Measurement Informed Route Selection

Nick Duffield[1], Kartik Gopalan[2], Michael R. Hines[2]
Aman Shaikh[1], Jacobus E. van der Merwe[1]

[1] AT&T Labs–Research {duffield,ashaikh,kobus}@research.att.com
[2] Binghamton University {kartik,mhines}@cs.binghamton.edu

1 Motivation

Popular Internet applications exhibit subtle dependencies on data path characteristics [3, 5]. On the other hand, various studies [2, 1] have shown that non-default Internet paths can have dramatically different quality characteristics compared to default paths. Unfortunately, the existing Internet routing infrastructure is not able to benefit from these observations. However, recently proposed routing architectures [4] open up the possibility to ease these constraints by allowing route selection to be more dynamic and to be informed by information outside the routing protocol. Motivated by these observations, in the work presented here, we explore the possibility of informing route selection based on measured path properties. Specifically, we observe that from their vantage points in the Internet topology, Tier-1 and other large ISPs typically have multiple possible routes that can be used to reach the majority of destination prefixes on the Internet. By monitoring routing information we track the availability of alternative paths available from a Tier-1 ISP. At the same time we perform detailed measurements of loss and delay to a large number of Internet destinations and characterize various properties of these alternate paths to determine: (i) Whether there are significant differences in the properties of these different paths that would warrant its consideration as part of the route selection process. (ii) The stability of these properties over various timescales which would impact how they can be utilized and dictate the requirements of a measurement infrastructure that can provide such information. We believe this is the first such study involving a single Tier-1 ISP.

2 Measurement Informed Route Selection

The essence of our approach is depicted in Figure 2. Consider source s_1 and destination d_1. s_1 connects to network $AS0$ at ingress router i_1. Network $AS0$ has two paths available to reach destination d_1, through egress router e_2 and network $AS1$ and through egress router e_3 and network $AS2$. In order to determine the "best" path, we assume the existence of a measurement infrastructure, such that performance measurements are available for the paths across $AS0$ between the ingress router and the two egress routers, and from each of the egress routers to the destination. To simplify the notation, in what follows we will ignore the

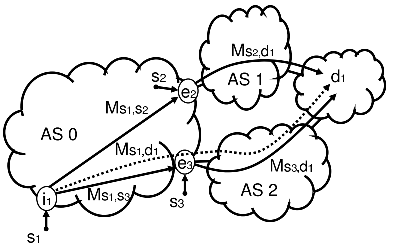

Fig. 1. Measurement Informed Route Selection: Composite measurements (s_1, s_2, d_1) and (s_1, s_3, d_1) and corresponding alternate paths.

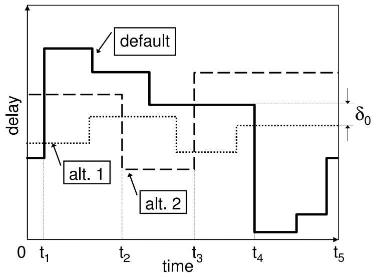

Fig. 2. Delay Measurement and Advantage: most recent measured delay on default and alternate paths. Advantage is $\geq \delta_0$ during interval $[t_1, t_4)$

ingress and egress router notation and assume that a measurement source is co-resident with each of the routers in $AS0$. The set of measurements between the ingress router and the two egress routers is thus denoted by $M_{s1,s2}$ and $M_{s1,s3}$ respectively, and that between the egress routers and the destination by $M_{s2,d1}$ and $M_{s3,d1}$. Given this information, a network equipped with the appropriate routing infrastructure [4] can select the "best" route between s_1 and d_1 by appropriate combination of measured characteristics along the two available composite measurement paths (s_1, s_2, d_1) and (s_1, s_3, d_1). In this abstract we present initial results of a measurement study in which we evaluate the benefit of such Measurement Informed Route Selection as compared with the default BGP route selection as observed in a Tier-1 ISP.

3 Methodology

Composite Performance Metrics. Given loss and delay (λ_1, δ_1) on the path (s_1, s_2) and (λ_2, δ_2) on (s_2, d_1), the composite transmission rate for the composite path (s_1, s_2, d_1) is the product $1 - \lambda = (1 - \lambda_1)(1 - \lambda_2)$. The composite delay is the sum $\delta = \delta_1 + \delta_2$. The composite metric for loss $\lambda_{(s_1,s_2,d_1),t}$ at time t is the composite of the most recent measurements $\lambda_{(s_1,s_2),t}$ and $\lambda_{(s_2,d_1),t}$ on the internal and external segments, and similarly for delay δ.

Performance Advantage of Alternate Routes. A *route trajectory* of a source-destination (SD) pair (s, d) specifies for each time t a source $\sigma(t)$ which is s if the default path is used, and $s' \neq s$ if the alternate path (s, s', d) is to be used. The *loss advantage* of using a route trajectory σ is $\mathcal{L}(\sigma(t), t) = \lambda_{(s,d),t} - \lambda_{(s,\sigma(t),d),t}$, i.e., the difference between the most recent performance metric on the default path and the alternate path $(s, \sigma(t), d)$. (Here $\lambda_{(s,s,d),t}$ simply denotes $\lambda_{(s,d),t}$.) The delay advantage $\mathcal{D}(\sigma(t), t)$ is similarly defined.

Available Performance Advantage. The available performance advantage represents a baseline advantage that would be obtained by a routing policy that enabled instant selection of the best performing path whenever a new direct or composite measurement becomes available. Figure 2 illustrates delay measurements for a default path and two alternates. In each case, the curve shows the value of the most recent measurement on that path, and is hence a right continuous step function, the measurements occurring at the steps. In the interval $[t_1, t_2)$, alternate 1 has the best most recent performance measurement; in $[t_2, t_3)$ alternate 2 is best; in $[t_3, t_4)$ alternate 1 is best again; prior to t_1 and after t_4 the default route is best. Note the available delay advantage is positive, i.e., there is benefit in using a non-default path, only in $[t_1, t_4)$.

Temporal Performance Advantage Metric. We analyze the duration of *runs* of performance advantage, i.e., maximal time intervals in which the available performance advantage for a given SD pair exceeds a given level. Longer runs are more useful, since the payoff for switching routes is longer lived. If runs are shorter than the typical settle down time after route changes, the period of advantage would be over before its benefit could be utilized. Our delay run performance statistic is the *time fraction* $F_D(\tau, \delta)$: the fraction of the measurement interval that a given SD pair spent in a delay advantage run of duration greater than τ and whose performance advantage exceeds δ, for each $\tau, \delta > 0$. For example, with reference to Figure 2, measuring over $[0, t_5]$, then for any $\tau < \min\{t_4 - t_3, t_3 - t_2, t_2 - t_1\}$ and $\delta < \delta_0$, $F_D(\tau, \delta) = (t_4 - t_1)/t_5$. The corresponding loss statistic $F_L(\tau, \lambda)$ is defined similarly.

4 Evaluation

Performance Measurement. The data we used was obtained by performing active measurements continuously over a 12 day period in April 2006. The Internet paths to 738 unique destinations were probed from 15 probe locations distributed throughout the backbone of a large Tier-1 ISP. The probe destinations were randomly chosen from a significantly larger set of known DNS server addresses; hence measurement probes traveled via a diverse set of paths through the Internet. We also conducted measurements between all pairs of sources. For each SD pair, the loss measure λ was the proportion of 100 packets for which no response was received. The delay measure δ was the median reported round-trip time reported, including infinite values for lost packets. We favored the median over the mean, since it is more robust to outlying values.

Probe Frequency. Over 94% of SD pairs had median interprobe time < 20 minutes. The median time between receipt of probes from *any* source was $<$ 2 minutes for over 97% of destinations. Thus composite performance data on alternate routes to a given destination is typically available within this timescale.

Median Performance. For each SD pair we calculated the median loss and delay across all probes. The median loss rate was 0 for about 92% of the pairs, with non-zero median loss rates distributed roughly uniformly between 0 and 1. About 95% of SD pairs had median delay less than 300ms. Between sources the median delay was roughly uniformly distributed between 0ms and 80ms.

Routing Data. We also determined BGP egress changes between every probe source and destination pair for the 12 day period, using BGP updates collected by a BGP Monitor from Route Reflectors in every PoP (Point of Presence) in the Tier-1 ISP. The updates allow us to determine the egress (the BGP next-hop)

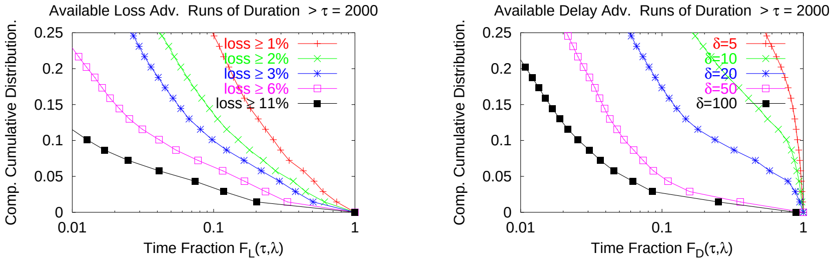

Fig. 3. Available Loss and Delay Advantage: CCDF of Time Fractions $F_L(\tau, \lambda)$, $F_D(\tau, \delta)$ over all SD pairs. Time fraction in which loss advantage exceeds threshold for $\tau = 2000$s. Left: loss. Right: delay.

used by the route reflector for any destination. We mapped each probe source to the nearest PoP. We then used the updates collected from a Route Reflector in that PoP to determine egress changes from the source to every probe destination.

Number of Egress Points. Our framework assumes alternate routes exist. To test this, we computed the distribution of number of distinct egress points which advertise a given destination. The maximum number of egresses seen per destination was 10. For most destinations the number of egresses was constant for most of the duration trace (about 90% of the time). For example, about 64% of the destinations spend most of the time advertised by at least 7 egresses.

Performance Advantage Metric. We found the distribution of $F_L(\tau, \lambda), F_D(\tau, \delta)$ over all SD pairs, for a range of parameters τ, λ, δ. Figure 3 illustrate these for runs lengths greater than $\tau = 2,000$s. The left plot shows the CCDF of $F_L(2000, \lambda)$ for loss advantages at least $(1\%, 2\%, 3\%, 6\%, 11\%)$. Consider 2% loss curve: about 15% of SD pairs spend at least 10% of the time in such runs lasting longer than 2,000s. The right plot shows $F_D(2000, \delta)$ for $\delta = (5, 10, 20, 50, 100)$ms. Consider the 20ms delay curve: about 17% of pairs spend about 10% of their time in such runs lasting longer than 2,000s.

Summary. Our initial results show the potential benefit of Measurement Informed Route Selection: (i) There is significant choice in terms of alternate paths to reach Internet destinations. (ii) Significant benefits in loss (at least 2%) and delay (at least 20ms) last for time periods that can be exploited by routing.

References

1. A. Akella, J. Pang, A. Shaikh, B. Maggs, and S. Seshan. A Comparison of Overlay Routing and Multihoming Route Control. In *Proc. ACM Sigcomm*, Sept. 2004.
2. D. G. Andersen, H. Balakrishnan, M. F. Kaashoek, and R. Morris. Resilient Overlay Networks. In *Proc. 18th ACM SOSP*, pages 131–145, Banff, Canada, Oct. 2001.
3. K.-T. Chen, P. Huang, G.-S. Wang, C.-Y. Huang, and C.-L. Lei. On the Sensitivity of Online Game Playing Time to Network QoS. IEEE Infocom, April 2006.
4. N. Feamster, H. Balakrishnan, J. Rexford, A. Shaikh, and J. van der Merwe. The Case for Separating Routing from Routers. FDNA Workshop, Aug 2004.
5. O. Tickoo, V. Subramanian, S. Kalyanaraman, and K. K. Ramakrishnan. LT-TCP: End-to-End Framework to Improve TCP Performance over Networks with Lossy Channels. IWQoS 2005, June 2005.