Partially Optimal Routing

Asu Ozdaglar

Daron Acemoglu and Ramesh Johari
Dept. of Economics, MIT and Dept. of MS&E, Stanford

September, 2006

Electrical Engineering and Computer Science Dept.

Massachusetts Institute of Technology

Motivation

- Most large-scale communication networks, such as the Internet, consist of interconnected administrative domains.
- Increasing interest to allow end users to choose routes themselves.
  - Selfish Routing
- Administrative domains control the routing of traffic within their own networks.
- Obvious conflicting interests as a result:
  - Users care about end-to-end performance.
  - Individual network providers optimize their own objectives.
- The study of routing patterns and performance requires an analysis of Partially Optimal Routing:
  - End-to-end route selection selfish
    - Transmission follows minimum latency route for each source.
  - Network providers route traffic within their own network to achieve minimum intradomain latency.

Our Work

- A model of partially optimal routing.
- Implications for equilibrium routing patterns and network performance.
- Three Main Objectives:
  1. Investigate whether partially optimal routing (i.e., the presence of traffic engineering) improves the overall network performance.
     – Relation to Braess' Paradox
  2. Quantify performance losses of partially optimal routing relative to optimal routing for the overall network:
     – Price of Anarchy for partially optimal routing [Pigou], [Koutsoupias and Papadimitriou], [Roughgarden and Tardos].
  3. Understand the choice of routing policy by a single network provider.

Model

- A network $G = (V, A)$, with distinguished source and destination nodes $s, t \in V$.
- P denotes the set of paths from s to t.
- X units of flow are to be routed from s to t.
- Each link $j \in A$ has a latency function $l_j(x_j)$ that represents the delay as a function of the flow x_j on link j.
  - Assume $l_j(x_j)$ is strictly increasing and nonnegative.
- We call the tuple $R = (V, A, P, s, t, X, \mathbf{l})$ a routing instance.

Socially Optimal Routing

Given a routing instance $R = (V, A, P, s, t, X, \mathbf{l})$:

- We define the social optimum $\mathbf{x}^{SO}(R)$, as the optimal solution of:

$$\begin{aligned} \text{minimize} \quad & \sum_{j \in A} x_j l_j(x_j) \\ \text{subject to} \quad & \sum_{p \in P : j \in p} y_p = x_j, \quad j \in A, \\ & \sum_{p \in P} y_p = X, \quad y_p \geq 0, \quad p \in P. \end{aligned}$$

- Given a routing instance R and a feasible flow $\mathbf{x}(R)$, we denote the total latency cost at $\mathbf{x}(R)$ by:

$$C(\mathbf{x}(R)) = \sum_{j \in A} x_j(R) l_j(x_j(R)).$$

Selfish Routing

- When traffic routes "selfishly," all paths with nonzero flow must have the same total delay.
- The Wardrop equilibrium flow, $\mathbf{x}^{WE}(R)$, is the unique solution of:

$$\begin{aligned} \text{minimize} \quad & \sum_{j \in A} \int_0^{x_j} l_j(z)\, dz \qquad (1) \\ \text{subject to} \quad & \sum_{p \in P : j \in p} y_p = x_j, \quad j \in A, \\ & \sum_{p \in P} y_p = X, \quad y_p \geq 0, \quad p \in P. \end{aligned}$$

- It is well-known that a feasible solution $\mathbf{x}^{WE}$ of Problem (1) is a Wardrop equilibrium if and only if

$$\sum_{j \in A} l_j(x_j^{WE})(x_j^{WE} - x_j) \leq 0,$$

for all feasible solutions $\mathbf{x}$ of Problem (1).

Partially Optimal Routing

- Consider a subnetwork inside of G, denoted $G_0 = (V_0, A_0)$.
- Assume first that G_0 has a unique entry and exit point, denoted by $s_0 \in V_0$ and $t_0 \in V_0$. P_0 denotes paths from s_0 to t_0.
- We call $R_0 = (V_0, A_0, P_0, s_0, t_0)$ a subnetwork of $G : R_0 \subset R$.
- Given an incoming amount of flow X_0, the network operator chooses the routing by:

$$\begin{aligned} L(X_0) = \min \quad & \sum_{j \in A_0} x_j l_j(x_j) \\ \text{s.t.} \quad & \sum_{p \in P_0 : j \in p} y_p = x_j, \quad j \in A_0, \\ & \sum_{p \in P_0} y_p = X_0, \qquad y_p \geq 0, \quad p \in P_0. \end{aligned}$$

- Define $l_0(X_0) = L(X_0)/X_0$ as the effective latency of POR in the subnetwork R_0.

POR Flows

- Given a routing instance $R = (V, A, P, s, t, X, \mathbf{l})$, and a subnetwork $R_0 = (V_0, A_0, P_0, s_0, t_0)$ defined as above, we define a new routing instance $R' = (V', A', P', s, t, X, \mathbf{l}')$ as follows:

$$V' = (V \setminus V_0) \bigcup \{s_0, t_0\};$$

$$A' = (A \setminus A_0) \bigcup \{(s_0, t_0)\};$$

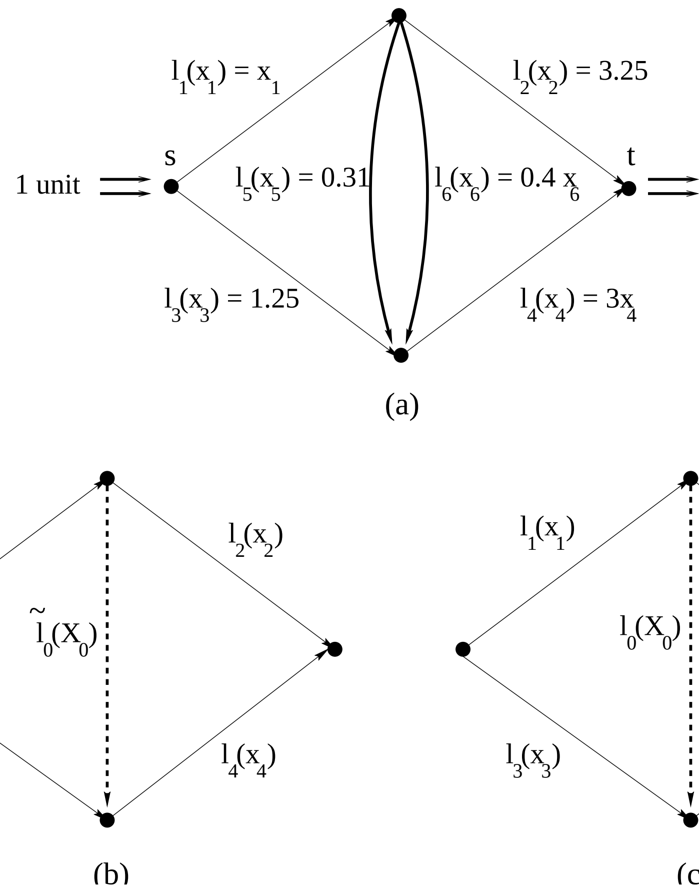
- $\mathbf{l}' = \{l_j\}_{j \in A \setminus A_0} \bigcup \{l_0\}$.

- We refer to R' as the equivalent POR instance for R with respect to R_0.

- The overall network flow in R with partially optimal routing in R_0, $\mathbf{x}^{POR}(R, R_0)$, is defined as:

$$\mathbf{x}^{POR}(R, R_0) = \mathbf{x}^{WE}(R').$$

Performance of Partially Optimal Routing



- Selfish Routing: Link flows $x_1^{WE} = 0.94$ and $X_0^{WE} = 0.92$, with a total cost of $C(\mathbf{x}^{WE}(R)) = 4.19$.
- Partially Optimal Routing: Link flows $x_1^{POR} = 1$ and $X_0^{POR} = 1$, with a total cost of $C(\mathbf{x}^{POR}(R)) = 4.25$,

Braess Paradox and POR Paradox

- Braess' Paradox: Consider a routing instance $R = (V, A, P, s, t, X, \mathbf{l})$. We say that *Braess' paradox* occurs in R if there exists another routing instance $R_m = (V, A, P, s, t, X, \mathbf{m})$, with a vector of strictly increasing, nonnegative latency functions, $\mathbf{m} = (m_j, j \in A)$, such that $m_j(x_j) \leq l_j(x_j)$ for all $x_j \geq 0$ and

$$C(\mathbf{x}^{WE}(R_m)) > C(\mathbf{x}^{WE}(R)).$$

- POR Paradox: Consider a routing instance $R = (V, A, P, s, t, X, \mathbf{l})$, and a subnetwork $R_0 = (V_0, A_0, P_0, s_0, t_0)$. We say that the *POR paradox* (partially optimal routing paradox) occurs in R with respect to R_0 if

$$C(\mathbf{x}^{POR}(R, R_0)) > C(\mathbf{x}^{WE}(R)).$$

Main Result

- Proposition: Consider a routing instance $R = (V, A, P, s, t, X, \mathbf{l})$ and a subnetwork $R_0 = (V_0, A_0, P_0, s_0, t_0) \subset R$. Assume that the POR paradox occurs in R with respect to R_0. Then Braess' paradox occurs in R.
  - *Proof Idea:* Uniformly lower the latency functions in the subnetwork R_0, such that the Wardrop effective latency of R_0 is given by l_0 (the effective latency of optimal routing within R_0).
- Corollary: Given a routing instance R, if Braess' paradox does not occur in R, then partially optimal routing with respect to any subnetwork always improves the network performance.
  - Milchtaich has shown that Braess' paradox does not occur in directed graphs where the underlying undirected graph has a *series-parallel* structure.
  - For a network with serial-parallel links, partially optimal routing always improves the overall network performance.

Price of Anarchy for Partially Optimal Routing

- Investigate the worst case efficiency loss of partially optimal routing with respect to socially optimal routing.
- We first recall the following key results in the analysis of selfish routing:
- Proposition [Roughgarden-Tardos (2002)]:

(a)

$$\inf_{R \in \mathcal{R}^{conv}} \frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{WE}(R))} = 0.$$

(b) Consider a routing instance $R = (V, A, P, s, t, X, \mathbf{l})$ where l_j is an affine latency function for all $j \in A$. Then,

$$\frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{WE}(R))} \geq \frac{3}{4}.$$

Furthermore, the bound above is tight.

Price of Anarchy for Partially Optimal Routing

- Proposition:Let $\mathcal{R}'$ denote a set of routing instances.

$$\inf_{\substack{R\in\mathcal{R}' \\ R_0\subset R}} \frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{POR}(R,R_0))} \leq \inf_{R\in\mathcal{R}'} \frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{WE}(R))}.$$

$$\inf_{\substack{R\in\mathcal{R}^{aff} \\ R_0\subset R}} \frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{POR}(R,R_0))} \geq \inf_{R\in\mathcal{R}^{conc}} \frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{WE}(R))}.$$

- Theorem:

(a)

$$\inf_{\substack{R\in\mathcal{R}^{conv} \\ R_0\subset R}} \frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{POR}(R,R_0))} = 0.$$

(b) Consider a routing instance $R = (V, A, P, s, t, X, \mathbf{l})$ where l_j is an affine latency function for all $j \in A$; and a subnetwork R_0 of R.

$$\frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{POR}(R,R_0))} \geq \frac{3}{4}.$$

Furthermore, the bound above is tight.

Price of Anarchy for Partially Optimal Routing

- *Proof of part (b):* The proof relies on the following two results:
- *Lemma:* Assume that the latency functions l_j of all the links in the subnetwork are nonnegative affine functions. Then, the effective latency of POR, $l_0(X_0)$, is a nonnegative concave function of X_0.
- *Proposition:* Let $R \in \mathcal{R}^{conc}$ be a routing instance where all latency functions are concave. Then,

$$\frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{WE}(R))} \geq \frac{3}{4}.$$

  Furthermore, this bound is tight.

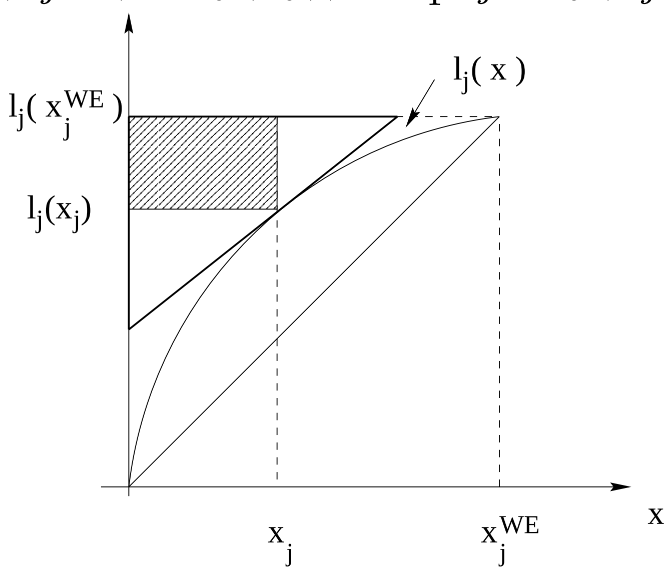
Price of Anarchy for Partially Optimal Routing

Proof of Proposition: From variational inequality representation of WE, for all feasible $\mathbf{x}$, we have

$$C(\mathbf{x}^{WE}) = \sum_{j\in A} x_j^{WE} l_j(x_j^{WE}) \leq \sum_{j\in A} x_j l_j(x_j^{WE}) = \sum_{j\in A} x_j l_j(x_j) + \sum_{j\in A} x_j (l_j(x_j^{WE}) - l_j(x_j)).$$

For all feasible $\mathbf{x}$, we have

$$x_j(l_j(x_j^{WE}) - l_j(x_j)) \leq \frac{1}{4} x_j^{WE} l_j(x_j^{WE}).$$



Bounds for Polynomial Latency Functions

- Given a class of latency functions $\mathcal{L}$, we define:

$$\beta(\mathcal{L}) = \sup_{l \in \mathcal{L},\ x \geq 0} \beta(l, x),$$

$$\beta(l, x) = \max_{z \geq 0} \frac{(l(x) - l(z))z}{l(x)x},$$

- Intuitively β is measure of the steepness of a class of latency functions: $\beta(\mathcal{L}^{aff}) = 0.25$, $\beta(\mathcal{L}^{quad}) = 0.385$.

- Theorem: Let $\mathcal{L}_d$ be a class of nonnegative polynomial latency functions of degree d. Consider a routing instance $R = (V, A, P, s, t, X, \mathbf{l})$ with $l_j \in \mathcal{L}_d$ for all $j \in A$, and a subnetwork R_0 of R. Then,

$$\frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{POR}(R, R_0))} \geq (1 - \beta(\mathcal{L}_d)).$$

  Furthermore, the bound above is tight.

Bounds for Polynomial Latency Functions

- Lemma: Let $R_0 = (V_0, A_0, P_0, s_0, t_0)$ be a subnetwork with polynomial latency functions of degree d. Then the effective latency $l_0(X_0)$ is given by

$$l_0(X_0) = \inf_{y \in \mathcal{Y}} \{f(X_0, y)\},$$

where $\mathcal{Y}$ is a nonempty compact set, $f(X_0, y)$ is a cont. function of y, and $\forall\ y \in \mathcal{Y}$, $f(\cdot, y)$ is a nonneg. polynomial of degree d.

- Lemma: Let $\mathcal{L}_s$ be a class of nonnegative latency functions which is closed under scaling by a constant $k \leq 1$. Let

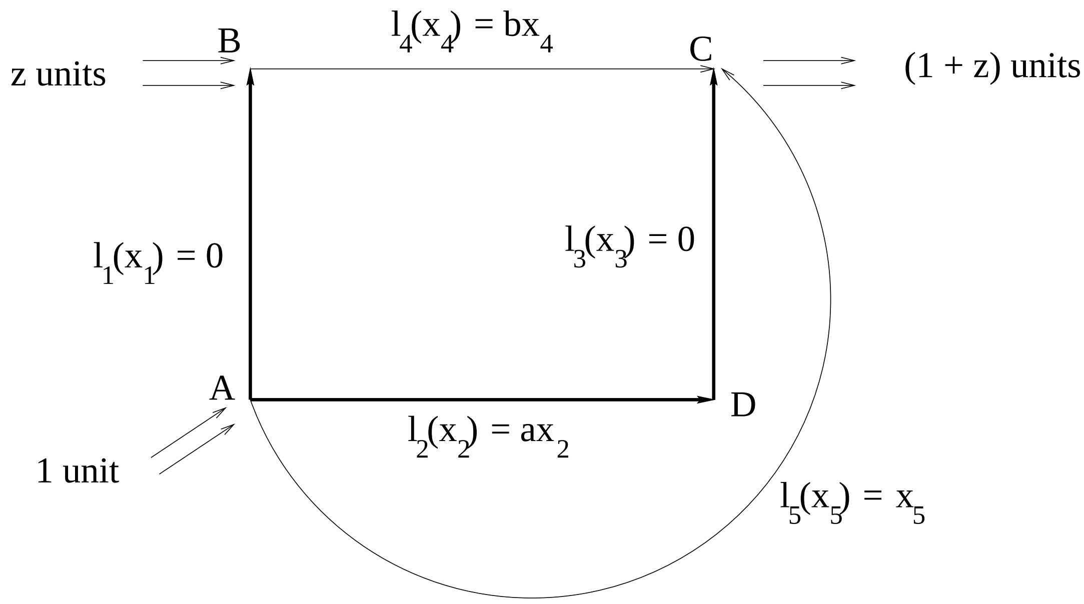
$$l_j(x) = \inf_{z \in \mathcal{Z}_j} \{f(x, z)\}, \qquad \forall\ j \in A,$$

where $\mathcal{Z}_j$ is a compact set; for each x, $f(x, \cdot)$ is a continuous function of z; and for each $z \in \mathcal{Z}_j$, $f(\cdot, z) \in \mathcal{L}_s$. Then:

$$\frac{C(\mathbf{x}^{SO}(R))}{C(\mathbf{x}^{WE}(R))} \geq (1 - \beta(\mathcal{L}_s)).$$

Subnetworks with Multiple Entry-Exit Points

- Even for linear latencies, efficiency loss of partially optimal routing can be arbitrarily high.

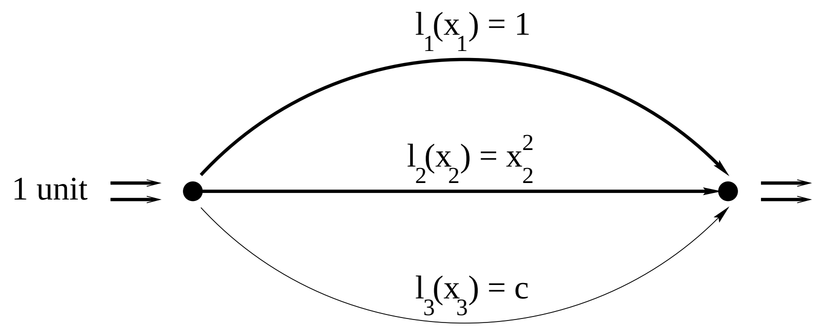


- Social Optimum: $x^{SO} = (0, \frac{1}{1+a}, \frac{1}{1+a}, z, \frac{a}{1+a})$.
- POR: $x^{POR} = (\frac{1-bz}{1+b}, 0, 0, \frac{1+z}{1+b}, \frac{b+bz}{1+b})$.
- For a fixed $b > 0$, as $a \to 0$ and $z \to 0$,

$$C(x^{SO}) \to 0, \qquad C(x^{POR}) \to \frac{b}{1+b} > 0,$$

Subnetwork Performance: Traffic Engineering

- We consider a model where a subnetwork can choose a routing policy to achieve the minimum latency within its subnetwork.



- Selfish Routing: $\sqrt{c}$ units of traffic is routed through the subnetwork, leading to a total cost of $C(\mathbf{x}^{WE}) = c$, and a subnetwork cost of $C_{G_0}(\mathbf{x}^{WE}) = c\sqrt{c}$.
- POR: Entire traffic is routed through the subnetwork, leading to $C(\mathbf{x}^{POR}) = C_{G_0}(\mathbf{x}^{POR}) = 1 - \frac{2}{3\sqrt{3}}$.
- For $c\sqrt{c} < 1 - \frac{2}{3\sqrt{3}}$, we have

$$C_{G_0}(\mathbf{x}^{POR}) > C_{G_0}(\mathbf{x}^{WE}).$$

Traffic Engineering for Parallel Link Topology

- Consider a network consisting of parallel links with d units of traffic.
- Suppose there are $N+1$ providers each owning a subset of links.
- Consider a local ("partial equilibrium") analysis for the routing choice within subnetwork 0.
- Represent network provider i, for $i = 1, \ldots, N$, by a single link with effective latency l_i (reflecting the intradomain routing policy of i)
- l_0: effective latency of optimal routing within subnetwork 0.
- $\tilde{l}_0$: effective latency of selfish routing within subnetwork 0.
- The routing policy choice of provider 0 can be parametrized by $\delta \in [0, 1]$, leading to an effective latency of

$$m_0(x, \delta) = (1 - \delta)\, l_0(x) + \delta \tilde{l}_0(x).$$

Traffic Engineering for Parallel Link Topology

- $l_R(x)$: effective latency of Wardrop routing x units on links $1, \ldots, N$.
- The optimization problem of subnetwork 0 then is:

$$\min_{0 \le x_0 \le d, \delta \in [0,1]} \left[(1-\delta) l_0(x_0) + \delta \tilde{l}_0(x_0) \right] x_0$$

$$\begin{aligned}
\text{s.t.} \quad (1-\delta)\, l_0\,(0) + \delta \tilde{l}_0\,(0) &\ge l_R\,(d)\,, && \text{if } x_0 = 0; \\
(1-\delta)\, l_0\,(d) + \delta \tilde{l}_0\,(d) &\le l_R\,(0)\,, && \text{if } x_0 = d; \\
(1-\delta)\, l_0\,(x_0) + \delta \tilde{l}_0\,(x_0) &= l_R\,(d - x_0)\,, && \text{if } 0 < x_0 < d.
\end{aligned}$$

- If $\tilde{l}_0(0) \ge l_R(d)$, optimal solution is $\delta = 1$, $x_0 = 0$.
- If $\tilde{l}_0(d) \le l_R(0)$, optimal solution is $\delta = 0$, $x_0 = d$.
- Otherwise, the optimization problem for subnetwork 0 reduces to:

$$\min_{x_0 \in [x_0^{MIN}, x_0^{MAX}]} \min \left\{ x_0 l_R(d - x_0), d l_0(d) \right\}$$

where

$$\tilde{l}_0(x_0^{MIN}) = l_R(d - x_0^{MIN}); \qquad l_0(x_0^{MAX}) = l_R(d - x_0^{MAX}).$$

Conclusions

- First extension of the classical traffic routing models to capture traffic engineering.
- Interesting global and subnetwork performance results.
- Extensions to subnetworks with multiple entry-exit points.
- General equilibrium analysis for subnetwork routing policy choice.
- Other objectives for subnetworks: profit maximization.