Quantum Packet for the Next Generation Network/ISDN3

Ray Y. W. Lam, Henry C. B. Chan, Hui Chen, Tharam S. Dillon, Victor O. K. Li, and Victor C. M. Leung

Abstract: **This paper proposes a novel method for transporting various types of user traffic effectively over the next generation network called integrated services digital network 3 (ISDN3) (or quantum network) using quantum packets. Basically, a quantum packet comprises one or more 53-byte quanta as generated by a "quantumization" process. While connection-oriented traffic is supported by fixed-size quantum packets each with one quantum to emulate circuit switching, connectionless traffic (e.g., IP packets and active packets) is carried by variable-size quantum packets with multiple quanta to support store-and-forward switching/routing. Our aim is to provide frame-like or datagram-like services while enabling cell-based multiplexing. The quantum packet method also establishes a flexible and extensible framework that caters for future packetization needs while maintaining backward compatibility with ATM. In this paper, we discuss the design of the quantum packet method, including its format, the "quantumization" process, and support for different types of user traffic. We also present an analytical model to evaluate the consumption of network resources (or network costs) when quantum packets are employed to transfer loss-sensitive data using three different approaches: cut-through, store-and-forward and ideal. Close form mathematical expressions are obtained for some situations. In particular, in terms of network cost, we discover two interesting equivalence phenomena for the cut-through and store-and-forward approaches under certain conditions and assumptions. Furthermore, analytical and simulation results are presented to study the system behavior. Our analysis provides valuable insights into the design of the ISDN3/quantum network.**

Index Terms: **ATM, ATM adaptation layer (AAL), integrated services digital network (ISDN), ISDN3, next-generation networks, quantumization, quantum packets.**

I. INTRODUCTION

A key issue in networking is how to packetize data so that it can be effectively and efficiently transferred over a wide-area network. Ever-increasing connectionless data traffic, the integration of the link and network layers, and the development of active networks make the study of a unified packetization method for the next generation network a matter of considerable interest.

In the last quarter of a century, we have seen two generations of integrated services digital network (ISDN) employed in the development of telecommunications infrastructure. The first generation of ISDN (ISDN1) was introduced in the 1980s to integrate narrowband voice, data, and video services in digital telephone systems [1]. The second generation of ISDN (ISDN2), known as ATM, was developed in the 1990s to support broadband traffic using cell switching [2]–[4]. In ISDN1, both fixed-size and variable-size frames were employed for packetization purposes. Basically, fixed-size frames were used to multiplex user traffic in pre-defined channels [5]. Although this minimizes overheads and transfer delays, it operates only under a synchronous architecture so it is not flexible. Another disadvantage of using fixed-size frames is that their inflexible format makes them less extensible and less effective in controlling priority. To transport asynchronous traffic more effectively, frame relay was introduced to enable efficient packet switching by means of variable-size frames [6], [7]. As the packet[1] size is variable, it is necessary to specify where a packet or frame ends. A special flag is included in the frames for this purpose. However, like other conventional link layer protocols, this creates a burden due to the use of "stuffing bits" [2]. Although we can achieve very high payload utilization by employing variable-size frames, the use of variable sizes generally makes it more difficult to process them at high speed. Moreover, it is not desirable with respect to priority control as high-priority packets may be delayed by large-sized low-priority packets. ISDN2 provides a mechanism for emulating circuit switching using packet switching. This is accomplished by using small fixed-size packets (i.e., ATM cells) with a header to specify the associated "circuit" for switching purposes [2], [3]. The use of cells both facilitates the development of high-speed switches and provides better priority control. ATM also defines a framework for specifying various service classes by using ATM adaptation layers (AALs). However, because it is connection-oriented (CO), it is less effective in supporting connectionless (CL) traffic.

At the network layer, the most popular packetization method is to use variable-size Internet datagrams [2], [8]. In this case, the packet length is specified in the packet header so as to identify the end of a packet. To ensure correctness of the header (e.g., the packet length information is correct), some error checking bits are included. Similar to frame relay frames, Internet datagrams are less effective in the aspects of priority control and ease of processing. While packets conventionally only transport data, they can also carry programs that can be run at active network nodes [9], [10]. As each active packet can be run independently,

Manuscript received December 18, 2003; approved for publication by Dan Keun Sung, Division III Editor, August 04, 2005.

This project has been supported by the Competitive Earmarked Research Grant of the Research Grants Council, the Hong Kong Special Administrative Region, China, under account number PolyU 5085/00E. We would like to thank the anonymous reviewers for providing us with valuable comments for improving this paper.

R. Y. W. Lam, H. Chen, and T. S. Dillon were with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China, T. S. Dillon is now with the Curtin University of Technology, Kent Street, Bentley, Perth, Australia, email: yingwail@ece.ubc.ca, huic@ece.ubc.ca, Tharam.Dillon@cbs.curtin.edu.au.

H. C. B. Chan is with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China, email: cshchan@comp.polyu.edu.hk.

V. O. K. Li is with the Department of Electrical and Electronic Engineering, the University of Hong Kong, Hong Kong, China, email: vli@eee.hku.hk.

V. C. M. Leung is with the Department of Electrical and Computer Engineering, the University of British Columbia, Canada, email: vleung@ece.ubc.ca.

---

[1] In this paper, the terms "frames," "cells," and "packets" are used in an interchangeable manner.

1229-2370/08/$10.00 © 2008 KICS

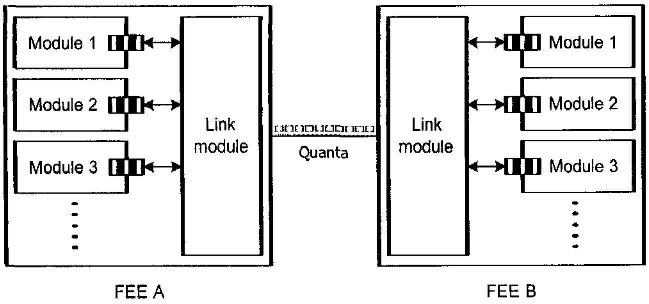


Fig. 1. FEE architecture.

Table 1. Comparison of ISDN1, ISDN2, ISDN3, and the Internet.

	ISDN1 (Circuit-switched ISDN and frame relay)	ISDN2 (ATM)	ISDN3 (Quantum network)	Internet
Base network	Connection-oriented	Connection-oriented	Connectionless	Connectionless
Transmission unit	Frame	Cell	Quantum packet	Datagram
Traffic forwarding device	Switch	Switch	FEE	Router
Network nature	Non-active with little intelligence	Non-active with some intelligence	Active with high intelligence	Non-active with some intelligence
Basic design philosophy	Integrate narrowband traffic in the digital telephone system	Integrate narrowband and broadband traffic over a cell-switching network	Integrate both network traffic and network functions over the same network	Provide best-effort service

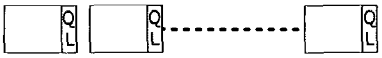
new services can be offered dynamically and highly customized services can be supported more easily. Related active network projects include Janos for providing a Java-based operating system for active network nodes [11] and NetScript for developing a programming tool for active networks [12].

Although the AAL packetization method has been employed extensively as a way of supporting different types of data traffic over ATM networks, it may not be effective in addressing the various needs of the next generation network because of the CO nature of the ATM. Therefore it would be desirable to develop a more flexible and extensible packetization framework to support both CO and CL data traffic at different layers of the network model. This paper seeks to contribute to this important goal in four ways. First, we propose a novel packetization method called the quantum packet for the next-generation network called ISDN3 (or quantum network) [13]. This packetization method can provide frame/datagram-like services while enabling cell-based multiplexing. Moreover, it is generally backward compatible with ATM cells (i.e., ATM cells can be viewed as one type of quantum packet—that is, a quantum packet with one quantum). Second, as an extension to or a generalization of the AAL process, we propose a "quantumization" process (also called AAL+) to support both cut-through (e.g., ATM-based) and store-and-forward (e.g., MPLS-based) traffic in a unified manner. The proposed method can also be used to encapsulate active data traffic. Third, we formulate an analytical model for evaluating the consumption of network resources when quantum packets are used to transfer loss-sensitive traffic. In particular, some close form mathematical expressions are derived. Finally, we present analytical and simulation results to study the system behavior. With the advent of data-oriented applications (e.g., Internet applications), there is an increasing need for the transport not only of time-sensitive but also of loss-sensitive traffic over a wide-area network. Our analysis provides valuable insights into the design of the next-generation network for these applications.

The remaining sections of the paper are organized as follows. Section II gives an overview of ISDN3. Section III presents the details of the quantum packet method. Section IV presents a theoretical model related to the consumption of network resources. Section V discusses the analytical and simulation results. Section VI concludes this paper.

II. OVERVIEW OF QUANTUM NETWORK/ISDN3

ISDN1 and ISDN2 were actively developed during, respectively, the 1980s and 1990s. As we enter the 2000s, it is an appropriate time to initiate research into the next generation of ISDN, which is referred to as ISDN3. In particular, we propose the realization of ISDN3 through the quantum network. Apart from ATM, the quantum network is inspired by a number of networking technologies and research projects, including multiprotocol label switching (MPLS) [14]–[16], A/I Net [17] and IthACI [18], various active network projects [9]–[12], and Internet2 [19]. In essence, our aim is to integrate these networking technologies using a common framework. While both ISDN1 and ISDN2 were built over a connection-oriented architecture, the explosive growth of the Internet and the emergence of active network services may require us to rethink this design principle. In our view, the integration of the ATM, Internet and active network will form the basis for ISDN3. Table 1 compares the main features of ISDN1, ISDN2, ISDN3, and the Internet. The base network for ISDN3 is connectionless. This facilitates the provision of different network services: Whether connection-oriented, connectionless or active. Motivated by the recent developments in integrating layer 2 and layer 3 technologies [14]–[18], [20]–[23], the traditional link and network layers integrate to form a forwarding layer with different partitions that support diverse network functions such as switching, routing and active network functions. We call this horizontal partitioning, which complements the traditional vertical layering network model. Essentially, ISDN3 combines not only various types of traffic but also a spectrum of network functions, thus creating a truly integrated network. As will be explained later, it employs a novel quantum packet method that supports frame/datagram-like services and active network capabilities while enabling cell-based multiplexing. Basically, a quantum packet comprises one or multiple quanta. It can be used to effectively support not only cut-through switching but also store-and-forward switching/routing. Quantum packets are generated by a "quantumization" process that is explained in the next section. This packetization process can be viewed as a generalized AAL process or an AAL+ process. To integrate different network services, a modular traffic forwarding device called the forwarding engine (FEE) is proposed. The proposed name (FEE) reflects what it does (i.e., forwarding packets) but not how it does it. This is because a wide range of forward-



Cut-through (m quantum packets each with one quantum)

Store-and-forward (one quantum packet with m quanta)

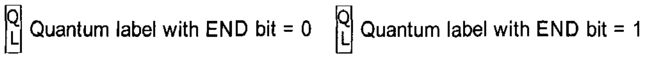


Fig. 2. Quantum packets with one and multiple quanta.

ing options are available. The implementation details of FEEs are outside the scope of this paper. Here, we give a schematic overview in Fig. 1. As will be explained in Section III, a FEE forwards quantum packets using the type-merging mechanism, which is similar to virtual circuit (VC)-merging. This means that a FEE can be built based on the architecture of existing VC-merging switches/routers, such as the ones proposed in [16] and [17]. Inspired by the architectures in [16] and [17], we outline a possible approach here. As shown in Fig. 1, different network modules can be installed in a FEE as necessary. The link module is for linking with other FEEs in the network. It mainly consists of incoming and outgoing ports with buffers for storing quantum packets. When a FEE receives a quantum from an incoming port, the quantum will be transferred to the corresponding module for processing according to its type (i.e., similar to [17]). For example, upon receiving an MPLS-based packet, it will be passed to the MPLS switching module. If an active packet is received, it will be processed by the active network module. When a new type of network service is introduced in the future, the respective network module can be added easily within this flexible framework. After processing (e.g., to determine the next node), a quantum will be sent to the required outgoing port in the link module for transferring to the next network node. In the case of cut-through switching service, the quantum can be transferred to the next network node independently. This is similar to ATM cell switching. To support store-and-forward switching/routing services, each quantum of a quantum packet is first buffered at the outgoing port according to its type (i.e., similar to the approaches used in [16] and [17] for supporting VC-merging). Upon receiving all the quanta of a quantum packet, the quanta can then be transferred to the next network node.

III. QUANTUM PACKETS

The quantum packet method, to be presented in this section, is an attempt to combine the advantage of frames and cells. In other words, it seeks to integrate the packetization methods used in ISDN1 and ISDN2. In general, a quantum packet consists of one or more quanta of 53 bytes as shown in Fig. 2. It functions like an extended ATM cell or a quantized frame relay frame. In particular, an ATM cell can be viewed as the simplest type of quantum packet—a quantum packet with one quantum. Moreover a quantum packet can contain multiple units (quanta) for transporting store-and-forward data traffic more effectively based on a type-merging mechanism similar to VC-merging [16], [17]. As will be discussed later, the quantum packet method is flexible and efficient in terms of packetization. Quantum packets are created in three steps using the following general "quantumization" process (see Fig. 3):

Step 1. Starting with a higher-layer data unit, an optional packet-specific header PH (h bytes) and an optional packet-specific trailer PT ($t+g$ bytes), which includes some padding bits (g bytes), are attached. The padding bits are needed to ensure that the resultant packet can be segmented into an integral number of units.

Step 2. Segment the packet as generated from step 1 into discrete units of l (to be computed later) bytes. To each unit, an optional quantum-specific header QH (a bytes) and an optional quantum-specific trailer QT (b bytes) are added. If required, it is also possible to subdivide the QH and QT into the inner and outer parts as follows:
   - QH is subdivided into inner and outer QHs: QH_i (a_i bytes) and QH_o (a_o bytes), where $a = a_i + a_o$.
   - QT is subdivided into inner and outer QTs: QT_i (b_i bytes) and QT_o (b_o bytes), where $b = b_i + b_o$.

Step 3. Finally, attach a compulsory 1-byte quantum label to each unit as generated from step 2. Also, two modes of operation, namely the cut-through mode and store-and-forward mode, are available. For the cut-through mode, m quantum packets are formed, each having one 53-byte quantum. In this case, each quantum is forwarded independently. For the store-and-forward mode, a quantum packet with m 53-byte quanta is formed. These quanta must be forwarded together.

This quantumization process is similar to the well-known AAL process [2] but there are two major differences:

- First, it is more flexible because all the headers/trailers are optional. Only the 1-byte quantum label is compulsory. This means that, as shown later, specific headers/trailers can be defined for each type of quantum packet to support different packetization needs. In other words, it can better cater for the requirements of next-generation networks.
- Second, unlike AAL, which assumes a connection-oriented (circuit-switching-like) framework, it provides both cut-through and store-and-forward operations to better support not only connection-oriented but also connectionless data traffic. As shown later, it can be used to support active network traffic as well.

A quantum packet (comprising one or more 53-byte quanta) is backward compatible with ATM cells. In other words, the AAL process can be viewed as a subset or special case of the quantumization process. Returning to the parameters g and l, it is not difficult to see that once the parameters $\{h, t, a, b\}$ are defined, we have $l = 52 - a - b$ and $g = \lceil \frac{M+h+t}{l} \rceil \times l - (M + h + t)$ where M is the length of the higher-layer data unit in bytes. As a result, $m = \lceil \frac{M+h+t}{l} \rceil$ 53-byte quanta are generated.

The 1-byte quantum label is used to identify the essential information about a quantum. In the case of store-and-forward traffic, the label is also used to link the corresponding quanta of a quantum packet through the type-merging multiplexing mechanism (described later). To a certain extent, a quantum label functions like a compacted ATM cell header. As shown in Fig. 4, it contains the following fields:

1. Synchronization (SYN): The 1-bit SYN field is reserved for

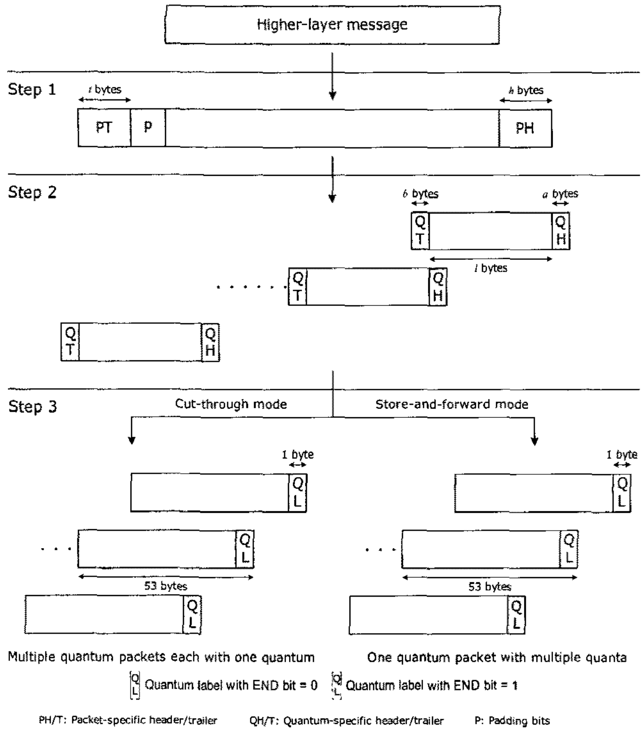


Fig. 3. Quantumization (AAL+) process.

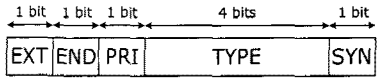


Fig. 4. Quantum label format.

synchronization purposes.

2. Type of quantum packet (TYPE): The 4-bit TYPE field identifies the quantum packet type. Table 2 shows a list of possible types. More types can be introduced if required.
3. Priority (PRI): The 1-bit PRI field indicates that the corresponding quantum should be processed at a higher priority than other quanta of the same type. Note that another priority policy can be defined for different types of quanta.
4. Packet end (END): Similar to the AUU bit in an ATM cell header, the 1-bit END field indicates whether the quantum is the last one in a quantum packet. Therefore, for store-and-forward traffic, the END field is set to zero except for the END field in the last quantum, which is set to one. The END field is always set to 1 in cut-through traffic because each quantum is transmitted independently.
5. Extension (EXT)—this bit is reserved for future uses.

In summary, in designing quantum packets, we do not assume using a particular forwarding technique (e.g., switching). The quantum label only indicates the packet type and whether there are more quanta to follow. According to the TYPE bits, the quanta are passed to the respective modules in the FEE for processing. Therefore, an ATM cell, an MPLS-based packet and an active packet are passed to the respective modules for processing. The proposed solution is very flexible and extensible because new modules and TYPEs can easily be added in the future. This design also allows each packet to be processed according to the most appropriate forwarding mechanism.

Table 2. TYPEs for different network services.

TYPE	Service	Forwarding Mode
0000 / 1111	Reserved for network control	N/A
0001	Layer 2 switching (ATM-like)	Cut-through
0010	Layer 3 routing	Store-and-forward
0011	Active services	Store-and-forward
0100	Layer 3 switching (MPLS-like)	Store-and-forward
0101 – 1110	Future uses	N/A

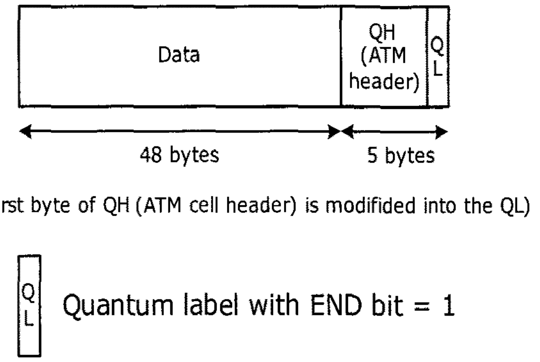


Fig. 5. Quantum packet for layer 2 (ATM) switching.

We now explain how the quantum packet method can be used to flexibly and effectively support different types of traffic. For CO (i.e., ATM-based) traffic, the above quantumization process can cover any one of the AAL processes. Essentially, the CSPDU header and trailer correspond to PH and PT, respectively. The SAR-PDU header (if any) and the last four bytes of the ATM cell header form the QH. More specifically, the SAR-PDU header is the QH_i and the last 4 bytes of the ATM cell header is the QH_o. Finally, the first byte of the ATM cell header is assumed to be modified into a quantum label as shown in Fig. 5. Alternatively, ATM cells can be turned into quantum packets (i.e., each having one quantum) by rewriting the first byte of the header as the required quantum label. The impact of this should be small because the first byte of an ATM cell at the Network/Network Interface (NNI) is only part of the VPI, which can be modified easily. In the case of CO traffic, the cut-through mode is used (i.e., each quantum is switched independently). Note that CO quantum packets are well-suited to transferring time-sensitive traffic (e.g., voice/video). In other words, they can be used to support voice/video services. In essence, like ATM cells, these CO quantum packets can provide circuit-switching-like services. The voice/video information is carried in the quantum payload and the virtual circuit identifier is written into the QH. For instance, to support voice services with silent suppression, a voice quantum is created periodically when the voice station is active (i.e., during a talkspurt). The voice quantum is transferred using the cut-through mode. Similarly, variable bit rate video traffic can also be supported. As each CO

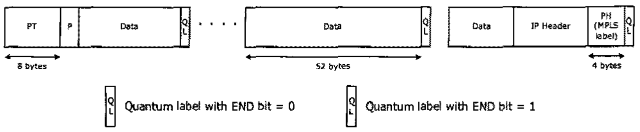


Fig. 6. Quantum packet for layer 3 (MPLS) switching.

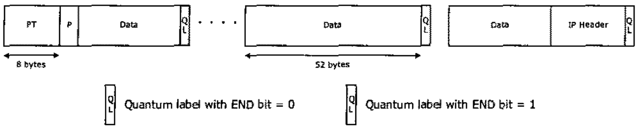
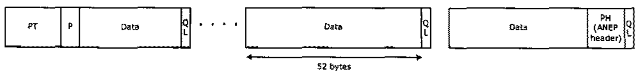
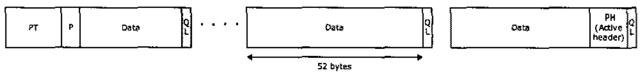


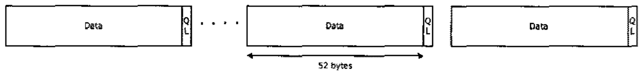
Fig. 7. Quantum packet for IP routing.



An active packet with the ANEP header



An active packet with programming code embedded into the packet-specific header



An active packet with programming code embedded into the packet itself

Fig. 8. Quantum packet for active data traffic.

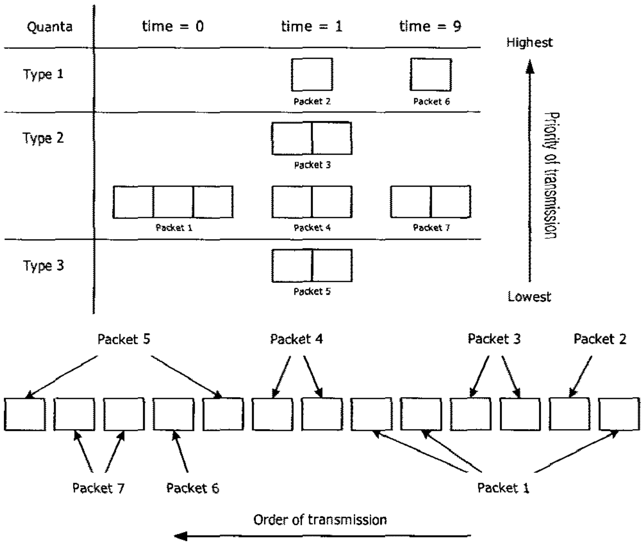


Fig. 9. Multiplexing of different types of quantum packets.

quantum can be processed independently or on a cut-through basis, voice/video traffic can be transferred very efficiently. Moreover, the priority mechanism used by quantum packets enables CO quanta to be transferred at a high priority, thus minimizing the effect of other types of quanta on its transfer delay (see the performance analysis in Section V).

For the support of MPLS-based IP traffic, an MPLS label can be regarded as a PH. Furthermore, a PT can be attached for error checking and other control functions. No QH and QT are needed. Essentially, the quanta of a quantumized MPLS packet are transferred based on the MPLS label in the first quantum (see Fig. 6). In this case, a FEE functions like an MPLS switch in the horizontal partitioning network model. Note that a quantumized MPLS packet can be transferred directly through the physical layer. Similarly, a quantumized IP datagram can be formed as shown in Fig. 7. Again QH and QT are not required. A quantumized IP packet can also be transferred directly through the physical layer based on the IP address in the first quantum.

For active data packets, a number of packetization options are available as shown in Fig. 8. In the first option, the ANEP header as described in [24] can be attached as a PH header together with a predefined PT. Again, no QH and QT is needed. The active data packet is processed according to the information in the ANEP header. In the second option, the ANEP header can be replaced with an active header carrying a program. Finally, an active application can create a whole active packet. We assume that an integral number of quanta is generated and the quantum labels are embedded inside the packet as comments (i.e., they can be ignored when the packet is executed).

Quantumization is thus very flexible and extensible. It is designed to support different types of traffic using a unified packetization framework. Note that the payload size of CL packets can be maintained at 52 bytes (except the first and last quanta). Compared to 48/53 = 91% as in ATM, we can see that the proposed quantum packet method can improve the payload utilization for CL traffic to 52/53 = 98%.

Motivated by VC-merging, we propose a "type-merging" method for multiplexing different types of quantum packets according to the TYPE and PRI fields in the quantum label. In general, different types of quantum packets can be multiplexed together over the same link provided that the quanta of a quantum packet are not interleaved with the quanta of another packet of the same type and of the same priority level. This requirement also applies to a FEE. Fig. 9 presents an example to illustrate how the multiplexing rule works. Note that type 1 traffic is assumed to have the highest priority so the first quantum of packet 2 is transmitted at a higher priority than the second quantum of packet 1. As packet 3 has a higher priority than packet 1 (i.e., the PRI bit is set), it is also processed first. Support for quantum packet switching/processing should lead to many new and interesting research issues. For example, due to the similarity between type-merging and VC-merging, a VC-merging switch such as the one proposed in [16] can generally be used to support quantum packet switching.

IV. ANALYTICAL MODEL

In this section, we present an analytical model to evaluate the consumption of network resources when quantum packets are used to transport loss-sensitive data from a sender to a receiver through network nodes using three different schemes: Cut-through, store-and-forward, and ideal. The analysis gives us valuable insights into the behavior of the quantum network in general.

The analytical model is described as follows. A sender and a receiver are connected by n network nodes numbered 1 to n as shown in Fig. 10. A loss-sensitive quantum packet with m

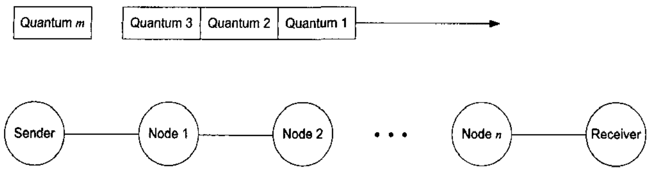


Fig. 10. Network resource model.

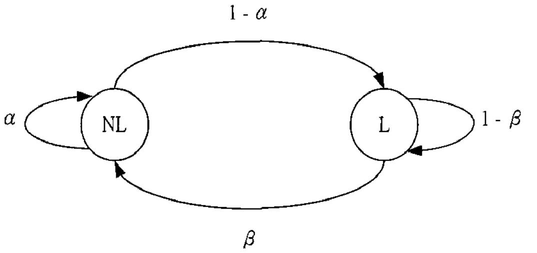


Fig. 11. Burst loss model.

quanta is to be transmitted from the sender to the receiver. In the meantime, we assume that m is the same for all cases in order to derive some interesting relationships. Later we will express m in terms of the length of the higher-layer data unit. At each node, a quantum is transmitted successfully with probability p (i.e., discarded with probability $q = 1-p$ at the outgoing interface, e.g., due to buffer overflow). For each successful hop-by-hop quantum transmission, one unit of network resources is consumed. In general, quanta may be lost in a burst due to network congestion. For example, if the previous quantum is dropped, the current quantum is more likely to be discarded too. We consider a burst loss model to capture this behavior as shown in Fig. 11. There are two states, "not lost" (NL) and "lost" (L), indicating whether the previous quantum is not lost or lost, respectively. The initial state is set at NL. If a quantum is successfully transmitted, the state will be NL and the probability that the next quantum can be successfully transmitted too is α. Otherwise, the state will become L and the probability of successfully transmitting the next quantum is β. This means that at the NL and L states, $p = \alpha$ and $p = \beta$, respectively. Denote by $P(\mathrm{L})$ and $P(\mathrm{NL})$ the probabilities that the system state is L and NL, respectively. We can write

$$P(\mathrm{NL}) = \alpha P(\mathrm{NL}) + \beta P(\mathrm{L}) \tag{1}$$

$$P(\mathrm{NL}) + P(\mathrm{L}) = 1. \tag{2}$$

The objective of the following analysis is to evaluate the average cost C of successfully transmitting a packet from a sender to a receiver in three different cases. Note that the transmission of the loss-sensitive packet is considered successful only if all its m quanta reach the receiver. If not, the whole packet or part of its quanta needs to be retransmitted.

Case 1 (cut-through): In this case, each quantum is sent independently. Hence if a transmission of the packet is unsuccessful (i.e., one or more quanta are lost at the intermediate nodes), the sender must retransmit all of the m quanta. This resembles the AAL5 approach. Note that in this case, it is assumed that lost quanta cannot be identified because no quantum sequence number is included.

Case 2 (store-and-forward): In this case, each node will not forward any quanta to the next node until all the m quanta are received. If a node cannot receive any of the quanta, all m quanta will be retransmitted from the sender. Also, an incomplete packet will not be forwarded to the next node but will be discarded.

Case 3 (ideal): This case is similar to Cases 1 and 2 except that the sender retransmits only the discarded quanta. Note that Case 3 can be considered ideal, but it is generally impractical as it requires additional processing (e.g., a quantum sequence number is required to identify lost quanta). Here the quantum sequence number is assumed to be embedded in the QH.

When $\alpha = \beta = p$, each quantum can be regarded as having been independently discarded. In this situation (i.e., independent loss), we can formulate a close form solution for each Case (1, 2, and 3). As shown later, the close form expressions are validated by simulation results. Furthermore, we will present simulation results for the general burst loss model.

Finding 1: *To transfer a loss-sensitive quantum packet with* m *quanta successfully through* n *nodes by using the cut-through approach, the average cost is* $\frac{mp(1-p^n)}{p^{nm}(1-p)}$ *under the independent loss model (i.e., each quantum is discarded independently with probability* $1-p$ *at each node).*

For each trial, the probability of transmitting the packet successfully from the sender to the receiver (i.e., no discarded quantum) is p^{nm}. The associated cost is nm. If one or more quanta are discarded at the intermediate nodes, the transmission is regarded as having failed, which occurs with probability $1-p^{nm}$. In this case, the associated average cost is denoted as C_{fail}. Suppose that the packet requires x trials before successfully reaching the receiver. Obviously x follows a geometric distribution given by $(1-p^{nm})^{x-1}p^{nm}$, $x = 1, 2, \cdots$. The average cost of x trials is $(x-1)\,C_{\text{fail}} + nm$. The average cost C is the expected value of $(x-1)\,C_{\text{fail}} + nm$, which can easily be found to be

$$\begin{aligned} C &= E\left[(x-1)\,C_{\text{fail}} + nm\right] \\ &= \left(\frac{1}{p^{nm}} - 1\right) C_{\text{fail}} + nm. \end{aligned} \tag{3}$$

A "failed" packet transmission means that one or more quanta are discarded. Let y be the number of lost quanta in a failed transmission and C_{lost} be the average cost due to a lost quantum. We have

$$\begin{aligned} &P\left(y \text{ out of } m \text{ quanta lost} \,|\, \text{failed transmission}\right) \\ &= \frac{P\left(y \text{ out of } m \text{ quanta lost, failed transmission}\right)}{P\left(\text{failed transmission}\right)} \\ &= \frac{P\left(y \text{ out of } m \text{ quanta lost}\right)}{P\left(\text{failed transmission}\right)} \\ &= \frac{\binom{m}{y}(p^n)^{m-y}(1-p^n)^y}{1-p^{nm}}. \end{aligned} \tag{4}$$

If y quanta are lost, the associated average cost is $yC_{\text{lost}} + (m-y)n$. Since a failed transmission discards at least one quantum and at most m quanta, we have

$$\begin{aligned} C_{\text{fail}} &= \sum_{y=1}^{m} \frac{\binom{m}{y}(p^n)^{m-y}(1-p^n)^y}{1-p^{nm}}\left[yC_{\text{lost}}+(m-y)n\right] \\ &= \frac{m(1-p^n)}{1-p^{nm}}(C_{\text{lost}}-n)+nm. \end{aligned} \tag{5}$$

The detailed derivation of (5) is given in Appendix 1.

A quantum may be discarded at any of the intermediate nodes. Given that a quantum is discarded, the probability that it will be discarded at the zth node is $\frac{p^{z-1}(1-p)}{1-p^n}$ and the associated cost is $z-1$. Note that if a quantum is discarded at the zth node, the quantum must have been successfully transmitted through the preceding nodes before being discarded at node z. Using the above argument, C_{lost} can be found by summing all possible values of $z=1,2,\cdots,n$ as follows:

$$\begin{aligned} C_{\text{lost}} &= \sum_{z=1}^{n}\left[\frac{p^{z-1}(1-p)}{1-p^n}\right](z-1) \\ &= \frac{(n-1)p^{n+1}-np^n+p}{(1-p)(1-p^n)}. \end{aligned} \tag{6}$$

The detailed derivation of (6) is given in Appendix 2.

Having found C_{lost} and C_{fail}, the average cost can then be calculated by using (3):

$$\begin{aligned} C=&\left(\frac{1}{p^{nm}}-1\right)\frac{m(1-p^n)}{1-p^{nm}}\left[\frac{(n-1)p^{n+1}-np^n+p}{(1-p)(1-p^n)}-n\right] \\ &+\left(\frac{1}{p^{nm}}-1\right)nm+nm \\ =&\frac{mp(1-p^n)}{p^{nm}(1-p)}. \end{aligned} \tag{7}$$

The detailed derivation of (7) is given in Appendix 3.

Finding 2: *Under the same assumptions as stated in Finding 1, the average cost of the store-and-forward approach is* $\frac{mp(1-p^{nm})}{p^{nm}(1-p^m)}$.

The average cost is still given by (3) but C_{fail} takes on a different value. In this case, all m quanta are transmitted together. Given that the packet transmission fails (i.e., one or more of the quanta is/are discarded), the probability that it fails at node y is $\frac{p^{m(y-1)}(1-p^m)}{1-p^{nm}}$. Note that the packet must be successfully transmitted from node 1 to node $(y-1)$ and then discarded at node y. Let $C(y)$ be the associated average cost, we have

$$C_{\text{fail}}=\sum_{y=1}^{n}\frac{p^{m(y-1)}(1-p^m)}{1-p^{nm}}C(y). \tag{8}$$

If the transmission fails at the yth node, at least one and at most m quanta are discarded. Let z be the number of discarded quanta and hence $m-z$ quanta can be successfully transmitted. Each of the transmitted and discarded quanta costs respectively y and $y-1$ units of resources. Thus, the resultant cost is $(m-z)y+z(y-1)=my-z$. As any z of the m quanta may be discarded at node y, we have

$$\begin{aligned} C(y) &= \sum_{z=1}^{m}\frac{\binom{m}{z}p^{m-z}(1-p)^z}{1-p^m}(my-z) \\ &= my-\frac{m(1-p)}{1-p^m} \end{aligned} \tag{9}$$

where $\binom{m}{z}p^{m-z}(1-p)^z/(1-p^m)$ is the probability that z quanta are discarded given that at least one quantum is lost. The detailed derivation of (9) is given in Appendix 4.

Combining (8) and (9), we have

$$\begin{aligned} C_{\text{fail}} &= \sum_{y=1}^{n}\frac{p^{m(y-1)}(1-p^m)}{1-p^{nm}}\left[my-\frac{m(1-p)}{1-p^m}\right] \\ &= \frac{m\left[np^{(n+1)m}-p^{nm+1}-np^{nm}+p\right]}{(1-p^{nm})(1-p^m)}. \end{aligned} \tag{10}$$

The detailed derivation of (10) is given in Appendix 5.

The average cost C can then be calculated by using (3) to obtain the following expression:

$$\begin{aligned} C =& \left(\frac{1}{p^{nm}}-1\right)\left\{\frac{m\left[np^{(n+1)m}-p^{nm+1}-np^{nm}+p\right]}{(1-p^{nm})(1-p^m)}\right\} \\ &+nm \\ =& \frac{m\left[np^{(n+1)m}-p^{nm+1}-np^{nm}+p\right]}{p^{nm}(1-p^m)}+nm \\ =& \frac{mp(1-p^{nm})}{p^{nm}(1-p^m)}. \end{aligned} \tag{11}$$

Finding 3: *Under the same assumptions as stated in Finding 1, the average cost of the ideal case is* $\frac{mp(1-p^n)}{p^n(1-p)}$.

In the ideal case, as each quantum is transmitted independently, we can treat each quantum as a separate packet. The average cost of transmitting a quantum can be found by setting $m=1$ in (3). Having found the average cost of transmitting one quantum, we can calculate the average cost of transmitting m quanta by multiplying by m:

$$C=m\left[\left(\frac{1}{p^n}-1\right)C_{\text{lost}}+n\right]. \tag{12}$$

Combining expressions (6) and (12), we obtain the expression for the average cost C as follows:

$$\begin{aligned} C &= m\left[\left(\frac{1}{p^n}-1\right)\frac{(n-1)p^{n+1}-np^n+p}{(1-p)(1-p^n)}+n\right] \\ &= m\left[\frac{(n-1)p^{n+1}-np^n+p}{p^n(1-p)}+n\right] \\ &= \frac{mp(1-p^n)}{p^n(1-p)}. \end{aligned} \tag{13}$$

Finding 4 (Two equivalence phenomena): *If the loss probability is small such that the second and higher order terms can be ignored, both the cut-through and store-and-forward approaches produce the same average cost of* $\frac{(1-q)nm}{(1-nmq)}$, *recalling that* $q=1-p$.

Substituting $p = 1 - q$ into (7) and (11) and ignoring the second and higher order terms, we obtain respectively for the cut-through approach

$$C = \frac{m(1-q)\left[1-(1-q)^n\right]}{(1-q)^{nm}q} \approx \frac{m(1-q)\left[1-(1-nq)\right]}{(1-nmq)\,q} = \frac{(1-q)nm}{(1-nmq)} \tag{14}$$

and for the store-and-forward approach

$$C = \frac{m\,(1-q)\left[1-(1-q)^{nm}\right]}{(1-q)^{nm}\left[1-(1-q)^m\right]} \approx \frac{m\,(1-q)\left[1-(1-nmq)\right]}{(1-nmq)\left[1-(1-mq)\right]} = \frac{(1-q)nm}{(1-nmq)}. \tag{15}$$

That is, the average costs are the same (under the assumption of ignoring the higher order terms) and depend on $n \times m$, the minimum possible cost of transferring the quantum packet through a lossless network. To summarize, if the loss probability is very small (which is likely to be valid in future networks) and the independent loss model is assumed, we have two interesting equivalence phenomena in terms of the network cost:

- Equivalence Phenomenon I: Transporting loss-sensitive quantum packets by means of the cut-through and store-and-forward approaches produces the same average cost[2] (provided that second and higher order terms are ignored). This indicates that although the cut-through and store-and-forward approaches are based on two different forwarding principles, they are equivalent from the point of view of consuming network resources under certain situations.
- Equivalence Phenomenon II: For both the cut-through and store-and-forward approaches, transporting m quanta through n nodes is equivalent to transporting n quanta through m nodes because the average cost depends on $n \times m$ (e.g., $3 \times 5 = 5 \times 3$). This indicates that transferring a packet to a more distant destination is as costly as sending a longer packet.

For completeness, we have also computed the first order approximation for the ideal case and all the second order approximations as shown in Table 3 in comparison with the exact expressions. The first order approximations for different schemes are included in Figs. 13 and 14 as well.

In the above calculations, we assume that m is the same for all cases. In practice, m depends on the length of the higher layer message M and the packetization parameters (i.e., $m = \left\lceil \frac{M+h+t}{52-a-b} \right\rceil$ as discussed previously). In the following analysis, we assume that the parameters for each case are as follows:

Case 1: $t = 8$, $a = 4$, $h = b = 0$, and hence $m = \left\lceil \frac{M+8}{48} \right\rceil$ (i.e., similar to AAL5).

Case 2: $t = 8$, $h = a = b = 0$, and hence $m = \left\lceil \frac{M+8}{52} \right\rceil$.

Case 3: $h = t = 4$, $a = 6$, $b = 2$, and hence $m = \left\lceil \frac{M+8}{44} \right\rceil$ (i.e., similar to AAL3/4).

Fig. 12 illustrates the value of m for different cases when M is varied. The staircase curve is due to the ceiling operation. We can see that when the store-and-forward approach is used, the number of resultant quanta is lowered significantly compared with the cut-through forwarding method. Moreover, due to the additional overhead, m for Case 3 is the largest.

Table 3. Values of C and m for different cases.

Case	C			m
	Exact expression	1st order approximation	2nd order approximation	
1	$\frac{mp(1-p^n)}{p^{nm}(1-p)}$	$\frac{(1-q)nm}{(1-nmq)}$	$\frac{(1-q)\left[1-\frac{(n-1)q}{2}\right]nm}{\left[1-nmq+\frac{nm(nm-1)q^2}{2}\right]}$	$\left\lceil \frac{M+8}{48} \right\rceil$
2	$\frac{mp(1-p^{nm})}{p^{nm}(1-p^m)}$	$\frac{(1-q)nm}{(1-nmq)}$	$\frac{(1-q)\left[1-\frac{(nm-1)q}{2}\right]nm}{\left[1-nmq+\frac{nm(nm-1)q^2}{2}\right]\left[1-\frac{(m-1)q}{2}\right]}$	$\left\lceil \frac{M+8}{52} \right\rceil$
3	$\frac{mp(1-p^n)}{p^n(1-p)}$	$\frac{(1-q)nm}{(1-nq)}$	$\frac{(1-q)\left[1-\frac{(n-1)q}{2}\right]nm}{\left[1-nq+\frac{n(n-1)q^2}{2}\right]}$	$\left\lceil \frac{M+8}{44} \right\rceil$

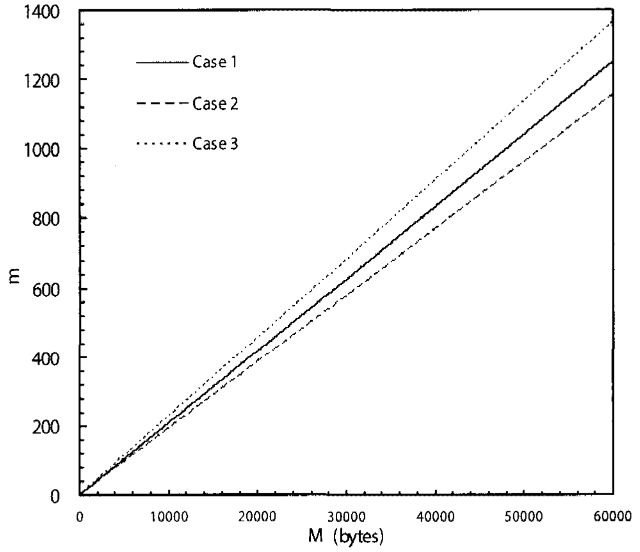


Fig. 12. Number of corresponding quanta (m) vs. message size (M).

V. RESULTS AND DISCUSSION

In this section, we first apply simulation results to verify the correctness of our analytical model by setting α and β to be p. Therefore, P(NL) and P(L) equal p and $1 - p$, respectively. In this situation, the probability of transmitting each quantum is the same regardless of the system state (i.e., independent loss). Later, we will also simulate the general burst loss model and present simulation results to evaluate the quantum packet method for transporting CO and CL traffic.

To validate the two phenomena, we have performed some simulations using the same m for the three cases. Fig. 13 validates the Equivalence Phenomenon I. As can be seen, when p is large, the average costs of the cut-through and store-and-forward

---

[2] Note that the two average costs are not exactly the same but it is found that the difference occurs in the higher order terms.

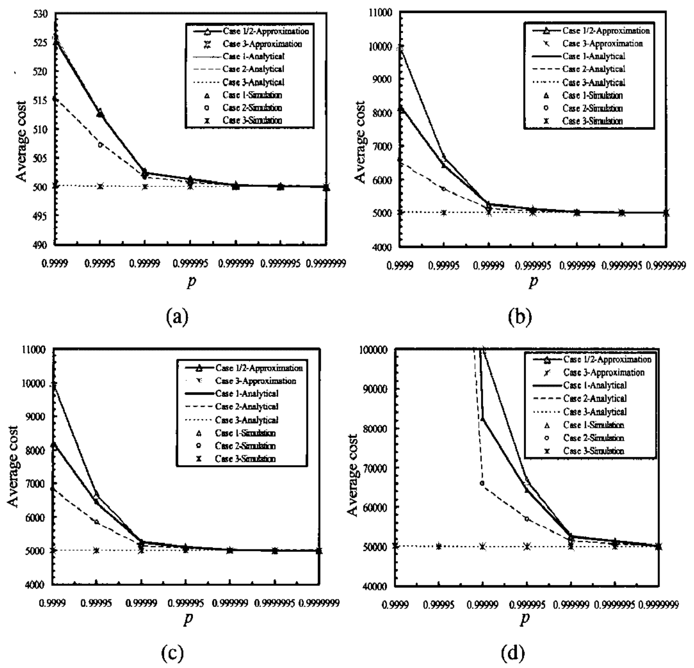


Fig. 13. Average cost vs. p for different n and m: (a) $n = 5$, $m = 100$, (b) $n = 50$, $m = 100$, (c) $n = 5$, $m = 1{,}000$, and (d) $n = 50$, $m = 1{,}000$.

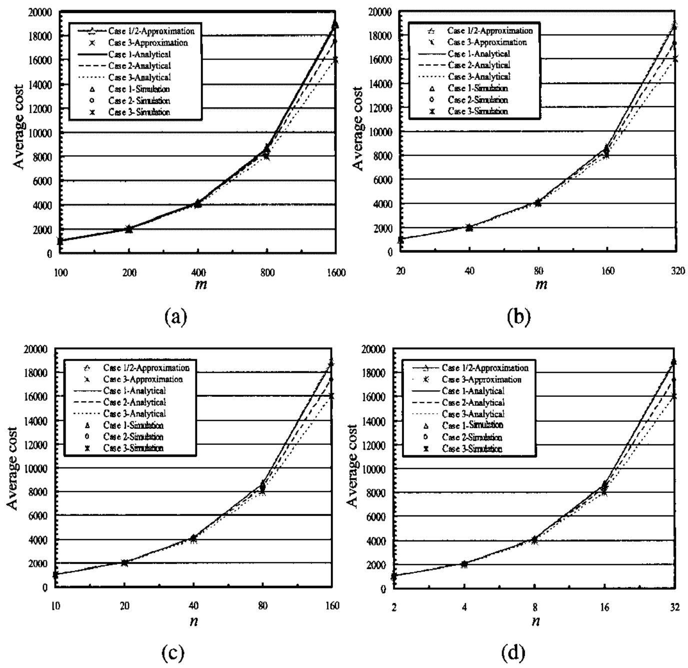


Fig. 14. Average cost when p is large: (a) $p = 0.99999$, $n = 10$, (b) $p = 0.99999$, $n = 50$, (c) $p = 0.99999$, $m = 100$, and (d) $p = 0.99999$, $m = 500$.

approaches are almost the same. In fact, under this situation, the average costs are very close to the ideal average costs (i.e., those of Case 3). Fig. 14 validates the Equivalence Phenomenon II. In the four sub-plots of Fig. 14, $n \times m$ are set to be the same and q is set to a low value (i.e., p is high). It can be seen that the resultant average costs are almost identical, which agrees with (14) and (15).

Next, we assume that a message of M bytes is encapsulated based on the aforementioned packetization parameters (i.e., m is different in each of the cases). To facilitate the comparison, we set the base parameters as M = 400, n = 10, and p = 0.999. In the following analysis, we compare the average cost of the three cases by varying each of the parameters in turn.

As shown in Fig. 15, when M is small (below 400 bytes), the average costs of Cases 1 and 3 are about the same. The average cost of Case 2 is the lowest due to the larger payload size (fewer quanta are required) as compared with the other cases. As M increases, the average cost of each case increases at a different rate. For Cases 1 and 2, the larger the value of n is, the more steeply the average cost increases. Specifically, the average cost of Case 1 increases more dramatically than that of Case 2. As indicated by expression (13), the average cost of Case 3 increases approximately linearly as M increases when n and p are constant. It can be seen that the store-and-forward approach can produce a much lower average cost in some situations. For example, when n = 40 and M = 12,800, the average cost of Case 2 is only one-twentieth of that of Case 1.

Fig. 16 shows the effect of varying n. As explained earlier, Case 2 is the best performer when M is small. For larger M, the average cost of Case 1 is the most sensitive to the increase of n. The average cost of Case 2 increases less dramatically. The average cost of Case 3 is the least sensitive to the increase of n.

Fig. 17 shows the effect of varying p and M while keeping n constant. It can be seen that when p is large, the average costs of all three cases are similar, particularly when n and M are both small. The effect of the small payload of Case 3 becomes apparent when p is large, making the average cost the highest among all the cases. On the other hand, the average cost of Case 2 is the lowest owing to the largest payload size. When p decreases by an order of magnitude, the situation is quite different. For Cases 1 and 2, the smaller the value of p, the more steeply the average cost increases as M increases. However, the average cost of Case 3 is insensitive to a change in p.

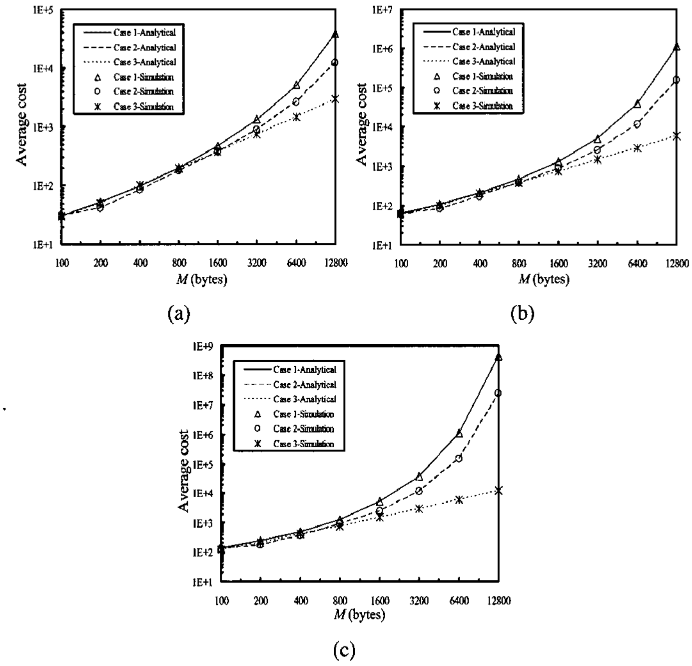


Fig. 15. Average cost vs. M for different n: (a) $p = 0.999$, $n = 10$, (b) $p = 0.999$, $n = 20$, and (c) $p = 0.999$, $n = 40$.

Fig. 18 shows the effect of varying p and n while keeping M constant. Again, the average costs of Cases 2 and 3 are re-

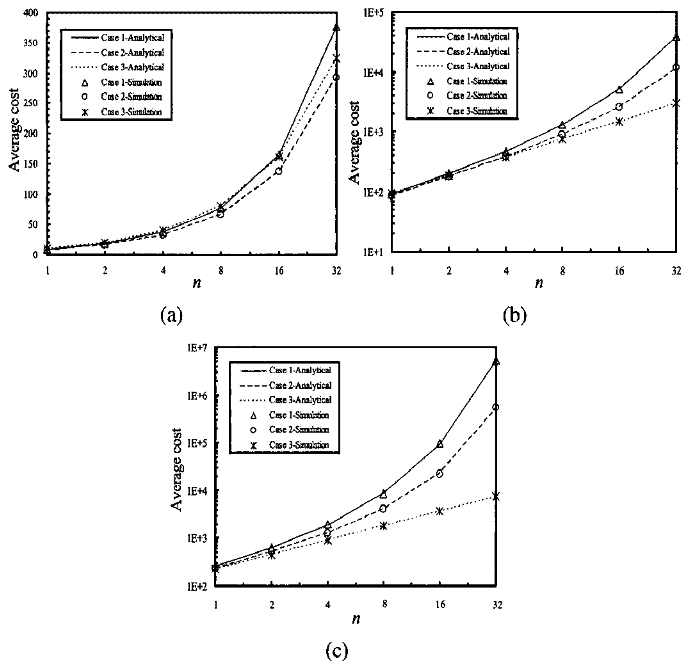


Fig. 16. Average cost vs. n for different M: (a) $p = 0.999$, $M = 400$, (b) $p = 0.999$, $M = 4,000$, and (c) $p = 0.999$, $M = 10,000$.

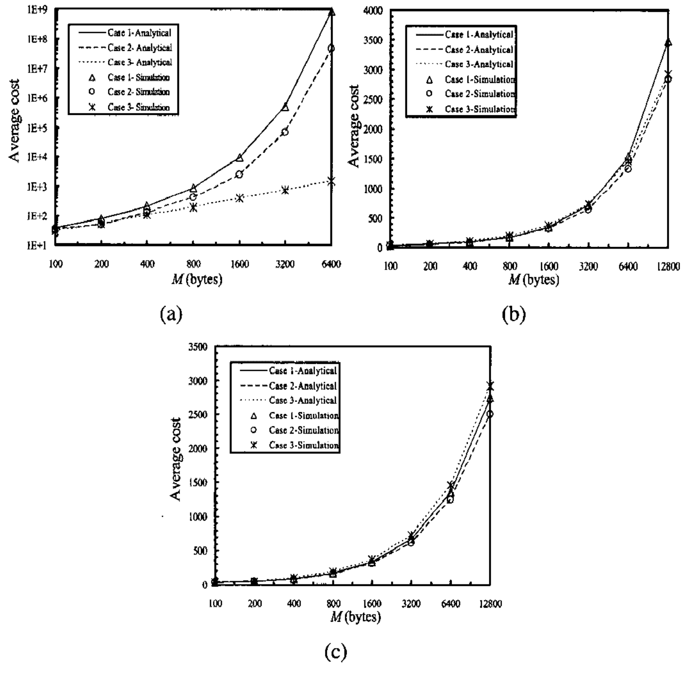


Fig. 17. Average cost vs. M for different p: (a) $p = 0.99$, $n = 10$, (b) $p = 0.9999$, $n = 10$, and (c) $p = 0.99999$, $n = 10$.

spectively the lowest and the highest when p is large. When p is substantially lowered, the average cost of Case 1 increases significantly, particularly when n is large. However, the average costs of Cases 2 and 3 are less affected.

From the above results, we can see that the cost consumption of the store-and-forward approach is even lower than that of the ideal case (i.e., Case 3) when n and M are both small and when p is large. This is due to the larger payload size, which means that fewer quanta are produced in the former approach. It is interesting to study which combination of these three parameters leads to Case 2 performing better than Case 3. To do this, we need to solve the following inequality:

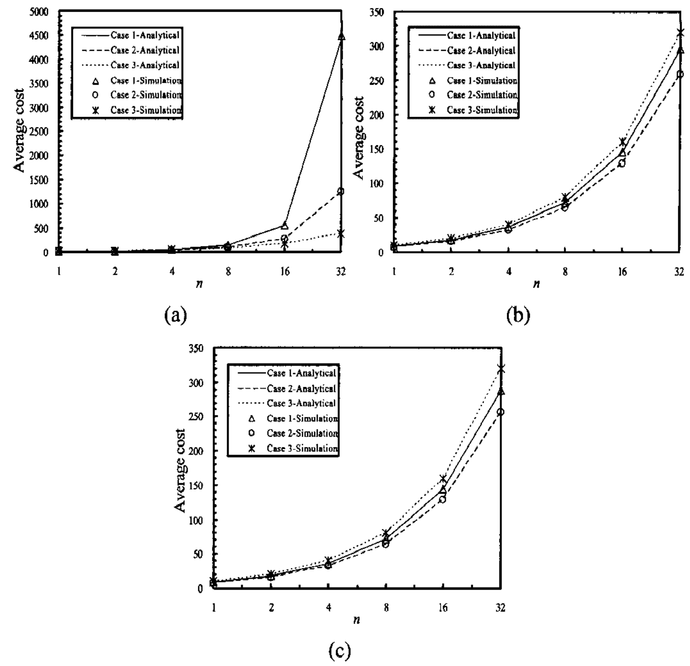


Fig. 18. Average cost vs. n for different p: (a) $p = 0.99$, $M = 400$, (b) $p = 0.9999$, $M = 400$, and (c) $p = 0.99999$, $M = 400$.

$$\frac{(1-q)nm_1}{(1-nm_1q)} < \frac{(1-q)\,nm_2}{(1-nq)} \tag{16}$$

where $m_1 = \frac{M+8}{52}$ and $m_2 = \frac{M+8}{44}$.

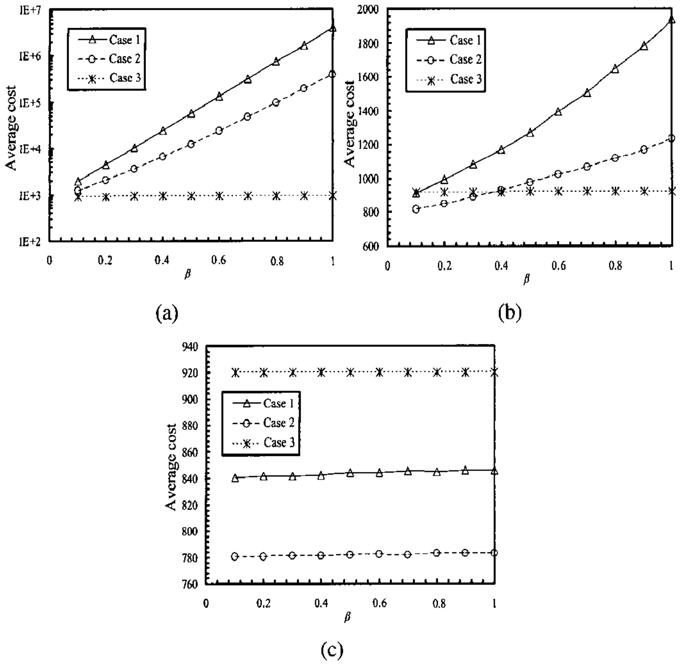
Note that for simplicity we use the first order approximation equations (see Table 3) to calculate C for Cases 2 and 3. As an approximation, we skip the ceiling operation for the calculation of m_1 and m_2. Therefore, we have

$$\frac{m_1}{(1-nm_1q)} < \frac{m_2}{(1-nq)} \Rightarrow nq(M-36) < 8. \tag{17}$$

When (17) is satisfied, the store-and-forward approach performs better than the ideal approaches.

After verifying the simulation model against the analytical model, we can further apply the simulation model to study the effect when burst loss is considered. As mentioned earlier, β is the probability that the current quantum can be successfully transmitted when the system state is L. Therefore, the smaller the value of β, the higher the chance that quanta are lost in a burst. We set $n = 10$ and $M = 4,000$ and vary β. The corresponding α is calculated using expressions (1) and (2) for a particular β and P(NL).

Fig. 19 shows the average cost when we vary β and P(NL). The average cost of Case 3 is constant irrespective of β. Since only the lost quanta are retransmitted, the average cost in Case 3 depends only on P(L) and not on the burstiness. When P(NL) is small, the average cost of Cases 1 and 2 increases sharply as β increases. Note that for a quantum packet, losing one quantum is the same as losing all m quanta since in either case it must be retransmitted. A smaller β means that the chance of losing consecutive quanta of the same packet is higher. However, to keep P(NL) the same, α is larger. This means fewer packets lose quanta. Hence, retransmission is less frequent if β



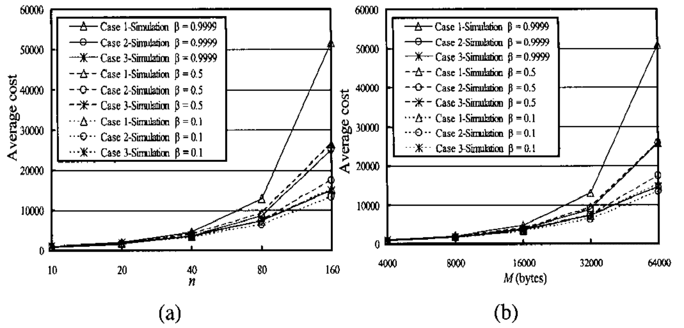
(a) (b)

(c)

Fig. 19. Average cost vs. β for different P(NL): (a) $P(\mathrm{NL}) = 0.99$, (b) $P(\mathrm{NL}) = 0.999$, and (c) $P(\mathrm{NL}) = 0.99999$.

is smaller. On the other hand, when β increases (i.e., the burstiness of loss decreases), quantum dropping spreads across different packets. Therefore, more packets lose quanta and retransmission becomes more frequent. This explains why the average costs of Cases 1 and 2 are larger as β increases. When P(NL) increases, the average costs of the two cases in general decrease and increase less sharply as β increases. When P(NL) is very large, the average cost is insensitive to a change in β. Again, the average cost of Case 3 is the highest when P(NL) is very large. It can be seen that the average costs of both of the cut-through and store-and-forward approaches are smaller in burst-loss situations than in independent-loss situations. This is because, in the burst-loss situation, quanta tend to be lost in a burst so fewer packets are affected. Note that whether a packet loses one or more quanta, all the other quanta in the same packet must be retransmitted. In other words, independent loss in fact produces a worse result for the same loss ratio. Hence, it is of interest to study the independent loss model. As shown in Fig. 20, it is found that doubling n produces a similar effect as doubling M when P(NL) is large. In other words, in terms of network resource consumption and according to the above model, transmitting a longer packet is similar to transmitting a packet across a larger network. Again this validates the second equivalence phenomenon.

Although the focus of this paper is on analyzing the consumption of network resources for loss-sensitive (CL) traffic, we have used simulations to analyze the delay and packet loss of CO and CL traffic. The simulation model is outlined as follows. We assume that the network has 10 nodes (i.e., 11 links) and the speed of each link is 150 Mbps. There is a buffer in each network node. The buffer can hold 200 quanta and it is shared by CO and CL traffic. A sender sends either CO (e.g., voice/video) or CL (e.g., data or IP) traffic, which is packetized based on the quantum packet method. We consider the worst case situation in which all packets are transmitted to the last network node. For the following analysis, we evaluate the performance of one of the stations at the first node. The CO traffic comprises quantum packets each with one quantum, similar to ATM cells. The source model for CO traffic is based on the widely used bi-state Markov model, which has mean active and idle durations of 1 and 1.35 seconds, respectively [25]. In each active period, quanta are generated at 64 kbps. Note that the bi-state Markov model can also be employed to model video traffic (i.e., using mini-sources) (see [26] for details). Hence, it is a general CO traffic model. The CL traffic is transferred by quantum packets with multiple quanta. As an example, it is assumed that the IP packets for the CL traffic have 1024 bytes. The IP packets are generated at an average data rate of 200 kbps. The inter-packet arrival times are given by an exponential distribution. A simple TCP protocol is employed to retransmit lost packets [27]. Basically, if a sender cannot receive an acknowledgement within a time-out period, the packet will be retransmitted. The time-out period is doubled following each retransmission failure. The CO traffic is transferred using the cut-through mode. Due to the time-sensitive nature of the CO traffic, it is transferred at a higher priority than the CL traffic. The CL traffic is transferred using either the cut-through (i.e., Case 1) or the store-and-forward (i.e., Case 2) approach.



(a) (b)

Fig. 20. Comparison of average cost when n and M are varied: (a) $P(\mathrm{NL}) = 0.9999,\ M = 4,000$ and (b) $P(\mathrm{NL}) = 0.9999,\ n = 10$.

Fig. 21 shows the average packet transfer delay of the CL traffic when the number of CL and CO stations at each node is varied. Note that the total number of each type of stations should be tenfold. From Fig. 21, as expected, the delay for the cut-through approach is lower when there are a small number of stations. However, in general, the difference between the cut-through and store-and-forward approaches is not large. When there are more stations, the delay for the cut-through approach increases dramatically because it generates much unnecessary traffic. Recall that in the cut-through approach, if a quantum is discarded, the network still needs to transfer other quanta of the same packet. This generates unnecessary traffic in the network, thus increasing the overall average transfer delay. On the other hand, the store-and-forward is relatively less affected because incomplete packets are discarded in order to minimize unnecessary traffic.

Fig. 22 shows the packet loss ratio for the CL traffic for the two approaches. It can be seen clearly that the cut-through approach gives a higher packet loss ratio because much unnecessary traffic is generated thus making the buffers more congested. In other words, while quanta can be transferred faster,

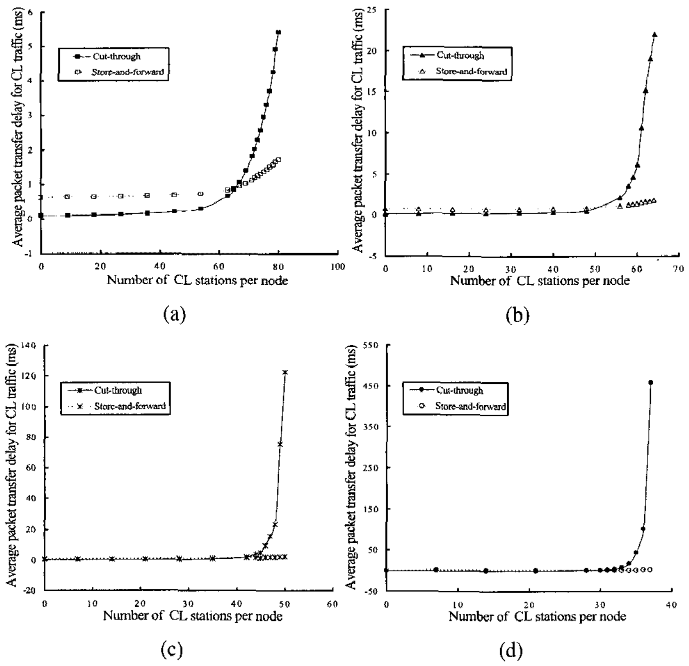


Fig. 21. Average packet transfer delay for CL traffic: (a) 0 CO stations per node, (b) 100 CO stations per node, (c) 200 CO stations per node, and (d) 300 CO stations per node.

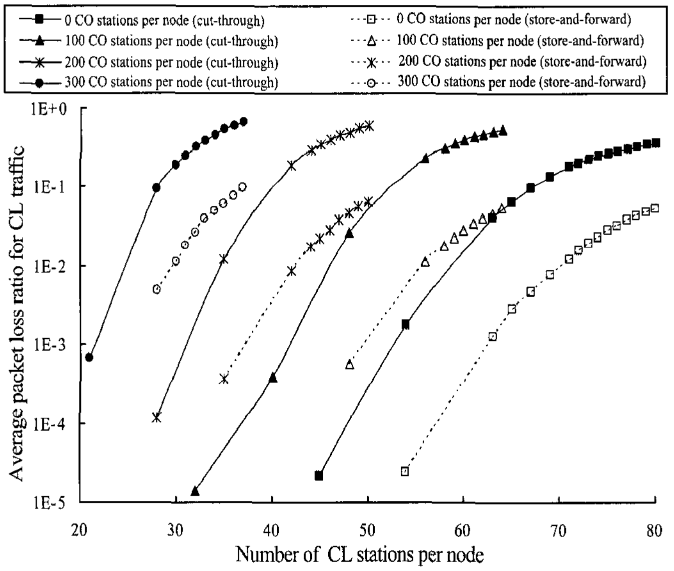


Fig. 22. Average packet loss ratio for CL traffic.

the network nodes need to handle more quanta, particularly when there are more stations.

Fig. 23 confirms that the average transfer delay for CO traffic remains almost constant because it can be transferred at a higher priority. This demonstrates the desirable result that the spare capacity can be used to transfer loss-sensitive CL traffic (e.g., IP packets) without affecting the performance of the time-sensitive CO (e.g., voice and video) traffic.

In summary, this analysis compares the cut-through and store-and-forward approaches. The results reflect the traffic load and buffering cost at the network nodes. Specifically, if the network cost is high, it means that the network nodes handle more packets (i.e., higher traffic load). Furthermore, if the delay is high, it means that the network nodes buffer the packets for a longer time (i.e., larger buffering cost). Obviously, the cut-through approach is well-suited for transporting time-sensitive traffic because quanta can be transferred as quickly as possible and it is not necessary to retransmit lost quanta. In the case of loss-sensitive (CL) traffic, when there are a small number of stations, the cut-through approach can produce a lower transfer delay but the difference is generally not very significant as compared to the store-and-forward approach. However, when there are more stations, the average transfer delay for the cut-through approach increases dramatically because much unnecessary traffic is generated. In other words, although quanta are transferred faster and the buffering cost is lower, more quanta must be processed. In general, the store-and-forward approach is preferred for CL traffic and in some situations can produce significant network cost savings. In particular, when n and M are small or when p or $P(\text{NL})$ is large, the average cost of the store-and-forward approach is the lowest of the three cases. The store-and-forward approach is also generally better where quanta are lost in a burst. When the network loss is very small, however, both the cut-through and store-and-forward approaches have a similar network cost. While the store-and-forward approach results in a slightly higher transfer delay when there are a small number of stations, it can produce a significantly lower transfer delay when there are more stations. This is because incomplete packets can be discarded rather than transferred. In other words, although the store-and-forward approach requires more buffer spaces to hold the packets, the traffic load to be handled can actually be lower. With advances in router technologies, it should now be cost-effective to build high-speed store-and-forward routers, including those with VC-merging/type-merging functions (see for example [16] and [28] and other papers in that special issue). Furthermore, this analysis demonstrates the need to develop a unified packetization method (i.e., the proposed quantum packet mechanism) for CO (cut-through) traffic and CL (store-and-forward) traffic.

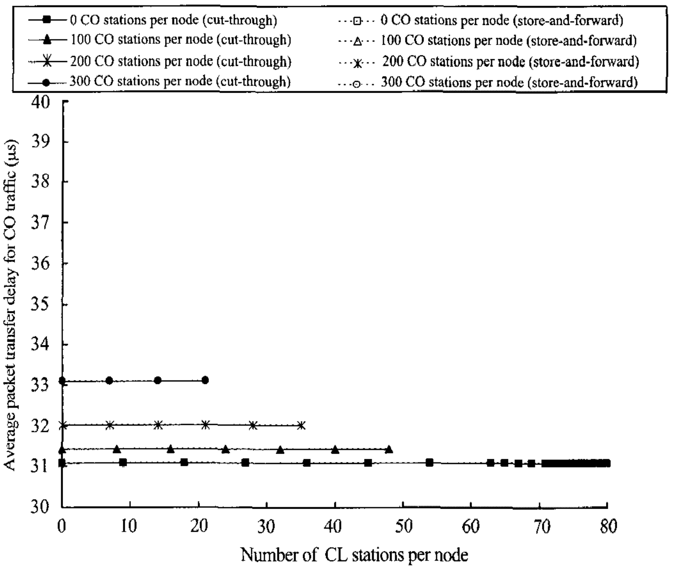


Fig. 23. Average packet transfer delay for CO traffic.

VI. CONCLUSION

We have proposed a novel quantum packet method for supporting various types of traffic over the next-generation network called ISDN3. The method provides frame/datagram-like services while enabling cell-based multiplexing. Quantum packets are created by a quantumization process, which can be viewed as an extended or generalized AAL process. However, unlike ATM, which employs a connection-oriented framework, the proposed quantumization process is more flexible and extensible. It can be employed to support MPLS, ATM and active network traffic based on a general packetization framework, thus enabling the network to provide a wide range of services in an integrated manner. The framework can also be extended to satisfy future packetization needs. We have presented a theoretical analysis that compares the network cost when loss-sensitive quantum packets are forwarded using three schemes: cut-through, store-and-forward, and ideal. Close form mathematical expressions have been obtained for some situations. Analytical and simulation results have been presented to show the system behavior. Through the mathematical model, we have discovered two interesting equivalence phenomena for the cut-through and store-and-forward approaches.

APPENDICES

Appendix 1: Detailed Derivation of (5)

$$\begin{aligned}
&C_{\text{fail}}\\
&= \sum_{y=1}^{m} \frac{\binom{m}{y}(p^n)^{m-y}(1-p^n)^y}{1-p^{nm}}\left[yC_{\text{lost}}+(m-y)n\right]\\
&= \frac{1}{1-p^{nm}}\sum_{y=1}^{m}\binom{m}{y}(p^n)^{m-y}(1-p^n)^y\, yC_{\text{lost}}\\
&\quad +nm-\frac{1}{1-p^{nm}}\sum_{y=1}^{m}\binom{m}{y}(p^n)^{m-y}(1-p^n)^y\, ny\\
&= \frac{m(1-p^n)}{1-p^{nm}}(C_{\text{lost}}-n)+nm.
\end{aligned}$$

Appendix 2: Detailed Derivation of (6)

$$\begin{aligned}
C_{\text{lost}} &= \sum_{z=1}^{n}\left[\frac{p^{z-1}(1-p)}{1-p^n}\right](z-1)=\sum_{z=0}^{n-1}\left[\frac{p^z(1-p)}{1-p^n}\right]z\\
&= \frac{1-p}{1-p^n}\sum_{z=0}^{n-1}p^z z=\frac{1-p}{1-p^n}\sum_{z=0}^{n-1}p\times p^{z-1}z\\
&= \frac{(1-p)p}{1-p^n}\sum_{z=0}^{n-1}\frac{d(p^z)}{dp}=\frac{(1-p)p}{1-p^n}\frac{d}{dp}\sum_{z=0}^{n-1}p^z\\
&= \frac{(1-p)p}{1-p^n}\frac{d(\frac{1-p^n}{1-p})}{dp}\\
&= \frac{(1-p)p}{1-p^n}\times\frac{(n-1)p^n-np^{n-1}+1}{(1-p)^2}\\
&= \frac{(n-1)p^{n+1}-np^n+p}{(1-p)(1-p^n)}.
\end{aligned}$$

Appendix 3: Detailed Derivation of (7)

$$\begin{aligned}
C &= \left(\frac{1}{p^{nm}}-1\right)\frac{m(1-p^n)}{1-p^{nm}}\left[\frac{(n-1)p^{n+1}-np^n+p}{(1-p)(1-p^n)}-n\right]\\
&\quad +\left(\frac{1}{p^{nm}}-1\right)nm+nm\\
&= \frac{m(1-p^n)}{p^{nm}}\left[\frac{(n-1)p^{n+1}-np^n+p}{(1-p)(1-p^n)}-n\right]+\frac{nm}{p^{nm}}\\
&= \frac{m(1-p^n)}{p^{nm}}\left[\frac{(n+1)p-n-p^{n+1}}{(1-p)(1-p^n)}\right]+\frac{nm}{p^{nm}}\\
&= \frac{m}{p^{nm}}\left[\frac{(n+1)p-n-p^{n+1}}{(1-p)}+n\right]=\frac{mp(1-p^n)}{p^{nm}(1-p)}.
\end{aligned}$$

Appendix 4: Detailed Derivation of (9)

$$\begin{aligned}
C(y) &= \sum_{z=1}^{m}\frac{\binom{m}{z}p^{m-z}(1-p)^z}{1-p^m}(my-z)\\
&= my-\frac{1}{1-p^m}\sum_{z=1}^{m}\binom{m}{z}p^{m-z}(1-p)^z z\\
&= my-\frac{m(1-p)}{1-p^m}.
\end{aligned}$$

Appendix 5: Detailed Derivation of (10)

$$\begin{aligned}
C_{\text{fail}} &= \sum_{y=1}^{n}\frac{p^{m(y-1)}(1-p^m)}{1-p^{nm}}\left[my-\frac{m(1-p)}{1-p^m}\right]\\
&= \sum_{y=1}^{n}\frac{p^{m(y-1)}(1-p^m)}{1-p^{nm}}my-\frac{m(1-p)}{1-p^m}\\
&= \frac{m(1-p^m)}{1-p^{nm}}\sum_{y=1}^{n}p^{m(y-1)}y-\frac{m(1-p)}{1-p^m}\\
&= \frac{m(1-p^m)}{1-p^{nm}}\sum_{y=1}^{n}\frac{d(p^{my})}{dp^m}-\frac{m(1-p)}{1-p^m}\\
&= \frac{m(1-p^m)}{1-p^{nm}}\times\frac{d}{dp^m}\sum_{y=1}^{n}p^{my}-\frac{m(1-p)}{1-p^m}\\
&= \frac{m(1-p^m)}{1-p^{nm}}\times\frac{d}{dp^m}\left[\frac{1-p^{(n+1)m}}{1-p^m}-1\right]-\frac{m(1-p)}{1-p^m}\\
&= \frac{m(1-p^m)}{1-p^{nm}}\times\frac{np^{(n+1)m}-(n+1)p^{nm}+1}{(1-p^m)^2}\\
&\quad -\frac{m(1-p)}{1-p^m}\\
&= \frac{m\left[np^{(n+1)m}-(n+1)p^{nm}+1\right]}{(1-p^{nm})(1-p^m)}-\frac{m(1-p)}{1-p^m}\\
&= \frac{m\left[np^{(n+1)m}-p^{nm+1}-np^{nm}+p\right]}{(1-p^{nm})(1-p^m)}.
\end{aligned}$$

REFERENCES

[1] G. C. Kessler and P. V. Southwick, *ISDN: Concepts, Facilities and Services*, Signature Ed., McGraw Hill, 1998.

[2] L. L. Peterson and B. S. Davie, *Computer Networks: A Systems Approach*, 3rd Ed., Morgan Kaufmann, 2003

[3] O. Kyas and G. Crawford, *ATM Networks*, 1st Ed., Prentice Hall, 2002.

[4] M. Kawarasaki and B. Jabbari, "B-ISDN architecture and protocol," *IEEE J. Sel. Areas Commun.*, vol. 9, no. 9, pp. 1405–1415, Dec. 1991.

[5] U. D. Black, *Emerging Communications Technologies*, 2nd Ed., Prentice Hall, 1997.

[6] W. J. Goralski, *Frame Relay for High-Speed Networks*, 1st Ed., John Wiley & Sons, 1999.

[7] M. Rahnema, "Frame relaying and the fast packet switching concepts and issues," *IEEE Network*, vol. 5, no. 4, pp. 18–23, July 1991.

[8] A. Rodriguez, J. Gatrell, J. Karas, and R. Peschke, *TCP/IP Tutorial and Technical Overview*, 7th Ed., Prentice Hall, 2001.

[9] K. L. Calvert, S. Bhattacharjee, E. Zegura, and J. Sterbenz, "Directions in active networks," *IEEE Commun. Mag.*, vol. 36, no. 10, pp. 72–78, Oct. 1998.

[10] D. L. Tennenhouse, J. M. Smith, W. D. Sincoskie, D. J. Wetherall, and G. J. Minden, "A survey of active network research," *IEEE Commun. Mag.*, vol. 35, no. 1, pp. 80–86, Jan. 1997.

[11] P. Tullmann, M. Hibler, and J. Lepreau, "Janos: A java-oriented OS for active network nodes," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 3, pp. 501–510, Mar. 2001.

[12] S. da Silva, Y. Yemini, and D. Florissi, "The NetScript active network system," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 3, pp. 538–551, Mar. 2001.

[13] R. Y. W. Lam, *Quantum Packet for Integrated Services Data Network—the Third Generation ISDN.* M.Phil. Thesis, Department of Computing, The Hong Kong Polytechnic University, 2003.

[14] U. D. Black, *MPLS & Label Switching Networks*, 2nd Ed., Prentice Hall, 2002.

[15] E. Rosen, A. Viswanathan, and R. Callon, "Multiprotocol Label Switching Architecture," RFC 3031, Jan. 2001.

[16] I. Widjaja and A. I. Elwalid, "Performance issues in VC-merge capable switches for multiprotocol label switching," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 6, pp. 1178–1189, June 1999.

[17] C. -T. Lea, C. -Y. Tsui, B. Li, L. C. -Y. Kwan, S. K. -M. Chan, and A. H. -W. Chan, "A/I Net: A network that integrates ATM and IP," *IEEE Network*, vol. 13, no. 1, pp. 48–55, Jan./Feb. 1999.

[18] I. Andrikopoulos, G. Pavlou, P. Georgatsos, N. Karatzas, K. Kavidopoulos, J. Rothig, S. Schaller, D. Ooms, and P. Van Heuven, "Experiments and enhancements for IP and ATM integration: The IthACI project," *IEEE Commun. Mag.*, vol. 39, no. 5, pp. 146–155, May 2001.

[19] — [Online]. Available: http://www.internet2.edu/

[20] K. Takagi, M. Nichihara, and K. Murakami, "A proposal of new data link protocol, universal data link, for high-speed IP/ATM-hybrid and IP-centric transport networks," in *Proc. ICC 2001*, Helsinki, Finland, vol. 4, June 2001, pp. 1259–1263.

[21] K. Shiomoto, M. Uga, M. Omotani, and S. Shimizu, "Design of IP+ATM switch router architecture for best-effort and guaranteed services in the next generation public data networks," in *Proc. ICC 2001*, Helsinki, Finland, vol. 1, June 2001, pp. 89–94.

[22] M. Bocci and J. Guillet, "ATM in MPLS-based converged core data networks," *IEEE Commun. Mag.*, vol. 41, no. 1, pp. 139–145, Jan. 2003.

[23] C. Metz, "Layer 2 over IP/MPLS," *IEEE Internet Comput.*, vol. 5, no. 4, pp. 77–82, July–Aug. 2001.

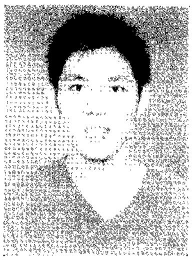
[24] D. Scott Alexander, B. Braden, C. A. Gunter, A. W. Jackson, A. D. Keromytis, G. J. Minden, and D. Wetherall, "Active Network Encapsulation Protocol (ANEP)," RFC Draft, July 1997. [Online]. Available: http://www.cis.upenn.edu/ switchware/ANEP/docs/ANEP.txt.

[25] P. T. Brady, "A model for generating ON-OFF speech patterns in two-way conversation," *Bell Syst. Tech. J.*, vol. 48, no. 7, pp. 2445–2472, Sept. 1969.

[26] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson, and J. D. Robbins, "Performance models of statistical multiplexing in packet video communications," *IEEE Trans. Commun.*, vol. 36, no. 7, pp. 834–844, July 1988.

[27] P. Karn and C. Partridge, "Improving round-trip time estimates in reliable transport protocols," *IEEE ACM Trans. Comput. Syst.*, vol. 9, no. 4, pp. 364–373, Nov. 1991.

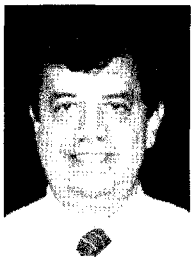
[28] H. C. B. Chan, H. M. Alnuweiri, and V. C. M. Leung, "A framework for optimizing the cost and performance of next-generation IP routers," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 6, pp. 1013–1029, June 1999.



Ray Y. W. Lam received his B.A. (first class honors) and M.Phil. degrees, both in Computing, from The Hong Kong Polytechnic University, Hong Kong in 2000 and 2003 respectively. During 2003, he was a research assistant in the Department of Computing at the same university. Currently, he is a Ph.D. candidate in the Department of Electrical and Computer Engineering at the University of British Columbia, Canada. His research interests include multimedia content delivery over wireless access and mesh networks and development of next-generation networks.

Henry C. B. Chan received his B.A. and M.A. degrees from the University of Cambridge, England and his Ph.D. degree from the University of British Columbia, Canada. From October 1988 to October 1993, he worked with Hong Kong Telecommunications Limited primarily on the development of networking services in Hong Kong. Between October 1997 and August 1998, he worked with BC TEL Advanced Communications on the development of high-speed networking technologies and ATM-based services. Currently, he is an Associate Professor in the Department of Computing at The Hong Kong Polytechnic University. His research interests include networking/communications, wireless networks, Internet technologies and electronic commerce. He has authored/co-authored a textbook on electronic commerce published by John Wiley and Sons, a book chapter in the Internet Encyclopedia (Wiley) and over 50 journal/conference papers. He is a Member of IEEE and ACM. He is currently the Chair of the IEEE Hong Kong Section Computer Chapter.

Hui Chen received his B.S. and M.S. degrees in Computer Science from Nanjing University, China, in 1999 and 2002, respectively. He is a Ph.D. candidate in the Department of Electrical and Computer Engineering at the University of British Columbia, Canada. His research topic is modeling and cross-layer optimization of packet-switched third generation wireless networks. From 2000–2003, he worked as a Research Assistant in the Department of Computing of The Hong Kong Polytechnic University. His research interests include wireless networks and electronic commerce.



Tharam S. Dillon received the B.E. (Hons) and Ph.D. from Monash University in Australia. He is currently a Distinguished Research Professor with the Digital Ecosystems and Business Intelligence Institute, Curtin University of Technology, Perth, WA, Australia. His current research interests include Web semantics, ontologies, Internet computing, e-commerce, hybrid neurosymbolic systems, neural nets, software engineering, database systems, and computer networks. He is the author or coauthor of more than 650 papers published in international and national journals and is the author of five books. He is the Editor-in-Chief of the International Journal of Computer Systems Science and Engineering as well as the Engineering Intelligent Systems. He is the Co-Editor of the Journal of Electric Power and Energy Systems.

Victor O. K. Li received S.B., S.M., E.E., and Sc.D. degrees in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology, Cambridge, Massachusetts, in 1977, 1979, 1980, and 1981, respectively. He joined the University of Southern California (USC), Los Angeles, California, USA in February 1981, and became Professor of Electrical Engineering and Director of the USC Communication Sciences Institute. Since September 1997 he has been with the University of Hong Kong, Hong Kong, where he is an Associate Dean (Research) of Engineering, and Chair Professor of Information Engineering at the Department of Electrical and Electronic Engineering. He has also served as Managing Director of Versitech Ltd., the technology transfer and commercial arm of the University, and on various corporate boards.

His research is in information technology, including all-optical networks, wireless networks, and Internet technologies and applications. Sought by government, industry, and academic organizations, he has lectured and consulted extensively around the world. He chaired the Computer Communications Technical Committee of the IEEE Communications Society 1987–1989, and the Los Angeles Chapter of the IEEE Information Theory Group 1983–1985. He co-founded the International Conference on Computer Communications and Networks (IC3N), and chaired its Steering Committee 1992–1997. He also chaired various international workshops and conferences, including, most recently, IEEE INFOCOM 2004 and IEEE HPSR 2005. He has served as an editor of IEEE Network, IEEE JSAC Wireless Communications Series, and Telecommunication Systems. He also guest edited special issues of IEEE JSAC, Computer Networks and ISDN Systems, and KICS/IEEE Journal of Communications and Networking. He is now serving as an editor of ACM/Springer Wireless Networks and IEEE Communications Surveys and Tutorials. He has been appointed to the Hong Kong Information Infrastructure Advisory Committee by the Chief Executive of the Hong Kong Special Administrative Region (HKSAR). He is a part-time member of the Central Policy Unit of the Hong Kong Government. He has also served on the Innovation and Technology Fund (Electronics) Vetting Committee, the Small Entrepreneur Research Assistance Programme Committee, the Engineering Panel of the Research Grants Council, and the Task Force for the Hong Kong Academic and Research Network (HARNET) Development Fund of the University Grants Committee. He was a Distinguished Lecturer at the University of California at San Diego, at the National Science Council of Taiwan, and at the California Polytechnic Institute. He has also delivered keynote speeches at many international conferences. He has received numerous awards, including, most recently, the PRC Ministry of Education Changjiang Chair Professorship at Tsinghua University, Beijing, the UK Royal Academy of Engineering Senior Visiting Fellowship in Communications, the Outstanding Researcher Award of the University of Hong Kong, the Croucher Foundation Senior Research Fellowship, and the Order of the Bronze Bauhinia Star, Government of HKSAR, China. He was elected an IEEE Fellow in 1992. He is also a Fellow of the HKIE and the IAE.

Victor C. M. Leung received the B.A.Sc. (Hons.) degree in electrical engineering from the University of British Columbia (U.B.C.) in 1977, and was awarded the APEBC Gold Medal as the head of the graduating class in the Faculty of Applied Science. He attended graduate school at U.B.C. on a Natural Sciences and Engineering Research Council Postgraduate Scholarship and obtained the Ph.D. degree in electrical engineering in 1981. From 1981 to 1987, He was a Senior Member of Technical Staff at Microtel Pacific Research Ltd. (later renamed MPR Teltech Ltd.), specializing in the planning, design and analysis of satellite communication systems. He also held a part-time position as Visiting Assistant Professor at Simon Fraser University in 1986 and 1987. In 1988, he was a Lecturer in the Department of Electronics at the Chinese University of Hong Kong. He returned to U.B.C. as a faculty member in 1989, where he is currently a Professor and the holder of the TELUS Mobility Industrial Research Chair in Advanced Telecommunications Engineering in the Department of Electrical and Computer Engineering. He is a member of the Institute for Computing, Information and Cognitive Systems at U.B.C. He also holds a Guest Professor appointment at Jilin University, China. He has authored/co-authored more than 300 technical papers in international journals and conference proceedings. His research interests are in the areas of architectural and protocol design and performance analysis for computer and telecommunication networks, with applications in satellite, mobile, personal communications and high speed networks. He is a registered professional engineer in the Province of British Columbia, Canada. He is a Fellow of IEEE and a voting member of ACM. He serves on the editorial boards of the IEEE Transactions on Wireless Communications, the IEEE Transactions on Vehicular Technology, the IEEE Transactions on Computers, and the International Journal of Sensor Networks. He has guest-edited several journal special issues, and served on the technical program committee of numerous international conferences. He is the General Chair of QShine 2007 in Vancouver, BC, and Symposium Chair for Next Generation Mobile Networks in IWCMC 2007 in Honolulu, Hawaii and IWCMC 2006 in Vancouver, BC. He was a General Co-chair of ACM MSWiM 2005 in Montreal, Canada, and a TPC Vice-chair of IEEE WCNC 2005 in New Orleans, USA.