On the Potential of Flow-Based Routing in Multihomed Environments

Vlad Manilici ^{1,2} Petros Zerfos ³ Andreas Wundsam^{1,2} Robert Sombrutzki^{1,4} Pablo Vidales¹ Anja Feldmann 1,2 Jatinder Singh 5

Bericht-Nr. 2009-02 ISSN 1436-9915

Abstract

The data rates provisioned by broadband Internet access connections continue to fall short of the requirements posed by emerging applications. Yet the potential of statistical multiplexing of the last mile broadband connections remains unexploited even as the average utilization of these connections remains low. Despite recent work in this area [15, 20], two key questions remain unanswered: a) What is the attainable benefit of broadband access sharing? and b) How much of this benefit is realizable given real-world constraints? In this work we quantify the attainable benefit of a multihomed broadband access environment by proposing and evaluating several flow-based access sharing policies using a custom flow-based simulator. We then analyze how much of the performance benefit is lost due to real-world constraints by migrating from simulations to a test-lab environment employing a wireless network. Our results show that in today's broadband Internet access scenarios, a significant reduction in download times by up to a factor of 3 is achievable.

Introduction



Figure 1.1: WLAN Broadband sharing

Driven by the wide adoption of wireless-enabled broadband access connections and the fact that time-based tariffs have all but disappeared, a new, opportunistic, *lowmobility* communication paradigm [16, 17] has emerged over the past few years, in which residential users *share* their broadband access links over wireless 802.11 connections with neighbors and other nomadic urban users. Since the average utilization of a residential Internet connection still remains low [13], statistical multiplexing of user traffic across multiple broadband access connections over a shared wireless medium can be used to accommodate nomadic users, increase the peak data rates and satisfy high instantaneous traffic demand of bandwidth-intensive applications.

Consider the scenario of Figure 1.1 depicting a community of users accessing the Internet via DSL links provisioned by different ISPs and sharing an ad-hoc wireless LAN. User 1 may experience congestion on her downlink connection due to serving a large download from *mobile* user Ext. 1. However, if User 3's Internet connection has spare bandwidth, User 1 may be able to redirect her traffic to User 3's connection via multihoming.

Realities of current network infrastructure design and operation, such as inability of end users to reserve public IP addresses and lack of ISP support for packet striping, limit the extent to which statistical multiplexing can be leveraged for bandwidth aggregation. To overcome these limitations, several multihoming mechanisms, e.g., [15, 20], have been proposed to route one user's flows via a wireless shared medium through another user's broadband connection. Complementary to these efforts, our work investigates the potential rather than a mechanism for multi-homing and attempts to answer a more fundamental question: what is the *attainable* benefit of flow-based re-routing in multihomed environments?

To answer this question, we employ hybrid simulations and real testbed experiments in various settings, using both synthetic and measurement-based traces for representing user traffic. We developed a flow-level simulator, *FlowSim*, which simulates various flow routing algorithms based on actual flow traces from real networks as input, and deployed a testbed using commodity wireless home gateways and DSL links. Synthetic Web traces generated using Surge [3] and residential access traces from the Munich Scientific Network (MWN) serve as traffic input. Our results show that for bandwidth intensive flows the average download time can be reduced by up to a factor of 3. This emphasizes the substantial potential benefit of flow multiplexing among broadband connections in community networks. We expect further improvements by considering different delay distributions via different providers. But this is beyond the scope of this paper. To summarize, this work makes the following contributions:

- A flow-based simulator and an experimental setup for evaluating multi-homing: reproducible results are obtained and the difference between ideal and progressively realistic experiments is quantified.
- Several multi-homing strategies, ranging from idealistic ones, in which full knowledge of links and flows is assumed, to a practical and realizable one that is suitable for implementation on a router.
- An evaluation of the benefits and potential of the various flow routing strategies in multihomed environments using synthetic and real traffic loads.

Related Work

The areas of local community networking and connectivity sharing have been well explored, and some enterprise solutions [1, 7, 8] have even been deployed in the real world. Settings where neighbors share a DSL connection are common, although not contractually allowed by most ISPs [4]. Bundling connectivity by utilizing multiple links in parallel is often used in access networks as well as backbones, where it is commonly known by the term "trunking" or "channeling". Trunking assumes that the links are terminated at the same devices.

Closely related to flow management, MAR [15] provides a framework for leveraging the diversity of wireless access providers and channels to provide improved wireless data performance for the commuters via a wireless multihomed device that can be deployed in moving vehicles. The packet scheduling architecture requires the presence of a proxy server as a concentrator and support from the ISP side.

The MultiNet project [6] proposes a multihomed single-card device by means of virtualization. In MultiNet the wireless card is continuously switched across multiple networks. In contrast to physically accessing multiple networks, our work brings forward the idea of routing flows across the network in order to achieve better performance.

Solutions for exploiting multiple paths between hosts at higher network layers have been proposed as well. At the transport layer, pTCP [10] is an extension of TCP that allows for bandwidth aggregation in multihomed devices. The protocol has been evaluated through simulations. Its mechanism is based on packet buffering and appears to us to require a huge management overhead. It is unclear what the performance benefits in a real wireless environment would be.

Habib et al. [9] propose the resurrection of the session layer for striping data from a single connection over several links. Implementing the proposal, however, requires extensive changes at the OS or the application level, which is also a requirement for the Stream Control Transmission Protocol [18] or the IPv6 shim6 layer [2].

Recent work by Thompson et al. [20] evaluates a framework for end-host multihoming with flow routing based on RTT measurements and prediction of the flow sizes. However, the evaluation is limited to a proof of concept system consisting of two nodes and one specific flow routing methodology proposed by the authors. Another example of flow management is the work presented in [19]. The authors study the feasibility of improving performance by increasing path diversity through flow-based path switching. This work was evaluated using a wired experimental setup but no evaluation in wireless environments has been reported. Papadopouli and Schulzrinne [14] propose the integration of channels with largely different QoS characteristics for serving streams with adaptable bandwidth. Their prototype operates as a multicast application with variable bandwidth. Standard applications such as WWW and email are unable to take advantage of the proposed system. Lad et al. [11] propose a high-level architecture for sharing connectivity in a coalition peering without discussing its realization or presenting a performance evaluation.

Complementary to the above approaches, the work reported in this paper offers an evaluation of multihoming in real wireless environments. There are a plethora of papers that have reported on the difficulties posed by wireless environments and their impact on urban networks, see [17] or [5] to mention a few.

Setup

To evaluate the attainable benefit of flow-based routing in multihomed environments, we use a combination of simulated and real testbed experiments in various settings. We employ as traffic input user traces from the border router of the Munich Scientific Network to represent real user behavior and synthetic workloads generated by the Surge [3] Web traffic generator. The latter is required to capture the reaction of the application and users to potential performance improvements, which is difficult to assess via trace-based evaluations.

3.1 Flow Routing Strategies

In this paper we investigate the following basic flow routing strategies to obtain a baseline of the possible benefits. This will serve as the basis for developing even better strategies in future work:

- **Direct:** the current practice in residential broadband connections: every flow is routed on the direct link i.e., the DSL connection of the user originating the flow.
- **FatPipe:** an "ideal case", which bundles all DSL links into one fat pipe with bandwidth equal to the sum of the capacities of the individual connections. The only restriction is that no flow may exceed the bandwidth of the originating DSL connection. This strategy cannot be implemented in a real network and is only supported by the simulator.
- **MinLargeFlows:** assigns each new flow to the link which, at that time, carries the least number of flows that have already transmitted some number of bytes, say 8 KB. This strategy tries to minimize the number of bulky flows that share any of the DSL links.

Moving flows from one link to another is difficult as it requires either support by all end-systems or transparent movement of the TCP state. Therefore, we do not allow re-routing a flow once it has been assigned to a certain link.

3.2 Flow Routing Testbed: FlowRoute

A testbed for conducting realistic experiments has been deployed at our site, emulating the setup in Figure 1.1. Four nodes have been deployed on one floor in our building,

coexisting with other wireless networks, as in real residential scenarios. The distribution of the testbed nodes was planned to achieve wireless connectivity among them. Each node consists of a wireless router with IEEE 802.11a/b/g interfaces and a client machine which is also equipped with a wireless interface. The routers are directly connected to the Internet via 2 Mbps DSL lines and to the corresponding client machines using an Ethernet interface. The client based flow routing system FlowRoute allows us to evaluate the benefits of flow routing in real-world testbeds. Hence, effects not easily captured in a simulator, such as the impact of different shared interconnection mediums, wired vs. wireless Ethernet, etc., can be studied.

For some of the experiments we need the flexibility to tune the bandwidth of the access lines to the Internet. This functionality is provided by using the NistNet network emulator [12]. The NistNet emulator is installed on a separate machine that is accessible by all testbed nodes via the management network.

The software prototype that enables flow routing on the client machines, FlowRoute [21], consists of three main components: a preloaded shared library, called libConnect, which intercepts all client access to the Berkeley socket layer and delegates routing decisions. Routed, the unique central decision making module, has a global view on the link loads and makes the actual routing decisions based on the flow routing strategies previously described. The third component, called Proxy, is responsible for re-directing the flows between a client and its destination and for informing the routed about the load of all shared DSL connections.

3.3 Simulator: FlowSim

We have developed a simulator, called FlowSim, to evaluate in a scalable way the potential of flow routing strategies using flow-level traces as input, captured from real networks or testbed experiments. Popular packet level simulators such as ns-2 or SSFNet are less suitable for our purpose, given the large number of flows considered here. To validate the simulator, we reproduce the testbed experiments within FlowSim and compare the results.

We can use FlowSim to simulate network setups such as those shown in Figure 1.1, with various number of users, different capacities and delays for uplink and downlink directions and flow routing algorithm parameters. The interconnection capacity between DSL subscribers is not supposed to be a bottleneck and thus modeled as infinite. As such, the simulator cannot distinguish if the community is interconnected via a wireless or wired network. FlowSim consists of two main components: the router and the scheduler. The *router* identifies interactive and bandwidth-limited flows. Bandwidth-limited flows are further classified by the direction (called dominant direction) in which they exceed a threshold into downlink-, uplink- and both-limited flows. Once a flow has been classified, it is routed using one of the flow routing strategies.

The role of the *scheduler* is to distribute the available bandwidth among competing flows, while approximating the dynamics of the TCP protocol. In order to achieve reasonable performance, it uses a fluid TCP model operating on discrete time slots, rather than on a per event basis. For details and download see: *http://www.net.t-labs.tu-berlin.de/~vlad/FlowRoutingSoftware*.

The simulation results do not show a significant difference when using smaller time slots than 1/10 second, the current setting. The distribution of the available up and downstream bandwidth is done in a fair manner between all flows that use one

link. Each bandwidth-limited flow exercises a TCP-style slow start before it is able to attain its fair share of the bandwidth. Connect and close delays are modeled as well, while ignoring packet losses. Comparisons with results obtained from the testbed for the same set of traces used as input show that the approximations are reasonable. For each bandwidth-limited flow, its fair share of the bandwidth is computed based on its dominant direction(s). The bandwidth share for the non-dominant direction is then set in proportion to the transmitted bytes. The results are again scaled to the available bandwidth.



Figure 3.1: Download bandwidth on simulator and testbed for the MWN experiment without routing

Figure 3.1 plots the downlink bandwidth from an experiment on the testbed and on the simulator for the same configuration. On the X axis we show the time and on the Y axis the normalized bandwidth. Data from each of the 4 clients is plotted in a separate "swim lane". We superimpose the two curves in different shades, one for the testbed and one for the simulator. With the exception of few outliers and lags, the two curves match closely, validating our FlowSim simulation.

Results

In this section, we begin by analyzing an experiment which uses artificial Web workload. Next we replay real-world TCP traces in a second experiment. For both of them, we analyze the benefits of flow routing using the testbed and compare them with the performance predicted by the simulator.

4.1 Synthetic Web Workload

We use Surge [3] to generate a synthetic workload representative of Web traffic and updated its parameters to reflect characteristics (e.g., the median and mean HTTP object size) of the current Web workload mix, including Web 2.0 and P2P, as observed in the MWN traces. As Web server software, we use the popular apache2 Web server. To impose a reasonable load, we use four Surge instances per household (host). This Web workload results in an average utilization of 0.39 Mbps per DSL link.

4.1.1 Testbed Results

Figure 4.1 plots the complementary cumulative distribution function (CCDF) of the flow durations on a logarithmic X-axis for an experiment with NistNet and a wired interconnection network. One can clearly see the benefit of flow routing, since durations with flow routing are clearly shorter than without. Yet, we also see that there is some overhead imposed by flow routing for short flows.

To quantify the achievable benefit, we compute the ratio of the durations, more precisely, we compute duration(direct) / duration(routed). Larger values denote better performance under flow routing. Considering all flows, the mean benefit is 1.11 with a median of 0.90. At first, this might seem a bit disappointing. Yet, there are many short flows and they in this scenario all suffer from the overhead of our prototype flow routing implementation. But when we only consider flows with a duration larger than 0.5 seconds the mean improvement increased to 2.47 (median 1.83). This implies that there is a significant benefit for bulky flows.

We next switch from NistNet to using the actual 2Mbps DSL lines, with the wired interconnection network. We observe that the results improve slightly in comparison to the emulated network. For all flows, the mean benefit is 1.20 (median 0.95). Flows that last longer than 0.5 seconds experience an improvement of 2.49 (median 2.03). The differences are explainable by small inaccuracies in NistNet.



Figure 4.1: CCDF of flow durations with and without routing for the Surge Web workload

Finally, we use the well connected wireless network in 802.11a mode as our interconnection network within the community. Somewhat surprisingly, the overall results improve slightly when compared to using a wired network. The mean (median) improvements for longer flows are 2.61 (2.11) and for all flows 1.27 (1.0). This shows that the wireless network is not the bottleneck. We reroute only 75% of the flows which imposes a load of less then 5 Mbs on the wireless network. Overall, we see that significant improvements are possible by taking advantage of multihoming.

4.1.2 Simulator Results

We investigate the upper bound of the routing performance by using the simulator on the traces gathered from the Web workload experiments. This allows us to estimate the potential benefits and compare the performance prediction of the simulator against the performance improvements observed in the testbed. The simulator displays some deviations in flow duration, which are in part due to the simulator's assumption that bandwidth is shared fairly between flows. This assumption is known to work well on average but not on smaller time scales. As such, it is not surprising that the size of the deviations decrease as flow durations increase. In total, the ratio of the durations for bandwidth-limited flows in the simulation vs. the testbed experiment has a mean of 1.07 and a median of 0.97. This indicates that, while for individual flows the predicted performance does not agree with the performance seen in the experiment, the overall results of the simulation match the experiment quite well.

We then compare the performance as predicted by the simulator with the performance of the actual experiment using the MinLargeFlows routing policy. The simulator reports a mean(median) improvement for flows longer than 0.5 seconds of 2.59 (2.09). In the synthetic Web workload experiment with the same strategy, the inter-arrival times between Web connections are shorter than with the Direct policy, and the load is also higher. These predictions match the actually observed benefits of 2.47 very well. Interestingly, the predicted mean improvement by the simulator for FatPipe is only slightly better than for MinLargeFlows having 2.79 vs. 2.24 mean improvement. This confirms that the achievable benefit for this scenario is indeed limited by the traffic properties of this specific Web workload.

4.2 Trace-Based Experiments

4.2.1 Workload

Using traces of real traffic to drive simulations as well as testbed experiments allows repeatability of results under realistic loads. We use connection-level summaries of traffic traces captured at the border router of the Munich Scientific Network, MWN. The MWN provides 10 Gbps Internet upstream capacity to roughly 50,000 hosts at two major universities including student dormitories, with the daily traffic amounting to 3-6 TB. Ample bandwidth is available to all users of this network via high-speed local area networks, thus Internet usage is not impaired by any bandwidth limitations of the access links.

In a typical 24 hour workday trace from April 24th 2007 we identify approximately 37 million flows, out of which 21.1 million have incomplete connection information records, a typical component of todays traffic mix, such as SYN attempts (56%), rejected connections (27%), in progress connections (8%), and missing information (9%). For our experiments, we consider the remaining 15.9 million flows, through which 641 GB (182 GB) were transferred from (to) the Internet. These volumes correspond to an average bandwidth utilization of 60 Mbps (17 Mbps) downstream (upstream).

From this large collection of flow data we selected a subset of flows, referred to as *MWN* flow trace, that originated from IP addresses that are assigned to student residences, to ensure that we consider only genuine residential traffic. Students are provided Internet access only via 28 NAT gateways. The traffic via those NAT gateways imposes a load of about 0.34 Mbs on the upstream and 5.74 Mbs on the downstream. We assigned these 28 NAT gateways to our DSL links randomly. Using such a dense trace allows us to investigate times with peak traffic while keeping the durations of our experiments relatively short (30 minutes).

To quantify the performance of our flow re-routing against different traffic loads, we run experiments on the testbed using the full MWN trace in addition to a random subselection of 1/3 of the total number of flows (reduced set).

4.2.2 Performance Analysis

Similar to the Surge experiment, we concentrate on bulky flows longer than 0.5 seconds. Shorter flows display the overhead imposed by our prototype. Figures 4.2(a) and 4.2(b) show the scatter plot of flow durations. Each point represents a flow, with the duration of the direct routing policy on the X axis and the duration of the re-routed policy on the Y axis. The overlaid contour lines plot the density of measurement points in one region. More points below the x = y diagonal imply more flows that benefit from routing. The density peak for the experiment with the full MWN trace is significantly below the diagonal, whereas the reduced MWN trace has a peak at about the level of the diagonal.

The experiment achieves improvements in the flow durations in the former case by a mean factor of 3.02 (median 1.98) and in the latter case by a factor of 2.04 (median



(a) Full MWN trace

(b) 1/3MWN trace

Experiment	Policy	Mean	Median
Surge (Ether, Nist)	MinLargeFlows	2.47	1.83
Surge (Ether, DSL)	MinLargeFlows	2.49	2.03
Surge (Wlan, Nist)	MinLargeFlows	2.61	2.11
Surge (Simulator)	MinLargeFlows	2.59	2.09
Surge (Simulator)	FatPipe	2.79	2.24
1/3MWN (E., N.)	MinLargeFlows	2.04	1.02
MWN (Ether, Nist)	MinLargeFlows	3.02	1.98
MWN (Simulator)	MinLargeFlows	2.47	1.99
MWN (Simulator)	FatPipe	3.73	2.40

Figure 4.2: Scatter plot with superimposed contour

Figure 4.3: Bulky flow (>0.5 sec.) speed boost

1.02). When considering all flows including the short ones, the mean/median values are for the full MWN trace 2.21/1.12 and 1.17/0.94 for the reduced set. See Figure 4.3 for a summary.

We plot the logarithmic density (PDF) of the flow durations ratio: duration(direct)/duration(routed) in Figure 4.4. We note that the shift towards right (values larger than 1) is much more pronounced for the full trace experiment. This means that flow routing performs better in regions close to the link congestion. Figure 4.5 shows the CCDF of flow durations with and without flow routing. From the Figure, we observe that the flows exhibit shorter durations when flow routing is employed.

The experimental results on the testbed for the standard MinLargeFlows routing algorithm compare well with the simulation. The median ratio of flow durations is 1.98 on the testbed and 1.99 in the simulator. Mean ratio of 3.02 is better on the testbed than 2.47 achieved on the simulator. This result is partly explained by a larger jitter on the testbed. Evaluating the FatPipe policy on the simulator, we achieve flow duration improvement by a mean factor of 3.73 (median 2.40).

The effect of the shared interconnection medium, wireless or wired, has also been evaluated on the testbed. Average and median flow durations vary by less than 15%

between a wireless shared medium with good connectivity and a wired one. The DSL emulation through NistNet achieves less than 3% average and median variation in flow lengths. Figure 4.6 compares the density distribution of the flow durations when using DSL lines or NistNet for the Internet connection and Ethernet or Wireless for the testbed backbone.



Figure 4.4: Flow duration densities with different workloads



Figure 4.5: CCDF of flow durations with and without routing for the MWN trace-based workload



Figure 4.6: Density of flow durations on different network media for the MWN experiment with routing

Conclusion

Broadband access connection sharing using wireless 802.11 technologies is emerging as an alternative approach towards nomadic broadband connectivity. Flow-based routing in conjunction with multihoming plays a key role in enabling such access sharing scenarios, by increasing the multiplexing of broadband connections. For this reason, it becomes important to evaluate the potential benefits of flow-based routing in multihomed environments.

In this work, we evaluate several flow re-routing strategies using synthetic workloads and captured user traces as traffic input. Our encouraging results show improvements of up to a factor of 3 in the download times of bandwidth-limited flows, with very small overhead. Bandwidth-demanding flows and an unbalanced distribution of load among the users of the sharing community particularly benefit from the flowbased routing approach. More importantly, full knowledge of flow characteristics is not needed, as the performance of MinLargeFlows exhibits.

Bibliography

- [1] A. Akella, S. Seshan, and A. Shaikh. Multihoming performance benefits: An experimental evaluation of practical enterprise strategies. In *USENIX*, 2004.
- [2] M. Bagnulo and E. Nordmark. Shim6: Level 3 Multihoming Shim Protocol for IPv6. draft-ietf-shim6-proto-08 (work in progress), April, 2007.
- [3] P. Barford and M. Crovella. Generating representative web workloads for network and server performance evaluation. In *SIGMETRICS/PERFORMANCE*, 1998.
- [4] BTFon. WiFi for everyone. website: http://btfon.com.
- [5] V. Bychkovsky, B. Hull, A. K. Miu, H. Balakrishnan, and S. Madden. A Measurement Study of Vehicular Internet Access Using In Situ Wi-Fi Networks. In *MOBICOM*, 2006.
- [6] R. Chandra, V. Bahl, and P. Bahl. Multinet: Connecting to multiple ieee 802.11 networks using a single wireless card. In *INFOCOM*, 2004.
- [7] D. Goldenberg, L. Qiu, H. Xie, Y. R. Yang, and Y. Zhang. Optimizing cost and performance for multihoming. In *ACM SIGCOMM*, 2004.
- [8] F. Guo, J. Chen, W. Li, and T. cker Chiueh. Experiences in building a multihoming load balancing system. In *INFOCOM*, 2004.
- [9] A. Habib, N. Christin, and J. Chuang. Taking Advantage of Multihoming with Session Layer Striping. In *IEEE Global Internet*, 2006.
- [10] H.-Y. Hsieh and R. Sivakumar. A transport layer approach for achieving aggregate bandwidths on multi-homed mobile hosts. In ACM MOBICOM, 2002.
- [11] M. Lad, S. Bhatti, S. Hailes, and P. Kirstein. Coalition-Based Peering for Flexible Connectivity. In *PIMRC*, 2006.
- [12] Nistnet home page. http://www-x.antd.nist.gov/nistnet/.
- [13] A. M. Odlyzko. Data networks are lightly utilized, and will stay that way. *Review* of Network Economics, 2003.
- [14] M. Papadopouli and H. Schulzrinne. Connection sharing in an ad hoc wireless network among collaborating hosts. In NOSSDAV, 1999.
- [15] P. Rodriguez, R. Chakravorty, J. Chesterfield, I. Pratt, and S. Banerjee. Mar: a commuter router infrastructure for the mobile internet. In *MobiSys*, 2004.

- [16] N. Sastry, J. Crowcroft, and K. Sollins. Architecting citywide ubiquitous Wi-Fi access. In ACM HotNets, 2007.
- [17] M. Solarski, P. Vidales, O. Schneider, P. Zerfos, and J. P. Singh. An experimental evaluation of urban networking using IEEE 802.11 technology. In *Proc. IEEE OpComm*, 2006.
- [18] R. Stewart and C. Metz. SCTP: new transport protocol for TCP/IP. *IEEE Internet Computing*, 2001.
- [19] S. Tao, K. Xu, Y. Xu, T. Fei, L. Gao, R. Guerin, J. Kurose, D. Towsley, and Z.-L. Zhang. Exploring the performance benefits of end-to-end path switching. In *IEEE ICNP*, 2004.
- [20] N. Thompson, G. He, and H. Luo. Flow Scheduling for End-host Multihoming. In *IEEE Infocom*, 2006.
- [21] A. Wundsam. Connection sharing in community networks how to accomodate peak bandwidth demands. Diplomarbeit, TU Munich, Germany, April 2007.