

A Fundamental Property for Traffic Management in ATM Networks

Soung C. Liew¹

¹ Bellcore, 445 South St., Morristown, NJ 07962

Tony T. Lee^{1,2}

² Polytechnic University, 333 Jay St., Brooklyn, NY 11201

Abstract

It is desirable that the traffic-control policy at a network node depends only on the external traffic loads on the input and output links, but not on the detailed addressing or distribution of packets from inputs to outputs. It should be possible to guarantee the grade-of-service of an input-output connection by controlling the aggregate loads on the input and output. Switch nodes in which such a traffic-control policy is possible are said to have the property of the sufficiency of the knowledge of external loads (SKEL). The contributions of this paper are twofold: clarifying issues related to SKEL and establishing its feasibility for a generic switch node on a rigorous basis.

I Introduction

Asynchronous Transfer Mode (ATM) has been widely accepted as the transport and switching mechanism for future broadband networks [1]. In an ATM network, a message is divided into fixed-length packets called cells for both transport and switching. Obviously, the performance of an ATM network is closely tied to the performance of the packet switches used. While there have been many isolated studies of packet switch performance and ATM network performance, studies of the specific properties of switch nodes that lead to simple traffic management policies at the network level have been relatively few. We initiate in this paper a study along this direction by focusing on a fundamental property essential for simple traffic management.

Generally, the performance of a packet switch is determined by two factors: the loadings of the input and output ports, and the output-address distributions of the packets arriving at the inputs. The first is a macroscopic and external factor which can be controlled and monitored by traffic control mechanisms at the network level. The second factor is microscopic and internal to the switch, and it requires much more information to characterize. To reduce the complexity of traffic control and to free the network design-

ers from microscopic switch considerations, it is desirable from the network-level viewpoint to have simple traffic-management rules that depend only on the first factor. This is the motivation behind a property which we term the sufficiency of the knowledge of external loads (SKEL).

To explain SKEL, let us consider traffic characterization at a switch node. In the simplest form, the traffic can be described by a traffic matrix, $[\lambda_{ij}]$, where λ_{ij} is the traffic load from input i to output j . The loads of input i and output j are the sum of row i , $R_i = \sum_j \lambda_{ij}$, and the sum of column j , $S_j = \sum_i \lambda_{ij}$, respectively. SKEL is defined as follows:

Definition 1: *Property of the Sufficiency of the Knowledge of External Loads (SKEL)* - A node is said to have SKEL if one can set bounds on R_i (for all i) and S_j (for all j) to guarantee the grades-of-service of the underlying traffic while maintaining reasonable throughput. The specific values of the individual entries in the traffic matrix should not matter given the knowledge of the external loads.

The reason for including in the above definition the condition of maintaining reasonable throughput is to avoid the trivial case where bounds set to values close to zero. It is worth noting that SKEL is analogous to the role of sufficient statistics in the theory of hypothesis testing in that they both refer to the minimal information needed for decision making in a system. With reference to Fig. 1, SKEL implies an important feature: there is a "firewall" between traffic streams of two different input-output pairs in the sense that the grades-of-service of these two traffic streams are decoupled. One of the reasons that SKEL is desirable is that many routing, congestion control, and flow-control mechanisms (e.g., see [2]) for traditional packet networks implicitly assume SKEL. If SKEL can also be achieved in high-speed switch nodes, then many of these mechanisms can also be adapted for high-speed packet networks.

The feasibility of SKEL depends on the architecture of the switch used, and the context in which its meaning should be interpreted depends on the temporal traffic characteristics of the individual connections

in the node and the definition of performance or grade-of-service. To see that the feasibility of SKEL depends on the switch architecture, consider as an example of infeasibility the 4×4 banyan switching network [3] in Fig. 2. As far as throughput is concerned, there is no problem with routing the traffic described by the traffic matrix in Fig. 2(a). Figure 2(b) shows a different traffic matrix with the same total loads on each row and column, but two internal links are now overloaded. This example shows that one cannot easily predict the performance of the banyan switch based on external link loadings. Indeed, because of the possibility of internal overloading, one cannot achieve SKEL with the banyan switch while maintaining reasonable throughput.

The traffic load on a link describes the *average* rate of information flow. Generally, the traffic on a link consists of contributions from a number of connections. If each component connection generates only a small traffic volume relative to the aggregate traffic on the link, the average load can usually capture the temporal characteristics of the link traffic by virtue of statistical smoothing. If this is not the case and the link traffic is bursty [4], parameters in addition to the average load are needed to better characterize traffic. How to characterize bursty traffic and how to deal with it are still lively and unsettled research topics. Therefore, this paper will only focus on the former situation so as to demonstrate the role of SKEL in the simplest terms. We believe our approach can be extended to more sophisticated traffic types by including traffic descriptors in addition to average load.

Certainly, what link loading is acceptable is dependent on the performance measure assumed, be it packet-loss probability or delay. The feasibility of SKEL must be defined in terms of meeting one or more performance measures; in practice, all performance measures that are essential for meeting the required grades-of-service must be taken into account. This paper concentrates on the packet-loss probability of switches based on the Knockout principle [6], and the mean and variance of delay of the generic output-queued switch.

The remainder of this paper is organized as follows. Section II provides some definitions and preliminary discussion that lay the groundwork for later discussion. Section III considers a limited form of SKEL with respect to traffic at a tagged output, with focus on the packet-loss probability, mean delay, and variance of delay. Section IV builds on the results of Section III to derive the feasibility of SKEL for traffic of an input-output pair. Further remarks and conclusions are presented in Section IV.

II Preliminary

Let us define several traffic distributions precisely before continuing the discussion:

Definition 2: Uniform Global Traffic Distribution – A node is said to have uniform global traffic distribution if all the entries in the traffic matrix are the same. That is, $\lambda_{ij} = R_i/N = S_j/N = c$ for some constant $c \leq 1/N$ for all i and j .

Definition 3: Uniform Output Traffic Distribution – An output j is said to have uniform output traffic distribution if all entries in the column j of the traffic matrix are the same. That is, $\lambda_{ij} = S_j/N$ for all i , so that loads from the inputs are balanced.

Comment: Note that whereas “uniform global traffic distribution” is a property associated with the overall switch, “uniform output traffic distribution” is the property associated with a particular output.

Definition 4: Uniform Input Traffic Distribution – An input i is said to have uniform input traffic distributions if all entries in the row i of the traffic matrix are the same. That is, $\lambda_{ij} = R_i/N$ for all j , so that each packet on this input is equally likely to be destined for any output.

Comment: This is a property of a particular input. If all inputs have uniform input traffic distributions, then the the load at output j is $\sum_i R_i/N$. That is, output loads are equal although the input loads are not necessary so. Also, it is easy see that “uniform output distributions at all outputs” and “uniform input distributions at all inputs” implies “uniform global traffic distribution”, and vice versa.

Definition 5: Uniform Traffic Distribution – This term will be used loosely to refer to any of the above distributions.

The assumption of uniform global traffic distribution underpins the design of most packet switch architectures [5] – [12] as far as guaranteeing acceptable performance is concerned. An obvious question one might ask is whether this assumption is justified in view of the fact that realistic switch traffic most likely does not follow the pattern.

One way to show that SKEL is achievable in a switch is to demonstrate that the performance under any nonuniform traffic is better than or close to the performance under uniform traffic. This is elaborated below. Let P be the performance measure of interest. Suppose we are interested in P as applied to traffic from input i to output j . Then, with SKEL, P must be a function of R_i and S_j only, regardless of the underlying traffic distributions from inputs to outputs. In general P is a function of λ_{ik} for all k and λ_{lj} for

all l . It is clear that P is a function of only R_i and S_j if input i has uniform input traffic distribution and output j has uniform output traffic distribution, since $\lambda_{ik} = R_i/N$ for all k and $\lambda_{lj} = S_j/N$ for all j . If we can show that P under any nonuniform traffic distribution is either upper-bounded or close to P under the uniform traffic distribution, then P under the uniform traffic distribution can be used to predict and manage traffic from input i to output j . That is, SKEL can be achieved by showing that the results from uniform-traffic analysis are applicable, either directly or with minimal modifications, to all nonuniform traffic distributions.

III Performance at a Tagged Output

There have been many proposals for packet-switch architectures [6] – [11]. We will focus on switch designs based on the Knockout principle [6] in this paper. The performance measures that we will focus on are packet-loss probability, mean delay and variance of delay. This section concentrates on SKEL for a tagged output. Building on the results here, the next section considers SKEL for a tagged input-output pair

A. Packet-Loss Probability

Output contention in a packet switch arises when several packets destined for the same output port arrive together. The contention problem can be solved by allowing up to N packets (N is the number of inputs or outputs) to reach [5] any output simultaneously. Because of the possibility of simultaneous arrivals, a queue is needed at each output to buffer packets that can not be transmitted on the output link immediately.

Letting N arriving packets reach the same output at the same time will entail very high switch complexity, especially if N is large. To solve this problem, we may allow at most L packets to reach any output simultaneously; any extra packets will be dropped, resulting in a finite packet-loss probability. The Knockout Principle [6] states that provided L is sufficiently large, the packet-loss probability can be made arbitrarily small, and that L is largely independent of N so that it is not necessary to scale L in proportion to switch size. The previous analysis on the Knockout switch [5] assumes the uniform output traffic distribution in which $\lambda_{lj} = S_j/N$ for all l . We will prove the following statement regarding the robustness of the assumption in terms of SKEL:

Property 1 Robustness of Uniform Output Traffic Assumption: Suppose the load on a tagged output

j is fixed at $S_j = \rho$, then the maximum loss probability incurred by packets destined for the tagged output occurs under the uniform output traffic distribution.

Proof of Property 1: Let us denote the packet-loss probability for the tagged output by P_{loss} . Let $S_j = \rho < 1$, and $D_\rho = (\rho_1 \ \rho_2 \ \dots \ \rho_N)$ be a vector describing the load contributions from the different inputs, where ρ_i is the contribution from input i . That is, D_ρ is column j in the traffic matrix for tagged output j . Since ρ is assumed to be fixed, $\sum_{i=1}^N \rho_i = \rho$. Let us focus on a arbitrary time slot or switch cycle. We can interpret ρ_i as the probability that there is a packet arriving from input i to the tagged output. For uniform output traffic, $\rho_i = \rho/N$ for all i . In general, given a fixed total load ρ , D_ρ can be either uniform or nonuniform. Let us denote explicitly the uniform traffic vector by $U_\rho = (\rho/N \ \rho/N \ \dots \ \rho/N)$.

Consider a nonuniform-traffic vector D_ρ . In general, D_ρ can differ from U_ρ in many of its entries. However, to facilitate our proof, we can define a sequence of traffic distributions $D_\rho^{(1)} = D_\rho$, $D_\rho^{(2)} = (\rho_1^{(2)} \ \rho_2^{(2)} \ \dots \ \rho_N^{(2)})$, \dots , $D_\rho^{(M)} = (\rho_1^{(M)} \ \rho_2^{(M)} \ \dots \ \rho_N^{(M)}) = U_\rho = (\rho/N \ \rho/N \ \dots \ \rho/N)$ such that two adjacent distributions differ in only two of their entries. In particular, as we move from one distribution to the next (from left to right), a small amount of load, Δ_{lm} , is shifted from some input l to some input m , with the restriction that the resulting load from input l is still at least as large as the resulting load from input m . That is, if $l > m$, $D_\rho^{(i+1)} - D_\rho^{(i)} = (0 \ \dots \ 0 \ \Delta_{lm} \ 0 \ \dots \ 0 \ -\Delta_{lm} \ 0 \ \dots \ 0)$, where $\Delta_{lm} > 0$ and $\rho_l^{(i+1)} = \rho_l^{(i)} - \Delta_{lm} \geq \rho_m^{(i+1)} = \rho_m^{(i)} + \Delta_{lm}$. Such a sequence, although not unique, always exists. This is obvious if one considers an analogous situation where we have N buckets of water. By pouring water judiciously from buckets with more water to buckets with less water, one can always end up with equal volumes of water in all buckets. For instance, a strategy is to choose at each step one bucket with more than ρ/N units of water and one bucket with less, and pour water from the former to the latter until one of them has exactly ρ/N units of water. In $(N - 1)$ steps, all buckets would have equal volumes of water.

With the above concept, to show that $P_{loss}(U_\rho) > P_{loss}(D_\rho)$, we need only prove that P_{loss} increases as we move from one distribution to the next distribution in the sequence. Let a be the number of packet arrivals per time slot.

$$P_{loss} = \frac{1}{\rho} \sum_{k=L+1}^N (k - L) \Pr[a = k], \quad (1)$$

Let a_{lm} be the number of packet arrivals in a given time slot, excluding the arrivals from inputs l and m . Assuming arrivals from different inputs are uncorrelated,

$$\Pr[a = k] = \rho_l \rho_m \Pr[a_{lm} = k - 2] + \{\rho_l(1 - \rho_m) + (1 - \rho_l)\rho_m\} \Pr[a_{lm} = k - 1] + (1 - \rho_l)(1 - \rho_m) \Pr[a_{lm} = k], \quad (2)$$

where ρ_l and ρ_m are load contributions from inputs l and m , respectively. Substituting into (1) and simplifying,

$$P_{loss} = \frac{1}{\rho} \{\rho_l \rho_m \Pr[a_{lm} = L - 1] + (\rho_l + \rho_m) \sum_{k=L}^{N-2} \Pr[a_{lm} = k] + \sum_{k=L+1}^{N-2} (k - L) \Pr[a_{lm} = k]\}. \quad (3)$$

We can obtain from (3)

$$P_{loss}(D_\rho^{(i+1)}) - P_{loss}(D_\rho^{(i)}) = \frac{1}{\rho} \left\{ \Delta_{lm} [(\rho_l^{(i)} - \rho_m^{(i)}) - \Delta_{lm}] \right\} \Pr[a_{lm} = L - 1] > 0. \quad (4)$$

The above completes our proof. \square

Implications for Overall Switch Throughput:

Let us consider the implications of Property 1 for the throughput of the overall switch.

Property 2 Suppose the total offered load to the node, $\sum_i R_i = \sum_j S_j$, is fixed at λ , and that all inputs have uniform input traffic distributions, the minimum total throughput, or carried load, is achieved when the node has the uniform global traffic distribution.

Proof of Property 2: With uniform input traffic distributions at all inputs, the loads at all outputs are equal to λ/N . Furthermore, the P_{loss} at all outputs are equal since each input contributes an equal amount of load to each output. Now,

$$\text{Overall Switch Throughput} = \sum_j (1 - P_{loss}) S_j. \quad (5)$$

Since P_{loss} is maximized under uniform output traffic distribution, we conclude that the uniform global traffic distribution achieves the minimum overall switch throughput.

Property 3 Suppose the total offered load to the node, $\sum_i R_i = \sum_j S_j$, is fixed at λ , and that all outputs have uniform output traffic distributions, the maximum total throughput, or carried load, is achieved when the node has the uniform global traffic distribution.

Proof of Property 3: A sketch of the proof is given here. Under uniform output traffic distribution, the loads at all inputs are equal to λ/N . The outputs, however, may have different loads.

$$\text{Overall Switch Throughput} = \sum_j (1 - P_{loss}(S_j)) S_j, \quad (6)$$

where $P_{loss}(S_j)$ denotes the loss probability at output j under uniform output traffic. Now,

$$P_{loss}(S_j) = \frac{1}{S_j} \sum_{k=L+1}^N (k-L) \binom{N}{k} \left(\frac{S_j}{N}\right)^k \left(1 - \frac{S_j}{N}\right)^{N-k} \quad (7)$$

By means of differentiation, it is routine to show that $P_{loss}(S_j)$ is a convex \cap function of S_j . The overall switch throughput, being a sum of convex \cap function, is therefore convex \cap . If we set up a nonlinear optimization problem in which the overall switch throughput $= \sum_j (1 - P_{loss}(S_j)) S_j$ is to be maximized subject to the convex region $\sum_j S_j = \lambda$, we find that the switch throughput is maximized when S_j 's are all equal. That is, the switch throughput is maximized under the uniform global traffic distribution. \square

As depicted in Fig. 3, Property 2 and Property 3 show that the assumption of uniform input traffic distributions at all inputs and the assumption of uniform output traffic distributions at all outputs separate the achievable throughput into two region; the throughput of the former case is no less than the throughput of the latter case, with equality when the distributions in both cases are the uniform global traffic distributions.

B. Mean Delay

We now turn our attention to packet delay at the output queue. This subsection focuses on mean delay, and the next subsection considers variance of delay. For analysis of delay in this paper, we will assume a switch in which up to N packets can arrive together at an output [5, 10] to avoid consideration of the possibility of packet loss in the switch. This is a good approximation for situations in which the packet-loss probability due to switch design is small. In addition,

we will assume the output buffers to be sufficiently large so that packet loss due to buffer overflows is negligible. Henceforth, the uppercase letters will be used to denote the z -transform of the random variables denoted by the corresponding lowercase letters.

We will first derive some useful preliminary results. Recall that a is the number of arriving packets to the tagged output at the beginning of a time slot. Assuming packet arrivals from different inputs are independent, the z -transform of a , $\sum_i \Pr[a = i]z^i$, is

$$A(z) = \prod_{i=1}^N (1 - \rho_i + \rho_i z), \quad (8)$$

where ρ_i is the probability that there is an arriving packet from input i to the tagged output. By differentiating $A(z)$ successively and setting $z = 1$, we get

$$E[a] = A'(1) = \sum_{i=1}^N \rho_i = \rho \quad (9)$$

$$E[a(a-1)] = A''(1) = \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N \rho_i \rho_j \quad (10)$$

$$\begin{aligned} E[a(a-1)(a-2)] &= A'''(1) \\ &= \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N \sum_{\substack{k=1 \\ k \neq i, j}}^N \rho_i \rho_j \rho_k \quad (11) \end{aligned}$$

The following lemma is useful for later derivations:

Lemma 1 Given $\sum_i \rho_i = \rho$, $A''(1) \leq \frac{(N-1)\rho^2}{N}$ and $A'''(1) \leq \frac{(N-1)(N-2)\rho^3}{N^2}$. In other words, $A''(1)$ and $A'''(1)$ are maximum when $\rho_i = \rho/N$ for all i .

Proof of Lemma 1: We will prove the lemma for $A''(1)$. The proof for $A'''(1)$ is similar. As in the analysis for P_{loss} , we can define a sequence of traffic distributions $D_\rho^{(1)} = D_\rho$, $D_\rho^{(2)} = (\rho_1^{(2)} \ \rho_2^{(2)} \ \dots \ \rho_N^{(2)})$, \dots , $D_\rho^{(M)} = (\rho_1^{(M)} \ \rho_2^{(M)} \ \dots \ \rho_N^{(M)}) = U_\rho = (\rho/N \ \rho/N \ \dots \ \rho/N)$ through which we can move from any nonuniform traffic distribution to the uniform traffic distribution by shifting a fraction of load from one input to another input as we move from one distribution to the next in the sequence (see the preceding section). It is then sufficient to show that for two adjacent distributions, $D_\rho^{(i)}$ and $D_\rho^{(i+1)}$, with $\rho_j^{(i)} = \rho_j^{(i+1)}$ for all j except that $\rho_l^{(i+1)} = \rho_l^{(i)} - \Delta_{lm}$, $\rho_m^{(i+1)} = \rho_m^{(i)} + \Delta_{lm}$, and $\rho_l^{(i+1)} \geq \rho_m^{(i+1)}$, $A''(z = 1, D_\rho^{(i+1)}) > A''(z = 1, D_\rho^{(i)})$. This is simple:

$$A''(z = 1, D_\rho^{(i+1)}) - A''(z = 1, D_\rho^{(i)})$$

$$\begin{aligned} &= 2\{(\rho_l^{(i)} - \Delta_{lm} + \rho_m^{(i)} + \Delta_{lm}) \sum_{j \neq l, m} \rho_j^{(i)} \\ &\quad + (\rho_l^{(i)} - \Delta_{lm})(\rho_m^{(i)} + \Delta_{lm})\} \\ &\quad - 2\{(\rho_l^{(i)} + \rho_m^{(i)}) \sum_{j \neq l, m} \rho_j^{(i)} + \rho_l^{(i)} \rho_m^{(i)}\} \\ &= 2\Delta_{lm} \{\rho_l^{(i)} - \rho_m^{(i)} - \Delta_{lm}\} > 0 \quad (12) \end{aligned}$$

The above completes the proof of the lemma. \square

Let $b(j)$ be the number of packets in the output queue at the end of time slot j and $a(j+1)$ be the number of packet arrivals at the beginning of time slot $j+1$. Then,

$$b(j+1) = \max[b(j) - 1 + a(j+1), 0]. \quad (13)$$

We will assume arrivals to be independent of the state of the queue (i.e., $a(j+1)$ and $b(j)$ are independent) In the equilibrium state (as $j \rightarrow \infty$), the subscript j can be dropped, and it can be shown [5, 4] that

$$zB(z) = (z-1)P_0 + A(z)B(z), \quad (14)$$

where $P_0 = \Pr[a+b=0]$. Differentiating the above and setting $z = 1$, we find that $P_0 = 1 - A'(1)$. Differentiating Eqn. (14) twice and setting $z = 1$, we get

$$E[b] = B'(1) = \frac{A''(1)}{2(1-A'(1))} = \frac{A''(1)}{2(1-\rho)}. \quad (15)$$

Differentiating Eqn. (14) three times and setting $z = 1$ yields

$$\begin{aligned} E[b(b-1)] &= B''(1) \\ &= \frac{3A''(1)B'(1) + A'''(1)}{3(1-A'(1))} \quad (16) \end{aligned}$$

$$= \frac{A''(1)^2}{2(1-\rho)^2} + \frac{A'''(1)}{3(1-\rho)}. \quad (17)$$

With the above results, the following property can be easily proved:

Property 4 Robustness of Uniform Output Traffic Assumption: Suppose the load on a tagged output j is fixed at $S_j = \rho$, then the maximum mean delay at the tagged output occurs under the uniform output traffic distribution.

Proof of Property 4: Consider any arbitrary nonuniform output traffic distribution. Let d be the

delay experienced by an arbitrary packet from any input. By Little's Law, the expected waiting time of a packet is $B'(1)/\rho$ in units of time slot. Therefore,

$$E[d] = \frac{B'(1)}{\rho} + 1 = \frac{A''(1)}{2\rho(1-\rho)} + 1 \quad (18)$$

$$\leq \frac{(N-1)\rho}{2N(1-\rho)} + 1, \quad (19)$$

$$\approx \frac{\rho}{2(1-\rho)} + 1, \quad (20)$$

for large N . The property is established by observing that the second line is obtained from Lemma 1, and it is the mean delay under uniform output traffic. \square

The property also allows us to achieve SKEL for mean delay by deriving an upper bound which is independent of the individual values of ρ_i 's.

C. Variance of Delay

A similar property as the above can be established for the variance of delay at the tagged output:

Property 5 Robustness of Uniform Output Traffic Assumption: Suppose the load on a tagged output j is fixed at $S_j = \rho$, then the maximum variance of delay at the tagged output occurs under the uniform output traffic distribution.

Proof of Property 5: Consider an arbitrary packet arriving at the tagged output. There may be other packets that arrive together with the tagged packet. Let h be the number of these simultaneously-arriving packets that are placed in the FIFO (first-in-first-out) queue ahead of the tagged packet. The delay of the tagged packet is given by

$$d = b + h + 1 \quad (21)$$

By the assumption that the arrival process is independent of the queue state, we have

$$\text{Var}(d) = \text{Var}(b) + \text{Var}(h). \quad (22)$$

Now,

$$\begin{aligned} \text{Var}(b) &= B''(1) + B'(1) - B'(1)^2 \\ &= \frac{A''(1)^2}{4(1-\rho)^2} + \frac{A''(1)}{2(1-\rho)} + \frac{A'''(1)}{3(1-\rho)}, \end{aligned} \quad (23)$$

by substituting from (15) and (17). To find $\text{Var}(h) = E[h^2] - E^2[h]$, let us assume that arriving packets are loaded into the output queue in a random order without any input being favored. Let h_i be the value of

h conditioned on the tagged packet being originated from input i . We have,

$$E[h] = \frac{\sum_i \rho_i E[h_i]}{\rho} \quad (24)$$

$$E[h^2] = \frac{\sum_i \rho_i E[h_i^2]}{\rho} \quad (25)$$

To find $E[h_i]$ and $E[h_i^2]$, let a_i be the number of other packets that arrive simultaneously with the tagged packet. We first derive the conditional probability, $\Pr[h_i = k | a_i = j]$.

$$\Pr[h_i = k | a_i = j] = \begin{cases} \frac{1}{j+1} & \text{for } 0 \leq k \leq j \\ 0 & \text{otherwise} \end{cases} \quad (26)$$

From this, we get

$$H_i(z | a_i = j) = \frac{1}{j+1} \sum_{k=0}^j z^k = \frac{1}{j+1} \left(\frac{1-z^{j+1}}{1-z} \right). \quad (27)$$

Removing the conditioning on a_i ,

$$(1-z)H_i(z) = \sum_{j=0}^{N-1} \left(\frac{1-z^{j+1}}{j+1} \right) \Pr[a_i = j]. \quad (28)$$

Differentiating the above twice with respect to z and setting $z = 1$, we find

$$E[h_i] = H_i'(1) = A_i'(1)/2. \quad (29)$$

Differentiating (28) three times and setting $z = 1$, we have

$$E[h_i(h_i - 1)] = E[h_i^2] - E[h_i] = H_i''(1) = A_i''(1)/3. \quad (30)$$

Substituting (29) and (30) into (24) and (25), we get

$$\begin{aligned} E[h] &= \frac{\sum_i \rho_i A_i'(1)}{2\rho} \\ &= \frac{A''(1)}{2\rho}, \end{aligned} \quad (31)$$

by observing (10), and

$$\begin{aligned} E[h^2] &= \frac{\sum_i \rho_i A_i''(1)}{3\rho} + \frac{\sum_i \rho_i A_i'(1)}{2\rho} \\ &= \frac{A'''(1)}{3\rho} + \frac{A''(1)}{2\rho}, \end{aligned} \quad (32)$$

by observing (10) and (11). Therefore,

$$\begin{aligned} \text{Var}(d) &= \text{Var}(b) + E[h^2] - E^2[h] \\ &= \frac{A''(1)^2}{4(1-\rho)^2} + \frac{A''(1)}{2(1-\rho)} + \frac{A'''(1)}{3(1-\rho)} + \frac{A''(1)}{3\rho} \end{aligned} \quad (33)$$

$$\begin{aligned}
& + \frac{A''(1)}{2\rho} - \left(\frac{A''(1)}{2\rho} \right)^2 \quad (34) \\
\leq & \frac{(N-1)^2 \rho^4}{4N^2(1-\rho)^2} + \frac{(N-1)\rho^2}{2N(1-\rho)} \\
& + \frac{(N-1)(N-2)\rho^3}{3N^2(1-\rho)} + \frac{(N-1)(N-2)\rho^2}{3N^2} \\
& + \frac{(N-1)\rho}{2N} - \left(\frac{(N-1)\rho}{2N} \right)^2 \quad (35) \\
\approx & \frac{2\rho^3 - 5\rho^2 + 6\rho}{12(1-\rho)^2}, \quad (36)
\end{aligned}$$

for large N . Inequality (35), which depends only on the output load, is obtained by Lemma 1, and it is satisfied with equality when the output traffic distribution is uniform. Therefore, as with P_{loss} and $E[d]$, $\text{Var}(d)$ is also maximized under uniform output traffic. This completes the proof.

IV Performance of a Tagged Input-Output Pair

The previous section concerns performance at a tagged output. Packets to this output can originate from multiple input sources. We now focus on a tagged input, in addition to the tagged output, in order to derive performance of this tagged input-output pair.

A. Packet-Loss Probability

We will assume that random choices are made of the packets to be dropped when more than L packets arrive for the tagged output. Under uniform output traffic, the loss probability of packets from any particular input to the tagged output is the same as P_{loss} , the packet-loss probability of the tagged output. Under nonuniform output traffic, however, loss probability is higher for inputs with lower loads; furthermore, the loss probability of the the worst-case input can actually become higher than P_{loss} under uniform output traffic. Nevertheless, the worst-case loss probability of an input-output pair can be closely estimated by P_{loss} under uniform output traffic, as summarized below:

Property 6 Robustness of Uniform Output Traffic Assumption: The packet-loss probability of a tagged input-output pair can be bounded tightly from above by a quantity that depends only on the load of the tagged output. Furthermore, the packet-loss probability of the tagged output under uniform output traffic is a good estimate for this bound.

Proof of Property 6: Consider without loss of generality the packets from input 1 to the tagged output. Let Q denote the corresponding loss probability.

$$Q = \sum_{k=L}^{N-1} \frac{k+1-L}{k+1} \Pr[a_1 = k], \quad (37)$$

where a_1 is the number of packet arrivals excluding the arrival from input 1. It is easily seen from the above that the smaller the ρ_1 , the larger the Q , since the contention from the other inputs will be more intense due to their higher shares of the aggregate loading. This is also the reason why nonuniform output traffic may give rise to higher worst-case Q than uniform traffic may. Let us use the notation $P_{loss}(N, L, \rho_1, \rho_2, \dots, \rho_N)$ to denote explicitly the dependence of loss probability of the tagged output on switch size N , Knockout parameter L , and the input load distributions. Continuing from (37),

$$\begin{aligned}
Q &= \sum_{k=L}^{N-1} \frac{k+1-L}{k+1} \Pr[a_1 = k] \\
&\leq \frac{1}{L+1} \sum_{k=L}^{N-1} (k+1-L) \Pr[a_1 = k] \\
&\leq \frac{\rho}{L+1} \left\{ \frac{1}{\rho - \rho_1} \sum_{k=L}^{N-1} (k+1-L) \Pr[a_1 = k] \right\} \\
&= \frac{\rho}{L+1} P_{loss}(N-1, L-1, \rho_2, \dots, \rho_N) \quad (38) \\
&\leq \frac{\rho}{L+1} P_{loss}(N-1, L-1, (\rho - \rho_1)/(N-1), \\
&\quad \dots, (\rho - \rho_1)/(N-1)) \quad (39) \\
&\leq \frac{\rho}{L+1} P_{loss}(N-1, L-1, \rho/(N-1), \\
&\quad \dots, \rho/(N-1)). \quad (40)
\end{aligned}$$

Inequality (38) relates Q to P_{loss} of a system with $(N-1)$ input or output ports and Knockout parameter of $(L-1)$. Inequality (39) exploits the previous result that uniform traffic has the highest overall P_{loss} . The last line is simply implied by the fact that increasing the load will not decrease the P_{loss} . Note that the bound in (40) is independent of the details of the distribution $(\rho_1 \ \rho_2 \ \dots \ \rho_N)$, and therefore it can be used to achieve SKEL.

To show the tightness of the upper bound on Q , Table 1 compares this bound with P_{loss} under uniform traffic for various L , assuming $\rho = 0.9$ and $N \rightarrow \infty$. It should be understood that since the upper bound on Q applies to any traffic distribution, it can not be lower than the P_{loss} (for the same N and L), which is the specific Q under the uniform output traffic distribution. We see that for $L \geq 3$, the bound is very good

Table 1: Upper bound for Q and P_{loss} for various L , assuming $\rho = 0.9$ and $N \rightarrow \infty$.

L	Bounds on Q	P_{loss} under uniform traffic
1	4.50×10^{-1}	3.41×10^{-1}
2	1.02×10^{-1}	8.78×10^{-2}
3	1.98×10^{-2}	1.80×10^{-2}
4	3.24×10^{-3}	3.04×10^{-3}
5	4.56×10^{-4}	4.36×10^{-4}
6	5.60×10^{-5}	5.42×10^{-5}
7	6.09×10^{-6}	5.93×10^{-6}
8	5.93×10^{-7}	5.81×10^{-7}
9	5.23×10^{-8}	5.14×10^{-8}
10	4.20×10^{-9}	4.14×10^{-9}

because it is very close to P_{loss} . Since large enough L will be required to achieve low loss probability in practice, we conclude that that the P_{loss} obtained by the uniform-traffic analysis is good enough as an estimate for the worst-case Q under nonuniform traffic. Therefore either the bound itself or the P_{loss} under uniform traffic can be used to achieve SKEL.

□

B. Mean and Variance of Delay

The following properties applies to the mean delay of a tagged input-output pair. The proofs are rather similar to the proofs of preceding properties and will not be presented here due to space limitation.

Property 7 Robustness of Uniform Output Traffic Assumption: The mean delay of a tagged input-output pair can be bounded tightly from above by a quantity that depends only on the load of the tagged output. Furthermore, the mean delay of the tagged output under uniform output traffic is a good estimate for this bound.

Property 8 Robustness of Uniform Output Traffic Assumption: The variance of delay of a tagged input-output pair can be bounded tightly from above by a quantity that depends only on the load of the tagged output. Furthermore, the variance of delay of the tagged output under uniform output traffic is a good estimate for this bound.

V Conclusions

To simplify traffic control in a network, it is desirable that the traffic-control policy at a network node

depends only on the external traffic loads on the input and output links, but not on the detailed addressing or distribution of packets from inputs to outputs. Switch nodes in which such a traffic-control policy is possible are said to have the property of the sufficiency of the knowledge of external loads (SKEL).

Although many studies of network-traffic control assume SKEL implicitly, this property is by no means guaranteed automatically, and its feasibility depends on the switch architecture adopted. This paper has contributed in two ways: clarifying issues related to SKEL and establishing its feasibility for the generic output-queued switch on a rigorous basis.

To demonstrate SKEL, we must show that traffic performance depends only on external link loadings. The key is to show that the performance under any nonuniform traffic distribution from inputs to outputs is better than or close to the performance under the uniform traffic distribution. This paper has shown that for switches that adopt the Knockout principle [6], the packet-loss probability for packets destined for any tagged output is maximum when the traffic to the output originates uniformly from all the inputs. This means that the formula obtained from the uniform-traffic analysis serves as an upper bound on the packet-loss probability under nonuniform traffic. Therefore, this formula can be used for network-traffic control purposes without the need to worry about the consequence of traffic nonuniformity. We have also shown that the loss probability under uniform traffic closely approximates the worst-case loss probability of a traffic stream from any particular input to any particular output. In addition, the implications of uniform traffic for the overall switch throughput have been established.

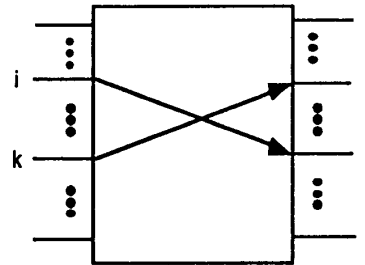
Similar results for mean and variance of delay have also been obtained. Although not presented in this paper, it is also possible to show that for a waiting system (or input-queued switch) [5, 10], SKEL for saturation throughput can be achieved in the sense that external link loading that does not result in queue saturation under uniform traffic will not do so under nonuniform traffic either.

In conclusion, although there have been many papers showing certain nonuniform traffic patterns may result in degraded performance in nonblocking switches (e.g., [13, 14]), this paper in contrast has shown that any nonuniform traffic pattern must result in performance close to or better than that of uniform traffic. These discrepancies are due to different assumptions and definitions of traffic-nonuniformity. We have adopted a network-level viewpoint in interpreting

switch-performance results. While the previous studies have allowed the external link loads to vary when comparing uniform traffic with nonuniform traffic, we have fixed the external link loads to be the same in both cases but allowed the distributions from inputs to outputs to vary. The intent of this additional restriction is to add meaning and fairness to the comparison so that network-level implications can be deduced. An outcome is the demonstration that link-loading control mechanisms based on uniform-traffic analysis are also robust even when the traffic is nonuniform.

References

- [1] CCITT, New Draft Recom. I. 150, B-ISDN ATM Layer Functionality and Specifications, *Committee XVIII*, Jan. 1990.
- [2] D. Bertsekas and R. Gallager, *Data Networks*, Prentice-Hall, 1987.
- [3] C. L. Wu and T. Y. Feng, *Tutorial: Interconnection Networks for Parallel and Distributed Processing*, IEEE Computer Society Press, 1984.
- [4] S. C. Liew and K. W. Lu, "Comparison of Buffering Strategies for Asymmetric Packet Switch Modules," *IEEE J. on Selected Areas in Communications*, vol 9, no 3, April 91, pp. 428-438.
- [5] M. Y. Karol, M. G. Hluchyj, and S. Morgan, "Input Versus Output Queueing on a Space-Division Packet Switch," *IEEE Trans. on Commun.*, vol. 35, no. 12, pp. 1347-1356, Dec. 1987.
- [6] Y. Yeh, M. Hluchyj, A. Acampora, "The Knock-out Switch: A Simple Modular Architecture for High-Performance Packet Switching," *IEEE J. on Selected Areas in Commun.*, vol. 5, no. 8, pp. 1274-1283, Oct. 1987.
- [7] J. Y. Hui and E. Arthur, "A Broadband Packet Switch for Integrated Transport," *IEEE J. on Selected Areas in Commun.*, vol. 5, no. 8, pp. 1264-1273, Oct. 1987.
- [8] T. T. Lee, "A Modular Architecture for Very Large Packet Switch," *IEEE Trans. on Commun.*, vol. 6, pp. 1455-1467, July 1990.
- [9] S. C. Liew and K. W. Lu, "A 3-stage Interconnection Structure Structure for Very Large Packet Switches," *International J. of Digital and Analog Communications*, vol. 2, pp. 303-316, 1989.
- [10] S. C. Liew and T. T. Lee, "Broadband Packet Switches based on Dilated Interconnection Networks," accepted for a future issue of *IEEE Trans. on Commun.*
- [11] S. C. Liew and T. T. Lee, " $N \log N$ Dual Shuffle-exchange Network with Error-correcting routing," *Conf. Record, IEEE ICC '92*.
- [12] Y. Oie *et al.*, "Effect of Speedup in Nonblocking Packet Switch," *Conf. Record, ICC '89*, vol. 1, pp. 410-415.
- [13] S.-Q. Li and M. J. Lee, "A Study of Traffic Imbalances in a Fast Packet Switch," *Conf. Record, ICC '89*, pp. 538-545.
- [14] J.S.-C Chen and T. E. Stern, "Throughput Analysis, Optimal Buffer Allocation, and Traffic Imbalance Study of a Generic Nonblocking Packet Switch," *IEEE J. on Selected Areas in Communications*, April 91, pp. 439-449.



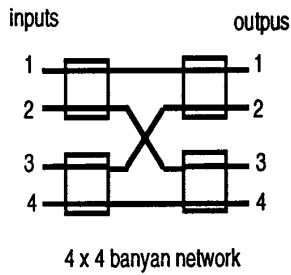
- With SKEL, the performance of the traffic stream from i to j can be guaranteed independently of the traffic stream from l to k .
- For the traffic stream from i to j , performance can be guaranteed by bounding the aggregate external loads at input i and output j , independently of the makeup of the aggregate loads (i.e., independently of the output address distribution of the input packets at i and the load contributions from the individual inputs to output j).

Figure 1. Implications of SKEL for the control of traffic streams from inputs to outputs

traffic matrix
(entry ij is load from
input i to output j):

$$\begin{bmatrix} 0.6 & 0 & 0 & 0 \\ 0 & 0 & 0.6 & 0 \\ 0 & 0.6 & 0 & 0 \\ 0 & 0 & 0 & 0.6 \end{bmatrix}$$

routes from inputs to outputs
based on traffic matrix on left

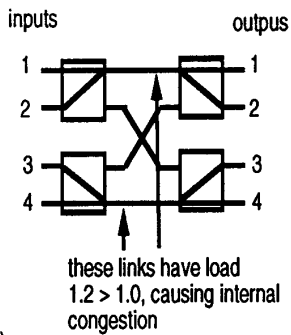


(a)

traffic matrix:

$$\begin{bmatrix} 0.6 & 0 & 0 & 0 \\ 0 & 0.6 & 0 & 0 \\ 0 & 0 & 0.6 & 0 \\ 0 & 0 & 0 & 0.6 \end{bmatrix}$$

routes from inputs to outputs
based on traffic matrix on left



(b)

Figure 2. (a) A traffic distribution that can be routed successfully by banyan network; (b) A traffic distribution that cannot be routed successfully by banyan network.

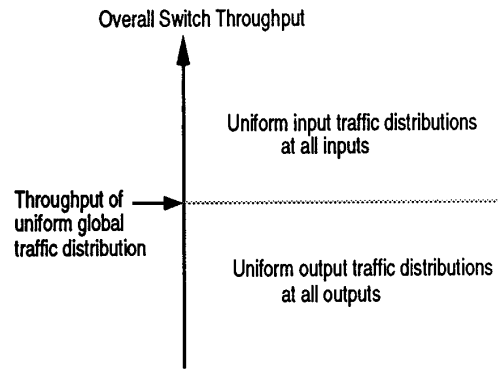


Figure 3. Relationships among different regions of uniform traffic distributions as for the overall switch throughput.