An Application of System Identification to Modeling End-to-End Packet Delay Dynamics of the Internet

Hiroyuki Ohsaki Masayuki Murata Hideo Miyahara

Graduate School of Engineering Science, Osaka University
1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan

(Phone) +81-6-6850-6588 (Fax) +81-6-6850-6589 (E-mail) oosaki@ics.es.osaka-u.ac.jp

1 Introduction

Understanding the end-to-end packet delay dynamics of the Internet is of crucial importance since (1) it directly affects the QoS (Quality of Services) of realtime services, and (2) it enables us to design an efficient congestion control mechanism for both realtime and non-realtime applications. In particular, for non-realtime applications, a delay-based approach for congestion control mechanisms, rather than a loss-based approach as in the current TCP (Transmission Control Protocol), has been proposed (e.g., [1, 2]). The main advantage of such a delay-based approach is, if properly designed, packet losses can be prevented by anticipating impending congestion from increasing packet delays.

For a long time, queueing theory has been extensively used as a powerful too to analyze both circuit-switched and packet-switched networks. In general, the queueing theory assumes stationarity of the network, and allows us to obtain several performance measures such as the average packet delay and the average packet loss probability. The stringent limitation of the queuing theory is its impossibility to analyze the *dynamical behavior* of the network. Several measurement-based studies suggest that the end-to-end packet delay of the Internet is quite dynamical [3, 4]. Another approach, being different from queueing theory, should therefore be taken to investigate the packet delay dynamics of the Internet.

In this paper, we propose a novel approach to modeling the packet delay dynamics of the Internet. The key idea of our approach is treating the network, seen by a specific source host, as a *black-box*, and modeling the packet delay dynamics using *system identification theory*, having been throughly used in the field of control engineering. The packet delay dynamics seen by a source host is modeled as a SISO (single-input and single-output) system, where the input to the system is a packet inter-departure time from the source host, and the output from the system is a round-trip time measured by the source host. In this paper, the ARX (Auto-Regressive Exogenous) model is used and its parameters are determined using system identification theory. Determination of the orders of the ARX model is also discussed.

2 Black-Box Model

As depicted in Fig. 1, the network seen by a specific source host is considered as a black-box. Our goal of this paper is to create a SISO model describing the packet delay dynamics seen by the source host: i.e., the relation between a packet sending process from the source host and its resulting packet delay. As the input to the system, we use a *packet inter-departure time* from the source host, i.e., the interval between two consecutive packet transmissions from the source host. Use of the packet inter-departure time is straightforward since it directly affects the end-to-end packet delay. On the contrary, as the output from the system, we use a *round-trip time* measured by the source host, i.e., the interval between a packet transmission and the receipt of the corresponding ACK packet. Instead of a one-way packet delay, a two-way round-trip delay is used as the output from the system, for building a model seen by the source host.

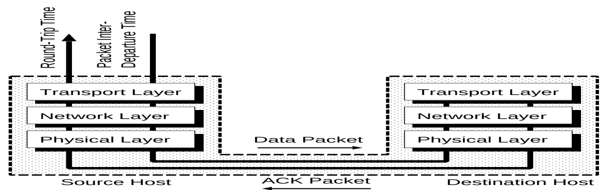

Figure 1: Network as SISO Model.

In this paper, the ARX model is used and its parameters are determined using system identification theory [5]. Letting $u(k)$ and $y(k)$ be the input and the output data at slot k, the ARX model is defend as

$$\begin{aligned} A(q)\,y(k) &= B(q)\,u(k-n_k)+e(k) \\ A(q) &= 1+a_1q^{-1}+\ldots+a_{n_a}q^{-n_a} \\ B(q) &= b_1+b_2q^{-1}+\ldots+b_{n_b}q^{-n_b+1} \end{aligned}$$

where $e(k)$ is unmeasurable disturbance (i.e., noise), and q^{-1} is the delay operator; i.e., $q^{-1}u(k) \equiv u(k-1)$. The numbers n_a and n_b are the orders of respective polynomials. The number n_k is the number of delays from the input to the output. In this paper, $u(k)$ and $y(k)$ correspond to the k-th packet inter-departure time and the k-th round-trip time, respectively. Refer to [5] for the detail of the ARX model.

The ARX model is a linear time-invariant model, so it cannot rigorously capture non-linearity of the packet delay dynamics. However, the ARX model is applicable in various control engineering problems since non-linear dynamical systems operating around the stable point can be well approximated by a linear system [6]. In the next section, we will investigate how accurately the packet delay dynamics of the Internet can be described by the ARX model.

3 Numerical Example

For determining parameters of the ARX model, sample input/output data are collected from a simple simulation experiment. The simulation model consists of 10 source–destination pairs and a single bottleneck node. A source host randomly generates a packet to the destination host. The packet inter-departure time follows the exponential distribution. The exponential distribution is used since it is one of ideal input signals for system identification [5]. A destination host sends an ACK packet back immediately after receiving a packet. The bottleneck node has a shared buffer of 100 packets, and processes incoming packets in the FIFO (First-In First-Out) order. Other simulation parameters are: the bottleneck bandwidth is 1.5 [Mbit/s], the packet size is 1,000 [byte], the propagation delay is 5 [ms], and the offered traffic load is 0.7.

We have collected the input/output data of 1500 packets for a single source host. We use 500 samples (i.e., $u(500)$–$u(999)$) for parameter determination, and 100 samples (i.e., $u(1200)$–$u(1299)$) for model validation (see below). Shown are the inter-departure time $u(k)$ in Fig. 2 and the round-trip time $y(k)$ in Fig. 3 recorded at a source host. As one might expect, positive correlation between the inter-departure time and the round-trip time can be observed. However, the round-trip time is disturbed by packets coming from other source hosts because of the shared FIFO buffer of the bottleneck node.

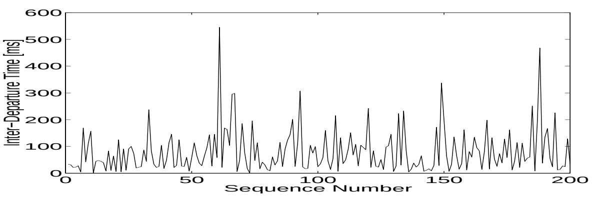

Figure 2: Input data: the packet inter-departure time.

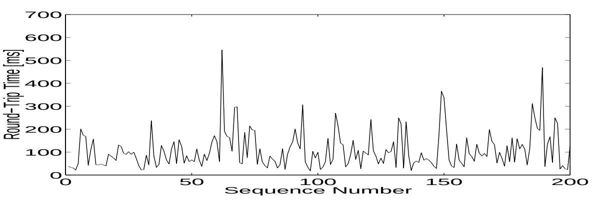

Figure 3: Output data: the round-trip time.

The system identification problem for the ARX model is formulated as a minimization problem, where the cost function is given by a loss function (i.e., the sum of squared prediction errors for all input/output data). This problem can be easily solved by a least square method [5]. Because of space limitation, only the results are included. Refer to [5] for the detail. For example, when n_a = 6, n_b = 2, and n_k = 1, parameters of the ARX model are identified as a_n = $\{$ 1.0, -0.21, -0.0064, 0.010, 0.031, -0.036, 0.044 $\}$ and b_n = $\{$ 0.67, -0.12 $\}$.

A comparison between the measured data and the model output is shown in Fig 4. The solid line is the round-trip time obtained from the simulation (i.e., $y(1200)$–$y(1299)$). The dotted line is the output of the ARX model, which is computed only from the packet inter-departure time used in the simulation (i.e., $u(1200)$–$u(1299)$). Note that parameters of the ARX model is determined from $u(500)$–$u(999)$ and $y(500)$–$y(999)$. One can find that these lines roughly coincide but slightly differ. This is probably because the measured round-trip time is disturbed by packets from other source hosts, which is unknown to the source host.

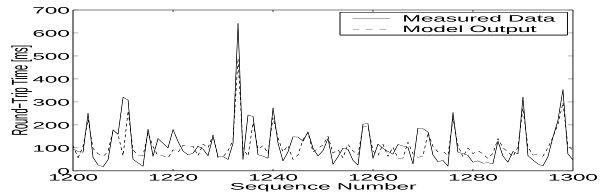

Figure 4: Comparison between measured data and model output.

Accuracy of the ARX model is affected by a choice of the orders, n_a and n_b, and the delay n_k. The loss function for different n_a's and n_b's ($1 \leq n_a, n_b \leq 20$) for n_k = 1 is plotted in Fig. 5. The smaller the loss function is, the less the prediction error of the ARX model becomes. This figure indicates a counter-intuitive result; i.e., a larger n_a does not always result in a better ARX model.

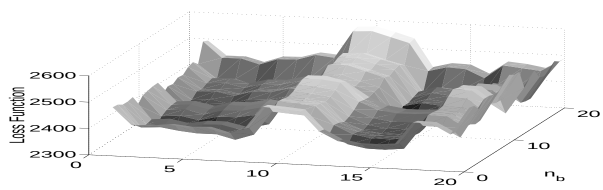

Figure 5: Loss function for different orders of the ARX model.

4 Conclusion

In this paper, we have proposed a novel approach to a black-box modeling of the packet delay dynamics of the Internet using system identification theory. With the input/output data obtained from a simulation, the ARX model for the packet delay dynamics has been created. We have also discussed determination of the orders of the ARX model.

As future work, effectiveness of the ARX model should be investigated for a through set of input/output data. In particular, using input/output data measured from the real Internet is of great interest.

References

[1] R. Jain, "A delay-based approach for congestion avoidance in interconnected heterogeneous computer networks," *Technical Report DEC-TR-566*, Apr. 1989.

[2] L. S. Brakmo and L. L. Peterson, "TCP Vegas: End to end congestion avoidance on a global Internet," *IEEE Journal on Selected Areas in Communications*, vol. 13, pp. 1465–1480, October 1995.

[3] V. Paxson, "End-to-end Internet packet dynamics," *N/A*, 1997.

[4] S. B. Moon, J. Kurose, P. Skelly, and D. Towsley, "Correlation of packet delay and loss in the Internet," *Technical Report 98-11*, Jan. 1998.

[5] L. Ljung, *System Identification — Theory for the User*. Englewood Cliffs, N.J.: Prentice Hall, 1987.

[6] J.-N. Juang, *Applied System Identification*. Englewood Cliffs, N.J.: Prentice Hall, 1994.