# Stability of the bipartite matching model

Ana Bušić

INRIA/ENS, Paris, France

joint work with Varun Gupta (Carnegie Mellon University) and
Jean Mairesse (CNRS/Université Paris 7)

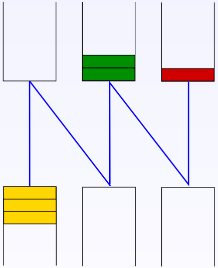
Edinburgh, June 2010

# Bipartite Matching Model

*Model with multiclass flavor but with customers/servers playing completely symmetrical roles.*

# Bipartite Matching Model

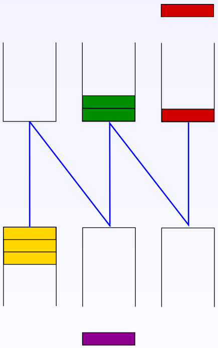
*Model with multiclass flavor but with customers/servers playing completely symmetrical roles.*



- Both “customers” and “servers” arrive into the system and depart from it.
- Discrete time: at each time step there is one customer and one server that arrive into the system, independently of the past.
- Instantaneous matchings according to a bipartite matching graph. Customers/servers that cannot be matched are stored in a buffer.

# Bipartite Matching Model

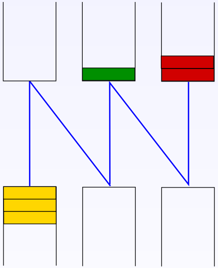
*Model with multiclass flavor but with customers/servers playing completely symmetrical roles.*



- Both "customers" and "servers" arrive into the system and depart from it.
- Discrete time: at each time step there is one customer and one server that arrive into the system, independently of the past.
- Instantaneous matchings according to a bipartite matching graph. Customers/servers that cannot be matched are stored in a buffer.

# Bipartite Matching Model

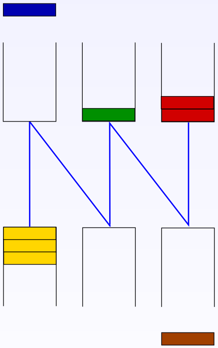
*Model with multiclass flavor but with customers/servers playing completely symmetrical roles.*



- Both "customers" and "servers" arrive into the system and depart from it.
- Discrete time: at each time step there is one customer and one server that arrive into the system, independently of the past.
- Instantaneous matchings according to a bipartite matching graph. Customers/servers that cannot be matched are stored in a buffer.

# Bipartite Matching Model

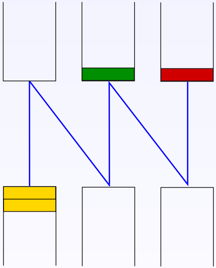
*Model with multiclass flavor but with customers/servers playing completely symmetrical roles.*



- Both "customers" and "servers" arrive into the system and depart from it.
- Discrete time: at each time step there is one customer and one server that arrive into the system, independently of the past.
- Instantaneous matchings according to a bipartite matching graph. Customers/servers that cannot be matched are stored in a buffer.

# Bipartite Matching Model

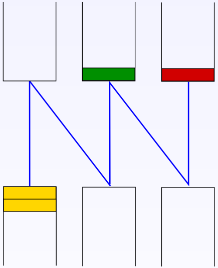
*Model with multiclass flavor but with customers/servers playing completely symmetrical roles.*



- Both “customers” and “servers” arrive into the system and depart from it.
- Discrete time: at each time step there is one customer and one server that arrive into the system, independently of the past.
- Instantaneous matchings according to a bipartite matching graph. Customers/servers that cannot be matched are stored in a buffer.

# Bipartite Matching Model

*Model with multiclass flavor but with customers/servers playing completely symmetrical roles.*



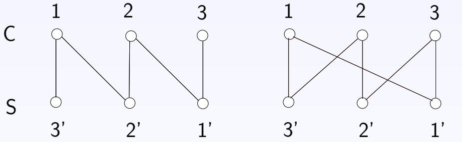
- Both "customers" and "servers" arrive into the system and depart from it.
- Discrete time: at each time step there is one customer and one server that arrive into the system, independently of the past.
- Instantaneous matchings according to a bipartite matching graph. Customers/servers that cannot be matched are stored in a buffer.

Fully specified by: a matching graph, a joint probability measure $\mu$ for arrivals of customers/servers and a matching policy.

# Bipartite Matching Model

Def. A bipartite matching structure is a quadruple $(C, S, E, F)$ where:

- $C$ (resp. $S$) finite set of customer (resp. server) types;
- $E \subset C \times S$ is the set of possible matchings;
- $F \subset C \times S$ is the set of possible arrivals.

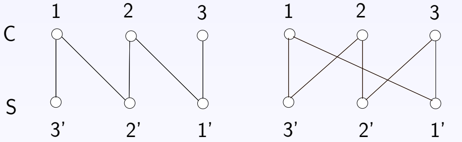



matching graph $(C, S, E)$ arrival graph $(C, S, F)$

# Bipartite Matching Model

Def. A bipartite matching structure is a quadruple $(C, S, E, F)$ where:

- $C$ (resp. $S$) finite set of customer (resp. server) types;
- $E \subset C \times S$ is the set of possible matchings;
- $F \subset C \times S$ is the set of possible arrivals.




matching graph $(C, S, E)$ arrival graph $(C, S, F)$

Def. A bipartite matching model is a triple $[(C, S, E, F), \mu, \textsc{Pol}]$ where $\mathrm{supp}(\mu) = F$, the marginals of $\mu$ satisfy: $\mathrm{supp}(\mu_C) = C$, $\mathrm{supp}(\mu_S) = S$, and $\textsc{Pol}$ is an admissible policy.

# Bipartite Matching Model

A matching policy $\textsc{Pol}$ is admissible if:

- Only the current state of the buffer is taken into account;
- Buffer-first assumption: priority is given to customers/servers that are already present in the buffer.

# Bipartite Matching Model

A matching policy POL is admissible if:

- Only the current state of the buffer is taken into account;
- Buffer-first assumption: priority is given to customers/servers that are already present in the buffer.

Discrete time Markov chain on:

- commutative (ex. Match the Longest, Random, Priorities)

$$\left\{(x,y) \in \mathbb{N}^C \times \mathbb{N}^S \ : \ \sum_{c\in C} x_c = \sum_{s\in S} y_s;\ \forall (c,s) \in E,\ x_c y_s = 0\right\} \quad (1)$$

- or non-commutative state space (ex. FIFO):

$$\left\{(u,v) \in \cup_{k\geq 0}(C^k \times S^k) \ : \ ([u],[v]) \text{ belongs to } (1)\right\},$$

where $[u]$ is the commutative image of $u$.

# Previous results

Introduced by Caldentey, Kaplan, and Weiss (2009):
FIFO and $\mu = \mu_C \times \mu_S$.

Related models:

- Constrained queueing networks (Tassiulas and Ephremides, 1992).
- Input-queued crossbar switches (McKeown, Mekkittikul, Anantharam, and Walrand, 1999).
- Models for call centers with "skills-based routing" (Gans, Koole, and Mandelbaum, 2003).

# Necessary stability conditions

Def. The model is said to be stable if the Markov chain has a unique and attractive stationary probability measure
(i.e. measure $\pi$ such that $\pi P = \pi$ and $\frac{1}{n}\sum_{k=1}^{n} \nu P^k$ converges weakly to $\pi$ for any initial measure $\nu$, where $P$ is the transition probability matrix).

For a matching graph $(C, S, E)$ we denote:

$$C(s) = \{c \in C \ : \ (c, s) \in E\}, \qquad S(c) = \{s \in S \ : \ (c, s) \in E\}.$$

Necessary conditions: If the model is stable then the marginals of $\mu$ satisfy

$$\textsc{NCond}: \quad \begin{cases} \mu_C(U) < \mu_S(S(U)), & \forall U \subsetneq C \\ \mu_S(V) < \mu_C(C(V)), & \forall V \subsetneq S \end{cases}$$

# Necessary stability conditions

Def. The model is said to be stable if the Markov chain has a unique and attractive stationary probability measure
(i.e. measure $\pi$ such that $\pi P = \pi$ and $\frac{1}{n}\sum_{k=1}^{n} \nu P^k$ converges weakly to $\pi$ for any initial measure $\nu$, where $P$ is the transition probability matrix).

For a matching graph $(C, S, E)$ we denote:

$$C(s) = \{c \in C \ : \ (c, s) \in E\}, \qquad S(c) = \{s \in S \ : \ (c, s) \in E\} \, .$$

Necessary conditions: If the model is stable then the marginals of $\mu$ satisfy

$$\textrm{NCond}: \quad \begin{cases} \mu_C(U) < \mu_S(S(U)), & \forall U \subsetneq C \\ \mu_S(V) < \mu_C(C(V)), & \forall V \subsetneq S \end{cases}$$

Prop. Given $[(C, S, E), \mu]$, there exists an algorithm of time complexity $O((|C| + |S|)^3)$ to decide if NCond is satisfied.

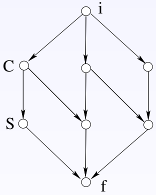
# Proof

Proof using network flow arguments:

$$\mathcal{N} = \left(C \cup S \cup \{i, f\}, E \cup \{(i, c), c \in C\} \cup \{(s, f), s \in S\}\right).$$

Lemma.

1. There exists a flow of value 1 in $\mathcal{N}$ iff $(\mu_C, \mu_S)$ satisfies $\text{NCOND}_{\leq}$ ($<$ replaced by $\leq$ in $\text{NCOND}$).
2. There exists a flow $T$ of value 1 such that $T(c, s) > 0$ for all $(c, s) \in E$ iff $(\mu_C, \mu_S)$ satisfies $\text{NCOND}$.



## Proof of Lemma 2

$\Rightarrow$ Follows easily from connectedness of the matching graph.

$\Leftarrow$ Fix $\eta$ such that $0 < \eta < 1/|E|$. A strictly positive flow of value $|E|\eta$:

$$T_\eta(x,y) = \begin{cases} \eta & \text{for } (x,y) = (c,s) \in E \\ |S(c)|\ \eta & \text{for } (x,y) = (i,c) \\ |C(s)|\ \eta & \text{for } (x,y) = (s,f)\,. \end{cases}$$

Define: $\widetilde{\mu}_C(c) = \frac{\mu_C(c) - |S(c)|\eta}{1-|E|\eta}, \quad \widetilde{\mu}_S(s) = \frac{\mu_S(s) - |C(s)|\eta}{1-|E|\eta}\,.$

For $\eta$ small enough, $\widetilde{\mu}_C$, $\widetilde{\mu}_S$ are probability measures satisfying $\mathrm{NC{\small OND}}$.

For $\widetilde{\mu}_C$, $\widetilde{\mu}_S$ there exists a flow $\widetilde{T}$ of value 1.

A strictly positive flow of value 1: $T = T_\eta + (1 - |E|\eta)\widetilde{T}\,.$

## Proof of Lemma 2

$\Rightarrow$ Follows easily from connectedness of the matching graph.

$\Leftarrow$ Fix $\eta$ such that $0 < \eta < 1/|E|$. A strictly positive flow of value $|E|\eta$:

$$T_\eta(x,y) = \begin{cases} \eta & \text{for } (x,y) = (c,s) \in E \\ |S(c)|\ \eta & \text{for } (x,y) = (i,c) \\ |C(s)|\ \eta & \text{for } (x,y) = (s,f)\,. \end{cases}$$

Define: $\widetilde{\mu}_C(c) = \frac{\mu_C(c) - |S(c)|\eta}{1-|E|\eta}, \quad \widetilde{\mu}_S(s) = \frac{\mu_S(s) - |C(s)|\eta}{1-|E|\eta}$ .

For $\eta$ small enough, $\widetilde{\mu}_C$, $\widetilde{\mu}_S$ are probability measures satisfying $\mathrm{NCond}$.

For $\widetilde{\mu}_C$, $\widetilde{\mu}_S$ there exists a flow $\widetilde{T}$ of value 1.

A strictly positive flow of value 1: $T = T_\eta + (1 - |E|\eta)\widetilde{T}$ .

Verification algorithm: The pair $(\mu_C, \mu_S)$ satisfies $\mathrm{NCond}$ iff the pair $(\widetilde{\mu}_C, \widetilde{\mu}_S)$ satisfies $\mathrm{NCond}$ for $\eta$ strictly positive and small enough.

Run $\mathrm{MaxFlow}$ on the input $(\mathcal{N}, \widetilde{\mu}_C, \widetilde{\mu}_S)$ by considering $\eta$ as a formal parameter "as small as needed".

# Stable structures

Def. A bipartite matching structure $(C, S, E, F)$ is stable if there exists a probability measure $\mu$ such that:

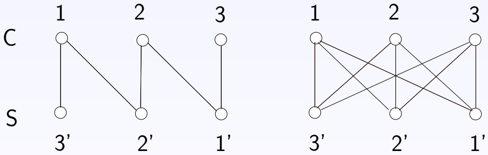
- $\text{supp}(\mu) = F$, $\text{supp}(\mu_C) = C$, $\text{supp}(\mu_S) = S$.
- $\mu_C$ and $\mu_S$ satisfy $\textsc{NCond}$.

# Stable structures

Def. A bipartite matching structure $(C, S, E, F)$ is stable if there exists a probability measure $\mu$ such that:

- $\text{supp}(\mu) = F$, $\text{supp}(\mu_C) = C$, $\text{supp}(\mu_S) = S$.
- $\mu_C$ and $\mu_S$ satisfy NCOND.

Example 1.



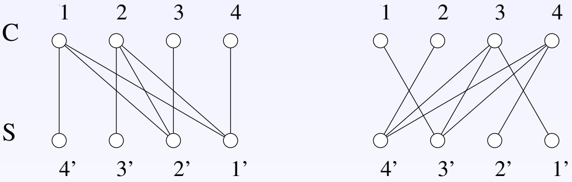

matching graph $(C, S, E)$ arrival graph $(C, S, F)$

Consider the product measure $\mu = \mu_C \times \mu_S$ with:
$\mu_C$ : $\mu_C(1) = \mu_C(2) = 2/5$, $\mu_C(3) = 1/5$,
$\mu_S$ : $\mu_S(1') = \mu_S(2') = 2/5$, $\mu_S(3') = 1/5$ .

Then $(\mu_C, \mu_S)$ satisfy NCOND.

# Stable structures

Example 2.




Consider any $\mu$ with $\mathrm{supp}(\mu) = F$. We have

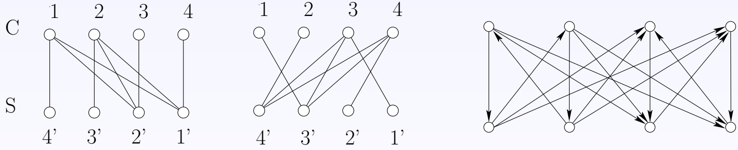
$$\mu_S(\{1', 2'\}) = \mu(3, 1') + \mu(4, 2') \leq \mu_C(\{3, 4\}),$$

which contradicts $\mathrm{NC{\small OND}}$ for $U = \{3, 4\}$.

# Connectivity properties

Consider a bipartite matching structure $(C, S, E, F)$. Associated directed graph: the nodes are $C \cup S$ and the arcs are

$$c \longrightarrow s, \quad \text{if } (c, s) \in E, \qquad s \longrightarrow c, \quad \text{if } (c, s) \in F\ .$$




matching graph $(C, S, E)$ arrival graph $(C, S, F)$ associated directed graph

# Connectivity properties

Thm. For a bipartite matching structure $(C, S, E, F)$ the following properties are equivalent:

1. There exists $\mu$ such that $\text{supp}(\mu) = F$, $\text{supp}(\mu_C) = C$, $\text{supp}(\mu_S) = S$ and $\mu$ satisfies $\textsc{NCond}$.
2. The associated directed graph is strongly connected.

# Connectivity properties

Thm. For a bipartite matching structure $(C, S, E, F)$ the following properties are equivalent:

1. There exists $\mu$ such that $\text{supp}(\mu) = F$, $\text{supp}(\mu_C) = C$, $\text{supp}(\mu_S) = S$ and $\mu$ satisfies $\text{NCond}$.
2. The associated directed graph is strongly connected.

Def. Property $\text{UTC}$ for the transition graph of the Markov chain: existence of a unique (terminal) strictly connected component with all states leading to it.

# Connectivity properties

Thm. For a bipartite matching structure $(C, S, E, F)$ the following properties are equivalent:

1. There exists $\mu$ such that $\text{supp}(\mu) = F$, $\text{supp}(\mu_C) = C$, $\text{supp}(\mu_S) = S$ and $\mu$ satisfies $\textsc{NCond}$.
2. The associated directed graph is strongly connected.

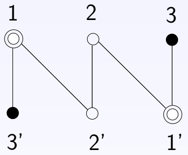
Def. Property $\mathrm{UTC}$ for the transition graph of the Markov chain: existence of a unique (terminal) strictly connected component with all states leading to it.

Thm. If the associated directed graph of $(C, S, E, F)$ is strongly connected, then any bipartite matching model $[(C, S, E, F), \mu, \textsc{Pol}]$ satisfies the property $\mathrm{UTC}$.

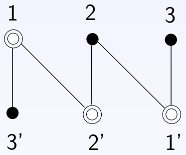
# State space decomposition

The state space can be decomposed into facets, defined only by the non-empty classes.

Def. A facet is an ordered pair $(U, V)$ such that: $U \subset C, V \subset S$ and $U \times V \subset (C \times S - E)$.




facet $(\{3\}, \{3'\})$




facet $(\{2, 3\}, \{3'\})$

For a facet $\mathcal{F} = (U, V)$, define:

$$C_\bullet(\mathcal{F}) = U, \quad C_\circledcirc(\mathcal{F}) = C(V), \quad C_\circ(\mathcal{F}) = C - (C_\bullet(\mathcal{F}) \cup C_\circledcirc(\mathcal{F}))$$
$$S_\bullet(\mathcal{F}) = V, \quad S_\circledcirc(\mathcal{F}) = S(U), \quad S_\circ(\mathcal{F}) = S - (S_\bullet(\mathcal{F}) \cup S_\circledcirc(\mathcal{F})).$$

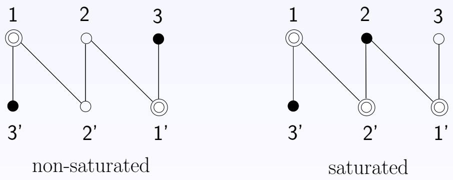
# Sufficient conditions

Conditions SCOND:

$$\mu_C(C_{\circledcirc}(\mathcal{F})) + \mu_S(S_{\circledcirc}(\mathcal{F})) \ > \ 1 - \mu(E \cap C_{\circ}(\mathcal{F}) \times S_{\circ}(\mathcal{F})), \ \forall \mathcal{F} \neq (\emptyset, \emptyset)$$

Def. A facet $\mathcal{F}$ is called saturated if $C_{\circ}(\mathcal{F}) = \emptyset$ or $S_{\circ}(\mathcal{F}) = \emptyset$.

SCOND $\Longrightarrow$ NCOND (considering only the saturated facets).




Prop. (Sufficient conditions) A bipartite model with probability $\mu$ satisfying SCOND is stable under any admissible matching policy.

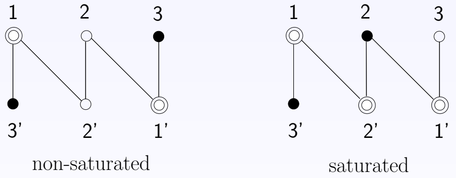
# Sufficient conditions

Conditions SCOND:

$$\mu_C(C_{\circledcirc}(\mathcal{F})) + \mu_S(S_{\circledcirc}(\mathcal{F})) \; > \; 1 - \mu(E \cap C_{\circ}(\mathcal{F}) \times S_{\circ}(\mathcal{F})), \; \forall \mathcal{F} \neq (\emptyset, \emptyset)$$

Def. A facet $\mathcal{F}$ is called saturated if $C_{\circ}(\mathcal{F}) = \emptyset$ or $S_{\circ}(\mathcal{F}) = \emptyset$.

SCOND $\Longrightarrow$ NCOND (considering only the saturated facets).




Prop. (Sufficient conditions) A bipartite model with probability $\mu$ satisfying SCOND is stable under any admissible matching policy.

Cor. Consider a bipartite graph in which any non-zero facet is saturated. For any admissible matching policy, the stability region is maximal.

# Match the Longest has maximal stability region

Match the Longest (ML) policy: a newly arriving customer of class $c$ is matched to a server in $S(c)$ with the largest buffer (similarly for newly arriving server).

Thm. For any bipartite graph, ML has a maximal stability region.

# Match the Longest has maximal stability region

Match the Longest (ML) policy: a newly arriving customer of class $c$ is matched to a server in $S(c)$ with the largest buffer (similarly for newly arriving server).

Thm. For any bipartite graph, ML has a maximal stability region.

Proof:

- Quadratic Lyapunov function: $L(x, y) = \sum_{c \in C} x_c^2 + \sum_{s \in S} y_s^2$.
- ML minimizes the value of this Lyapunov function at each step.

# Match the Longest has maximal stability region

Match the Longest (ML) policy: a newly arriving customer of class $c$ is matched to a server in $S(c)$ with the largest buffer (similarly for newly arriving server).

Thm. For any bipartite graph, ML has a maximal stability region.

Proof:

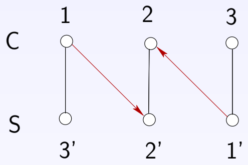
- Quadratic Lyapunov function: $L(x, y) = \sum_{c \in C} x_c^2 + \sum_{s \in S} y_s^2$.
- ML minimizes the value of this Lyapunov function at each step.
- Facet-dependent alternate policy. In a non-zero facet $\mathcal{F}$: the server $s \in S_{\circledcirc}(\mathcal{F})$ is matched to $c \in C_{\bullet}(\mathcal{F}) \cap C(s)$ with probability $P_{sc}^{\mathcal{F}}$. These probabilities can be chosen such that:

$$\forall c \in C_{\bullet}, \quad \sum_{s \in S(c)} \mu_S(s) P_{sc}^{\mathcal{F}} > \mu_C(c).$$

  (and symmetrically for customers)
- For this alternate policy stability can be shown using Foster-Lyapunov criterion.

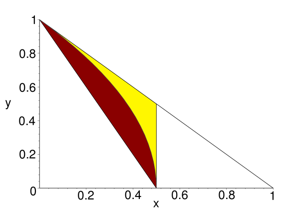
# Priorities and Match the Shortest are not always stable

Prop. NN model with either the MS policy or the PR (priority) policy such that customers of class 1 (resp. servers of class $1'$) give priority to servers of class $2'$ (resp. to customers of class 2):




For both policies, the stability region is not maximal.

Proof:
Consider $\mu_C = (1/3, 2/5, 4/15)$, $\mu_S = \mu_C$, and $\mu = \mu_C \times \mu_S$. The conditions $\mathrm{NC{\scriptsize OND}}$ are satisfied, but the Markov chain is transient (for MS or PR defined as above).



# Final remarks/questions

- ML has a maximal stability region. Is stability region is always maximal for some other policies (ex. FIFO and RANDOM)?
- MS and Priorities have non-maximal stability region. How to compute the stability region?
- Better sufficient conditions for stability, valid for all admissible policies?
- More general arrival assumptions?