Walking the tightrope : Responsive yet stable Traffic Engineering

Presented by Aparna Sundar

Problem Definition and current solutions

TE problem definition

Offline Methods – OSPF-TE , MPLS

Online Methods – MATE, TeXCP

Inadequecy of Offline methods

- Cannot react to real-time traffic reroutes.
- Load distribution not guarenteed to be optimal.
- Suboptimal reaction to failure.

Online methods

- Should react to real-time traffic demands and failures.
- Prior approaches – centralized, assuming a global oracle, lacking stability analysis. Eg MATE.
- TeXCP – distributed and stable.

Summary of Results

- For same traffic demands, TeXCP supports same utilisation and failure resiliance with a third of the capacity as traditional offline methods.
- Network utilization is always within a few percentage points of optimal value.
- Prefers shorter routes while trimming long routes that are not useful.

Big Picture

- Two Components
- Load Balancer : multiple paths delivering demands from ingress to egress router, moving traffic from over-utilized to under utilized paths.
- Closed loop Feed Back controller: collects network feedback at faster time scale than LB to ensure traffic stability.

Diagram, Explanation of terms

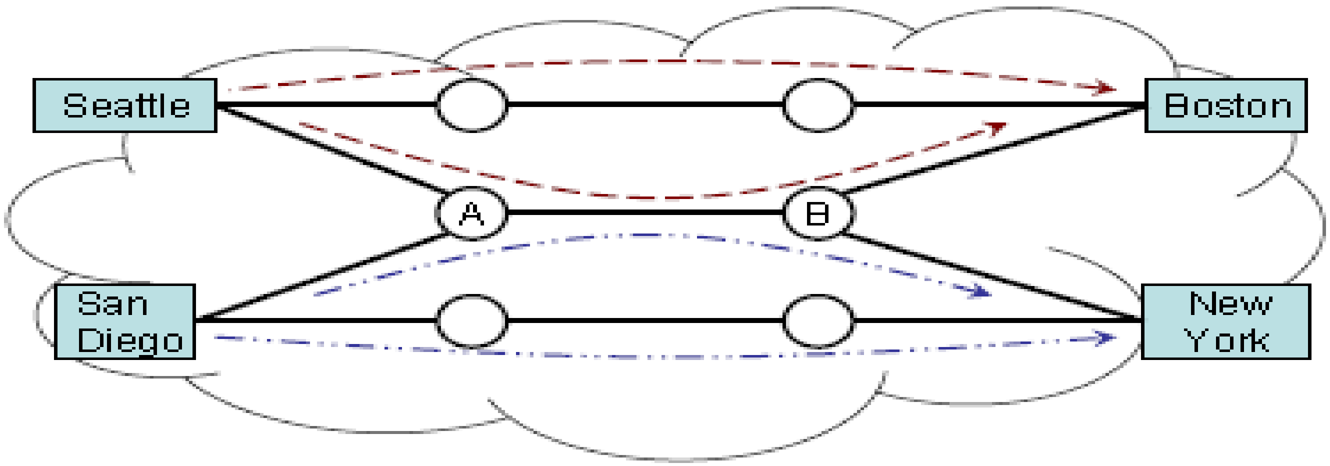

Figure 1: For each Ingress-Egress (IE) pair, there is a TeXCP agent at the ingress router, which balances the IE traffic across available paths in an online, distributed fashion.

Var	Definition
R_s	Total Traffic Demand of IE pair s
P_s	Set of paths available to IE pair s
r_{sp}	Traffic of IE pair s sent on path p. i.e., $R_s = \sum r_{sp}$
x_{sp}	Fraction of IE, s, traffic sent on path p, called path weight.
u_{sp}	The utilization of path p observed by IE pair s
u_l	The utilization of link l
C_l	The capacity of link l
P_l	Set of paths that traverse link l
$\overline{u}_s$	Weighted average utilization of paths used by IE pair s

Table 2: The variables most-used in the paper. All of these variables are functions of time.

LP Formulation at each IE pair

$$\min_{x_{sp}} \max_{l \in L} u_l, \tag{1}$$

subject to the constraints:

$$\begin{aligned} u_l &= \sum_s \sum_{p \in P_s, p \ni l} \frac{x_{sp} \cdot R_s}{C_l}, && (2) \\ \sum_{p \in P_s} x_{sp} &= 1, \quad \forall s, && (3) \\ x_{sp} &\geq 0, \quad \forall p \in P_s, \forall s. && (4) \end{aligned}$$

Path Selection,Probing Network state

- Path Selection:
- ISP picks set of K shortest paths that it can use.
- Probing Network state:
- Maintain probe timer, Tp, to maintain track of path utilization. Tp > RTT.
- Probe packet with updatable utilization field sent by ingress node. Egress node sends it back to app agent.
- Probe loss: estimate util to max(1,p u_sp)

Load Balancer

- Each agent maintains a decision timer, which fires every Td sec, > 5Tp.
- each time the agent,computes change in fraction of IE traffic sent on path p.
- At eqbm, x_sp is constant, traffic is conserved, no negative traffic possible,updates should decrease max utilization.

Load Balancer (contd)

$$\Delta x_{sp} = \begin{cases} \frac{r_{sp}}{\sum_{p'} r_{sp'}} \left(\overline{u}_s - u_{sp}\right) & \forall p, u_{sp} > u_{min} \\ \frac{r_{sp}}{\sum_{p'} r_{sp'}} \left(\overline{u}_s - u_{sp}\right) + \epsilon & p, u_{sp} = u_{min}. \end{cases}$$

- Intuition:path whose util is greater than avg shd dec its rate while path whose util isbelow avg should increase its rate.
- Change in traffic is proportional to current traffic on path (which is prop to util).
- Use of epsilon – to re-use or restart util on a path.

Preventing Oscillations , Managing Congestion

- Two agents working independantly may shift flow to link that was previously under-utilized.
- Solution (inspired by XCP)
- Compute aggregate feedback: $\Phi = \alpha \cdot T_p \cdot S - \beta \cdot Q,$
- Compute per IE flow feedback based on a Max-Min approach: $\Phi \geq 0 \;\; \Rightarrow \;\; \delta^+ = \frac{\Phi}{N},\, \delta^- = 0,$
- $\Phi < 0 \;\; \Rightarrow \;\; \delta^+ = 0,\, \delta^- = \frac{\Phi}{\phi_l},$
- Positive feedback added, negative multiplied.

Preventing Oscillations , Managing Congestion(contd)

- Sending feedback to agents using probe. $$g_{sp} = g_{sp} + \delta^{+} - \delta^{-} \times g_{sp},$$
- g_sp is allowed rate on path p.
- Actual rate = min(g_sp, x_sp R_s).
- Prefer to use shorter paths: Use weighted max-min fairness to push a preference for shorter paths.
- Heuristic : Shorter paths better for better network utilisations

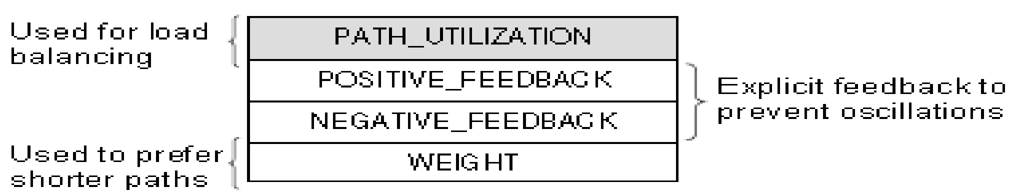

Figure 2: Probe Packet. Feedback is returned in two fields because Positive Feedback is additive while Negative Feedback is multiplicative.

Analysis

- Computation of explicit feedback for each pair, by load balancer, that leads to more stable per-IE flow rates and subsequently utilizations.
- Effect of feedback on network, “mostly” done by the time load balancer kicks into action ie,the explicit feedback brings path util to 90% of desired value, before the next time any of the load balancers need to make a decision.

Results and Comparision

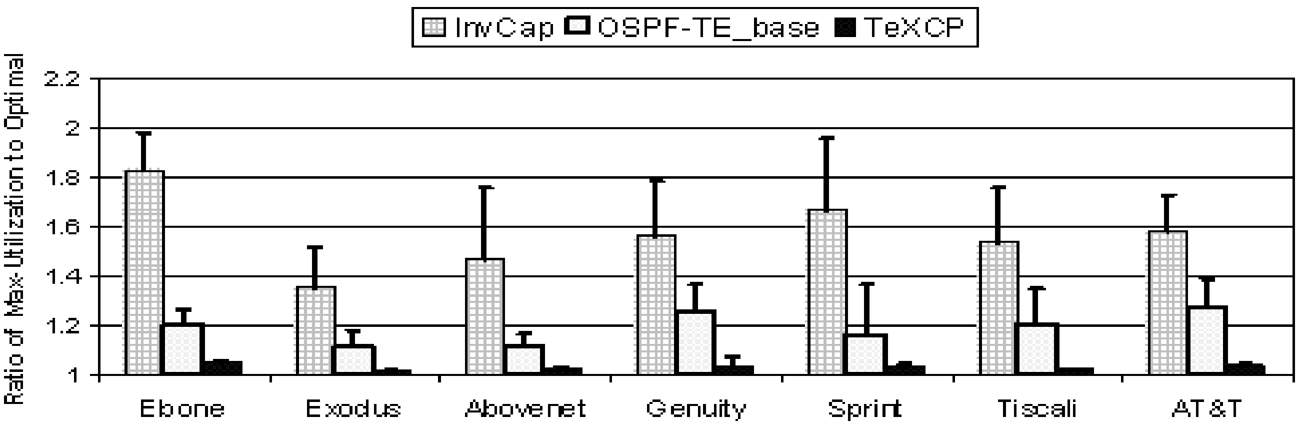

Figure 4: When traffic matches TM, TeXCP results in a max-utilization within a few percent of the optimal, and much closer to optimal than OSPF-TE or InvCap. Figure shows both average (thick bars) and maximum (thin bars) taken over 40 TMs.

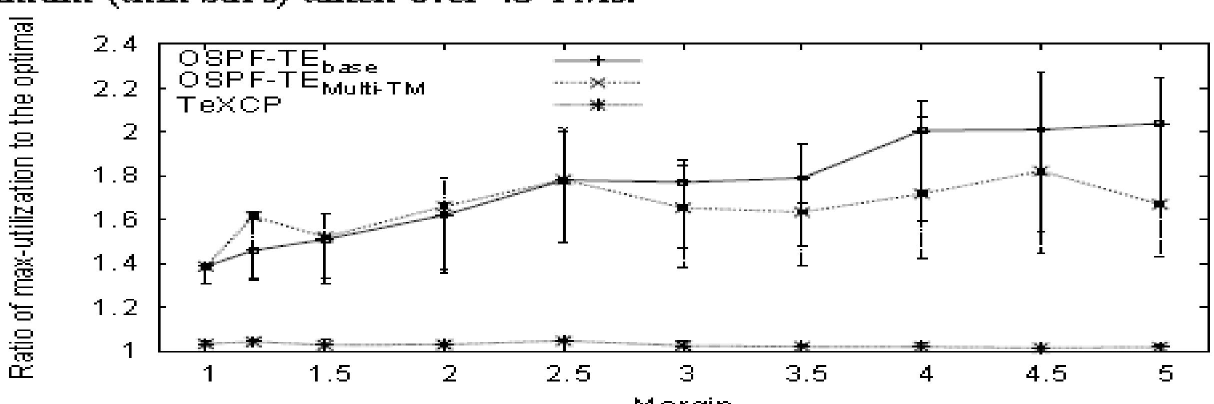

Figure 5: When traffic deviates from TM by a margin, TeXCP stays within a few percent of the optimal max-utilization; OSPF-TE$_{Base}$ and OSPF-TE$_{Multi-TM}$ lead to much larger max-utilization.

Results and comparision (contd)

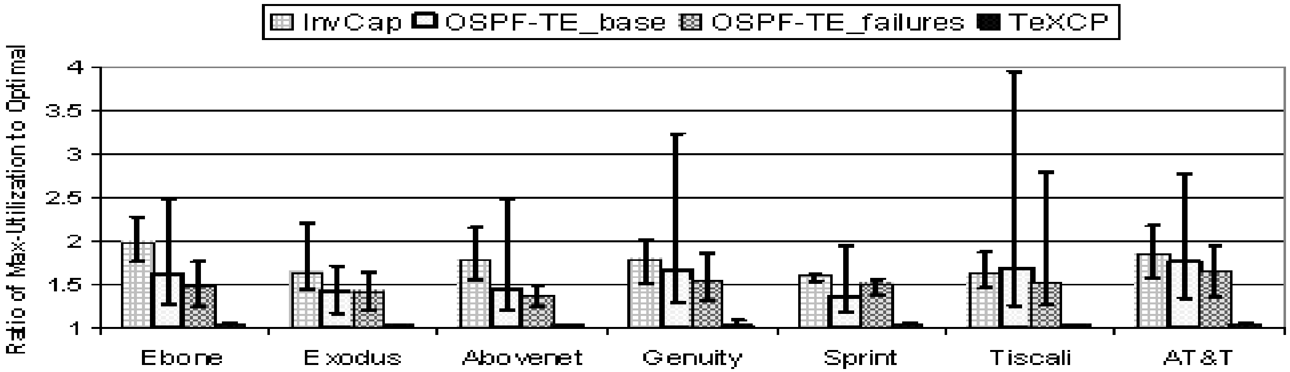

Figure 6: Under failures, TeXCP's max-utilization is within a few percent of the optimal; InvCap, OSPF-TE$_{Base}$, and OSPF-TE$_{Failures}$ become highly suboptimal. Figure shows the $90th$ percentile (thick) and maximums (thin) taken over multiple TMs.

Discussion

- Look at source-dest paths, instead of ingress-egress paths?
- Metric for network utilization
- $$Metric = \frac{\textbf{max-utilization}_{\,Tech.}}{\textbf{max-utilization}_{\,Oracle}}.$$
- Including estimate for egress-ingress links?