Regular Papers

Phase Transition Phenomena in Internet Traffic—Observations and Possible Causes

Kensuke Fukuda†

Abstract

This paper shows that the dynamics of Internet traffic can be viewed as phase transition phenomena between non-congested and congested phases. In the phase transition model, the traffic is characterized by self-similarity as the traffic density approaches the critical point, though away from the critical point, the traffic behavior is close to the traditional Poisson model. In this sense, the phase transition model naturally includes self-similar and Poisson characteristics. Investigating the essential features for reproducing phase transition dynamics from the network protocol mechanism, we found that the key roles of the dynamics come from the non-linearity of the buffers at the router and the effect of implicit cooperation between nodes produced by feedback control.

1. Introduction

On the Internet, data (representing text, voice, and movies) transferring between applications is split into packets that each has a header including a source and a destination address. A packet from a source travels to a destination via routers. A router along the path determines the next-hop router for incoming packets based on the destination address of the packet and its own routing table. Unlike in the traditional telephone network, in the Internet, packets from sources are multiplexed at routers; i.e., a packet from one source may be located between packets from others.

When the number of injected packets exceeds the capacity of the network, packets are temporarily stored at a buffer in a router, and in the worse case, they are dropped because the router's buffer has a finite size. Such a state is called congestion. Considering that a huge number of packets from many users are involved in the generation of congestion, a macroscopic view is better for understanding the dynamics of Internet traffic than a view focusing on the behavior of individual users' packets.

There have been many studies on statistical characterization of Internet traffic [1]. Historically, Internet traffic was characterized by a Poisson model, meaning that the mean and variance of the traffic are described by a single parameter. One of the important properties of the Poisson model is its memory-less nature, i.e., the current value of traffic volume is independent of past values. However, it was found that the fluctuation of traffic volume in the Internet is well modeled by self-similarity, deviating clearly from the Poisson model [2], [3]. The self-similarity is a scale-invariant property in which a burst still remains when the observed time scale changes. Comparing real traffic with a Poisson time series and a self-similar surrogate time series, it is visually apparent that real traffic has scale-invariance and is closer to a self-similar time series than to a Poisson one, as shown in **Fig. 1**. Theoretically, the power spectrum of a self-similar time series is characterized by a power law $S(f) \propto f^{-\beta}$ for $0 < \beta \leq 1$, where f is frequency and β is called the scaling exponent. On the other hand, the power spectrum of Poisson traffic is white noise, corresponding to $\beta = 0$. Thus, a self-similar time series has long-range correlation, meaning that the current value strongly depends on past values, though a Poisson time series is memory-less.

† NTT Network Innovation Laboratories
Musashino-shi, 180-8585 Japan
E-mail: fukuda@t.onlab.ntt.co.jp

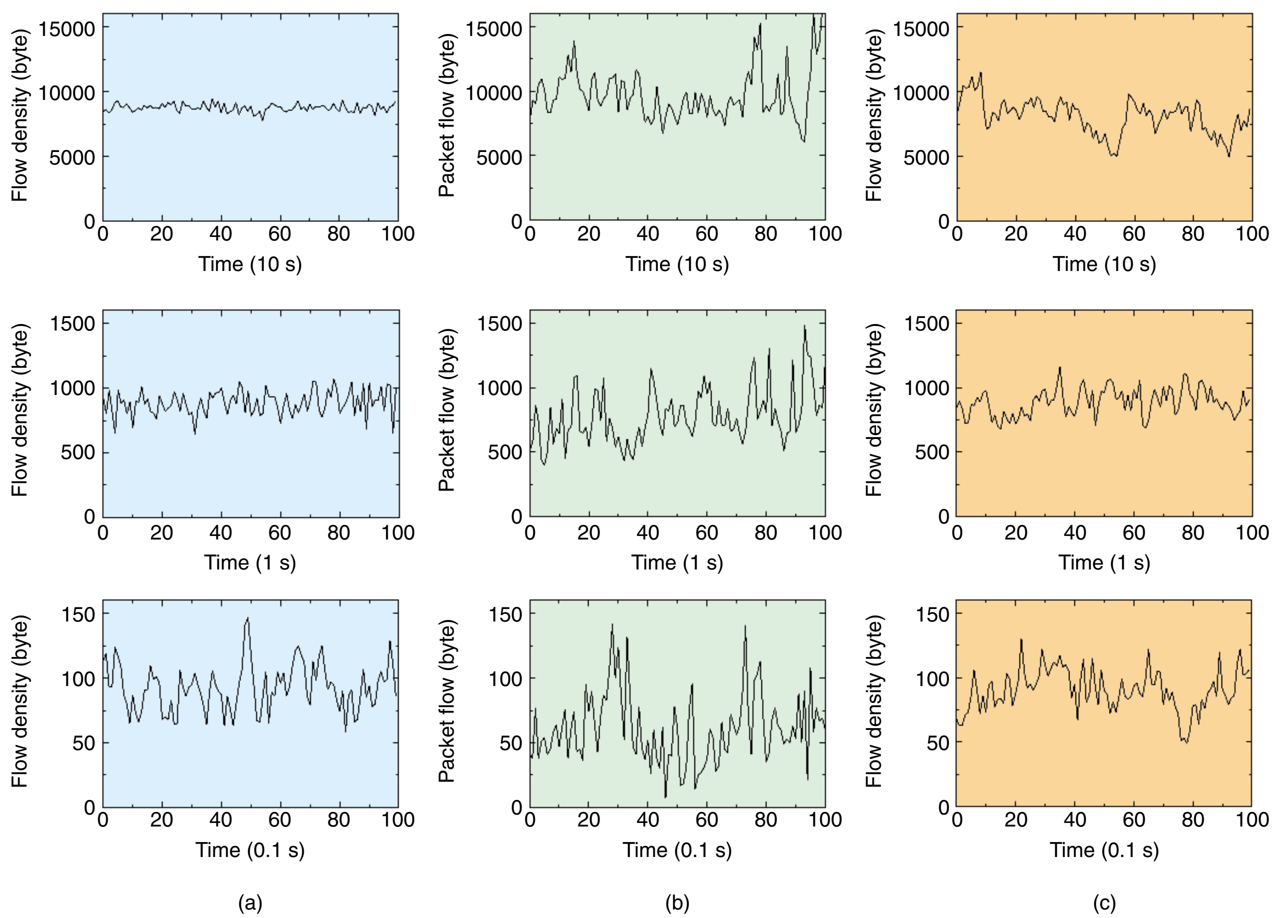

Fig. 1. Differences in statistical properties of traffic: (a) Poisson traffic, (b) real traffic, and (c) self-similar traffic. Upper graphs correspond to a coarser time scale.

2. Phase transition phenomena in Internet traffic

As I have explained, the temporal variation of network traffic exhibits self-similarity. However, we may ask a question: Is the self-similarity observable independent of the status of the network? In other words, when the network is less-congested or highly congested, does the traffic still show long-range correlation?

To answer this question, the rest of this section reveals that the traffic fluctuation is characterized by phase transition phenomena between non-congested and congested phases [4]. Phase transition phenomena are widely known in statistical physics and are defined as phenomena where a macroscopic parameter (order parameter) of the system changes greatly when a microscopic parameter (control parameter) approaches an uncertain critical point [5]. One well known example is that a magnet loses its magnetic power at the critical temperature. One of the most important properties of the phase transition is that self-similarity appears at the critical point. In Internet traffic, the control parameter corresponds to the aggregated traffic density (i.e., level of congestion). When the traffic density is low or too high, the fluctuation in traffic density is temporally non-correlated, though it becomes self-similar at the critical point between non-congested and congested phases. Thus, the phase transition model naturally includes the traditional Poisson model and the self-similar model.

Figure 2 shows the cumulative distribution ($P(L)$) of congestion duration L for some real Internet traffic, indicating how much congestion statistically appears. Here, the congestion duration is the number of consecutive congestion time steps multiplied by the bin size of traffic. A congestion time step is defined as a time step whose traffic density exceeds a certain threshold.

The three plots in the figure show different time periods in a day, indicating the different values of the

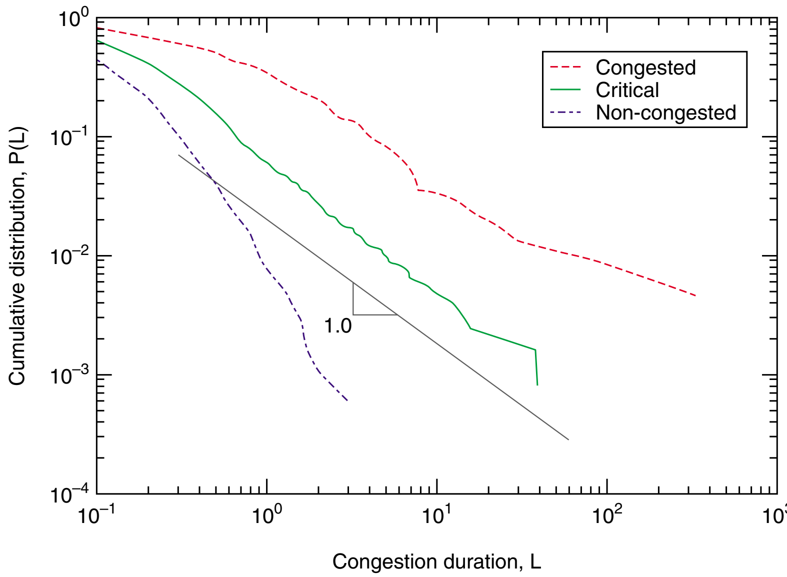

Fig. 2. Cumulative distribution of congestion duration in real traffic.

control parameter: non-congested (night), critical (day), and congested (day). When the network is less congested, the plot decays exponentially $P(L) \propto e^{-L}$. Thus, each congestion duration is short and there is a typical size. On the other hand, in the congested phase, the plot consists of a small number of short-lived congestion events, and one much longer one (near the observation time scale), because most time steps can be regarded as congestion time steps. In particular, at the critical point between two phases, the plot approximately obeys a power law $P(L) \propto L^{-\gamma}$, where γ is the scaling exponent for the congestion duration distribution. This power law demonstrates that there is no typical congestion duration. Thus, analyzing more traffic lets one observe a longer congestion duration. In addition, it is known that the scaling exponent γ corresponds to the scaling exponent β for the power law in the power spectrum density. This means that the traffic becomes self-similar at the critical point. Moreover, Ref. [4] confirmed that the correlation duration of the traffic tends to diverge at the critical point, supporting the idea that the traffic has long-range correlation only at the critical point. From the observational results, we can conclude that the traffic is not always characterized by self-similarity, depending on the value of the control parameter. Thus, phase transition phenomena are more suitable for characterizing Internet traffic.

Having explained that the statistical properties of traffic change in time by the value of the control parameter, one more question may arise: How does the control parameter behave? This corresponds to the pattern of the coarse-grained view of congestion. Intuitively, the value of the control parameter depends on the level of the human activity: daytime is congested and nighttime is not congested. It is reported that the cumulative distribution of the periods when the control parameter is stable follows a power law, indicating that the control parameter changes rapidly most of the time, though it keeps roughly the same value for a long time [6]. However, interestingly, the scaling exponent is larger during the day than during the night. This explains why the control parameter's stable periods are shorter during the day than during the night. Thus, the control parameter itself fluctuates slowly in time, and the corresponding traffic changes in a statistical manner.

3. Origin of self-similarity

This section explains two possible causes of the phase transition phenomena from the viewpoint of network protocols: the non-linearity of buffers [7] and implicit cooperation between communicators [8].

3.1 Non-linearity of buffers

One of the most basic components of the network is a buffer for packet transfer at a router or a switch. When packets arrive at a router, they are usually stored in first-in first-out (FIFO) order. As a packet passes through the router, a non-linear delay is generated because the router merges the packets from other

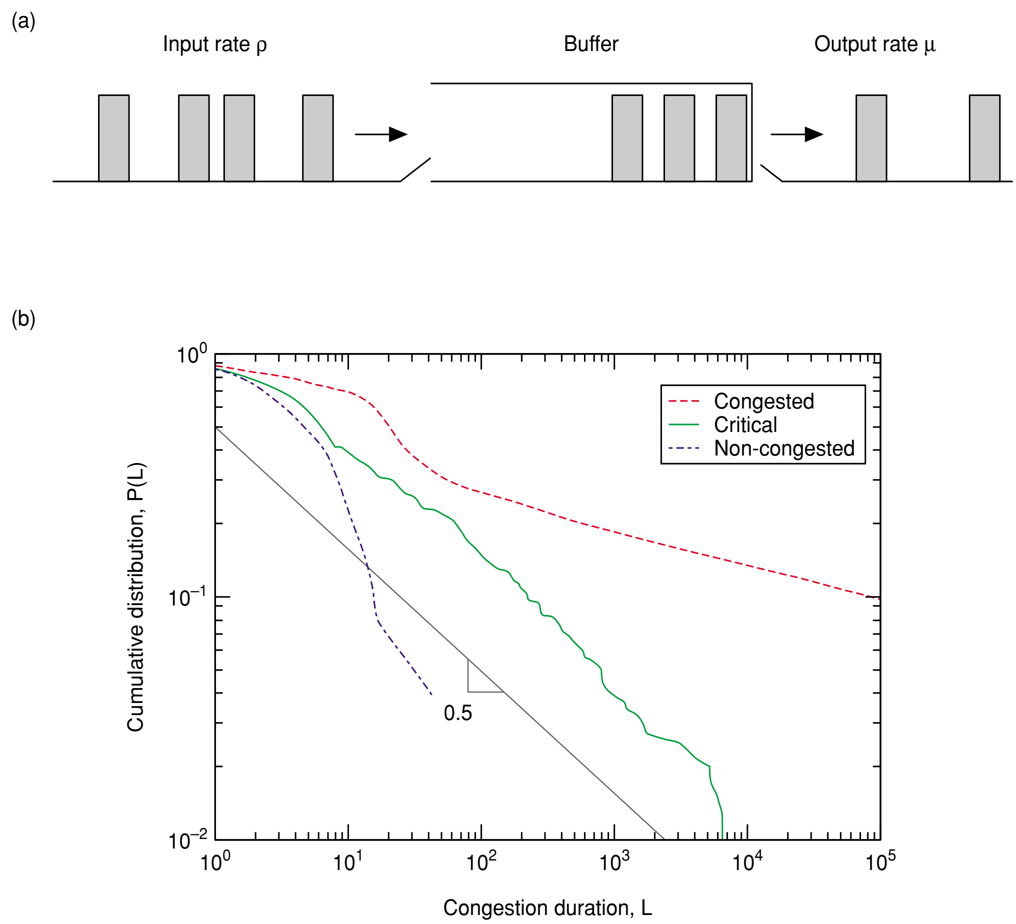

Fig. 3. Non-linearity of the buffer.

sources.

When packet input rate (ρ) to the buffer is smaller than packet output rate (μ) from the buffer ($\rho < \mu$), the number of the packets in the buffer is almost zero (See **Fig. 3**.). Conversely, for $\rho < \mu$, the buffer size increases in time and consequently diverges if the buffer capacity is infinite. However, in a real network, the buffer capacity is finite, so packets are dropped.

If the input time series has a packet arrival behavior that follows an exponential distribution for $\rho < \mu$, then the output time series has short-time correlation because the input traffic pattern does not change at the router. On the other hand, for $\rho > \mu$, the output time series stays at the maximum output rate. So, what kind of behavior can be observed at the critical rate $\rho \approx \mu$? It is important to understand that the buffer size at the router does not always diverge at the critical point for a finite time series. At the critical rate, we can observe that most buffer sizes are still small, and self-similarity appears in the output time series. Figure 3 shows the cumulative distribution of the congestion duration for the simple buffer simulation. It is visually apparent that the plot is approximately a power law at the critical rate, though it is close to exponential for $\rho < \mu$. Thus, we conclude that the buffer model is essential to generate the phase transition phenomena. However, interestingly, the value of the exponent of the power law is 0.5 at the critical point in the buffer model. The value of the exponent for the buffer model is determined by the random walk of the buffer size: at the critical rate, the fluctuation of buffer size becomes a random walk. It is theoretically known that the scaling exponent of the power law in the power spectrum for the random walk is 0.5. However, as we observed in Fig. 2, the scaling exponent for real traffic at the critical point is close to 1.0. Therefore, the simple buffer model alone is still inadequate to fully explain real network behavior.

3.2 Implicit cooperation between communicators

The dynamics of traffic in the IP layer is modeled by the simple buffer model. However, in the Internet,

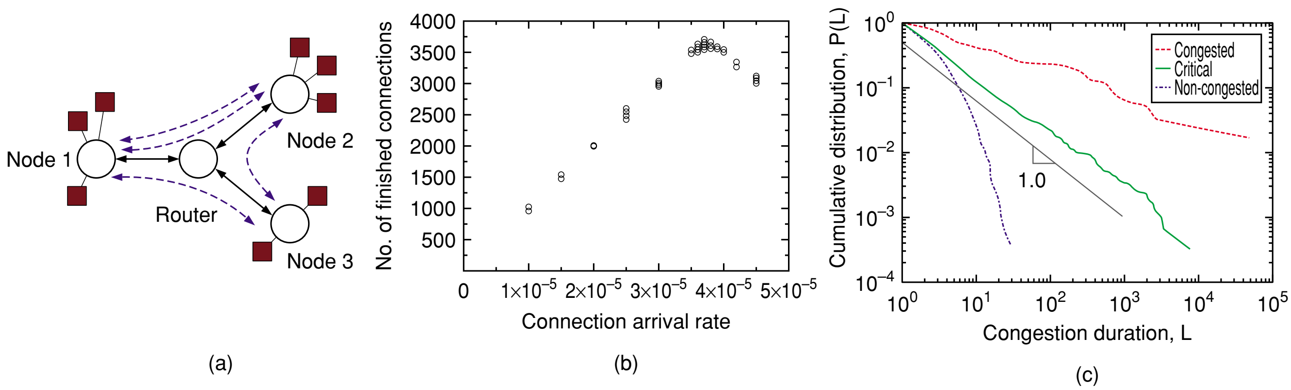

Fig. 4. Contribution of phase transition to TCP: (a) simulation topology, (b) performance, and (c) cumulative distribution of congestion duration.

higher protocol layers are actually responsible for traffic control, which affects the statistical properties of traffic. Here, we focus on the contribution of the transport layer on the generation of the phase transition and self-similar traffic.

TCP (transmission control protocol) [9] is the current *de facto* standard transport protocol for end-to-end traffic control. For example, over 90% of the traffic in Fig. 1 is TCP traffic. TCP itself is now a complicated protocol because several major features have been added to improve network performance. The main mechanisms of TCP are for (1) sending packets with appropriate timing for the current status of the network (congestion/flow control) and (2) guaranteeing reliability against packet loss (which would require retransmission). TCP is strictly standardized, so it is hard to analyze the system as a whole due to the complexity of the algorithm. Therefore, analyses have focused on the above two components.

In the rest of this section, for the first mechanism, we analyze the minimal model of the transport protocol to reproduce the phase transition and self-similar traffic by simulation.

Figure 4(a) shows the network topology used for our simulation. It consists of one router and three nodes. In this simulation, each connection between a sender and a receiver is established on the nodes and terminated after a fixed number of packets have been sent correctly. For example, there are two connections between nodes 1 and 2, one connection between 2 and 3, and one connection between 1 and 3 in the figure. The flow control of the connection is a stop-and-wait protocol, which is a simplified TCP. In this protocol, the sender is permitted to send the next packet only after receiving from the receiver a packet acknowledging receipt of the previous packet. Namely, there is only one packet for a given connection in the network at a certain instance. The control parameter of this system is connection arrival rate ρ, and the arrival pattern has no temporal correlation (i.e., it is random). Note that neither the number of packets to be sent nor the connection arrival pattern has long-range correlation. Therefore, like the single buffer model: there is no temporal correlation in input. Furthermore, because the retransmission algorithm is omitted, the buffer size of the router and nodes is set to infinity.

Figure 4(b) shows the number of properly finished connections as the control parameter changes. It is visually apparent that up to $\rho \approx 3.75 \times 10^{-5}$, this number increases linearly, though it rapidly decreases for $\rho > 3.75 \times 10^{-5}$. The critical rate between the two phases is $\rho \approx 3.75 \times 10^{-5}$.

Figure 4(c) shows the cumulative distribution of congestion duration using the time series of the number of packets passing through a link. In the non-congested phase, the plot decays exponentially, and one large congestion event covers the link in the congested phase. As expected, the plot is approximately a power law at the critical point. However, interestingly, the value of the exponent of the power law is close to 1.0 rather than to 0.5. From this figure, we can conclude that the stop-and-wait algorithm is a good model for reproducing the traffic.

In the stop-and-wait algorithm, the transmission rate of packets naturally depends on the round trip time between the sender and receiver. The round trip time is the sum of the transfer delay and the waiting time at the routers. Although the transfer delay is independent of the degree of network congestion, the waiting time is affected by packets from other sources. Therefore, changing the packet transmission rate based on the round trip time leads to implicit cooperation with other sources and improves the performance of the system as a whole. Moreover, this cooperation generates a power law with exponent 1.0 in the reproduced traffic at the critical point.

Another question is whether this stop-and-wait model is truly a minimum model. We checked the CBR (constant bit rate) algorithm instead and found that although the traffic causes a phase transition, the scaling exponent of the power law at the critical point is close to 0.5. This means that the statistical properties of CBR-controlled traffic can be explained by the single buffer model. In the stop-and-wait model, a topology with more than three nodes results in the same statistical behavior as one with three nodes. However, the traffic between a pair of nodes is categorized into the same class as for the CBR result, even though the traffic is controlled by the stop-and-wait algorithm. This is because the stop-and-wait protocol is symmetric between the sender and the receiver. Therefore, the dynamics of two traffic sources can be viewed as that of one traffic source, which has no mechanism for generating the delay. Thus, it is essential that the round trip time between two nodes is determined by other sources. Of course, from the viewpoint of the third node, the traffic relating to the third node is similarly affected by others.

Concerning the retransmission mechanism, exponential backoff, which selects the waiting time for retransmission based on the random time followed by the exponential function, plays an important role in determining the statistical properties of reproduced traffic [10]. Furthermore, it is reported that even two TCP sources can generate self-similarity through the retransmission mechanism [11]. Similarly, it has been pointed out that even in a different protocol layer, packet collision and exponential backoff in Ethernet (CSMA/CD (carrier sense multiple access with collision detection)) are essential for generating the same type of self-similar traffic as the real traffic [12].

4. Efficiency of the network

From the viewpoint of phase transition phenomena, it is clear that the critical point is the most efficient point for the network as a system. Thus, a control in which the control parameter always stays near the critical point is a good strategy. Intuitively, from the users' viewpoint, when the network is less congested, the traffic naturally increases because users do not experience any stress using the network. On the other hand, when the network is congested, users hesitate to generate more traffic because of the stress. Similarly, from the system's viewpoint, to adapt the traffic rate to the current network status, TCP behaves aggressively in a less-congested link, though it decreases the rate in a congested link. Moreover, at the macroscopic level, a link that is usually congested will be replaced by a faster link. Thus, it is plausible that the network status we usually encounter tends to stay around the critical point based on observations that self-similarity is observed in several type of links. However, in a real wide-area network, the time period when the control parameter is stable also fluctuates in time, meaning that it does not always stay at the critical point.

Thus, future work includes developing a method for planning the network and for controlling the traffic to improve the network performance from the phase transition viewpoint. As is clear from the definition, the critical point is the most efficient point for controlling the network. Therefore, controlling the amount of traffic so that it approaches the critical point is a promising strategy. One difficulty for planning and controlling is that it is still unclear whether the efficiency of the network as a whole system is maximum even when a local link is controlled near the critical point.

5. Conclusion

In this paper, I showed from a real data analysis that the dynamics of Internet traffic can be viewed as a phase transition phenomenon between non-congested and congested phases. From the phase transition viewpoint, the traffic is characterized by self-similarity only at the critical point, though the traffic in the non-congested phase is modeled as traditional Poisson traffic. We also investigated the essential features for reproducing phase transition dynamics from the network protocol mechanism and found that the key roles of the dynamics are (1) the non-linearity of the buffers at the router and (2) implicit cooperation between nodes produced by feedback control. Thus, considering the control and performance evaluation of networks, we should assume self-similarity in the reproduced traffic even with a Poisson input traffic, when the total performance of the network is optimal.

References

[1] L. Kleinrock, "Queueing Systems," John Wiley & Sons, New York, 1975.

[2] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Willson, "On the Self-Similar Nature of Ethernet Traffic," IEEE/ACM Transactions on Networking, 2, pp. 1-15, 1994.

[3] V. Paxson and S. Floyd, "Wide-Area Traffic: The Failure of Poisson

Modeling," IEEE/ACM Transactions on Networking, 3, pp. 226-244, 1995.

[4] M. Takayasu, H. Takayasu, and K. Fukuda, "Dynamic Phase Transition Observed in Internet Traffic Flow," Physica A, 277, pp. 248-255, 2000.

[5] H. E. Stanley, "Introduction to Phase Transition and Critical Phenomena," Oxford University Press, New York, 1971.

[6] K. Fukuda, L. A. N. Amaral, and H. E. Stanley, "Similarities between communication dynamics in the Internet and autonomic nervous system, EuroPhysics Letters, Vol. 62, pp. 189-195, 2003.

[7] M. Takayasu, A. Tretyakov, K. Fukuda, and H. Takayasu, "Phase Transition and 1/f noise in the Internet Traffic Transport," Proc. of Traffic and Granular Flow 97, pp. 55-76, 1997.

[8] K. Fukuda, M. Takayasu, and H. Takayasu, "Minimal model of Internet traffic," Proc. of Traffic and Granular Flow 01, pp. 389-400, 2001.

[9] J. Postel, "Transmission Control Protocol," Request for Comments 793, 1981.

[10] A. Feldmann, A. C. Gilbert, P. Huang, and W. Willinger, "Dynamics of IP Traffic: A Study of the Role of Variability and the Impact of Control," Proc. of SIGCOMM 99, pp. 301-313, 1999.

[11] A. Veres and M. Boda, "The Chaotic Nature of TCP Congestion Control," Proc. of IEEE INFOCOM 2000, pp. 1715-1723, 2000.

[12] K. Fukuda, M. Takayasu, and H. Takayasu, "Origin of Critical Behavior in Ethernet Traffic," Physica A, 297, pp. 289-301, 2000.

Kensuke Fukuda

Research Scientist, Ubiquitous Network Service System Laboratory, NTT Network Innovation Laboratories.

He received the Ph.D. degree in computer science from Keio University, Tokyo in 1999. He joined NTT Network Innovation Laboratories in 1999. In 2002, he was a visiting scholar in Boston University. His research interests are the dynamics of Internet traffic/routing and scientific aspects of networks. He is a member of the Association for Computing Machinery.