Reverse Engineering TCP/IP

Steven Low
EAS, Caltech

Joint work with:
Li, J. Wang, Pongsajapan, Tan, Tang, M. Wang

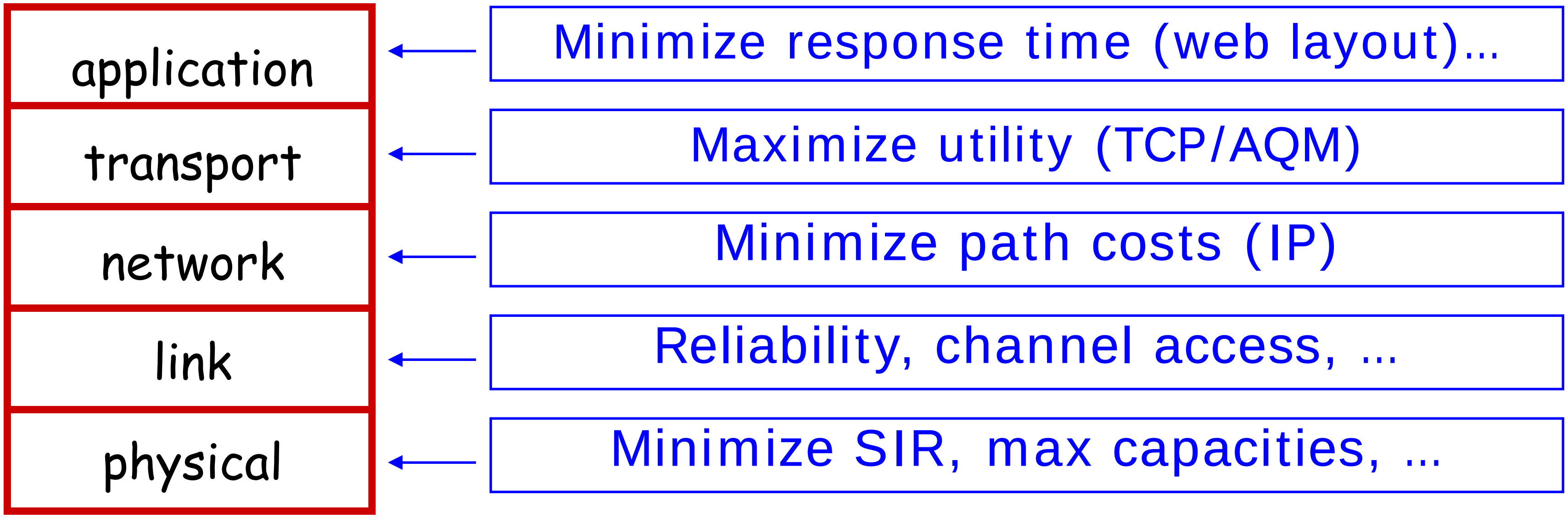
Outline

- Background
  - Layering as optimization decomposition
  - Reverse engineering TCP
- Reverse engineering TCP/IP
  - Delay insensitive utility
  - Delay sensitive utility
  - How bad is single-path routing

J. Wang, Li, Low, Doyle. ToN, 2005
Pongsajapan, Low, Infocom 2007
M. Wang, Tan, Tang, Low, pre-print, 2009

Layering as optimization decomposition

- Each layer designed separately and evolves asynchronously
- Each layer optimizes certain objectives



application ← Minimize response time (web layout)...

transport ← Maximize utility (TCP/AQM)

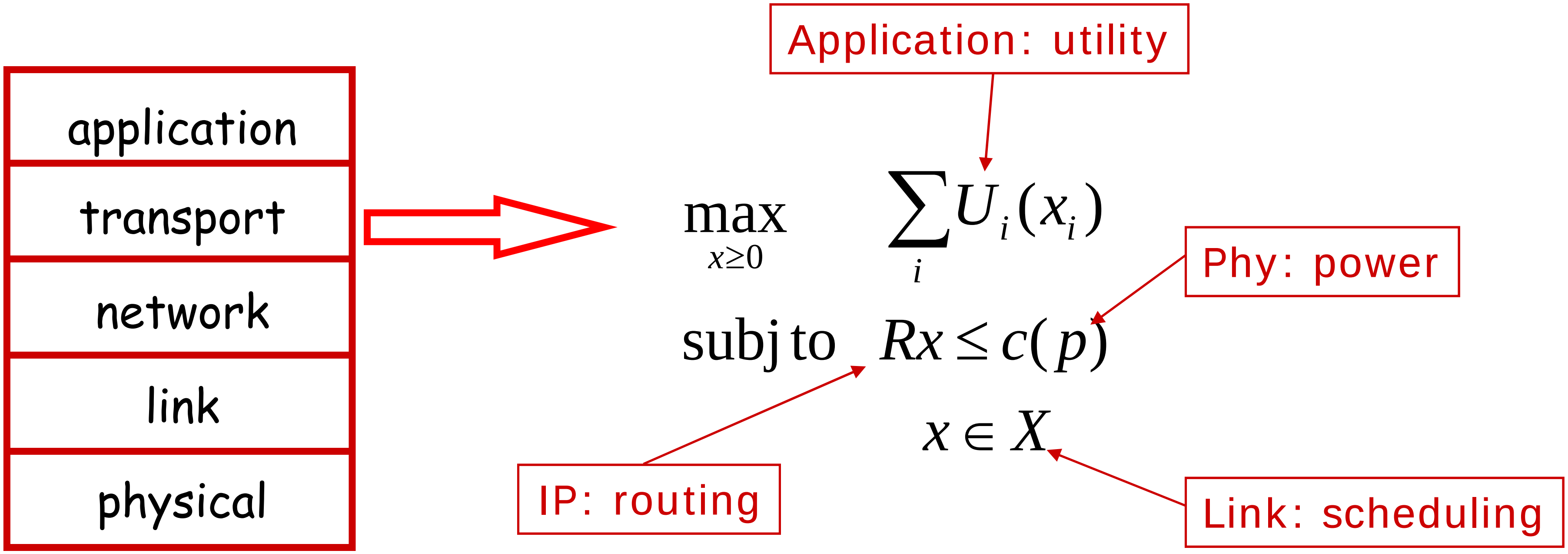
network ← Minimize path costs (IP)

link ← Reliability, channel access, ...

physical ← Minimize SIR, max capacities, ...

Layering as optimization decomposition

- Each layer is abstracted as an optimization problem
- Operation of a layer is a distributed solution
- Results of one problem (layer) are parameters of others
- Operate at different timescales



$$
\begin{aligned}
\max_{x\geq 0} \quad & \sum_i U_i(x_i) \\
\text{subj to} \quad & Rx \leq c(p) \\
& x \in X
\end{aligned}
$$

A wireless example

Application objective

Network objective

$$\begin{aligned} \max_{x \geq 0} \quad & \sum_i U_i(x_i) + \sum_l V_l(w_l) \\ \text{subj to} \quad & R(G)\, x \leq c(w, \mathbf{P}) \\ & x \in C(\mathbf{P}) \end{aligned}$$

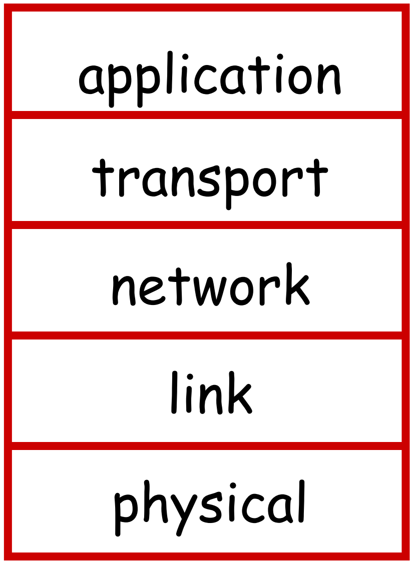
IP: optimize route given network graph G

Rate also constrained by interaction of coding mechanism & ARQ

Link: maximize channel capacity given link resources w and desired error probability P

Layering as optimization decomposition

- □ Each layer is abstracted as an optimization problem
- □ Operation of a layer is a distributed solution
- □ Results of one problem (layer) are parameters of others
- □ Operate at different timescales



application

transport

network

link

physical

1) Understand each layer in isolation, assuming other layers are designed nearly optimally

2) Understand interactions across layers

3) Incorporate additional layers

4) Ultimate goal: entire protocol stack as solving one giant optimization problem, where individual layers are solving parts of it

Layering as optimization decomposition

□ Network	generalized NUM
□ Layers	subproblems
□ Layering	decomposition methods
□ Interface	functions of primal or dual vars

application

transport

network

link

physical

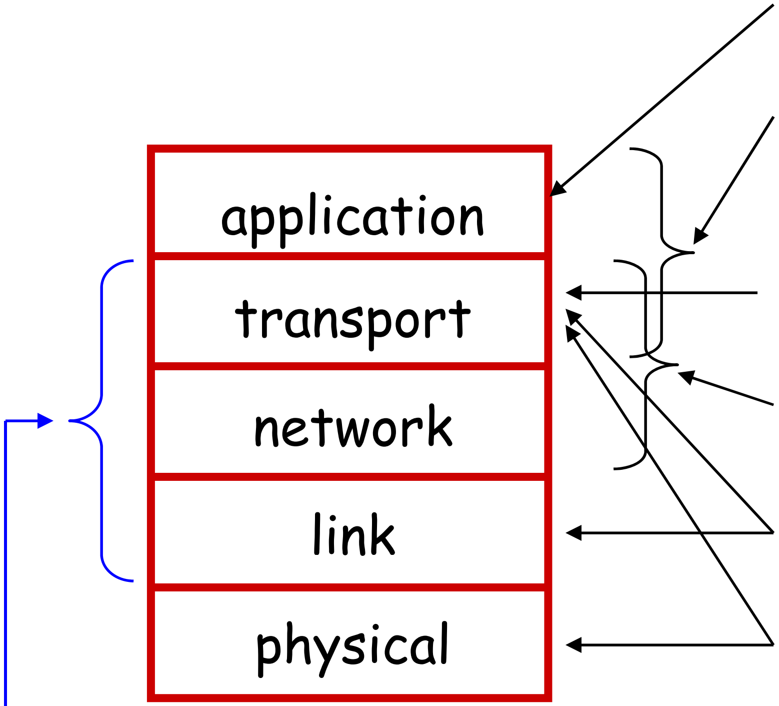
1) Understand each layer in isolation, assuming other layers are designed nearly optimally

2) Understand interactions across layers

3) Incorporate additional layers

4) Ultimate goal: entire protocol stack as solving one giant optimization problem, where individual layers are solving parts of it

Examples



Optimal web layer: Zhu, Yu, Doyle ’01

HTTP/TCP: Chang, Liu ’04

TCP: Kelly, Maulloo, Tan ’98, ……

TCP/IP: Wang et al ’05, ……

TCP/MAC: Chen et al ’05, ……

TCP/power control: Xiao et al ’01,
Chiang ’04, ……

Rate control/routing/scheduling: Eryilmax et al ’05, Lin et al ’05, Neely, et al ’05, Stolyar ’05, Chen, et al ’05

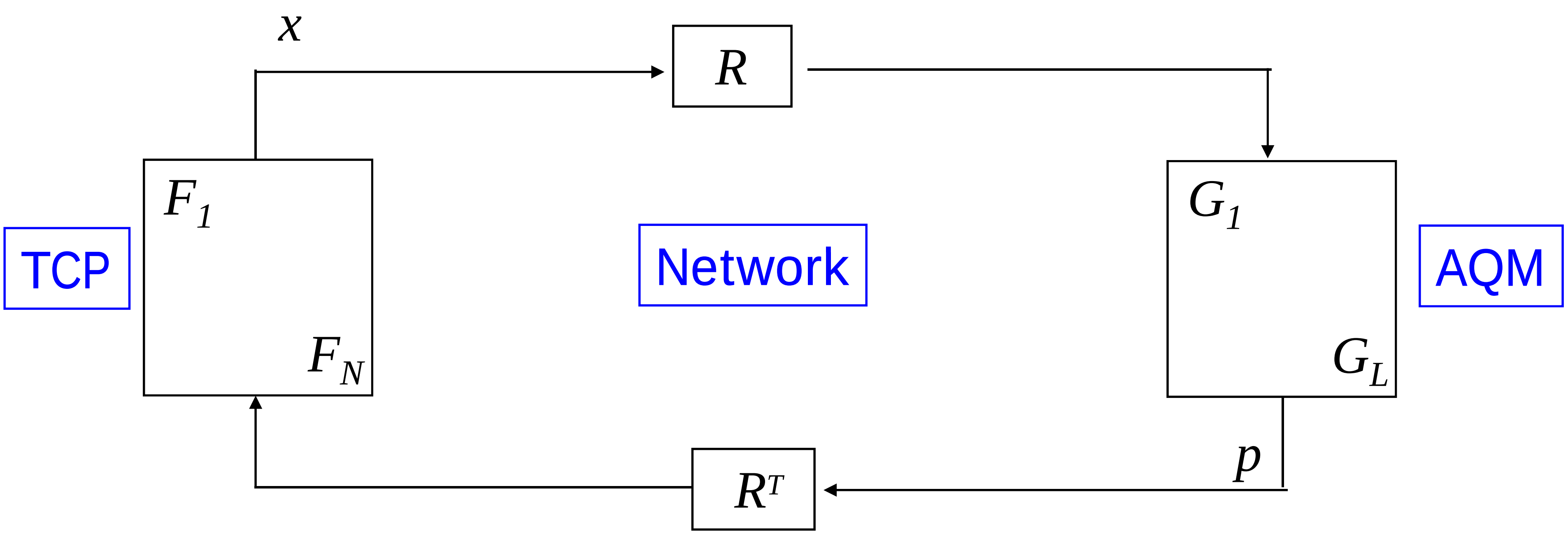
Survey in Proc. of IEEE, 2006

Outline

- □ Background
  - Layering as optimization decomposition
  - Reverse engineering TCP
- □ Reverse engineering TCP/IP
  - Delay insensitive utility
  - Delay sensitive utility
  - How bad is single-path routing

J. Wang, Li, Low, Doyle. ToN, 2005
Pongsajapan, Low, Infocom 2007
M. Wang, Tan, Tang, Low, pre-print, 2009

Network model: general



$$R_{li} = 1 \quad \text{if source } i \text{ uses link } l$$

IP routing

$$x(t+1) = F\,(R^T p(t),\ x(t))$$

Reno, Vegas, FAST

$$p(t+1) = G\,(p(t),\ Rx(t))$$

DropTail, RED, …

Network model: example

Reno:

Jacobson
1989

```
for every RTT
{   W += 1   }                 (AI)
for every loss
{   W := W/2   }               (MD)
```

$$x_i(t+1) = \overset{\boxed{\text{AI}}}{\overset{\downarrow}{\frac{1}{T_i^2}}} - \frac{x_i^2}{2}\sum_l R_{li}\,p_l(t) \longleftarrow \boxed{\text{MD}}$$

$$p_l(t+1) = G_l\left(\sum_i R_{li}\,x_i(t),\ p_l(t)\right) \longleftarrow \boxed{\text{TailDrop}}$$

Network model: example

FAST:

Jin, Wei, Low
2004

```
periodically
{
        $W := \frac{\text{baseRTT}}{\text{RTT}} W + \alpha$
}
```

$$x_i(t+1) = x_i(t) + \frac{\gamma_i}{T_i}\left(\alpha_i - x_i(t)\sum_l R_{li}\, p_l(t)\right)$$

$$p_l(t+1) = p_l(t) + \frac{1}{c_l}\left(\sum_i R_{li}\, x_i(t) - c_l\right)$$

- How to characterize equilibrium of TCP

$$\begin{aligned} x^* &= F\,(R^T p^*,\ x^*) \\ p^* &= G\,(p^*,\ Rx^*) \end{aligned}$$

$$R_{li} = 1 \quad \text{if source } i \text{ uses link } l \longleftarrow \boxed{\text{IP routing}}$$

$$x(t+1) = F\,(R^T p(t),\ x(t)) \longleftarrow \boxed{\text{Reno, Vegas, FAST}}$$

$$p(t+1) = G\,(p(t),\ Rx(t)) \longleftarrow \boxed{\text{DropTail, RED, ...}}$$

Duality model of TCP

- □ TCP

$$x^* = F(R^T p^*, x^*)$$

$$p^* = G(p^*, Rx^*)$$

- □ Equilibrium (x^*,p^*) primal-dual optimal:

$$\max_{x\geq 0} \sum U_i(x_i) \qquad \text{subject to} \quad Rx \leq c$$

  - F determines utility function U
  - G guarantees complementary slackness
  - p^* are Lagrange multipliers

Kelly, Maloo, Tan 1998
Low, Lapsley 1999

Uniqueness of equilibrium

- x^* is unique when U is strictly concave
- p^* is unique when R has full row rank

Duality model of TCP

- □ TCP

$$x^* = F(R^T p^*,\ x^*)$$

$$p^* = G(p^*,\ Rx^*)$$

- □ Equilibrium *(x*,p*)* primal-dual optimal:

$$\max_{x \geq 0} \sum U_i(x_i) \qquad \text{subject to} \quad Rx \leq c$$

  - F determines utility function U
  - G guarantees complementary slackness
  - p^* are Lagrange multipliers

Kelly, Maloo, Tan 1998
Low, Lapsley 1999

The underlying concave program also
leads to simple dynamic behavior

Duality model of TCP

- Equilibrium (x^*,p^*) primal-dual optimal:

$$\max_{x\geq 0} \sum U_i(x_i) \qquad \text{subject to} \quad Rx \leq c$$

Mo & Walrand 2000:

$$U_i(x_i) = \begin{cases} \log x_i & \text{if} \quad \alpha = 1 \\ (1-\alpha)^{-1} x_i^{1-\alpha} & \text{if} \quad \alpha \neq 1 \end{cases}$$

- $\alpha = 1$: Vegas, FAST, STCP
- $\alpha = 1.2$: HSTCP
- $\alpha = 2$: Reno
- $\alpha = \infty$: XCP (single link only)

Duality model of TCP

- □ Equilibrium *(x*,p*)* primal-dual optimal:

$$\max_{x\geq 0} \sum U_i(x_i) \qquad \text{subject to} \quad Rx \leq c$$

Mo & Walrand 2000:

$$U_i(x_i) = \begin{cases} \log x_i & \text{if} \quad \alpha = 1 \\ (1-\alpha)^{-1} x_i^{1-\alpha} & \text{if} \quad \alpha \neq 1 \end{cases}$$

- ■ $\alpha = 0$: maximum throughput
- ■ $\alpha = 1$: proportional fairness
- ■ $\alpha = 2$: min delay fairness
- ■ $\alpha = \infty$: maxmin fairness

Some implications

- □ Equilibrium
  - Always exists, unique if R is full rank
  - Bandwidth allocation independent of AQM or arrival pattern
  - Can predict macroscopic behavior of large scale networks
- □ Counter-intuitive throughput behavior
  - Fair allocation is not always inefficient
  - Increasing link capacities do not always raise aggregate throughput

[Tang, Wang, Low, ToN 2006]

- □ FAST TCP
  - Design, analysis, experiments

[Jin, Wei, Low, Hegde, ToN 2007]

Outline

- □ Background
  - Layering as optimization decomposition
  - Reverse engineering TCP
- □ Reverse engineering TCP/IP
  - Delay insensitive utility
  - Delay sensitive utility
  - How bad is single-path routing

For joint congestion control and **multipath** routing:
Gallager (1977), Golestani & Gallager (1980), Bertsekas, Gafni & Gallager (1984), Kelly, Maulloo & Tan (1998), Kar, Sarkar & Tassiulas (2001), Lestas & Vinnicombe (2004), Kelly & Voice (2005), Lin & Shroff (2006), He, Chiang & Rexford (2006), Paganini (2006)

Motivation

Primal:
$$\max_{x\geq 0} \sum_i U_i(x_i) \quad \text{subject to} \quad Rx \leq c$$

Dual:
$$\min_{p\geq 0} \left(\sum_i \max_{x_i\geq 0} \left(U_i(x_i) - x_i \sum_l R_{li} p_l \right) + \sum_l p_l c_l \right)$$

Motivation

Primal: $\boxed{\max_{R}} \max_{x\geq 0} \sum_i U_i(x_i) \quad \text{subject to} \quad Rx \leq c$

Dual: $\min_{p\geq 0} \left(\sum_i \max_{x_i\geq 0} \left(U_i(x_i) - x_i \boxed{\min_{R_i} \sum_l R_{li} p_l} \right) + \sum_l p_l c_l \right)$

Shortest path routing!

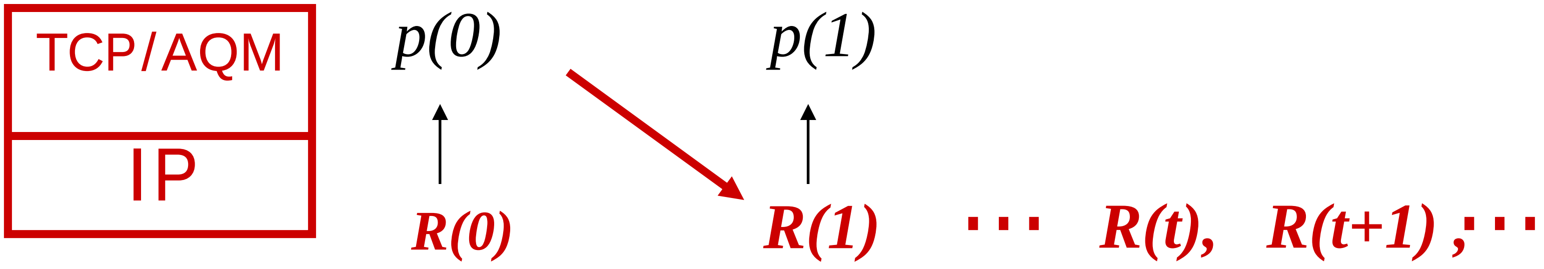
Can TCP/IP maximize utility?

Assumptions

- Two timescales
  - TCP converges instantly
  - Route changes slowly
- Single-path shortest path routing $R(t)$
  - Link cost: $p_l(t) + b\ \tau_l$

prop delay

queueing delay



Assumptions

- Two timescales
  - TCP converges instantly
  - Route changes slowly
- Single-path shortest path routing $R(t)$
  - Link cost: $p_l(t) + b\ \tau_l$

prop delay

queueing delay

will only consider $b=0$ or $b=1$

TCP/IP dynamic model

TCP

$$x(t) = \arg\max_{x\geq 0} \sum_i U_i(x_i)$$

$$\text{subject to} \quad R(t)x \leq c$$

AQM

$$p(t) = \arg\min_{p\geq 0} \sum_i \left(\max_{x_i\geq 0} U_i(x_i) - x_i \sum_l R_{li}(t)\, p_l \right) + \sum_l c_l p_l$$

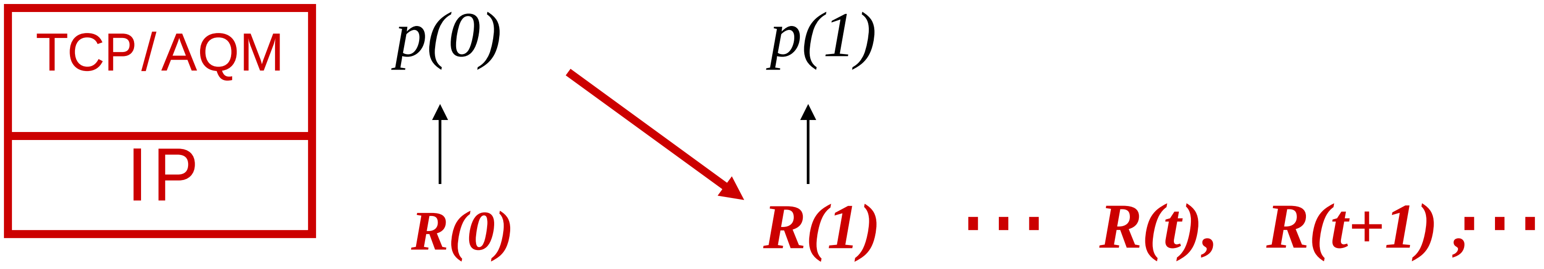
slow timescale

IP

$$R_i(t+1) = \arg\min_{R_{li}} \sum_l R_{li}(\overbrace{p_l(t) + b\tau_l}^{\text{Link cost}})$$

Reverse engineering TCP/IP

- Does equilibrium routing R_b exist ?
- How to characterize R_b?
- Is R_b stable ?
- Can it be stabilized?



TCP/AQM

IP

$p(0)$ $p(1)$

$R(0)$ $R(1)$ $\cdots$ $R(t)$, $R(t+1)$,$\cdots$

Delay insensitive utility: $b=0$

Theorem

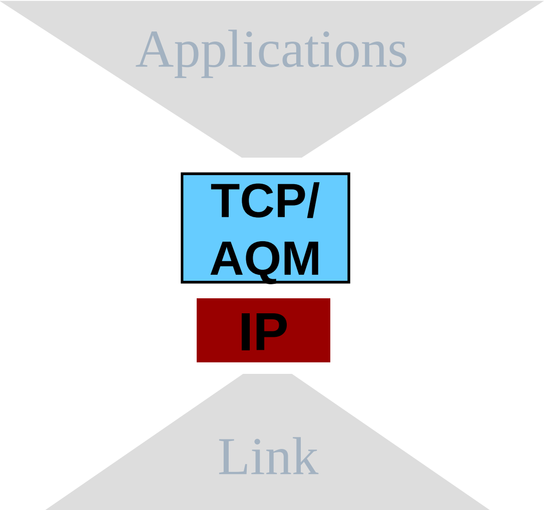
If $b=0$, R_b exists & solves NUM iff zero duality gap

- Shortest-path routing is optimal with congestion prices
- No penalty for not splitting

$$\text{Primal:}\quad \max_{R}\ \overbrace{\max_{x\geq 0}\ \sum_i U_i(x_i) \quad \text{subject to} \quad Rx \leq c}^{\text{Kelly's problem solved by TCP}}$$

$$\text{Dual:}\quad \min_{p\geq 0}\ \left(\sum_i \max_{x_i\geq 0} \left(U_i(x_i) - x_i \min_{R_i} \sum_l R_{li}\, p_l \right) + \sum_l p_l c_l \right)$$

Delay insensitive utility: *b=0*



$$\overset{\boxed{\text{IP}}}{\max_{R}}\ \overset{\boxed{\text{TCP-AQM}}}{\max_{x\geq 0}}\ \sum_i U_i(x_i)$$

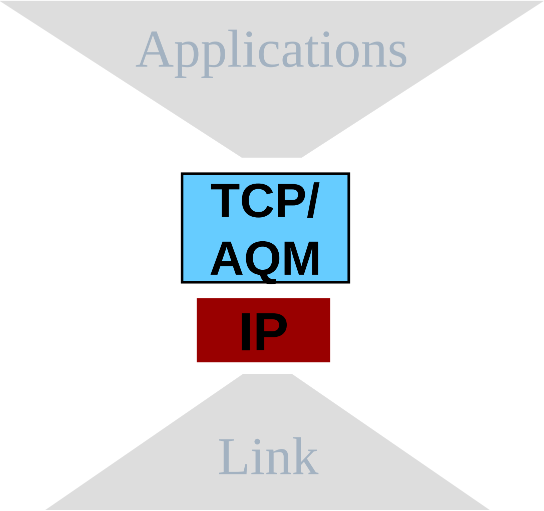
$$\text{subject to}\quad Rx \leq c$$

TCP/IP (with fixed *c*):

- Equilibrium of TCP/IP exists iff zero duality gap
- NP-hard, but subclass with zero duality gap is P
- Equilibrium, if exists, can be unstable
- Can stabilize, but with reduced utility

Nonzero duality gap: complexity, cost of not splitting

Delay insensitive utility: *b=0*



$$\overset{\boxed{\text{IP}}}{\downarrow}\ \overset{\boxed{\text{TCP-AQM}}}{\downarrow}$$

$$\max_{R}\ \max_{x\geq 0}\ \sum_i U_i(x_i)$$

$$\text{subject to}\quad Rx \leq c$$

BUT...

- *b* is never zero in practice
- If *b>0* then there are networks for which equilibrium routings exist but do not maximize any delay insensitive utility function

Outline

- □ Background
  - Layering as optimization decomposition
  - Reverse engineering TCP
- □ Reverse engineering TCP/IP
  - Delay insensitive utility
  - Delay sensitive utility
  - How bad is single-path routing

J. Wang, Li, Low, Doyle. ToN, 2005
Pongsajapan, Low, Infocom 2007
M. Wang, Tan, Tang, Low, pre-print, 2009

Delay sensitive utility: *b=1*

$$U_i\left(x_i, d_i\right) = V_i\left(x_i\right) - x_i d_i$$

$$\text{Round-trip prop delay} \longrightarrow d_i = \sum_l R_{li}\tau_l \longleftarrow \text{Link prop delay}$$

- Round trip propagation delay depends on R
- Delay sensitive utility function
  - Utility from throughput … balanced by
  - Disutility from delay

Delay sensitive utility: *b=1*

Theorem

If $b=1$, R_b exists & solves NUM iff zero duality gap

- Shortest-path routing is optimal
- No penalty for not splitting

Primal: $\max_{R} \max_{x\ge 0} \sum_i U_i(\underline{x_i, d_i})$ subject to $Rx \le c$

Dual: $$\min_{p\ge 0} \left(\sum_i \max_{x_i \ge 0} \left(U_i(\underline{x_i, d_i}) - x_i \min_{R_i} \sum_l R_{li}(p_l + \tau_l) \right) + \sum_l p_l c_l \right)$$

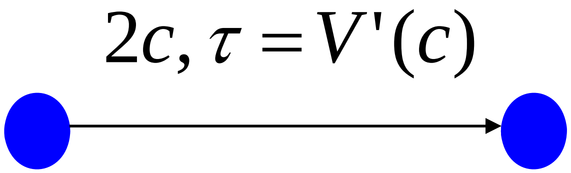
Counter-intuitive behavior

With delay sensitive utility

- Bottleneck links can be under-utilized

There exist networks such that the TCP/IP equilibrium (x^*, p^*, R^*) is in the interior:

$$R^*x^* < c$$



Equilibrium rate: $x^* = c < 2c$

$$U(x,d) = V(x) - x\tau$$

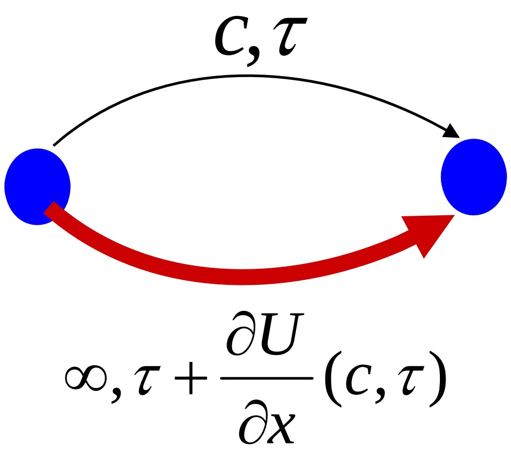
$$\frac{\partial U}{\partial x}(c,\tau) = V'(c) - \tau \;=\; 0$$

Counter-intuitive behavior

With delay sensitive utility

- Extra paths that will be utilized by delay-insensitive utility functions may not

It is sub-optimal to use the long path, even when traffic is allowed to distribute over multiple paths



$$U(x, d) = V(x) - x\tau$$

Equilibrium routing: use short path <u>only</u>

Counter-intuitive behavior

Any delay sensitive utility that a TCP/IP equilibrium maximizes necessarily possesses one of 3 “strange” properties

- The specific utility $U(x,d)=V(x)-x\tau$ has two of the 3
- In contrast to joint congestion control and multi-path routing

Counter-intuitive behavior

$\mathcal{B}$ must have at least one of the following three properties:

1) $\exists U(x,d) \in \mathcal{B}, d > 0$ so that $U(x,d)$ is not strictly increasing in x.
2) $\forall U_1(x,d) \in \mathcal{B}$, $\forall \epsilon > 0$, we have $U_2(x,d) := U_1(x+\epsilon, d)$ is not in $\mathcal{B}$.
3) $\exists U(x,d) \in \mathcal{B}$, $D > 0$ such that $f(d) := M(U,d)$ is finite and discontinuous for all $d > D$.

$$M(U,d) := \lim_{c \to \infty} U(c,d)$$

Routing stability

Given any network, suppose

- link cost: $ap_l(t) + \tau_l$
- $0<a<a_\#$ is small enough

If every SD pair has unique min prop delay path, then TCP/IP is asymptotically stable

Routing stability

Given any network, suppose

- link cost: $ap_l(t) + \tau_l$
- $0<a<a_\#$ is small enough

Otherwise, consider a modified network in which every SD pair has a unique min delay path, but

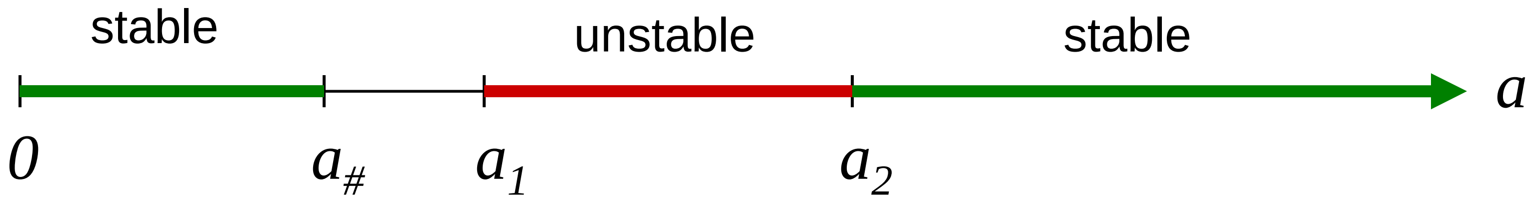
- link cost: $p_l(t)$

Then the two networks have the same equilibrium and stability properties

Routing stability

For any delay sensitive or insensitive utility function, there exists a network such that decreasing a can destabilize TCP/IP

- link cost: $ap_l(t) + \tau_l$



Outline

- Background
  - Layering as optimization decomposition
  - Reverse engineering TCP
- Reverse engineering TCP/IP
  - Delay insensitive utility
  - Delay sensitive utility
  - How bad is single-path routing

J. Wang, Li, Low, Doyle. ToN, 2005
Pongsajapan, Low, Infocom 2007
M. Wang, Tan, Tang, Low, pre-print, 2009

Multi-path routing

- Source can split its total rate into multiple paths

$$\text{Total source rate}: \quad x_i = \left(x_{i1}, \ldots, x_{ik_i}\right)$$

$$i\text{'s rate on path } j: \quad x_{ij}$$

$$\text{multi-path}: \quad \| x_i \|_1 = \sum_j x_{ij}$$

$$\text{single-path}: \quad \| x_i \|_\infty = \max_j \; x_{ij}$$

Multi-path routing

$$\text{Multi-path}: \quad \max_{R} \; \max_{x \geq 0} \; \sum_i U_i(\| x_i \|_1)$$
$$\text{subject to} \quad Rx \leq c$$

$$\text{Single-path}: \quad \max_{R} \; \max_{x \geq 0} \; \sum_i U_i(\| x_i \|_\infty)$$
$$\text{subject to} \quad Rx \leq c$$

Total source rate: $x_i = \left(x_{i1}, \ldots, x_{ik_i} \right)$

i's rate on path j: x_{ij}

multi-path: $\| x_i \|_1 = \sum_j x_{ij}$

single-path: $\| x_i \|_\infty = \max_j \; x_{ij}$

Multi-path routing

$$\text{Multi-path:} \quad \max_{R} \quad \max_{x \geq 0} \quad \sum_i U_i(\| x_i \|_1)$$

$$\text{subject to} \quad Rx \leq c$$

$$\text{Single-path:} \quad \max_{R} \quad \max_{x \geq 0} \quad \sum_i U_i(\| x_i \|_\infty)$$

$$\text{subject to} \quad Rx \leq c$$

For multi-path routing

- Joint routing and congestion control is a concave program (polynomial-time solvable)
- Zero duality gap
- Upper bounds max utility of single-path TCP/IP

Multi-path routing

$$\text{Multi-path:} \quad \max_{R} \quad \max_{x\geq 0} \quad \sum_i U_i(\| x_i \|_1)$$
$$\text{subject to} \quad Rx \leq c$$

$$\text{Single-path:} \quad \max_{R} \quad \max_{x\geq 0} \quad \sum_i U_i(\| x_i \|_\infty)$$
$$\text{subject to} \quad Rx \leq c$$

For single-path TCP/IP:

- No longer concave program; primal is NP-hard
- Non-zero duality gap in general
- Zero gap iff TCP/IP equilibrium exists
- Duality gap = cost of not splitting

Multi-path routing

$$\text{Multi-path:} \quad \max_{R} \quad \max_{x \geq 0} \quad \sum_i U_i(\| x_i \|_1)$$

$$\text{subject to} \quad Rx \leq c$$

$$\text{Single-path:} \quad \max_{R} \quad \max_{x \geq 0} \quad \sum_i U_i(\| x_i \|_\infty)$$

$$\text{subject to} \quad Rx \leq c$$

Theorem

- For any multi-path solution *(R, x)*, there is a multi-path solution *(R', x')*
  - That uses no more than $N+L$ paths
  - Achieves the same utility

Multi-path routing

$$\text{Multi-path}: \quad \max_{R} \ \max_{x\geq 0} \ \sum_{i} U_i(\| x_i \|_1)$$

$$\text{subject to} \quad Rx \leq c$$

$$\text{Single-path}: \quad \max_{R} \ \max_{x\geq 0} \ \sum_{i} U_i(\| x_i \|_\infty)$$

$$\text{subject to} \quad Rx \leq c$$

Theorem

- Duality gap is upper bounded by

$$\min(L, N) \max_i \rho_i$$

$$\rho_i = \max_{y\in[0,M^i]} (U^i(y) - U^i(y/K^i))$$

Multi-path routing

$$\text{Multi-path:}\quad \max_{R}\ \max_{x\geq 0}\ \sum_i U_i(\| x_i \|_1)$$

$$\text{subject to}\quad Rx \leq c$$

$$\text{Single-path:}\quad \max_{R}\ \max_{x\geq 0}\ \sum_i U_i(\| x_i \|_\infty)$$

$$\text{subject to}\quad Rx \leq c$$

Corollary

- For Vegas/FAST $U_i(x_i) = \alpha_i \log x_i$ duality gap is bounded by

$$\min(L, N)\ \max_i \alpha_i \log K_i$$

Conclusion & open issues

- Summary
  - Equilibrium of TCP/IP can be interpreted as maximizing network utility over rates & routes

- How to reconcile TCP utility maximization and TCP/IP utility maximization?
  - Given routing, TCP utility is increasing in throughput
  - With TCP/IP, this is no longer the case
- In general, can/how we regard layering as optimization decomposition?