Internet Pricing

- Until April 30, 1995, the primary US backbone for the www was the NSFNET. NSF paid about $11.5 million per year to operate NSFNET (80% for leasing fiber optic lines and routers)
- When it was shut down, the question of Internet pricing arose.
- What is Internet Pricing? Charging end-users served by ISPs
  - Charging companies (including ISPs) to connect to the backbone.
- Why charge for the Internet?
  - Cost recovery (& profit) (The cost of Connectivity 2014)
    - This is how Internet speed and price in the U.S. compares to the rest of the world
  - Efficiency & congestion reduction
  - Service levels
- Cost Recovery & Profit
  - Nearly all pricing that is currently done is for this reason
  - Companies: connecting to the backbone
    - Fixed fee for fixed bandwidth
    - T1 (4.5 mbps) ~ $230 to $1,000 per month depending on other services
    - T3 (45 mbps; composed of 672 voice and data channels operating at 64 kbps; essentially 28 T1 lines) ~ $500-2,000 per month
    - “committed information rate”: fee for max guaranteed flow (“Bustable billing rate”)
  - Users: connection to ISP
    - Fixed fee (Verizon, Comcast (Xfinity), AT&T, etc)
    - Fixed fee + usage
• Congestion ("Value") pricing
  – Because of externalities, Internet resources are not allocated properly
  – Service times play some role in "rationing" or "allocating" capacity
    • not efficient because users do not bear the full cost of their usage
    • Hence, there should be some sort of "value" pricing
  – The Demand-side of Internet Services
    • The demand for Internet services clearly depends on the time it takes for requests to be completed & on the cost.
    • Let’s assume that people pay a fixed fee (price does not depend on # of requests)
      – The only influence on demand is the performance.
      – Performance of the system depends on the # of requests made
  • Our Goal:
    – To figure out where we wind up given the interaction btwn demand and performance
    – To see if usage-dependent fees are warranted
• How do people behave?

T - Service time (function of # requests)

R - Inverse “demand” function
(given # of requests what must the “price” be)

How would we regulate requests if we could?

We would Maximize the difference between consumer surplus and the total cost
(assuming service is being provided efficiently)

Intuition: If each person paid what they were willing to pay, the area would be the revenue. So that must be what it is worth.
So we need to solve the following problem:

\[
\text{Max } \{ S(Q) = \int_0^Q R(q) \, dq - T(Q) \cdot Q \} \]

Differentiating (using Liebniz’s Rule for differentiating the definite integral)

\[
dS/dQ = R(Q) - T(Q) - Q \cdot dT(Q)/dQ
\]

Setting derivative = 0
\[
dS/dQ = 0 \Rightarrow R(Q) = T(Q) + Q^* \cdot dT(Q)/dQ
\]

Note that:
- If \( Q = 0 \), then \( Q \cdot dT(Q)/dQ = 0 \)
- If \( Q > 0 \), then \( Q \cdot dT(Q)/dQ > 0 \)

So, \( T(Q) + Q \cdot dT(Q)/dQ > T(Q) \)

So the demand, \( Q \) is higher than optimal, \( Q^* \)
But if we charge \( Q \cdot dT(Q)/dQ \) for \( Q \) requests, then we will get optimal usage, \( Q^* \)

We can also charge a usage invariant price \( P \) if we know \( T \) and \( R \).

**Implications:**
- It’s clear that everyone’s cost goes up if you charge.
- But the revenues are a transfer (i.e. one person loses but another gains), so they aren’t a bad thing.
- In total, “society” is better off.
Timing of Requests

• Consider how people choose when to make requests
• Consider a simple deterministic model that illustrates the point. Specifically consider a deterministic queuing model.

• Notation:
  – \( r(t) \) - request rate at time \( t \)
  – \( q(t) \) - queue size at time \( t \)
  – \( S \) - service rate of the system (requests per unit time)
  – \( N \) - total # of requests in time segment \([0, t]\)
  – \( C(t) \) - “cost” of making a request at time \( t \)
  – \( T(t) \) - service time at time \( t \)

• In a deterministic model, the service time is simply
  \[ T(t) = \frac{q(t)}{S} \]
• Assuming that the “cost” is dominated in time-units and people do not like to wait to make a request.

• Specifically, assuming the first time a request can be made is at $t = 0$:
  \[ C(t) = T(t) + \gamma t \]

Evolution of the Queue

given a request rate function, $r$, the size of the queue is simply:

\[ q(t) = q(0) + \int_0^t r(w) \, dw - S \, t \quad t \in [0, \tau] \]

How do People Behave?

Let’s consider the equilibrium situation where no one has an incentive to change their behavior.

\[ C(t) = C(0) \quad \text{for every } t \in [0, \tau] \quad \Rightarrow q(t) / S + \gamma t = q(0) / S \]
Introducing the expression for \( q(t) \):

\[
\frac{1}{S} \left[ q(0) + \int_0^t r(w) \, dw - S \, t \right] + \gamma \, t = \frac{q(0)}{S}
\]

\[\Rightarrow \int_0^t r(w) \, dw = (1-\gamma) \, S \, t \quad : \text{want to find function, } r(w), \text{ that satisfies this eq.}\]

Differentiating both sides with respect to \( t \):

\[
d/dt \left\{ \int_0^t r(w) \, dw \right\} = d/dt \left\{ (1-\gamma) \, S \, t \right\}
\]

\[\Rightarrow r(t) = (1-\gamma) \, S\]

To find \( t \) just observe that there are no “gaps” in requests. So, \( t = N/S \)

Finally, to find \( q(0) \) we use the fact that \( C(t) = C(0) \)

\[\Rightarrow q(0) = \gamma \, N \quad (\text{we know there has to be a bulk departure b/c we are in Equilib.})\]

Summary:

- \( q(0) = \gamma \, N \)
- \( r(t) = (1-\gamma)S \) (\( \gamma \) must be less than 1)
- \( q(t) = \gamma \, N - \gamma \, S \, t \quad t \in [0, N/S] \)
- Equilibrium cost = \( \gamma \, N /S \) (experienced by all)
- Total cost = \( \gamma \, N^2 /S \)
How would we Schedule people if we could?

If we could regulate when people make requests, we would want to minimize: **total cost**

That would be done by spreading requests evenly at the rate of $S$ over the period $[0, N/S]$. Then there would be no queue at all.

In this case: $q(0) = 0$, $r(t) = S$

Total Cost$_{optimal} = \int_0^{N/S} r(t) \, dt = \int_0^{N/S} S \, \gamma \, t \, dt = S \, \gamma \int_0^{N/S} t \, dt = S \, \gamma \left(1/2\right)(N/S)^2$

Therefore:

Total Cost$_{optimal} = (1/2S) \, \gamma \, N^2$

Total Cost$_{equilibrium} = (1/S) \, \gamma \, N^2$

So, we should price access to avoid “bunching up”
Charging Helps by:

• Charge a fee equal to the queuing time
• The people will voluntarily spread themselves out (at no increase in total cost)
• The revenue is a transfer and can be “returned” in a variety of ways, like increasing the capacity.