Towards Stochastic Flow-Level Network Modeling: Performance Evaluation of Short TCP Flows

Fabien Geyer*†, Stefan Schneele*, Georg Carle†

*Airbus Group Innovations
Dept. TX4CP
D-81663 Munich, Germany
{fabien.geyer,stefan.schneele}@eads.net

†Technische Universität München
Institut für Informatik, I-8
D-85748 Garching b. München, Germany
carle@in.tum.de

***Abstract*—We present in this paper a stochastic flow-level network model for the performance evaluation of IP networks with multiple bottlenecks supporting short-lived and long-lived TCP flows. Flow-level network models are efficient at estimating the mean bandwidth of TCP flows in various topologies, but they are generally limited to the study of infinite flows. This paper extends such models in order to evaluate short-lived flows alternating between idle and active periods where data of random size is transferred. We study the interaction between multiple flows and derive mean bandwidths, durations of file transfer or average number of active flows. We first study a single bottleneck, and then extend our analysis to networks with multiple bottlenecks as well as the effect of slow-start. We apply our results to various networks and assess the accuracy of our approach by comparing our analytical results with results of the discrete event simulator ns-2.**

***Index Terms*—Performance evaluation, TCP, Quality of Service**

I. Introduction

Models of TCP behavior have been proposed and studied for more than a decade. Such models are crucial for network planning and resources allocation, and need to take into consideration the current use of network protocols. The purpose of this paper is to provide a framework for the evaluation of short-lived TCP flows on Ethernet networks with multiple bottlenecks. While the study of statistical bandwidth sharing of TCP and elastic flows is not new, previous models are often limited to the study of networks with a single bottleneck or use unrealistic model of TCP bandwidth sharing. By basing our work on flow-level network modeling, we are able to extend our results to more detailed systems, namely packet scheduling algorithms, fixed-rate flows, as well as effects of TCP acknowledgments.

By short-lived or ON/OFF TCP flows, we mean flows alternating between OFF periods of random duration, and ON periods with transfers of data of random size. A flow becomes idle only once it has finished transferring its message. Our aim is to predict the throughput of individual flows, average durations for transferring a file, average numbers of concurrent users on the studied links, as well as other parameters such as average round-trip times, drop probabilities and queue sizes.

Flow-level network modeling is based on previous effort on TCP packet-level models, where the throughput of TCP is modeled as a function of loss probability and round-trip time, such as the so-called PFTK model [1]. Using those models with models of queue behavior, flow-level models have been proposed, such as the early work presented in [2], [3] or [4].

Our contribution is the extension of traditional flow-level network modeling as presented in [5], initially limited to infinite flows, to the study of ON/OFF TCP flows. We aim at modeling the following features: key properties of TCP like slow-start and congestion-avoidance phases, distributions of file sizes and think time, network parameters such as drop probability and RTT, as well as applicability to multiple bottleneck topologies. Numerical results of our model are compared with *ns-2* simulations.

II. Related work

While the majority of flow-level models are dedicated to the study of infinite TCP flows, some extensions have already been proposed to include short-lived flows. The model introduced in [3] is limited to flows without any idle period. [6] introduced briefly a way to analyze short-lived TCP flows, but it is limited to the single bottleneck case and does not account for the slow-start phase of TCP flows. [7] used a similar ON/OFF model as the one used here, but used a formulation based on a Markov chain for modeling the behavior of TCP. More recently [8] proposed a closed form formula for the distribution of the throughput obtained by an ON/OFF source, but with a limitation to the single bottleneck case where all flows share the same RTT and drop probability. [9] studied a similar problem and proposed a simplified formula for the saturated regime. More generally on the notion of statistical bandwidth sharing of elastic flows, [10] presented a survey of various models, based on queuing theory or various notions of fairness.

III. Model for ON/OFF TCP flows

We define the set of studied TCP flows as $\mathcal{F} = \{F_1, \ldots, F_N\}$. An ON/OFF flow F_i has two states, ON and OFF, which corresponds to the sample space $\Omega_{F_i} = \{\omega_{F_i}^{\text{ON}}, \omega_{F_i}^{\text{OFF}}\}$. We denote by $St(F_i)$ the state of flow F_i, and $\Pr\left(St(F_i) = \omega\right)$ the probability that F_i is in state ω.

We define $\mathcal{S}$ as the sample space or set of all possible combinations of states of the different studied flows. We define $\mathcal{C}$ as an event of the sample space $\mathcal{S}$. Each combination $\mathcal{C}$

978-1-4799-3780-6/14/$31.00 ©2014 IEEE

has a probability noted $\Pr(\mathcal{C})$. Using those properties, we characterize the studied probability space as:

$$\mathcal{S} = \prod_{F_i \in \mathcal{F}} \Omega_{F_i} \tag{1}$$

$$\mathcal{C} = \{(F_1, \omega_{F_1}), \ldots, (F_N, \omega_{F_N})\} \tag{2}$$

$$\Pr(\mathcal{C}) = \prod_{(F_i,\omega_k)\in\mathcal{C}} \Pr\left(St(F_i) = \omega_k\right) \tag{3}$$

Using this probability space, we determine the performances according to the combinations, and then use the low of total probability to derive the mean performance measures.

A. Evaluation of the bandwidth of the flows for each possible combination $\mathcal{C}$

In this first step, we use flow-level modeling to determine the throughput of each flow as if the flows were infinite, for every possible combination of flow states. For the purpose of this paper, we introduce here only a basic flow-level model and refer to our previous work [5] for a more details.

The studied network consists of *servers*, which represent the different queues and links. Each flow F_i has a bandwidth model $T^{\omega}_{F_i}$ which is a function of the round-trip time of the flow RTT_{F_i} , the drop probability p_{F_i}, and the state of the flow ω. The packets of flow F go through a path of servers, noted here $S_{F_i} = \{s_k\}$, which are characterized by a drop probability p_k, a queue size q_k, a maximum output bandwidth c_k and a delay D_k. The queue size and the drop probability of a server are functions of its traversing flows.

Based on those definitions, the flow-level network model is described by the following set of equations. The drop probability p_{F_i} experienced by the flow F_i is:

$$p_{F_i} = 1 - \prod_{s\in S_F} (1 - p_s) \tag{4}$$

The round-trip time RTT_F of a packet of flow F with packet size $psize$ is:

$$RTT_{F_i} = \sum_{s\in S_{F_i}\cup\overline{S_{F_i}}} (q_s + psize)\cdot c_s + D_s \tag{5}$$

with S_{F_i} the path of servers of the flow from the source node to the destination node, and $\overline{S_{F_i}}$ the path of servers of the flow from the destination node to the source node. The output bandwidth B^{out}_s of a server s is defined by the aggregated bandwidth of $\mathcal{F}_s$, the set of flows traversing the server:

$$B^{out}_s = \sum_{F_i\in\mathcal{F}_s} \left(T_{F_i}(RTT_{F_i}, p_{F_i}) \cdot \prod_{k\in\mathcal{U}(S_{F_i},s)} (1 - p_k) \right) \tag{6}$$

where $\mathcal{U}(S_{F_i}, s)$ corresponds to the path of servers of the flow F_i up to the server s (included). This bandwidth must satisfy:

$$B^{out}_s \leq c_k \tag{7}$$

A fixed point evaluation is then used on Equations (4) to (7) for a numerical evaluation.

Once the general flow-level framework characterized, we define the bandwidth models $T^{\omega}_{F_i}$ corresponding to each state ω of the flow. For the idle state, we have $T^{\text{OFF}}_{F_i} = 0$. For the active state, we use the formula derived in [1], often referred as the approximated PFTK formula:

$$\min\left(\frac{W_{max}MSS}{RTT}, \frac{MSS}{RTT\sqrt{\frac{2bp}{3}}+T_0\min\left(1,3\sqrt{\frac{3bp}{8}}\right)p(1+32p^2)}\right)$$

For each combination of flow states in $\mathcal{S}$, we use the flow-level network model defined by Equations (4) to (7) to derive the steady-state bandwidth of each flow F_i. Let $\rho(F_i|\mathcal{C})$ be the throughput of flow according to the combination $\mathcal{C}$.

B. Evaluation of the probabilities $\Pr(\mathcal{C})$

We derive here the expression of $\Pr(\mathcal{C})$, the probability of a combination of flow states $\mathcal{C}$. The size of the data transferred during an ON state of flow F_i has distribution function H_{F_i} with mean $1/\mu_{F_i} < \infty$. The duration of an OFF state has distribution function G_{F_i} with mean $1/\lambda_{F_i} < \infty$. Let $\Delta(\omega)$ be the mean duration of state ω. The long-run probability that F_i is in state ω is noted $\Pr\left(St(F_i) = \omega\right)$ and is specified by:

$$\Pr\left(St(F_i) = \omega\right) = \lim_{t\to\infty} \frac{\text{total time in state } \omega \text{ by } t}{t} \tag{8}$$

$$= \frac{\Delta(\omega)}{\sum_{\omega_j\in\Omega_i}\Delta(\omega_j)} \tag{9}$$

Using the law of total probability, we derive the mean throughput of a flow according to its state:

$$\rho\left(F_i|St(F_i) = \omega\right) = \sum_{\{\forall\mathcal{C}\in\mathcal{S}|St(F_i)=\omega\}} \frac{\Pr(\mathcal{C})\cdot\rho(F_i|\mathcal{C})}{\Pr(St(F_i) = \omega)} \tag{10}$$

where $\Pr(\mathcal{C})$, the probability of combination $\mathcal{C}$, has already been defined in Equation (3). According to our two-states model, the mean duration of each state is then:

$$\Delta(\omega^{\text{OFF}}_{F_i}) = 1/\lambda_{F_i} \tag{11}$$

$$\Delta(\omega^{\text{ON}}_{F_i}) = \frac{1/\mu_{F_i}}{\rho\left(F_i\middle|St(F_i) = \omega^{\text{ON}}_{F_i}\right)} \tag{12}$$

Equations (9) to (12) and Equation (3) are coupled and we use a fixed-point evaluation to find the equilibrium of the system. As presented in [11], $\Pr\left(St(F_i) = \omega\right)$ depends on the distributions G_{F_i} and H_{F_i} only through their means.

C. Results of the topology

We derived in Sections III-A and III-B the throughput of each flow according to the combination of flow states $\mathcal{C}$, as well as the probability of having each combination $\Pr(\mathcal{C})$. Using the law of total probability, we obtain the mean bandwidth $\overline{\rho(F)}$ of each flow as:

$$\overline{\rho(F)} = \sum_{\omega\in\Omega_F} \left(\rho(F|St(F) = \omega)\cdot\Pr(St(F) = \omega)\right) \tag{13}$$

Similarly, we derive the probability of having n active flows and the mean number of active flows $\overline{\mathcal{A}}$:

$$\Pr(n \text{ flows active}) = \sum_{\{\forall\mathcal{C}|\mathcal{A}(\mathcal{C})=n\}} \Pr(\mathcal{C}) \tag{14}$$

$$\overline{\mathcal{A}} = \sum_{n=0}^{N} n\cdot\Pr(n \text{ flows active}) \tag{15}$$

Using the same method, other mean performance measures are computed, such as mean round-trip times or drop probabilities. Note also that distributions of performance measures can be derived using the probability of each combination.

D. Inclusion of TCP slow-start

We noted earlier that we did not take into account the *slow-start* phase of a TCP flow. To overcome this issue, we extend our two-states flow model to three states: idle ω^{OFF}, active in slow-start $\omega^{\text{ON}_{SS}}$ and active in congestion-avoidance $\omega^{\text{ON}_{CA}}$, so that the state sample space of a flow is now: $\Omega = \{\omega^{\text{OFF}}, \omega^{\text{ON}_{SS}}, \omega^{\text{ON}_{CA}}\}$.

In order to describe the behavior of a TCP flow in slow-start, we use the results from [12], often referred as the CSA model. This model is used to determine the duration and amount of data transferred during slow-start. The time spent in the slow-start state $\omega^{\text{ON}_{SS}}$ is the sum of three durations: the three-way TCP handshake ($E[L_h]$ in [12]), the exponential growth phase (T_{exp}), and the time needed to recover from the first packet loss which ends the slow-start ($E[T_{loss}]$ in [12]). We then derive the complete duration of the $\omega_F^{\text{ON}_{SS}}$ state and it associated throughput:

$$\Delta(\omega_F^{\text{ON}_{SS}}) = E[L_h] + T_{exp} + E[T_{loss}] \tag{16}$$

$$\rho\left(F \middle| St(F) = \omega_F^{\text{ON}_{SS}}\right) = \frac{E[d_{ss}]}{\Delta(\omega_F^{\text{ON}_{SS}})} \tag{17}$$

with T_{exp}, the time spent in the exponential phase:

$$T_{exp} = \begin{cases} RTT \log_\gamma\left(\frac{E[d_{ss}](\gamma-1)}{w_1} + 1\right) & \text{if } W_{ss} \leq W_{max} \\ RTT\left[\log_\gamma\left(\frac{W_{max}}{w_1}\right) + 1\right] & \text{otherwise} \end{cases}$$

Note that T_{exp} is different from $E[T_{ss}]$ in [12]. In [12] it is assumed that when the TCP window reaches its maximum window size W_{max}, the TCP window will remain constant and all the remaining data will be transfered. We differ here by saying that we switch to the congestion-avoidance state to account for the interaction with other flows.

The $\omega_F^{\text{ON}_{CA}}$ state has to account for the data that was already transferred during the $\omega^{\text{ON}_{SS}}$ state, so that if there is still data to be transfered (*i.e.* $E[d_{ss}] < 1/\mu$):

$$\Delta(\omega_F^{\text{ON}_{CA}}) = \frac{1/\mu_F - E[d_{ss}]}{\rho\left(F \middle| St(F) = \omega_F^{\text{ON}_{CA}}\right)} + RTT_F \tag{18}$$

Otherwise $\Delta(\omega_F^{\text{ON}_{CA}}) = 0$. Now that the three-states flow model is defined, we use the same method as in Section III to solve numerically our system.

IV. Experimental evaluation

We evaluate in this section the accuracy of our model by comparing analytical results with results of simulations made with *ns-2*. We focus here on the evaluation of the topology presented in Figure 1, with switches using a drop-tail policy.

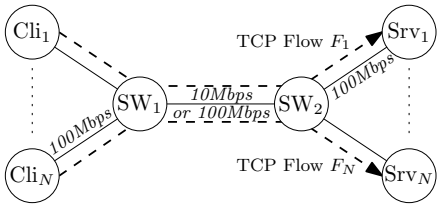

Fig. 1. Dumbbell topology with N TCP sources and N TCP destinations. Properties of the links (latency and drop probability) vary through the different use cases studied here.

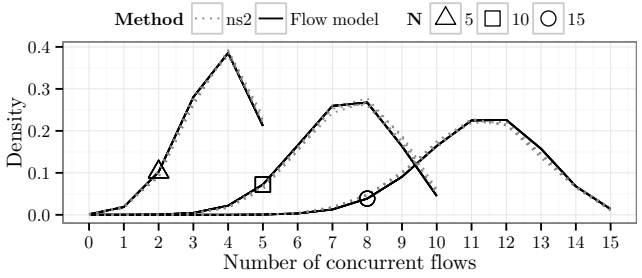

Fig. 2. Density of concurrent flows with 5, 10 or 15 TCP sources, and with different distributions for the file size and idle time

A. Insensitivity to the distributions

In order to show the insensitivity of the model to the distributions G and H, we simulated several distributions for the file size and the idle time. The link between SW_1 and SW_2 is set to 1.5Mbps with a 150ms delay in each way. The mean file size transferred by the sources is set to 200kBytes, and the mean idle to 5s. We use the two following distributions for the file size and the idle time: exponential and Pareto with shape parameter of 2.5. We simulate three cases where the number of sources N is set to 5, 10 or 15, which makes it a total of 12 scenarios.

We evaluate here the density of number of concurrent flows in the topology, which is presented in Figure 2. The distribution type has indeed a low influence on the number of concurrent flows as we see little to no differences between the different runs of the *ns-2* simulations. We see that the model accurately describes the steady-state performance of the system with regards to the number of concurrent flows, regardless of the distribution.

B. Ethernet dumbbell topology with two classes

To illustrate our framework with flows of heterogeneous properties, we simulate two types of clients: n flows of class 1 with mean file size of 10MB and mean idle time of 10s, and one flow of class 2 with a mean file size of 30MB and a mean idle time of 1s. Both classes follow a Pareto distribution of shape 2.5 for the file size, and an exponential distribution for the idle time. Results regarding the mean bandwidth of each flow are presented on Figure 3. The results of the model are in accordance with the results of the simulation.

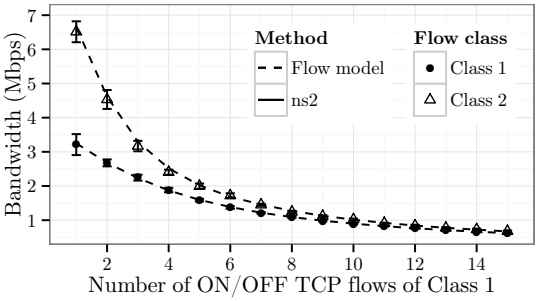

Fig. 3. Mean bandwidth of individual flows according to the number of simultaneous flows in the dumbbell topology presented in Figure 1. Error bars for the simulation results correspond to a 95% confidence interval.

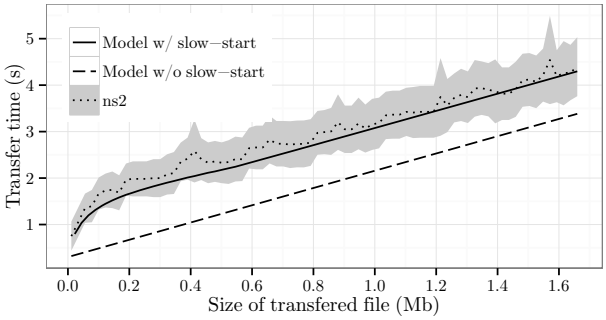

Fig. 4. Comparison between model and simulation regarding the time needed to transfer a file on the dumbbell topology, with an RTT of 300ms and a drop probability of $5 \cdot 10^{-3}$ in both directions. The gray ribbon corresponds to a 95% confidence interval for the *ns-2* simulations.

C. Slow-start and three-states flow model evaluation

We study here the impact of the slow-start algorithm on the accuracy. The latency between SW_1 and SW_2 is set to 150ms (each way), with a drop probability of $5 \cdot 10^{-3}$ following a Bernoulli model (each way). We first measure the time needed to transfer a file between 1kb and 1.6Mb when only one TCP flow is in the topology ($N = 1$). The results are presented on Figure 4. The gray ribbon on the figure corresponds to a 0.95 confidence interval of the simulation results. As expected, the model including the slow-start part of the TCP algorithm produces better results than the model without.

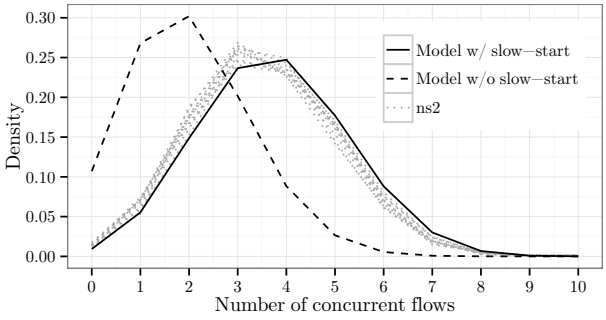

Fig. 5. Number of concurrent flows on the dumbbell topology with $N = 10$ ON/OFF TCP flows, a file size following a Pareto distribution of mean 340kb and shape 2.5, and an idle duration following an exponential distribution of mean 3s.

We evaluate the same topology, with $N = 10$ clients. The file size follows a Pareto distribution with mean 340kb and shape 2.5, which means that the slow-start phase will have a large impact on the performances. The OFF period follows an exponential distribution with mean 3s. We present in Figure 5 the number of concurrent flows in the topology. As expected, our model including the slow-start produces better results compared to the *ns-2* simulations than the model which only includes the congestion-avoidance phase.

V. Conclusion and Future Work

We presented in this paper a mathematical model for the performance evaluation of ON/OFF TCP flows using the well-studied flow-level modeling framework, which was initially developed for infinite TCP flows. Analytical results are shown to give realistic results compared to simulations. Our framework uses a simplistic ON/OFF TCP model for modeling network protocols, where TCP flows are unidirectional. We would like to extend it to more advanced models where bidirectional communications occur (client-server paradigm), as well as other models for modeling advanced network protocols. A second research direction would be to use known properties of the application layer, such as the knowledge that active period always start at the same time after an idle period, in order to reduce the number of studied combination of flow states.

References

[1] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose, "Modeling TCP Reno Performance: A Simple Model and Its Empirical Validation," *IEEE/ACM Trans. Netw.*, vol. 8, no. 2, pp. 133–145, Apr. 2000.

[2] E. Altman, K. Avrachenkov, and C. Barakat, "TCP Network Calculus: The case of large delay-bandwidth product." in *Proceedings of INFOCOM 2002*. IEEE, 2002, pp. 417–426.

[3] V. Firoiu, I. Yeom, and X. Zhang, "A Framework for Practical Performance Evaluation and Traffic Engineering in IP Networks," in *IEEE International Conference on Telecommunications*, 2001.

[4] R. Gibbens, S. Sargood, C. Van Eijl, F. Kelly, H. Azmoodeh, R. Macfadyen, and N. Macfadyen, "Fixed-Point Models for the End-to-End Performance Analysis of IP Networks," in *ITC specialist seminar*, 2000.

[5] F. Geyer, S. Schneele, and G. Carle, "Practical Performance Evaluation of Ethernet Networks with Flow-Level Network Modeling," in *Proceedings of the 7th International Conference on Performance Evaluation Methodologies and Tools (VALUETOOLS)*, ICST. ACM, Dec. 2013.

[6] T. Bu and D. Towsley, "Fixed Point Approximations for TCP behavior in an AQM Network," in *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 29, no. 1. ACM, 2001, pp. 216–225.

[7] C. Casetti and M. Meo, "A New Approach to Model the Stationary Behavior of TCP Connections," in *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 1. IEEE, 2000, pp. 367–375.

[8] F. Baccelli and D. R. McDonald, "A stochastic model for the throughput of non-persistent TCP flows," *Performance Evaluation*, vol. 65, no. 6–7, pp. 512 – 530, 2008.

[9] T. Bonald, P. Olivier, and J. Roberts, "Dimensioning high speed ip access networks," in *proceedings of the 8th International Teletraffic Congress (ITC 18)*, 2003, pp. 241–251.

[10] J. Roberts, "A survey on statistical bandwidth sharing," *Computer Networks*, vol. 45, pp. 319–332, 2004.

[11] D. P. Heyman, T. Lakshman, and A. L. Neidhardt, "A New Method for Analysing Feedback-Based Protocols with Applications to Engineering Web Traffic over the Internet," in *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 25, no. 1. ACM, Jun. 1997, pp. 24–38.

[12] N. Cardwell, S. Savage, and T. Anderson, "Modeling TCP Latency," in *Proc. of INFOCOM 2000*, vol. 3. IEEE, Mar. 2000, pp. 1742–1751.