j) \in A$, there is a job $J_k \in \Gamma$ that migrates from server i to server j. Consider the new coalition Γ' formed by all members of Γ except J_k. Then we have $\emptyset \neq \Gamma' \subset \Gamma$. Let Δ' be such a coalitional deviation of Γ' that is the same as Δ except without the involvement of J_k and the job(s) that migrate(s) to i (resp. j) in Δ will migrate to j (resp. i) in Δ'. Then we have $\mathrm{IR}(\Delta') = \mathrm{IR}(\Delta)$, contradicting the minimality of the deviation graph G according to Definition 1. □

The following observation is a result of Observation 3:

Observation 4 *Assume $i, j \in M'$. Hence $(i, j), (j, i) \notin A$ according to Observation 3. Let Δ' be the same as Δ except that any job that migrates to i (resp. j) in Δ will migrate to j (resp. i) in Δ'. Then $IR(\Delta') = IR(\Delta)$, and $G(\Delta')$ is minimal.*

3 Analysis of Minimal Deviation Graph

3.1 Auxiliary arc set $\widetilde{A}$ and node set W

To help our analysis, we will introduce in this subsection a special arc set $\widetilde{A} \subseteq A$ in the minimal deviation graph $G(\Delta)$ and three node sets that $\widetilde{A}$ determines: W_0, W_1 and $\widetilde{W}_1$.

For any node $i \in M$, denote $Q^+(i) := \{j \in M : (i,j) \in A\}$ and $Q^-(i) := \{j \in M : (j,i) \in A\}$. For notational convenience, for any node set $S \subseteq M$, we denote $Q^+(S) := \bigcup_{i \in S} Q^+(i)$ and $Q^-(S) := \bigcup_{i \in S} Q^-(i)$. With A replaced by $\widetilde{A}$ above, we similarly define $\widetilde{Q}^+(i)$, $\widetilde{Q}^-(i)$, $\widetilde{Q}^+(S)$ and $\widetilde{Q}^-(S)$.

Let us formally define $\widetilde{A}$ as follows. According to Lemma 3, $|Q^-(i)| \geq 1$ for any $i \in M$. For each $i \in M$ we pick up an arc from non-empty set $Q^-(i)$ to form an m-element subset $\widetilde{A} \subseteq A$. Then $\widetilde{A}$ possesses the following properties, where $b_i := |\widetilde{Q}^+(i)|$ for any $i \in M$:

$$|\widetilde{Q}^-(i)| = 1 \text{ for any } i \in M, \tag{1}$$

$$\sum_{i=1}^{m} b_i = |\widetilde{A}| = m. \tag{2}$$

If node set S is a singleton, then we will also use S to denote the singleton if no confusion can arise. Hence, due to (1) we will also use $\widetilde{Q}^-(i)$ to denote the single *element* of the corresponding set. Any arc set $\widetilde{A} \subseteq A$ that satisfies properties (1) and (2) is said to be *tilde-valid*. Immediately we have

Lemma 5 $W_0 := \{i \in M : b_i = 0\} \neq \emptyset$.

Proof. Suppose to the contrary that $W_0 = \emptyset$. Then any $b_i \geq 1$ in (2), which implies that $b_i = 1$ for any $i \in M$, so that $\widetilde{A}$ forms some node-disjoint directed cycles that span all nodes, contradicting Lemma 2. $\square$

Note that, from the formation of arc set $\widetilde{A}$, it is clear that $\widetilde{A}$ as a tilde-valid arc set may not be unique. However, among all possible choices of tilde-valid arc set $\widetilde{A} \subseteq A$, we choose one that has some additional properties in terms of minimum cardinalities of some combinatorial structures, which we shall use a sequence of three assumptions to describe. These assumptions are made without loss of generality due to the finiteness of the total number of tilde-valid arc sets. Similarly, for a given coalition Γ, we shall also choose our coalitional deviation Δ so that it has certain property (see Assumption 4).

Assumption 1 *Arc set $\widetilde{A}$ is tilde-valid and it minimizes $|W_0(\widetilde{A})|$.*

Let $\widetilde{W}_0 = Q^+(W_0)$. Then $\widetilde{W}_0 \neq \emptyset$ according to Lemmas 5 and 1. A node $i \in M$ is said to be *associated* with W_0 if it is *linked* to an element of $\widetilde{W}_0$ through a sequence of arcs in $\widetilde{A}$ and A in alternation. More formally, $i \in M$ is associated with W_0 if and only if, for some integer $k \geq 0$, there are nodes $\{i_0, \ldots, i_k, j_0, \ldots, j_k\} \subseteq M$ with $i = i_k$ and $j_0 \in \widetilde{W}_0$, such that

$$(i_0, j_0), \ldots, (i_k, j_k) \in \widetilde{A} \quad \text{and} \quad (i_0, j_1), \ldots, (i_{k-1}, j_k) \in A. \tag{3}$$

Note that in the above definition, if $i = i_k$ is associated with W_0, then i_0, ..., i_{k-1} used in (3) are each associated with W_0. Define

$$
\begin{aligned}
W_1 &:= \{i \in M : \text{ node } i \text{ is associated with } W_0\}, \\
\widetilde{W}_1 &:= \widetilde{Q}^+(W_1).
\end{aligned}
$$

Immediately we have $\widetilde{Q}^-(\widetilde{W}_0) \subseteq W_1$, which implies that

$$\widetilde{W}_0 \subseteq \widetilde{W}_1. \tag{4}$$

On the other hand, since $\widetilde{Q}^-(\widetilde{W}_0) \neq \emptyset$ according to (1), we have $W_1 \neq \emptyset$.

Lemma 6 *For any $i \in W_1$, $b_i = 1$. Furthermore, $Q^+(W_0 \cup W_1) = \widetilde{W}_1$.*

Proof. It is clear from the definition that $W_1 \cap W_0 = \emptyset$. Hence $b_i \geq 1$ for any $i \in W_1$. Assume for contradiction that $b_i \geq 2$ for some $i \in W_1$. Since i is associated with W_0, in addition to nodes $\{i_0, \ldots, i_k, j_0, \ldots, j_k\} \subseteq M$ satisfying (3), we have a node $h \in W_0$ such that $(h, j_0) \in A$ according to the definition of $\widetilde{W}_0$. Now we remove $k+1$ arcs $(i_0, j_0), \ldots, (i_k, j_k)$ from $\widetilde{A}$ and add $k+1$ new arcs $(h, j_0), (i_0, j_1), \ldots, (i_{k-1}, j_k)$ to $\widetilde{A}$. It is easy to see that the new set $\widetilde{A}$ still has properties (1) and (2). Additionally, under the new $\widetilde{A}$, all $\{b_k\}$ remain the same except two of them: b_h and b_i, with the former increased by 1 and the latter decreased by 1. Since $b_i \geq 2$ under the original $\widetilde{A}$, then $i \notin W_0$ under the new $\widetilde{A}$. Consequently, the new W_0 determined by the new $\widetilde{A}$ contains a smaller number of elements, contradicting Assumption 1 about the original $\widetilde{A}$.

To prove the second part of the lemma, we notice (4) and let $i \in W_1$ and $(i, j) \in A$. We show that $j \in \widetilde{W}_1$. In fact, since $|\widetilde{Q}^-(j)| = 1$ according to (1), we have a node $h \in M$ such that $(h, j) \in \widetilde{A}$. Now since i is associated with W_0, we conclude that h is also associated with W_0, which implies that $j \in \widetilde{W}_1$. Therefore, we have proved that $Q^+(W_0 \cup W_1) \subseteq \widetilde{W}_1$. The other direction of inclusion is apparent. $\square$

As a result of Lemma 6 and (1), mapping $\widetilde{Q}^+(\cdot)$ from W_1 *onto* $\widetilde{W}_1$ is a one-to-one correspondence and hence

$$|W_1| = |\widetilde{W}_1| > 0. \tag{5}$$

Let

$$W := W_0 \cup W_1 \cup \widetilde{W}_1.$$

3.2 Decomposition of node set X

Recall that, for any $i \in M$, a_i is the number of jobs assigned to server i in the NE assignment and $b_i = |\widetilde{Q}^+(i)|$ for a fixed arc set $\widetilde{A}$ satisfying Assumption 1. For a pair of integers a and b with $2 \leq a \leq m$ and $0 \leq b \leq a$, let $M_a^b := \{i \in M : a_i = a, b_i = b\}$. Then it is clear that

$$\bigcup_{2 \leq a \leq m} \bigcup_{0 \leq b \leq a} M_a^b = M. \tag{6}$$

As we will see in the next section, bounding the sizes of the sets M_2^2 and M_3^3 is vital in our establishment of the desired approximation bound. We therefore take a close look at the two sets by partitioning

$$X := M_2^2 \cup M_3^3$$

into a number of subsets, so that different bounding arguments can be applied to different subsets.

Notation 2 *Let $\widetilde{M_2^2} := \{\ell \in M_2^2 : \widetilde{Q}^+(\ell) \nsubseteq W\}$. For convenience, we reserve letter ℓ to exclusively index elements of $\widetilde{M_2^2}$ and let $\widetilde{Q}^+(\ell) = \{\ell_1, \ell_2\}$ with the understanding that it is always the case that $\ell_1 \notin W$.*

For any $\ell \in \widetilde{M_2^2}$, $\ell_1 \notin W$ implies $b_{\ell_1} \geq 1$ since $W_0 \subseteq W$. On the other hand, since arc $(\ell, \ell_1) \in \widetilde{A} \subseteq A$, we have $a_{\ell_1} = \delta^+(\ell_1) + 1 \geq b_{\ell_1} + 1$ according to Observation 3, which implies that ℓ_1 must belong to one of the following three mutually disjoint node sets:

$$\begin{aligned} Z_1 &:= \{i \in M \backslash W : b_i = 1\}, \\ Z_2 &:= \{i \in M \backslash W : b_i > 1, a_i > b_i + 1\}, \\ Z &:= \{i \in M \backslash W : b_i > 1, a_i = b_i + 1\}. \end{aligned}$$

Therefore, if we define

$$\left\{ \begin{array}{lcl} X_3 & := & \{\ell \in \widetilde{M_2^2} : \widetilde{Q}^+(\ell) \cap W = \emptyset\}; \\ X_4 & := & \{\ell \in \widetilde{M_2^2} \backslash X_3 : \ell_1 \in Z_1 \cup Z_2\}; \\ X_5 & := & \{\ell \in \widetilde{M_2^2} \backslash X_3 : \ell_1 \in Z, Q^+(\ell_1) \nsubseteq W, \delta^-(\ell_1) > 1\}; \\ X_6 & := & \{\ell \in \widetilde{M_2^2} \backslash X_3 : \ell_1 \in Z, Q^+(\ell_1) \nsubseteq W, \delta^-(\ell_1) = 1\}; \end{array} \right.$$

then we have

$$X_{11} := \widetilde{M_2^2} \setminus \bigcup_{k=3}^{6} X_k = \{\ell \in \widetilde{M_2^2} \backslash X_3 : \ell_1 \in Z, Q^+(\ell_1) \subseteq W\}.$$

Now let

$$\left\{ \begin{array}{lcl} X_1 & := & \{i \in X : \widetilde{Q}^+(i) \subseteq W\} \cup X_{11}; \\ X_2 & := & \{i \in M_3^3 : \widetilde{Q}^+(i) \nsubseteq W\}. \end{array} \right.$$

Clearly, $X_i \cap X_j = \emptyset$ $(1 \leq i \neq j \leq 6)$ and $X = \bigcup_{k=1}^{6} X_k$.

3.3 Bounding the size of node set X

3.3.1 Part 1

We continue our analysis of the node set X. Through a series of six lemmas, we establish that the number of nodes in X_t is at most $c_t|Y_t|$ $(t = 1, \ldots, 6)$, where $c_t \in \{1, \frac{1}{2}\}$ and $Y_1, \ldots, Y_6 \subseteq M \backslash X$ are mutually disjoint node sets to be defined below.

Lemma 7 *Let $Y_1 := W_1$. Then $|X_1| \leq |Y_1|$.*

Proof. Suppose to the contrary that $|X_1| > |Y_1|$, that is, $|X_1| > |W_1| = |\widetilde{W}_1|$ according to (5). Let $X'_1 \subsetneq X_1$ be a proper subset of $|\widetilde{W}_1| > 0$ elements. Define

$$\begin{aligned} K &:= W \backslash \widetilde{W}_1 \subseteq W_0 \cup W_1, \\ K' &:= \{k \in M \backslash W : \ Q^-(k) \cap X'_1 \neq \emptyset\}, \\ \Gamma' &:= \{J_j \in \Gamma : \ J_j \in \bigcup_{k \in \widetilde{M}} S_k\}, \end{aligned}$$

where $\widetilde{M} := X'_1 \cup K \cup K'$. Then $\Gamma' \neq \emptyset$ since $X'_1 \neq \emptyset$. We claim Γ' is a proper subset of Γ. To see this, let $i \in X_1 \backslash X'_1 \neq \emptyset$. Since $X \cap (W_0 \cup W_1) = \emptyset$ (Lemma 6), we have $i \notin K$. Observation 3 implies $i \notin K'$. Therefore, we have $i \notin \widetilde{M}$, i.e., $S_i \cap \Gamma' = \emptyset$, but $S_i \subseteq \Gamma$.

Note that $X'_1 \cap \widetilde{W}_1 = \emptyset$ (definition of $\widetilde{W}_1$ and Observation 3). On the other hand, since $|X'_1| = |\widetilde{W}_1|$, we can assume there is a one-to-one correspondence between the nodes (i.e., servers) of the two sets X'_1 and $\widetilde{W}_1$. Now let us define a new coalitional deviation Δ' of Γ', which is the same as Δ restricted on Γ' except that, if $J_j \in \Gamma'$ migrates in Δ to a server of $\widetilde{W}_1$, then let J_j migrate in Δ' to the corresponding server of X'_1.

We show that the improvement ratio of any job deviation in Δ' is at least the same as that in Δ, which then implies that $\mathrm{IR}(\Delta') \geq \mathrm{IR}(\Delta)$, contradicting the minimality of $G = G(\Delta)$ according to Definition 1. To this end, we only need to show that the new coalitional deviation Δ' takes place among the servers assigned with jobs of the coalition Γ', that is,

$$Q^+(\widetilde{M}) \subseteq W \cup K' = \widetilde{W}_1 \cup K \cup K', \tag{7}$$

so that benefit of any job deviation will not decrease due to the fact that *all* jobs on servers of X'_1 migrate out in Δ and hence in Δ' as well, leaving empty space for deviational jobs under Δ', which originally migrate to servers of $\widetilde{W}_1$ under Δ.

First we have $Q^+(K) \subseteq \widetilde{W}_1$ according to Lemma 6. On the other hand, it can be easily verified that $Q^+(X'_1) \subseteq W \bigcup K'$ according to the definition of K'. Now we show $Q^+(K') \subseteq W$, which then implies (7). In fact, for any $i \in K'$, noticing that $X'_1 \subseteq X_1$, according to the definitions of K' and X_1, we have $i \in Q^+(X_{11}) \backslash W$, which implies that $Q^+(i) \in W$ according to the definition of X_{11}. □

Since $\widetilde{Q}^+(i) \backslash W \neq \emptyset$ for any $i \in X_2$ according to the definition of X_2, we immediately have the following lemma thanks to Observation 3.

Lemma 8 *Let* $Y_2 := \bigcup_{i \in X_2} \widetilde{Q}^+(i) \backslash W$. *Then* $Y_2 \subseteq M'' \backslash W$ *and* $|X_2| \leq |Y_2|$.

Note that $|\widetilde{Q}^+(\ell)| = 2$ for any $\ell \in X_3$ and $\widetilde{Q}^+(i) \cap \widetilde{Q}^+(j) = \emptyset$ $(i \neq j)$ due to (1), which lead to the following lemma.

Lemma 9 *Let* $Y_3 := \bigcup_{\ell \in X_3} \widetilde{Q}^+(\ell)$. *Then* $Y_3 \subseteq M'' \backslash W$ *and* $2|X_3| \leq |Y_3|$.

The following lemma follows directly from the definition of X_4:

Lemma 10 *Let* $Y_4 := \bigcup_{\ell \in X_4} \widetilde{Q}^+(\ell) \backslash W$. *Then* $Y_4 \subseteq M'' \backslash W$ *and* $|X_4| \leq |Y_4|$. *For any* $j \in Y_4$, $b_j > 1$ *and* $a_j > b_j + 1$, *unless* $b_j = 1$.

At this point, we introduce the second additional assumption about $\widetilde{A}$ without loss of generality.

Assumption 2 *Arc set* $\widetilde{A}$ *is such that it* first *satisfies Assumption 1 and* then *minimizes* $|M_2^2(\widetilde{A})|$.

For any $\ell \in X_5$, since $\delta^-(\ell_1) > 1$ according to the definition of X_5, there is $j \in Q^-(\ell_1)\backslash\{\ell\}$. Then $j \notin W_0 \cup W_1$ (otherwise we would have $\ell_1 \in W$ according to Lemma 6). In fact, node j has the following property:

$$j \in M_2^1 \cap M' \backslash W. \tag{8}$$

To see this, consider replacing (ℓ, ℓ_1) with (j, ℓ_1) in $\widetilde{A}$ to form a new tilde-valid arc set $\widetilde{A}'$. It is easy to see that $\widetilde{A}'$ satisfies Assumption 1. However, with the new arc set $\widetilde{A}'$, ℓ is no longer a node in the new $M_2^2(\widetilde{A}')$, which implies that j has to become a node in $M_2^2(\widetilde{A}')$ in order not to contradict Assumption 2 with the original choice of $\widetilde{A}$, which in turn implies properties (8). Furthermore, since $j \in M_2^1$ and $(j, \ell_1) \in A\backslash\widetilde{A}$, there is no $k \neq \ell_1$ such that $(j,k) \in A\backslash\widetilde{A}$, which implies that $j \notin Q^{-1}(\widetilde{Q}^+(\ell'))\backslash\{\ell'\}$. Consequently, we have the following lemma.

Lemma 11 *Let* $Y_5 := \bigcup_{\ell \in X_5} Q^-(\widetilde{Q}^+(\ell)\backslash W)\backslash\{\ell\}$. *Then* $Y_5 \subseteq M_2^1 \cap M'\backslash W$ *and* $|X_5| \leq |Y_5|$.

3.3.2 Part 2

The following two structures in graph $G(\Delta)$ with tilde-valid arc set $\widetilde{A}$ play an important role in deriving our next lemmas:

$$\begin{aligned}
\Omega(\widetilde{A}) \;&:=\; \{i \in M' : \ i_1 = \widetilde{Q}^{-1}(i) \in \widetilde{Q}^+(i), \\
&\qquad \delta^-(i_1) = 1, \ \delta^+(i_1) = b_{i_1} > 1\}; \\
\Pi(\widetilde{A}) \;&:=\; \{(i, i_1, j) : \ i \in \widetilde{M_2^2}\backslash X_3, \ i_1 \in \widetilde{Q}^+(\ell) \cap \widetilde{Q}^-(j)\backslash W, \\
&\qquad i \neq j, \ \delta^-(i_1) = 1, \ \delta^+(i_1) = b_{i_1} > 1, \ j \in M'\}.
\end{aligned}$$

Note that each element in $\Omega(\widetilde{A})$ *represents* a directed 2-cycles of both arcs in $\widetilde{A}$ and each element in $\Pi(\widetilde{A})$ *is* a directed 2-path of both arcs in $\widetilde{A}$. In both cases of $\Omega(\widetilde{A})$ and $\Pi(\widetilde{A})$, the *interior* node i_1 has an in-degree $\delta^-(i_1) = 1$ and all its out-arcs are in $\widetilde{A}$. Our next result is based on the following further refinement of the tilde-valid arc set $\widetilde{A}$.

Lemma 12 *If* $\Omega(\widetilde{A}) \neq \emptyset$ *for some arc set* $\widetilde{A}$ *satisfying Assumption 2, then there exists an arc set* $\widetilde{A}'$ *such that, while it also satisfies Assumption 2, additionally,* $\Omega(\widetilde{A}')$ *is a* proper *subset of* $\Omega(\widetilde{A})$.

Proof. Assume $i \in \Omega(\widetilde{A})$ and let $i_1 = \widetilde{Q}^{-1}(i)$ be as in the definition of $\Omega(\widetilde{A})$. Then there must be a node $h \in Q^-(i)$ with $h \neq i_1$, since otherwise $\delta^-(i) = \delta^-(i_1) = 1$, which implies that there would be no directed path from any other nodes in $G(\Delta)$ to nodes i or i_1, contradicting Lemma 4. Therefore, the following set is not empty:

$$H_i := \{h \in M : \ (h,i) \in A, \ \text{either } \delta^-(h) > 1 \text{ or } (h,i) \notin \widetilde{A}\}. \tag{9}$$

Let $h \in H_i \neq \emptyset$. We define a new tilde-valid arc set

$$\widetilde{A}' := \{\widetilde{A} \backslash \{(i_1, i)\}\} \cup \{(h, i)\}. \tag{10}$$

It is easily seen that $i \in \Omega(\widetilde{A}) \backslash \Omega(\widetilde{A}')$ and $\Omega(\widetilde{A}') \cup \{i\} = \Omega(\widetilde{A})$. On the other hand, $\widetilde{A}'$ still satisfies Assumption 1 due to $b_{i_1} > 1$, and hence also satisfies Assumption 2 since $\delta^+(h) < a_h$ (which implies that $h \notin \Omega(\widetilde{A}) \cup \Omega(\widetilde{A}')$) according to Observation 3 (as no other node not in $M_2^2(\widetilde{A})$ can possibly become a member of $M_2^2(\widetilde{A}')$). $\square$

As a result of Lemma 12, we further refine our initial choice of $\widetilde{A}$ so that it satisfies the following assumption.

Assumption 3 *Arc set $\widetilde{A}$ is such that it first satisfies Assumption 2 and then lexicographically minimizes* $(|\Omega(\widetilde{A})|, |\Pi(\widetilde{A})|)$.

Corollary 13 *Any arc set $\widetilde{A}$ satisfying Assumption 3 must satisfy* $\Omega(\widetilde{A}) = \emptyset$. $\square$

An arc set $\widetilde{A}$ in graph $G(\Delta)$ that satisfies Assumption 3 is said to be *derived* from Δ. Without loss of generality, our coalitional deviation Δ is considered to have been chosen so that it satisfies the following assumption.

Assumption 4 *Coalitional deviation Δ defining minimal deviation graph $G(\Delta)$ is such that the arc set $\widetilde{A}$ derived from Δ gives lexicographical minimum* $V(\Delta) := (|W_0(\widetilde{A})|, |M_2^2(\widetilde{A})|, |\Omega(\widetilde{A})|, |\Pi(\widetilde{A})|)$.

Lemma 14 *Let minimal deviation graph $G(\Delta)$ with Δ satisfying Assumption 4 be given. For any $\ell \in X_6$, there is $j \neq \ell$, which we shall call a* company *of ℓ, such that $(\widetilde{Q}^-(\ell), j) \in \widetilde{A}$ and $j \in M'' \backslash W$.*

Proof. Given $\ell \in X_6$ and $\ell_1 = Q^+(\ell) \backslash W$. Since $a_{\ell_1} = b_{\ell_1} + 1$ according to the definition of X_6, we have $\delta^+(\ell_1) = b_{\ell_1}$ and hence $\widetilde{Q}^+(\ell_1) = Q^+(\ell_1)$ since $a_{\ell_1} = \delta^+(\ell_1) + 1$ according to Observation 3. Since $b_{\ell_1} > 1$ and $\widetilde{Q}^+(\ell_1) = Q^+(\ell_1) \not\subseteq W$ (again according to the definition of X_6), we let $j \in \widetilde{Q}^+(\ell_1) \backslash W$. Then $j \neq \ell$ since otherwise we would have $\ell \in \Omega(\widetilde{A})$, contracting Corollary 13 with our Assumption 3.

We claim $j \in M''$ and hence are done. Let us assume for a contradiction that $j \in M'$. Note that with $\{\ell, \ell_1, j\}$ replacing $\{i, i_1, j\}$ in the definition of $\Pi(\widetilde{A})$, we conclude that $\ell \in \Pi(\widetilde{A})$. Now let us define a new coalitional deviation Δ' so that its derived arc set $\widetilde{A}'$ gives a $V(\Delta') = (|W_0(\widetilde{A}')|, |M_2^2(\widetilde{A}')|, |\Omega(\widetilde{A}')|, |\Pi(\widetilde{A}')|)$ that is lexicographically smaller than $V(\Delta) = (|W_0(\widetilde{A})|, |M_2^2(\widetilde{A})|, |\Omega(\widetilde{A})|, |\Pi(\widetilde{A})|)$, a desired contraction to Assumption 4.

In fact, let Δ' be defined as in Observation 4 after node i has been replaced by ℓ in the statement of Observation 4. Denote A' as the arc set of the resulting minimal deviation graph $G(\Delta')$. Let $\widetilde{A}'$ be the natural result of $\widetilde{A}$ after the re-orientation from Δ and Δ', i.e., an arc in $\widetilde{A}$ pointing to ℓ (resp. j) will become an arc in $\widetilde{A}'$ pointing to j (resp. ℓ). Other arcs are the same for $\widetilde{A}$ and $\widetilde{A}'$. Apparently,

$$|W_0(\widetilde{A}')| = |W_0(\widetilde{A})|, \ |M_2^2(\widetilde{A}')| = |M_2^2(\widetilde{A})|.$$

On the other hand, if the value of $|\Omega(\widetilde{A}')|$ has increased from $|\Omega(\widetilde{A})|$, then clearly it must be the result of ℓ and/or j becoming element(s) of $\Omega(\widetilde{A}')$. In any such case (say, the former case for the sake of argument), based on the definition of $\Omega(\widetilde{A}')$, we can use the approach in Lemma 12 to find $h \in H_\ell$ as defined in (9) and perform an arc-swap as in (10) with i and i_1 replaced by ℓ and ℓ_1, respectively, to reduce $|\Omega(\widetilde{A}')|$ while maintaining the values of $|W_0(\widetilde{A}')|$ and $|M_2^2(\widetilde{A}')|$. For convenience, we still use $\widetilde{A}'$ to denote the tilde-valid arc set after such arc-swap(s) if needed. Consequently, we have

$$\Omega(\widetilde{A}') = \Omega(\widetilde{A}) = \emptyset.$$

However, we claim

$$|\Pi(\widetilde{A}')| < |\Pi(\widetilde{A})|, \tag{11}$$

a desired contradiction. To see inequality (11), we first note that (i) any 2-path in $\Pi(\widetilde{A})$ starting at $i \neq \ell, j$ is also a 2-path in $\Pi(\widetilde{A}')$, and vice versa, and (ii) any 2-path in $\Pi(\widetilde{A})$ (resp. $\Pi(\widetilde{A}')$) starting at ℓ (resp. j) must have the first arc (ℓ, ℓ_1) (resp. $(j, \widetilde{Q}^+(j)\backslash W)$, since $|\widetilde{Q}^+(j)\backslash W| = 1$ due to $j \in \widetilde{M_2^2}\backslash X_3$). On the other hand, the following can be easily observed:

1. If $(\ell, \ell_1, j') \in \Pi(\widetilde{A})$ $(j' \neq j)$, then $(\ell, \ell_1, j') \in \Pi(\widetilde{A}')$, and vice versa.
2. If $(j, j_1, j') \in \Pi(\widetilde{A})$ $(j' \neq \ell)$, then $(j, j_1, j') \in \Pi(\widetilde{A}')$, and vice versa.
3. $(\ell, \ell_1, j) \in \Pi(\widetilde{A})\backslash\Pi(\widetilde{A}')$, since $(\ell, \ell_1, j) \in \Pi(\widetilde{A}')$ would imply $(\ell_1, j) \in \widetilde{A}' \subseteq A'$ by definition of $\Pi(\widetilde{A}')$ and hence $(\ell_1, \ell) \in A$ by definition of A', which in turn implies that $(\ell_1, \ell) \in \widetilde{A}$ since $b_{\ell_1} = \delta^+(\ell_1)$ under $\widetilde{A}$. Consequently, we obtain $\ell \in \Omega(\widetilde{A})$, contradicting Corollary 13.
4. With similar reasons for $(\ell, \ell_1, j) \notin \Pi(\widetilde{A}')$, we have $(j, j_1, \ell) \notin \Pi(\widetilde{A}')$.

Therefore, overall $\Pi(\widetilde{A}')$ contains at least one element less than $\Pi(\widetilde{A})$ as indicated in points 3 and 4 above. □

Corollary 15 *Let* $X_{61} := \{j \in M\backslash X :$ *node* j *is a company of* $\ell\})$ *and* $Y_6 := \bigcup_{\ell \in X_6}((\widetilde{Q}^+(\ell)\backslash W) \cup X_{61}$. *Then* $Y_6 \subseteq M''\backslash W$ *and* $2|X_6| \leq |Y_6|$.

Proof. For any $\ell \in X_6$, if j is a company of ℓ, then j cannot be a company of $\ell' \in X_6$, $\ell' \neq \ell$, because of the uniqueness of $\widetilde{Q}^-(\ell)$, $\widetilde{Q}^-(\ell')$ and $\widetilde{Q}^-(j)$, which implies the inequality $2|X_6| \leq |Y_6|$. □

Now let us look at the six sets $Y_1, \ldots, Y_6$, defined in this subsection. According to Lemmas 7–11 and Corollary 15, we have

$$\begin{aligned} &Y_1 = W_1 \subseteq W\backslash X;\ Y_5 \subseteq M_2^1 \cap M'\backslash W \subseteq M\backslash X; \\ &Y_2, Y_3, Y_4, Y_6 \subseteq M''\backslash W \subseteq M\backslash X; \end{aligned}$$

and

$$\begin{aligned} &Y_t \subseteq \widetilde{Q}^+(X_t), t = 2, 3, 4; \\ &Y_6 \subseteq \widetilde{Q}^+(X_6) \cup \widetilde{Q}^+(M\backslash X). \end{aligned}$$

Consequently, noticing $\widetilde{Q}^+(j) \cap \widetilde{Q}^+(j) = \emptyset$, we conclude that

$$Y_t \cap X = \emptyset \text{ and } Y_t \cap Y_s = \emptyset,\ s \neq t,\ s, t \in \{1, \ldots, 6\}. \tag{12}$$

4 Establishment of Strong Stability

We are now ready to enter the final stage of establishing the desired approximability as stated in the following theorem.

Theorem 1 *For any minimal deviation graph $G(\Delta_m)$ involving m servers, its improvement ratio $IR(\Delta_m) \leq 5/4$.*

Proof. Let $r = \mathrm{IR}(\Delta_m)$. Denote $m_a^b = |M_a^b|$ for all possible pairs a and b: $2 \leq a \leq m$ and $0 \leq b \leq a$. Then according to (1) and (6), we have

$$\sum_{a=2}^{m}\sum_{b=0}^{a} m_a^b = m, \quad \text{and} \quad \sum_{a=2}^{m}\sum_{b=0}^{a} b\, m_a^b = m. \tag{13}$$

According to the definition of IR, we have $rL'_j \leq L_i$. Summing up these inequalities over all m arcs in $\widetilde{A}$ leads to

$$\sum_{j=1}^{m} rL'_j \leq \sum_{i=1}^{m} b_i L_i,$$

which implies that

$$r \leq \frac{\sum_{i=1}^{m} b_i L_i}{\sum_{i=1}^{m} L_i}. \tag{14}$$

According to Observations 1 and 2, $1 \leq L_i \leq a_i/(a_i-1) \leq 2$, which together with the definition of IR, implies that the right-hand side of (14) is between 1 and 2, which in turn implies that it is a decreasing function of L_i for which $b_i = 0$ or $b_i = 1$, and an increasing function of L_i for which $b_i \geq 2$. Therefore,

$$r \leq \frac{\sum_{a=2}^{m}\sum_{b=2}^{a} \frac{ab}{a-1} m_a^b + \sum_{a=2}^{m} m_a^1}{\sum_{a=2}^{m}\sum_{b=2}^{a} \frac{a}{a-1} m_a^b + \sum_{a=2}^{m} m_a^1 + \sum_{a=2}^{m} m_a^0},$$

which together with (13) implies that

$$r \leq \frac{\sum_{a=2}^{m}\sum_{b=2}^{a} \frac{b}{a-1} m_a^b + m}{\sum_{a=2}^{m}\sum_{b=2}^{a} \frac{1}{a-1} m_a^b + m}.$$

To show $r \leq 5/4$, it suffices to show

$$\frac{\sum_{a=2}^{m}\sum_{b=2}^{a} \frac{b}{a-1} m_a^b + m}{\sum_{a=2}^{m}\sum_{b=2}^{a} \frac{1}{a-1} m_a^b + m} \leq \frac{5}{4},$$

which is equivalent to

$$\sum_{a=2}^{m}\sum_{b=2}^{a} \frac{4b-5}{a-1} m_a^b \leq m,$$

or (due to (13))

$$0 \leq \sum_{a=2}^{m} m_a^1 + \sum_{a=2}^{m}\sum_{b=2}^{a}\left(b - \frac{4b-5}{a-1}\right) m_a^b,$$

that is

$$m_2^2 + \frac{1}{2}m_3^3 \leq \sum_{a=2}^{m} m_a^1 + \frac{1}{2}m_3^2 + \sum_{a=4}^{m}\sum_{b=2}^{a}\left(b - \frac{4b-5}{a-1}\right) m_a^b. \tag{15}$$

In what follows, we are to prove (15) above based on our bounds derived in Section 3.3. Since $X_2 \subseteq M_3^3$, the left-hand side of inequality (15) is at most

$$|X| - \frac{1}{2}|M_3^3| \leq \sum_{t=1}^{6} |X_t| - \frac{1}{2}|X_2|. \tag{16}$$

On the other hand, if we let

$$Y_t' := \{i \in Y_t : \; b_i = 1\} \;\text{ and }\; Y_t'' := Y_t \backslash Y_t', \text{ for } t = 2, 3, 4, 6,$$

which imply

$$\begin{aligned} Y_4'' = \{i \in Y_4 : \; b_i \geq 2, a_i \geq b_i + 2\} &\subseteq \bigcup_{2\leq b\leq a-2} M_a^b, \\ \bigcup_{t\in\{2,3,4,6\}} Y_t' \cup Y_1 \cup Y_5 &\subseteq \bigcup_{a\geq 2} M_a^1, \\ \bigcup_{t\in\{2,3,6\}} Y_t'' &\subseteq \bigcup_{2\leq b<a} M_a^b, \end{aligned}$$

then noticing the properties (12) and that

$$b - \frac{4b-5}{a-1} \geq \begin{cases} \frac{1}{2}, & \text{if } 2 \leq b < a; \\ 1, & \text{if } 2 \leq b \leq a-2, \end{cases}$$

we see that the right-hand side of inequality (15) is at least

$$\begin{aligned} &|Y_1| + |Y_5| + \sum_{t\in\{2,3,4,6\}} |Y_t'| + \frac{1}{2}\sum_{t\in\{2,3,6\}} |Y_t''| + |Y_4''| \\ &\qquad \geq |Y_1| + \frac{1}{2}|Y_2| + \frac{1}{2}|Y_3| + |Y_4| + |Y_5| + \frac{1}{2}|Y_6|, \end{aligned}$$

which is at least the right-hand side of inequality (16) according to Lemmas 7–11 and Corollary 15, which in turn ultimately leads to inequality (15). □

With Theorem 1 and the simple example of a coalitional deviation Δ_3 involving $m = 3$ servers with $\mathrm{IR}(\Delta_3) = 5/4$ presented in [7], we establish the following theorem.

Theorem 2 *In the m-server load balancing game ($m \geq 3$), any NE is a $(5/4)$-approximate SNE and the bound is tight.*

5 Concluding Remarks

By establishing a tight bound of 5/4 for the approximation of general NEs to SNEs in the m-server load balancing game for $m \geq 3$, we have closed the final gap for the study of approximation of general NEs to SNEs. However, as demonstrated by Feldman & Tamir [7] and by Chen [4], a special subset of NEs known as LPT assignments, which can been easily identified as NEs [9], do approximate SNEs better than general NEs. It is still a challenge to provide a tight approximation bound for this subset of NEs.

References

[1] Albers S., On the value of coordination in network design. *SIAM Journal on Computing* 38(6) (2009), 2273–2302.

[2] Andelman N., Feldman M., andMansour Y., Strong price of anarchy. In: *Proc. of 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, 2007, 189–198.

[3] Aumann R., Acceptable Points in General Cooperative n-Person Games. In: *Contributions to the Theory of Games IV, Annals of Mathematics* 40 (eds. R.D. Luce, A.W. Tucker), 1959, 287–324.

[4] Chen B., Equilibria in Load Balancing Games. *Acta Mathematica Applicatae Sinica* (English Series), 25(4) (2009), 723–736.

[5] Chen B. and Gürel S., Efficiency analysis of load balancing games with and without activation costs. *Journal of Scheduling* 15(2) (2012), 157–164.

[6] Czumaj A. and Vöcking B., Tight bounds for worst-case equilibria. *Proc. 13th Annual ACM-SIAM Symposium on Discrete Algorithms*, 413–420, 2002.

[7] Feldman M. and Tamir T., Approximate strong equilibrium in job scheduling games. *Journal of Artificial Intelligence Research* 36 (2009), 387–414.

[8] Finn G. and Horowitz E., A linear time approximation algorithm for multiprocessor scheduling. *BIT* 19(3) (1979), 312–320.

[9] Fotakis D., Kontogiannis S., Mavronicolas M., and Spiraklis P., The structure and complexity of Nash equilibria for a selfish routing game. *Proc. of the 29th International Colloquium on Automata, Languages and Programming*, 2002, 510–519.

[10] Koutsoupias E., Mavronicolas M., and Spirakis P., Approximate equilibria and ball fusion. *Theory of Computing Systems* 36(6) (2003), 683–693.

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

[12] Vöcking B., Selfish Load Balancing. *Algorithmic Game Theory* (eds.: Nisan N., Roughgarden T., Tardos É., and Vazirani V.V.), 517–542.