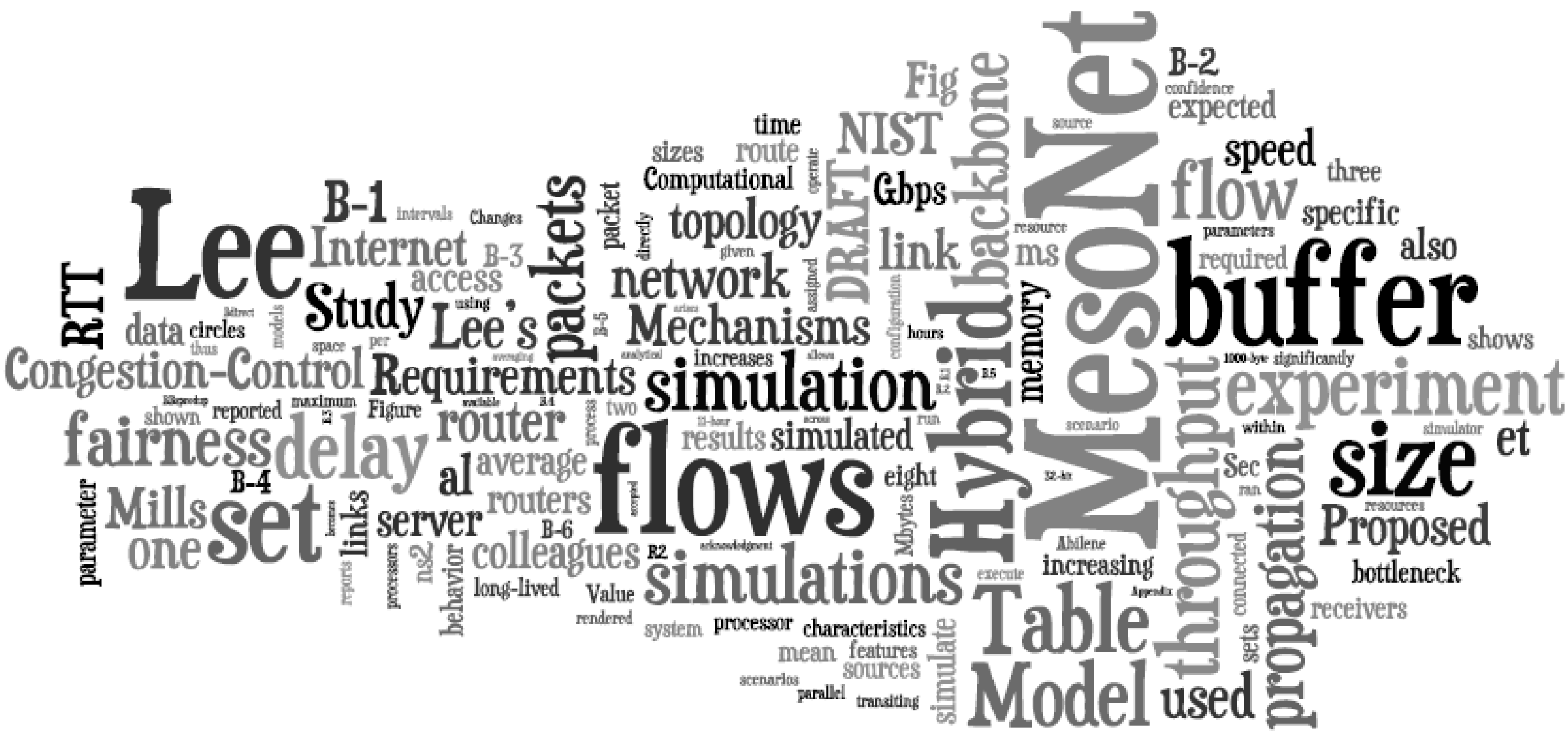
Appendix B – Computational Requirements: MesoNet vs. Hybrid Model



Appendix B Computational Requirements: MesoNet vs. Hybrid Model

Junsoo Lee and colleagues [71] observe that packet-level network simulations, such as *ns2* [79], require substantial computational resources for large-scale simulations and also entail so many parameters that it becomes difficult to understand the influence of specific factors on overall system performance. Lee also points out that aggregate fluid-flow models [e.g., 73] address these shortcomings but can only capture steady-state behaviors averaged over long time intervals. Lee describes a hybrid modeling framework that continuously approximates discrete variables by averaging over short intervals of time. Constraining the averaging interval allows generation of significant events, such as packet drops and related adjustments in congestion windows. Like MesoNet, Lee's hybrid framework aims to simulate a manageable parameter space and thereby illuminate the influence of specific factors on system behavior, while reducing computational requirements.

In this appendix, we use MesoNet to replicate a simulation experiment reported by Lee and colleagues [71]. The specific experiment conducted by Lee uses a hybrid model to simulate an 11-hour scenario involving 30 long-lived flows transmitting data across a subset of the Abilene topology. Lee reports that this scenario was infeasible using *ns2* because his available computer had only 512 Mbytes of memory, which proved insufficient. Replicating this experiment with MesoNet serves three purposes: (1) to illustrate that MesoNet can simulate a scenario found to be infeasible with a commonly used network simulator, (2) to show that MesoNet produces behavior similar to Lee's hybrid simulator (which was validated against predictions from a widely accepted analytical model) and (3) to compare computational requirements of MesoNet against reported computational requirements for Lee's hybrid model. In the process of achieving these objectives, we raise confidence in MesoNet and we demonstrate that hybrid network models hold promise as replacements for discrete-event network simulations.

We begin in Sec. B.1 by describing our experiment design. Where applicable, we identify and justify specific differences in the MesoNet experiment setup and the configuration used by Lee. In Sec. B.2, we outline how we executed the simulations and how we collected the required data. Next, in Sec. B.3, we present results regarding flow behavior. In Sec. B.4, we compare our findings with those reported by Lee. We conclude in Sec. B.5.

B.1 Experiment Design

The fundamental purpose of the experiment designed by Lee and colleagues [71] was to investigate the effect of buffer size on relative fairness among long-lived TCP flows that transit network routes with differing propagation delays and a shared bottleneck link. The expected result is that smaller buffer sizes allow propagation delay to be the dominant component of round-trip time (RTT), which implies that flows transiting longer paths should receive lower throughputs than flows transiting shorter paths. As buffer size increases, queuing delay becomes the dominant component of RTT, which implies that the throughput of all flows will come closer together. This expectation arises from a widely accepted analytical formula to predict TCP throughput, which generally underestimates the fairness ratio, as confirmed by *ns2* simulations with small network topologies. Lee and colleagues show that their hybrid model yields the expected behavior in a large network based on the original Abilene topology. We aim to show that MesoNet also exhibits the expected behavior in the same topology used by Lee. This will increase our confidence in MesoNet. We will also be able to compare resource requirements of MesoNet against reported requirements for the hybrid model.

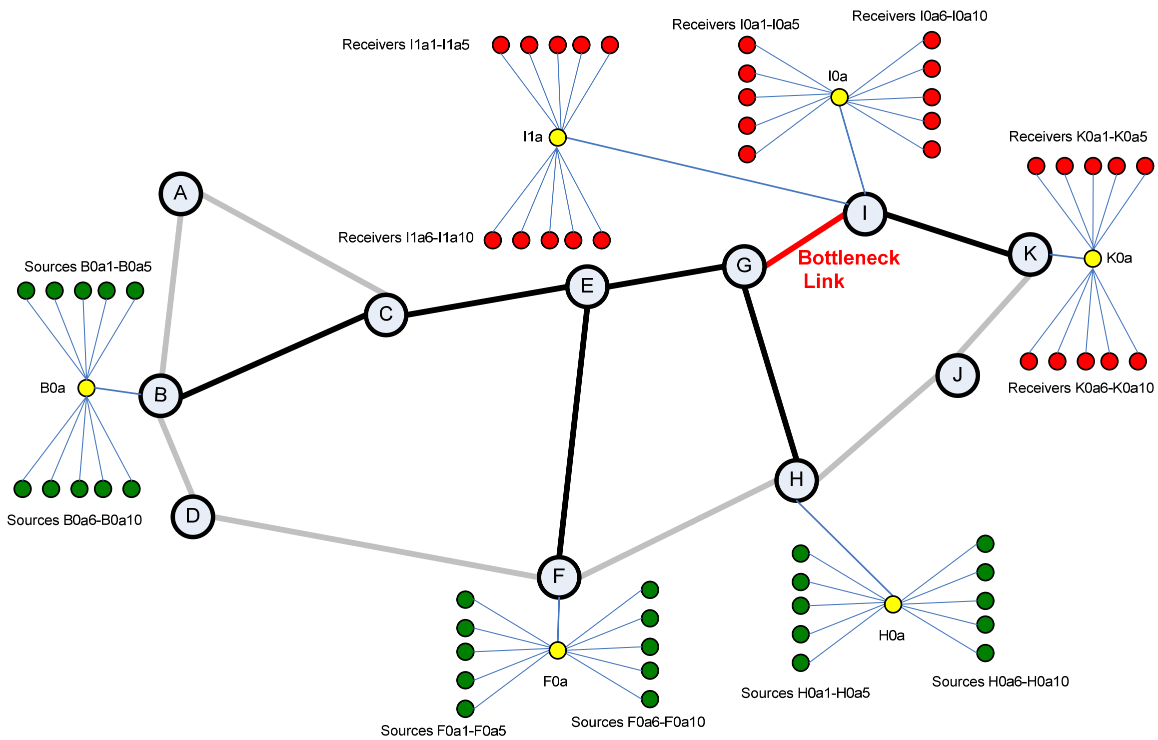


Figure B-1. Experiment Topology

Fig. B-1 shows the network topology we simulated. The backbone is derived from the original Abilene topology, as given by Lee [71]. The backbone consists of 11 routers (grey circles designated A-K in our topology) that each serve a different location within the United States. The backbone routers are connected by 14 bidirectional links. We assigned a propagation delay to each link, as specified in Table B-1. We used the same propagation delay for each direction on a given link (e.g., links A→B and B→A both have 17 ms propagation delays). We adopted the propagation delays used by Lee, except that we rounded to the nearest millisecond.

The seven grey links in Fig. B-1 are not used in this experiment because Lee focused on three sets of flows, where each set transits a different route and where the routes share a bottleneck link (G-I), rendered in red in Fig. B-1 (remaining links used by flows are shown in black). Flow sources are rendered as green circles in Fig. B-1 and flow receivers are rendered as red circles. As required by MesoNet, each source and receiver must be connected to an access router (yellow circles in Fig. B-1), and for this experiment each access router is connected directly to a backbone router (i.e., there are not Point of Presence routers in the topology). This differs from Lee's configuration, where sources and receivers connected directly to backbone routers.

Table B-1. One-Way Propagation Delay on Each Link in the Simulated Topology

source	destination	prop. delay (ms)
A	B	17
A	C	26
B	C	25
B	D	8
C	E	11
D	F	32
E	F	16
E	G	9
F	H	20
G	H	11
G	I	4
H	J	16
I	K	20
J	K	4

Table B-2 reports relevant characteristics for each set of simulated flows. The first set of flows has 10 sources under access router H0a. Each source transmits to one of 10 receivers located under access router I0a. In MesoNet, packets transiting access routers experience queuing delay but no propagation delay; a packet experiences propagation delay only when crossing backbone links. MesoNet sends data packets for these 10 flows over backbone route (H-G-I) and returns acknowledgments[1] over the reverse route (I-G-H); thus, the round-trip propagation delay between a data packet and its acknowledgment is 30 ms (twice the 15 ms propagation delay on the route). Similar information is provided for two additional sets of 10 flows. As the backbone route increases from two to three to five hops with each set of flows, relative propagation delay approximately doubles. Table B-2 highlights the bottleneck link shared by all flows.

Lee's experiment simulates backbone links operating at 10 Gbps. While Lee does not report the speed of simulated sources and receivers, we assume their speed is sufficient to achieve more than 10 Gbps when 30 flows are aggregated across the bottleneck link. Lee allows each flow to start at a random time, uniformly distributed

[1] Note that Lee's hybrid model does not specifically simulate acknowledgments. This represents another difference with MesoNet. Also, in MesoNet, packets have no specific size, so each acknowledgment consumes one packet of buffer space, which is also the buffer space consumed by each data packet.

over one second, and then the flows continue transmitting (as congestion permits) for just over 11 hours. Lee repeats this simulation six times, while increasing buffer sizes in increments of 25 x 10^3 (1000-byte) packets from 25 x 10^3 to 150 x 10^3.

Table B-2. Characteristics of Three Flow Sets Simulated in the Experiment

sets	# of flows	prop. delay	src/dest	route (symmetric)
set one	10	15 ms	H0a/I0a	H-G-I
set two	10	29 ms	F0a/I1a	F-E-G-I
set three	10	69 ms	B0a/K0a	B-C-E-G-I-K

Table B-3. MesoNet Parameter Settings for the Experiment

Parameter	Value
M	60 x 10^3
MI	660
R1	1250 p/ms
BBspeedup	1
R2	1
R3	1
Bdirect	1
QszAlg	Directly Set
Hfast	80
Flow Start	uniform (0..1s)

To match Lee's conditions, we assigned MesoNet parameters as specified in Table B-3. MesoNet assigns a speed to each router in the topology. Parameter **R1** specifies that backbone routers process 1250 packets/millisecond. Setting related parameters (**BBspeedup, R2, R3 and Bdirect**) to one ensures that all routers operate at the same speed. Assuming 1000-byte packets, each of the backbone and access routers then operate at 10 Gbps (1250 packets/milliseconds x 1000 milliseconds/second x 1000 bytes/packet x 8 bits/byte). We assigned sources and receivers to operate at (**Hfast** =) 80 packets/millisecond, which equates to a maximum of 640 Mbps (80 packets/milliseconds x 1000 milliseconds/second x 1000 bytes/packet x 8 bits/byte). When 30 flows cross the bottleneck, the potential demand of 19.2 Gbps (640 Mbps/flow x 30 flows) exceeds the available link capacity. We measured system state every (**M** =) 60 x 10^3 milliseconds (i.e., once a minute) and we run the simulation for (**MI** =) 660 measurement intervals (i.e., for 660/60 = 11 hours). We set the buffer size in each router directly to the appropriate value for each repetition: we vary buffers from 25 x 10^3 to 200 x 10^3 packets[2] in 25 x 10^3 packet increments. Table B-4 gives the domain view of the parameter settings shown in Table B-3.

Table B-4. Domain View of the Simulated Network Characteristics

Characteristic	Value(s)
Measurement Interval Size	60 seconds
Simulation Duration	11 hours/run
Backbone Router Speed	10 Gbps
Access Router Speed	10 Gbps
Router Buffer Sizes	25 x 10^3 – 200 x 10^3 packets
Maximum Host Speed	640 Mbps
Max. Link Demand on G-I	19.2 Gbps

For each simulation run, we make the same measurements taken by Lee. Specifically, we measure throughput fairness ($\mathbf{FR}_{i,j}$) and RTT fairness ($\mathbf{RR}_{i,j}$). In equations (1) and (2), *i* and *j* (*i* not equal to *j*) each denote a specific set of flows. Thus, we average either the throughput (1) or RTT (2) for each set and then take the ratio of each pair of sets, where the denominator is chosen from the set expected to have the lower value in a given pair.

$$\mathbf{FR}_{i,j} \equiv \frac{\mathbf{mean}(\mathbf{Throughput}_i)}{\mathbf{mean}(\mathbf{Throughput}_j)} \tag{1}$$

$$\mathbf{RR}_{i,j} \equiv \frac{\mathbf{mean}(\mathbf{SRTT}_i)}{\mathbf{mean}(\mathbf{SRTT}_j)} \tag{2}$$

[2] While Lee simulated only six buffer sizes, we simulate eight buffer sizes because we had access to a server with eight processors. We ran the eight simulations in parallel on the server.

B.2 Experiment Execution and Data Collection

We ran eight, parallel instances (one per buffer size) of the MesoNet simulator, where each instance ran within one 32-bit SLX process on one processor within a computation server, configured as shown in Table B-5. Table B-6 reports the computation and memory resources required for each simulation.

Table B-5. Configuration of Compute Server for Simulations

Property	Characteristics
Operating System	Microsoft Windows Server 2003 R2 x64 Edition SP2
Server	Dell Server PE6950
Server Memory	32 Gbytes
Processor Chip	Four Dual-Core AMD Opteron Processors 8222SE
Processor Speed	3 GHz
Total Processors	(4 x 2 =) 8
Simulation Environment	SLX 32-bit Version Release 2.3 (PR229)

Table B-6. Resource Requirements for Simulations

Buffer Size (packets)	Processor Hours	Memory (Mbytes)
25000	80.33	34
50000	80.41	41
75000	99.23	47
100000	123.21	55
125000	102.19	60
150000	110.79	67
175000	129.08	73
200000	122.39	80

Every minute, we measured the instantaneous (60-second) average throughput and smoothed RTT seen on each of the 30 long-lived flows. This enabled us to collect 660 samples per metric per flow over an 11-hour simulation. We then averaged (660 x 10 =) 6600 samples to generate a mean throughput for each set of 10 flows. We similarly obtained an average RTT for each set of flows. We used these averages to form the fairness ratios defined in equations (1) and (2).

B.3 Results

Fig. B-2 plots the changing RTT fairness ratios as buffer size increases. Fig. B-3 shows variation in throughput fairness. These two plots exhibit the expected convergence in fairness as buffer size increases. The curves for throughput fairness bump up slightly, as buffer size moves from 100 x 10^3 to 125 x 10^3 packets, before continuing the downward trend. This bump arises from a dip in average throughput for flow set number 3, coupled with a slight increase in average throughput for flow set number 2, as shown in Fig. B-4. We attribute these fluctuations to randomness arising from using a single repetition of the simulation to generate each set of data points.

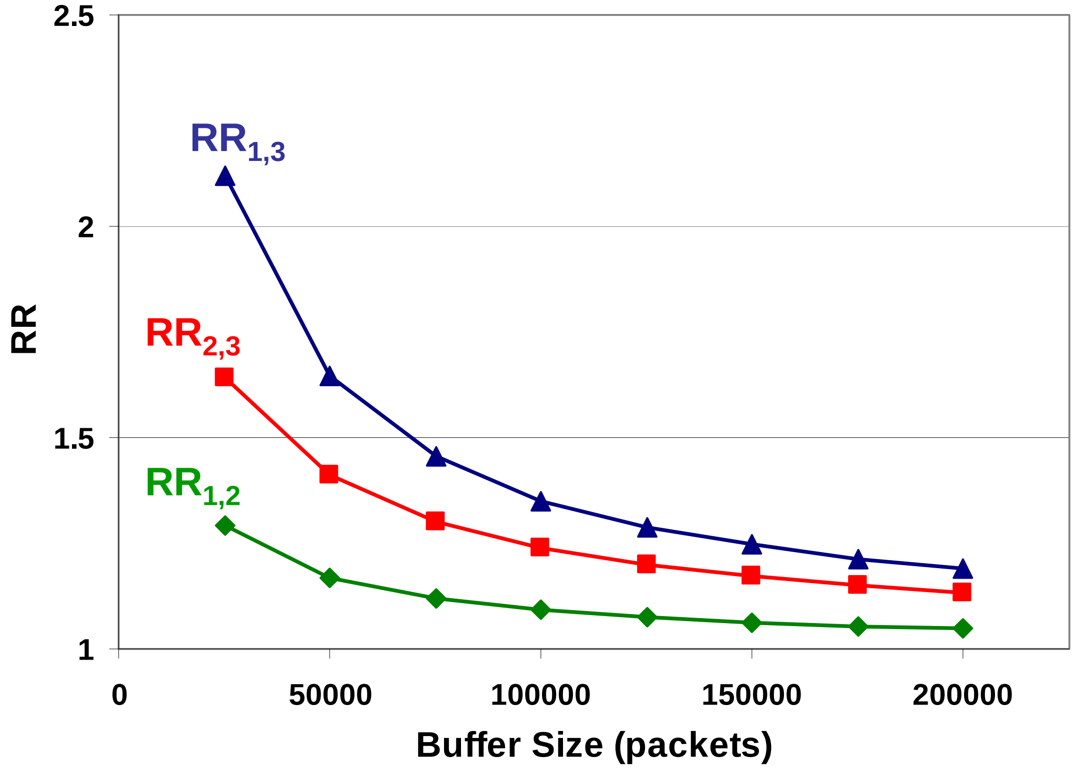


Figure B-2. Changes in RTT Fairness with Increasing Buffer Size

B.4 Discussion

As expected, mirroring the results of Lee and colleagues, RTT and throughput fairness converge with increasing buffer size. These results enhance our confidence in MesoNet. Further, Table B-6 shows that we can execute the required MesoNet simulations in under 100 Mbytes of memory, whereas Lee and colleagues found that they could not execute these simulations using *ns2* in a machine with 512 Mbytes of memory. On the other hand, running these MesoNet simulations took just under 5 ½ days, the time required by the maximum simulation run (buffer size of 175 x 10^3 packets). From reading the information provided by Lee and colleagues, we would expect the hybrid model, running all eight simulations in parallel, to complete in less than one day. This comparison of processing requirements shows that hybrid models have the potential to significantly accelerate simulation in scenarios such as the one here.

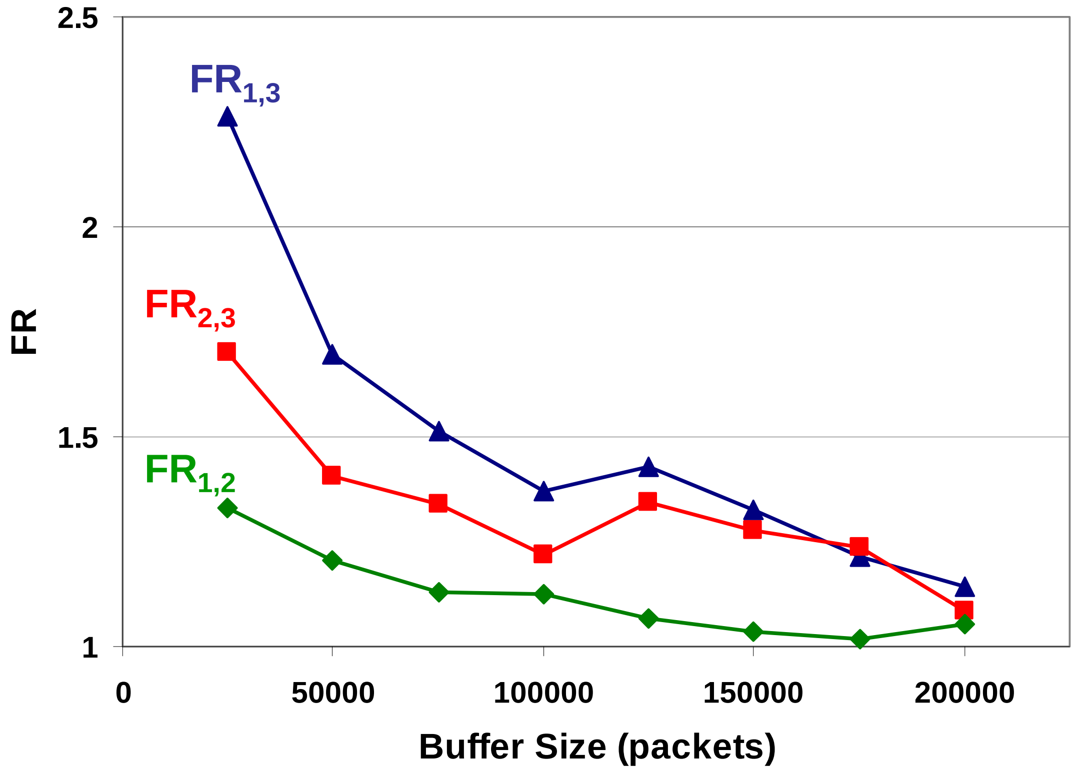


Figure B-3. Changes in Throughput Fairness with Increasing Buffer Size

When considering the use of a hybrid model for other scenarios, such as those described throughout this report, we note that Lee's model would need to be extended to include many features not currently present. Such features include: multiple routing tiers and router classes, arriving and departing flows, variety in flow types, many more measurements, connection establishment procedures, and support for arbitrary topologies. In principle, we expect that such features could be incorporated into a hybrid model. Further, we suspect that such a hybrid model would execute more swiftly that our MesoNet simulation. Confirming these hypotheses requires future work.

B.5 Conclusions

In this section, we used MesoNet to repeat an experiment conducted by Lee and colleagues. We compared the results obtained by Lee with MesoNet results, finding general agreement. We also demonstrated that MesoNet requires significantly fewer memory resources than *ns2*. Further, we showed the Lee's hybrid model could likely simulate scenarios involving long-lived flows at a rate more than five times faster than MesoNet, which relies on discrete-event simulation. Further work remains to extend Lee's hybrid model with features needed to conduct the full suite of experiments used in the remainder of our study. We believe hybrid modeling holds the promise of significantly reducing resource requirements for network simulations.

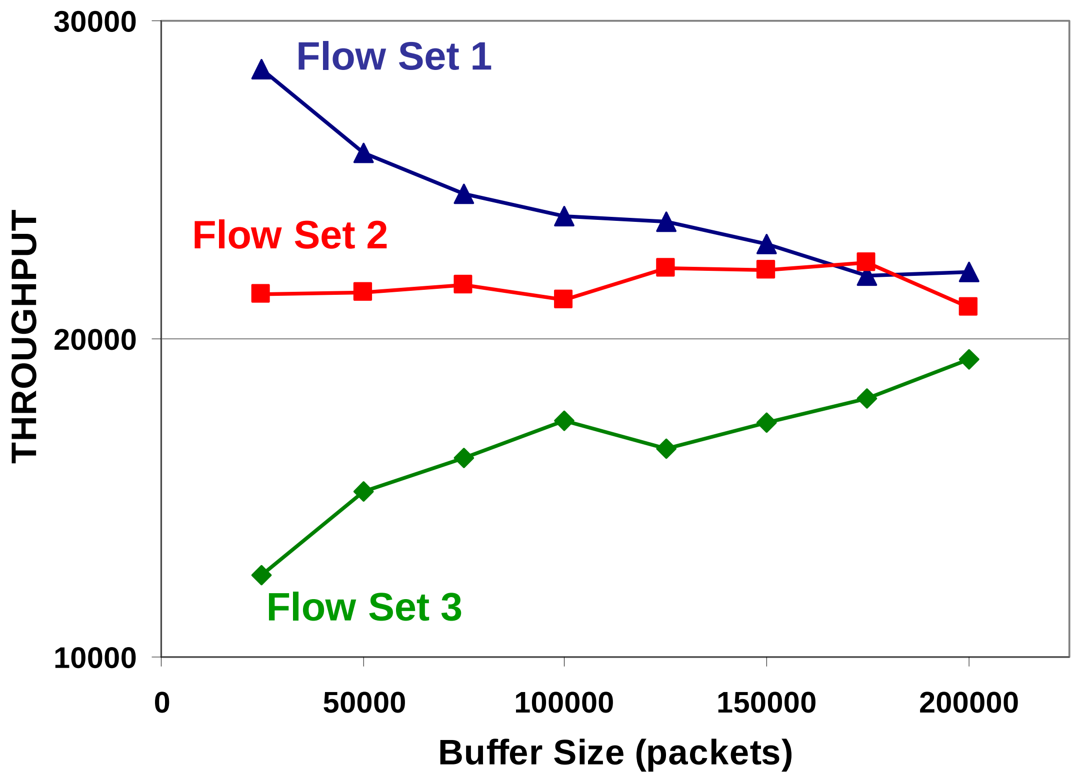


Figure B-4. Changes in Average Throughput with Increasing Buffer Size