High-Speed IP/IPsec Processor LSIs

● Tomokazu Aoki ● Teruhiko Nagatomo ● Kazuya Asano

(Manuscript received November 1, 2005)

In recent years, we have seen an increase in the speed of Internet access lines together with demands for concealed communications. To meet these demands, we have developed two LSIs, the MB86977 and MB86978, which are optimized for use in gateways such as broadband VPN routers for secure, high-speed communications. The MB86977 is a high-performance IP packet processing engine. After it has been programmed, this device can process bi-directional routing, Network Address Port Translation (NAPT), Point to Point Protocol over Ethernet (PPPoE), and filtering at a 100 Mb/s full-wire speed. The MB86978 is an IPsec accelerator LSI capable of processing packets inline on the transmission path. Once this device has been programmed, it can perform bi-directional IPsec processing at a 100 Mb/s full-wire speed. This paper describes the functions, performance, and mechanisms of these two LSIs. It then describes how these LSIs are superior to software processing solutions. Lastly, this paper introduces a reference board that can facilitate rapid development of broadband VPN routers.

1. Introduction

Since the introduction of the keyword "broadband," there has been a remarkable increase in demands from Internet service providers and users for higher Internet speeds. Even with Internet connections in the home and in Small Office Home Office (SOHO) environments, continuous broadband connections using Asymmetrical Digital Subscriber Line (ADSL) and fiber-optic lines are now the norm. Internet access-related demands will continue to increase, even though a fast, convenient Internet environment has already been provided. We are now at the beginning of the commercialization phase, in which these lines will be used to distribute digital content, image communication data, and broadcasts. Accordingly, a further increase in the quantity of IP packets flowing through the lines can be expected. Moreover, we can expect an increase in the number of packets that must be transmitted and received in real time, for example, audio packets and streaming data packets for moving images.

In this situation, an improvement in the processing ability of the gateway equipment that connects home and office LANs to the Internet is required.

In the past, CPU-based software processing was the main solution used in this gateway equipment. However, as the processing rate increases in proportion to the increase in transmission quantity and quality, the method of relying on CPU performance for these operations is beginning to fail.

In response, we developed the following 100 Mb/s full-wire-speed LSI solutions for smooth, bi-directional IP packet processing in gateway equipment: a high-speed IP packet processing engine called the MB86977 and a high-speed inline system IPsec processing engine called the MB86978.

This paper describes the functions and configurations of these two LSIs, their evaluation

results, and some application examples.

2. Problems with conventional solutions

The following high-load processing tasks are required of gateway equipment.

1) Routing processing

Routing is the process of moving data packets from source to destination.

This process is executed in the gateway, which is at the boundary between networks.

Processing an IP packet involves analyzing the IP header, reassigning the Media Access Control (MAC) address, Time to Live (TTL) subtraction, IP header checksum recalculation, and other operations.

2) Network Address Port Translation (NAPT)

This function is used to translate the IP addresses and Transmission Control Protocol (TCP)/User Datagram Protocol (UDP) port numbers in the IP packet to make effective use of IP addresses, which are now a scarce resource. It is also referred to as IP Masquerade.

3) PPP over Ethernet (PPPoE) connection

PPPoE is a protocol that enables Point to Point Protocol (PPP) connection to a provider's network in an Ethernet network environment. In order to use this protocol to connect to a network, PPPoE frame encapsulation/re-encapsulation processing is required.

4) Filtering

To prevent transmission to an internal LAN from an outside network, this function passes on or discards a received packet according to the IP address, protocol, port number, and other details.

5) IP Security Protocol (IPsec) transmission

IPsec is the most commonly used protocol for maintaining the security of Internet transmissions. It encrypts data at the IP protocol level. Specifically, IPsec performs IP packet encryption/decryption and authentication and IP packet encapsulation/decapsulation for each packet.

Entrusting all of these complex processing tasks to a CPU invites a reduction in transfer speed (transmission throughput) and a very high amount of jitter, which is a measure of the variability over time of latency across a network. If we take 100 Mb/s to represent 100%, an average broadband router that uses CPU solutions has a bi-directional routing throughput of less than 50%. Adding NAPT and filtering functions invites an even further reduction in performance. Moreover, if IPsec processing is also performed, the throughput falls below the 30% level, even for a unidirectional transmission. To lighten the transmission load, dedicated LSIs must perform these processing tasks.

3. LSI development principles

The transmission processing performed in gateway equipment is complex, but if we analyze the processing content, it can be broadly divided into the following two phases.

1) Determining the parameters to be used for transmission
2) Packet processing using a set procedure based on those parameters

Phase 1) is suitable for flexible CPU-based software processing, while Phase 2) is suitable for high-speed, LSI-based processing.

Accordingly, system performance can be improved by optimizing the assignment of these functions within network equipment. Moreover, to make these functions available for home-based use, we must work hard to reduce the unit price of the LSIs and configure systems that are as inexpensive as possible.

Based on these ideas, we decided on the following LSI development principles.

1) CPU connection method

We decided to adopt a commodity input/output interface for connection with the host CPU. Firstly, this approach is less expensive than when a dedicated interface such as PCI bus is incorporated in the LSI configuration. Secondly, a commodity interface enables CPU connectivity without restrictions, while providing the freedom to select a CPU performance and architecture to

suit the application.

2) Functions incorporated in LSIs

We decided to implement the routing, NAPT, PPPoE, and filtering functions in a single LSI as basic IP packet processing functions performed in the gateway.

On the other hand, we decided to implement the IPsec IP packet encryption/decryption, authentication, and encapsulation/decapsulation processing functions as well as the search functions for suitable key encryptions and algorithms that precede these functions (i.e., the IPsec Security Association [SA] search functions) in another LSI. Using the two LSIs, we designed a system in which the required parameters are set in the LSIs' internal tables by CPU-based software processing and then the LSIs' internal logic automatically performs the packet processing tasks according to those parameters. The MB86977's table settings for performing the NAPT function are shown in **Figure 1**, and the MB86978's table settings for performing the IPsec function are shown in **Figure 2**.

3) LSI internal table size

The LSIs' internal tables are mainly used for setting the parameters mentioned in 2) above. The unit used for setting a group of parameters within a table is referred to as an "entry." The size of the internal tables is proportional to the number of entries and is linked to the surface area and therefore to the cost of the LSIs. Conversely, if the number of entries is too small, the number of transmission connections handled by the LSIs becomes limited.

For the routing and NAPT functions, we therefore allocated 128 entries based on our estimation of peak-time Internet access in home and SOHO LAN environments (Figure 1). Our policy makes effective use of these 128 entries by deleting settings that are no longer required at termination of a TCP-IP connection, which is the basic protocol for controlling data communication on the Internet. Based on current usage of the PPPoE function, we decided that enough entries for four Internet Service Provider (ISP) connections would be sufficient (i.e., four entries). The default DROP policy is used for the filter, and we allocated 64 entries for performing passage settings as required. For the IPsec SA function for normal VPN router operation in home and SOHO environments, we allocated 128 entries (64 upstream and 64 downstream) (Figure 2).

Entry number	Protocol and IP version	IP address			TCP/UDP port number			MAC address		Interface	
		IP address of WAN side entity	IP address of LAN side entity	IP address used for NAPT	Port number of WAN side entity	Port number of LAN side entity	Port number used for NAPT	MAC address of WAN side entity	MAC address of LAN side entity	WAN side interface	LAN side interface
0	IPv4, TCP	192.168.101.4	10.109.121.44	192.168.100.10	80	2000	3000	00:e0:16:83:56:83	00:20:ed:93:b7:00	WAN	LANO
1											
⋮											
127											

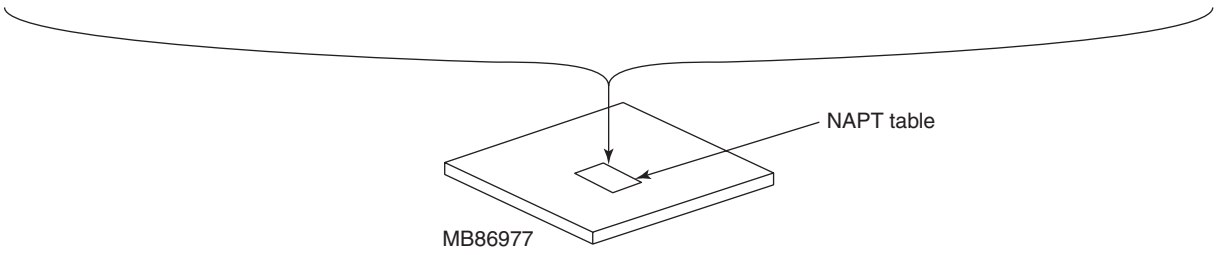


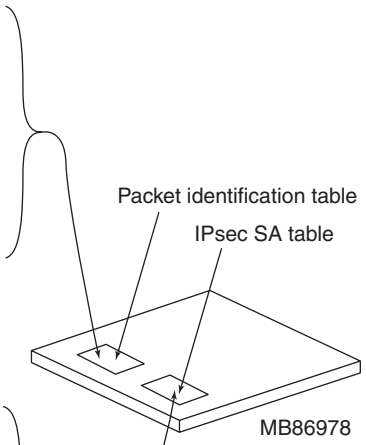
Figure 1
Example of MB86977 NAPT table settings.

Packet identification table settings

Entry number	Protocol and IP version	IP address		TCP/UDP port number		Packet control setting
		Destination IP address	Source IP address	Destination Port number	Source Port number	
0	IPv4, TCP	192.168.101.4	10.109.121.44	500	2000	SA ID = 1
1						
⋮						
63						

IPsec SA table settings

SA ID	Algorithm and mode setting	Key for EPS	Key for authentication	Tunnel header	ESP SPI	AH SPI	Lifetime
0							
1	AES, SHA-1	key data	key data	IP header	40	50	10 minutes
⋮							
63							



AH : Authentication Header
ESP: Encapsulating Security Payload
SPI: Security Parameters Index

Figure 2
Example of MB86978 IPsec table settings.

4) IP protocol version

We decided to ensure compatibility not only with the current mainstream IPv4 protocol, but also with the IPv6 protocol, which will be the next mainstream protocol.

4. LSI configuration and key technologies

In this section, we describe the configurations of the LSIs we developed and the key technologies we used to improve their performance.

4.1 Configuration of MB86977

The block diagram of the MB86977 is shown in **Figure 3**, and a photograph is shown in **Figure 4**.

The MB89677 has four Media Independent Interface (MII) ports for standard 10 base-T/100 base-TX Ethernet interfaces: one for a WAN (Internet) connection and three for LAN connections.

The packets received from the ports are temporarily stored in the switching and queuing block. When an Ethernet frame is received at one of the three LAN ports, its MAC address table is searched to see which of the three LANs it is addressed to. One of the LAN ports differs from the other two ports in that it is a Demilitarized Zone (DMZ) port when DMZ mode is enabled. When it is enabled, MAC address base switching between the DMZ port and the other two LAN ports is not performed. In the same way, because transmissions between the WAN and DMZ ports consist of packets moving between network segments in different broadcasting domains, MAC address base switching between the WAN and DMZ ports is not performed. Similarly, MAC address base switching between the WAN and LAN ports is not performed.

With this type of inter-segment transmission system, packets are transferred from the switch queue block to the IP packet processing block. The IP packet processing block searches for the table

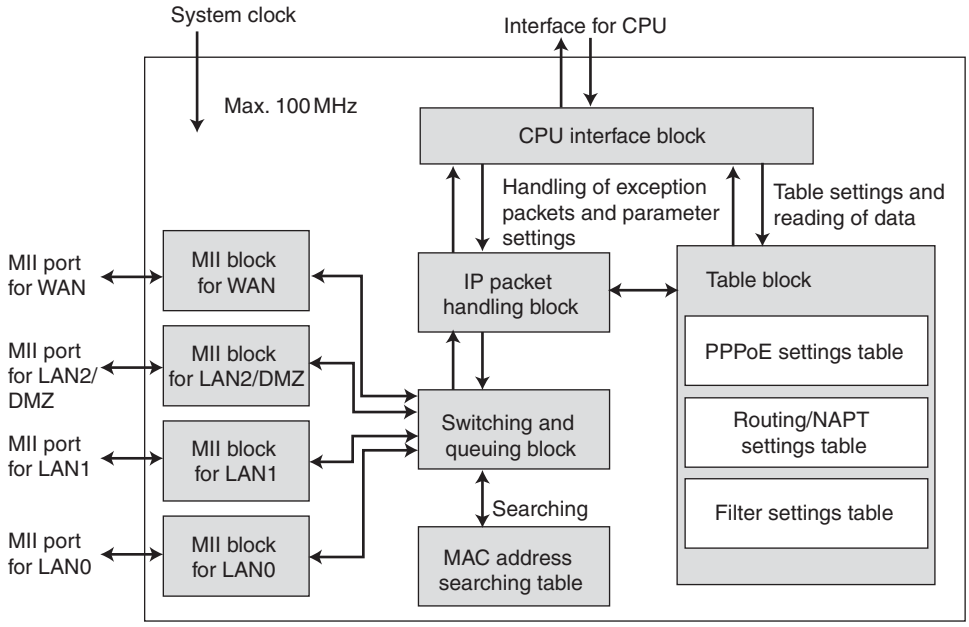


Figure 3
MB86977 block diagram.

block and extracts the parameters required for IP packet processing from each table. The IP packet processing block then performs routing/NAPT, filtering, and PPPoE packet processing based on these parameters. Then, the packet is returned to the switch queue block and output to the transfer destination MII interface. This configuration enables a series of these operations to be performed at a bi-directional, full-wire speed of 100 Mb/s when the clock is operating at 50 MHz.

There are three types of tables in the table block: one for setting the 128 connections of the mechanism used to perform the routing and NAPT functions, one for setting the four connections used for PPPoE connection, and one for the 64 filter settings for inputs that are output to a different broadcasting domain.

These tables can be set via the CPU interface block. Moreover, if the IP packet processing block cannot process packets due to insufficient table block settings or an unknown protocol, the packets can be sent to the CPU via the CPU interface block and processed by the CPU.

A summary of the MB86977 specifications is given in **Table 1**.

LQFP 208-pin

Figure 4
MB86977.

4.2 Configuration of MB86978

The block diagram of the MB86978 is shown in **Figure 5**, and a photograph is shown in **Figure 6**.

The MB86978 has two of the same MII ports that are used in the MB86977. One of these is a

Table 1
MB86977 specification summary.

Functions	IP routing/NAPT entries: 128
	Filter entries (inbound: 64, outbound: 64)
	PPPoE session entries: 4
	DMZ/Switching modes incorporated
	4-port, 10/100 Mb/s MII
	Commodity input/output interface for CPU connection
Performance	Bidirectional at 50 MHz clock operation
	100 Mb/s full wire transmission enabled
Operating frequency	Max. 100 MHz
Power consumption	600 mW (at 100 MHz operation)
Technology	0.18 µm, 5-layer CMOS process
Package	208-pin LQFP

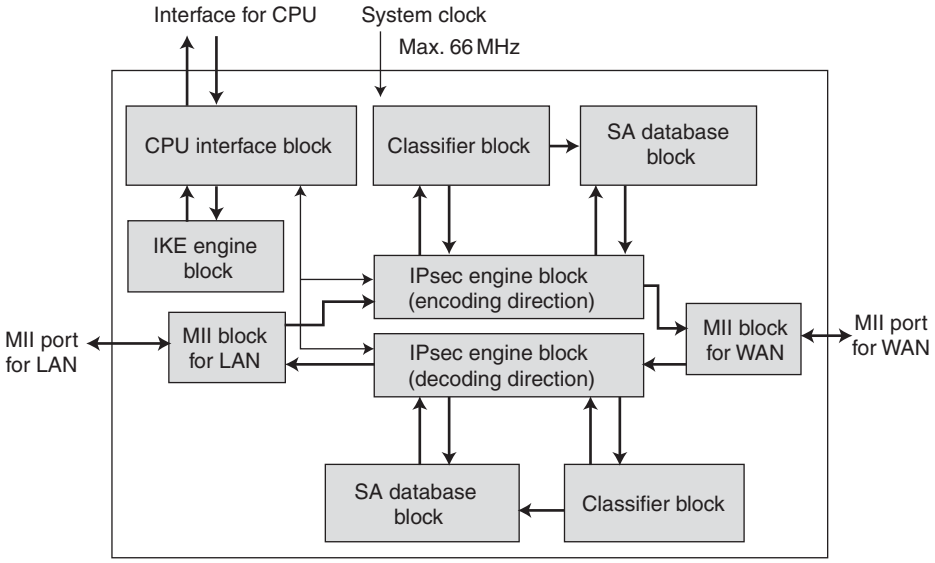
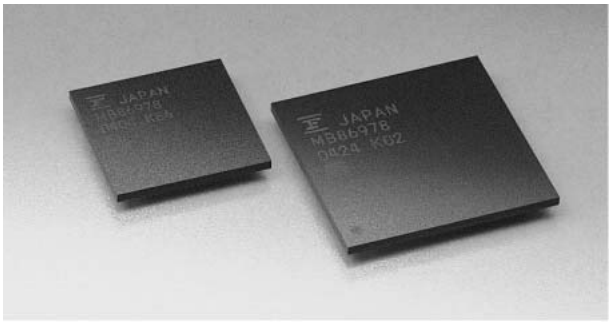


Figure 5
MB86978 block diagram.



FBGA 337-pin FBGA 288-pin

Figure 6
MB86978.

WAN interface, and the other is a LAN interface for connecting with the router functions.

The basic configuration consists of 1) a classifier block for identifying IP packets transferred between the LAN and WAN ports and then judging whether they are suitable for IPsec SA processing; 2) an SA database block for registering parameters (e.g., an encryption key, authentication key, algorithm specification, and tunnel header for encapsulation) when packets are judged suitable for IPsec SA processing; and 3) an IPsec engine block for obtaining parameters from the SA database block and performing encryption, authentication, encapsulation, and other IPsec processing tasks.

This IPsec engine block can operate independently in the LAN-to-WAN encoding direction and in the WAN-to-LAN decoding direction. The classifier and SA database blocks each have two 64-entry tables: one for packet identification and one for IPsec SA registration. When a packet is identified by referencing the packet identification table, the entry number corresponding to the IPsec SA table is decided (Figure 2). Then, the ground-to-ground number (number of IPsec transmissions with the corresponding network devices) can be established for the 64 entries.

When its clock is operating at 66 MHz, the MB86978 can judge whether IPsec SA processing is suitable for the packet, obtain parameters from the SA database block, and enable the series of IPsec processing operations to be performed at a full-wire speed of 100 Mb/s independently in the encoding and decoding directions. Moreover, because IPsec transmission requires that the CPU perform Internet Key Exchange (IKE) negotiation before IPsec SA establishment, we incorporated an accelerator for modulo-exponentiation operations to support the CPU and an encryption/authentication engine in the IKE engine block. The CPU can therefore access the IKE engine block via the CPU interface block to perform operations at high speed.

A summary of the MB86978 specifications is given in **Table 2**.

4.3 Key technologies for realizing high-speed packet and IPsec processing

In this subsection we describe the packet identification mechanism used to realize high-

Table 2
MB86978 specification summary.

Functions	Selectable tunnel/transport modes
	Ground-to-ground number: 64 entries
	DES/3DES, AES encryption compatibility
	SHA-1, MD5 authentication compatibility
	IKE engine incorporated
	2-port, 10/100 Mb/s MII
	Commodity input/output interface for CPU connection
Performance	Bidirectional at 66 MHz clock operation
	Full-wire processing enabled
Operating frequency	Max. 66 MHz
Power consumption	200 mW (at 66 MHz operation)
Technology	0.18 µm, 5-layer CMOS process
Package	288-pin FBGA or 337-pin FBGA

DES: Data Encryption Standard
AES: Advanced Encryption Standard
SHA-1: Secure Hash Algorithm 1
MD5: Message Digest 5

speed processing in the MB86977 and MB86978. We also describe the DES circuit for performing high-speed IPsec processing in the MB86978.

The packet identification mechanism is the key technology used in the MB86977 and MB86978 to achieve a full-wire speed of 100 Mb/s in a small chip surface area. The concept of this packet identification mechanism is shown in **Figure 7**.

Packets are received at 100 Mb/s and transferred to the packet identification mechanism in 8-bit segments. This mechanism consists of a protocol analysis circuit for identifying the identifiable data in the packet, an address control circuit for controlling the embedded memory address, and a unit comparison circuit for comparing the packet data with the reference data recorded in the embedded memory. The reference data is stored in the embedded memory in the form of a reference table. The data for several entries is recorded in 8-bit units in each line of the reference table. For example, the MB86977 records 128 bytes of data per line for the NAPT function.

The operation of this packet identification mechanism is shown in **Figure 8**.

As shown in the figure, the reference table contains the destination IP address and source IP address. The protocol analysis circuit compares, in 8-bit units, the data for comparison in the packet that was received with the reference data in each entry. In the case of an IPv4 packet, when the 8-byte segments of the source IP address and destination IP address are compared, each matching entry becomes valid. If there are multiple matching entries, the entry with the lowest number is selected.

This comparison in 8-bit units is completed in one system clock cycle in the MB86977 and two system clock cycles in the MB86978. The packet identification speed is 800 Mb/s at 100 MHz in the MB86977 and 266 Mb/s at 66 MHz in the MB86978. For the MB86978, a processing performance of 200 Mb/s is sufficient. However, because the MB86977 has four ports, it needs a

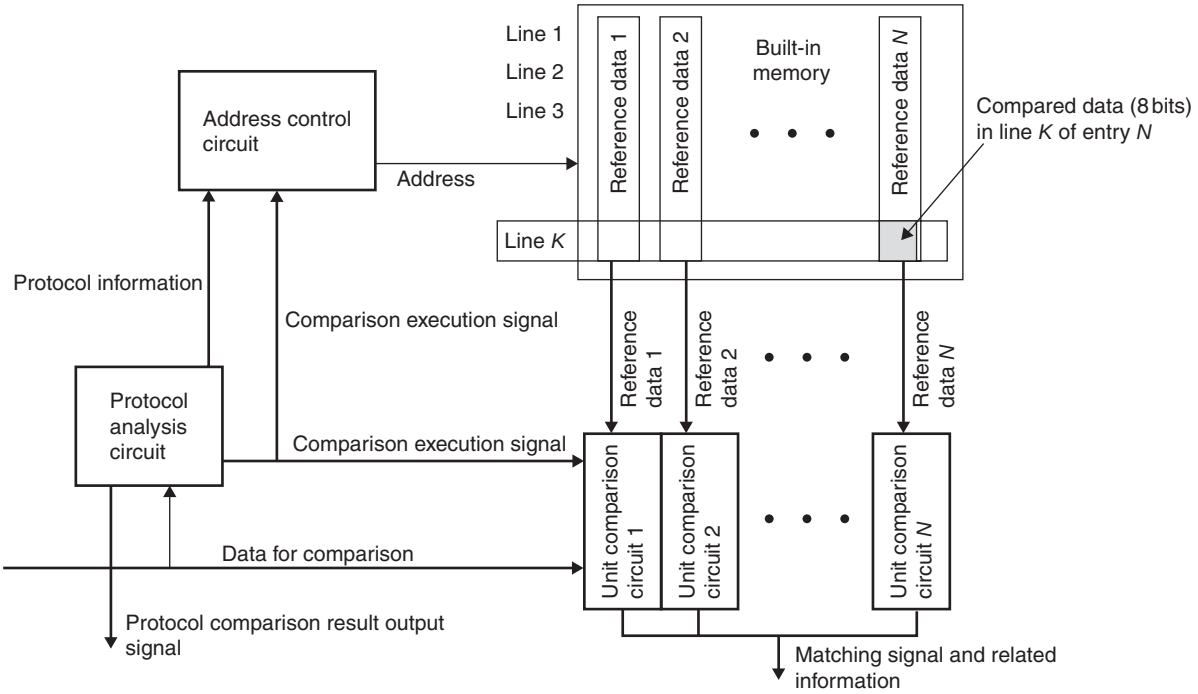


Figure 7
Concept of packet identification mechanism.

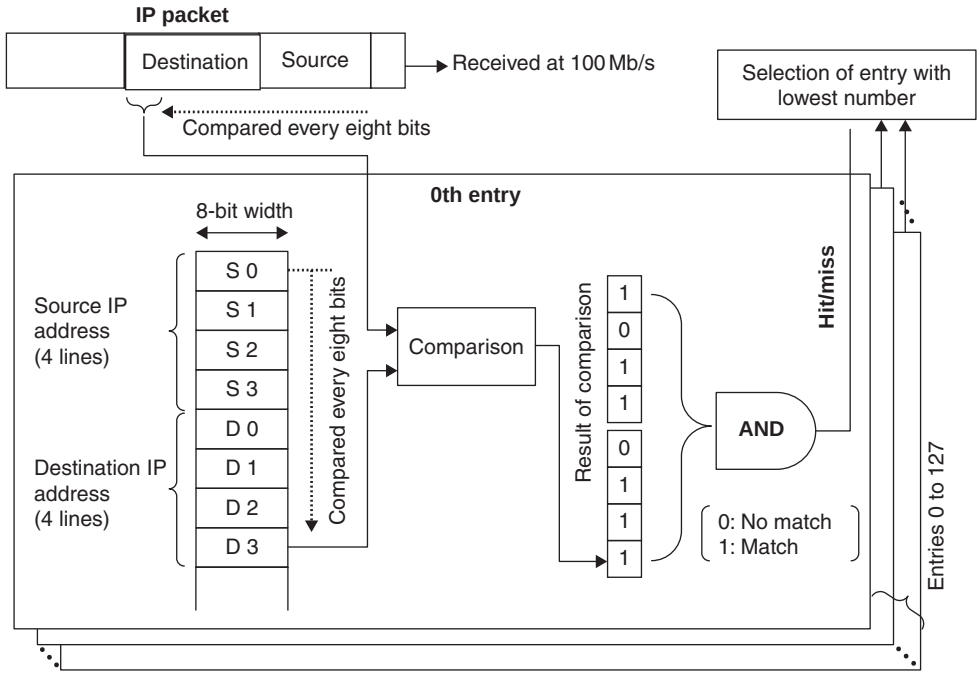


Figure 8
Operation of packet identification mechanism.

minimum processing performance of 400 Mb/s to enable two sets of bi-directional, full-wire transfer operations to be carried out. The MB86977 therefore has a higher processing performance.

Moreover, we also incorporated one more device in the MB86978 to increase the speed of the DES circuit. As shown in **Figure 9**, to perform 64-bit processing in 16 rounds, a DES circuit is generally configured to perform this processing in 16 system clock cycles. However, a bottleneck occurs when a 3DES circuit is used. To solve this problem, we configured the DES circuit in the MB86978 as shown in **Figure 10** to enable processing in four system clock cycles.

As shown in **Table 3**, this gave the encryption and authentication macro blocks sufficient performance to achieve full-wire speed.

5. Reference system

In this section, we introduce a reference system for developing broadband VPN routers using the MB86977 and MB86978.

For the users' convenience, we developed a reference system board (MB86978DK001) containing the MB86977 and MB86978 LSIs. The system's block diagram is shown in **Figure 11**, and a photograph is shown in **Figure 12**. The host CPU is an ARM9TDMI processor, and the MB86977 WAN port and MB86978 LAN ports are directly connected via the MIIs. The MB86978 performs IPsec processing, and the MB86977 performs the other router processing functions. This configuration enables bi-directional, full-wire-speed WAN and LAN IPsec transmissions.

6. Evaluation results

In this section, we describe the evaluation results for throughput and jitter under high-load conditions, which will be important considerations when these LSIs are used in broadband VPN routers.

6.1 Throughput evaluation

We used Spirent Communications' SmartBits

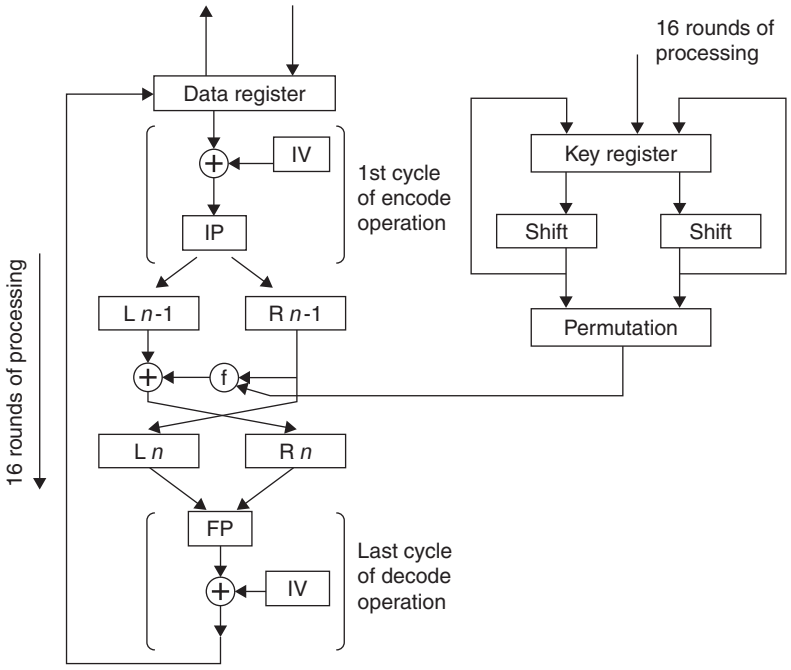


Figure 9
Conventional DES circuit.

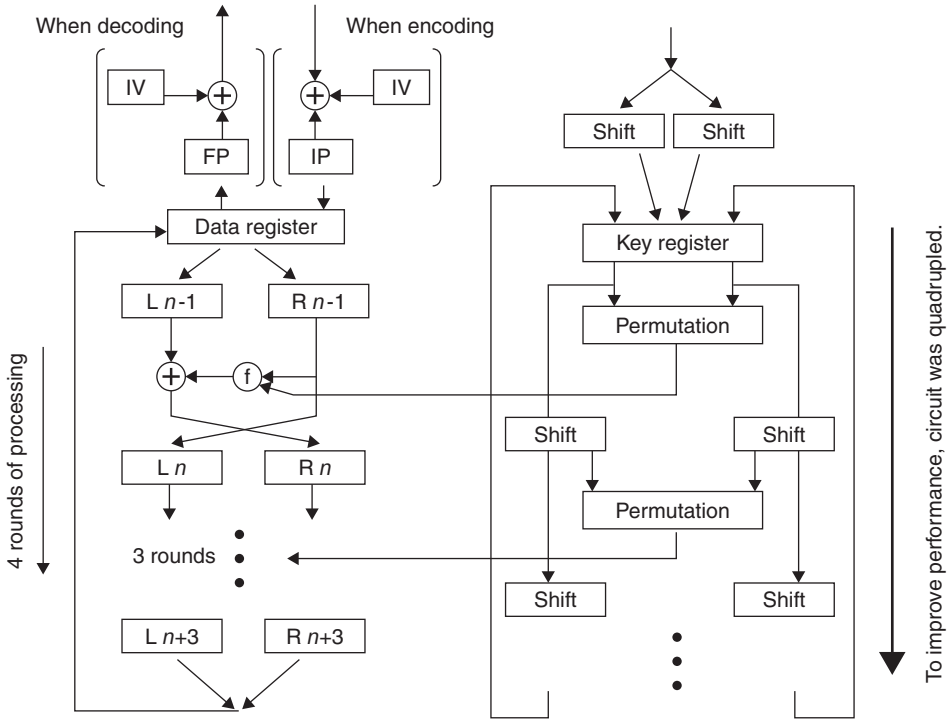


Figure 10
MB86978 DES circuit.

Table 3
Encryption and authentication macro block performance.

Macro name	DES	AES	SHA-1	MD5
Performance (Mb/s)	457	985	492	589

note: Values are for 100 MHz operation.

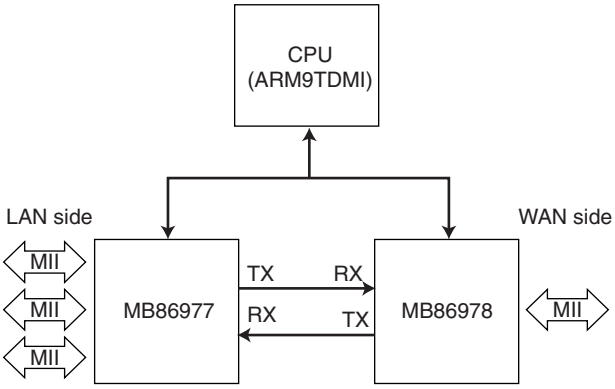


Figure 11
MB86978DK001 block diagram.

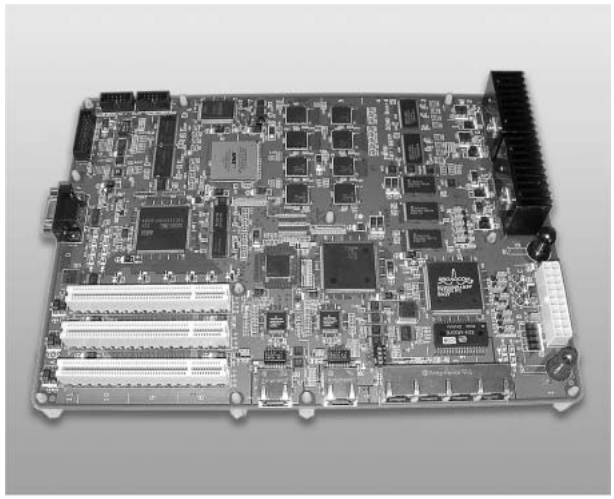


Figure 12
MB86978DK001.

network performance analysis system to measure the IP packet throughput performance. The MB86978DK001 board's theoretical throughput values and measured throughput results are shown in **Table 4**.

In the IPsec encryption direction, the theoretical values apply because an Authentication Header (AH), Encapsulating Security Payload (ESP) header, ESP trailer, and authentication data are appended to the packet processing results in accordance with the encryption and authentication modes. In the decryption direction, because these appended items are all deleted, the throughput is always 100%. The theoretical values in both directions indicate the wire-speed upper limits.

Figure 13 shows a throughput comparison with VPN equipment containing a look-aside architecture IPsec accelerator LSI that uses a PCI bus for packet transfer, encryption, and decryption. Systems using this type of LSI cannot achieve full-wire speed because they have an inadequate routing performance and PCI bus throughput.

From our measurements, the maximum packet delay is 250 μs for the MB86977 and 180 μs for the MB86978.

For IP networks that use Voice over Internet Protocol (VoIP) for IP telephony, a delay between the listener and caller that does not exceed 50 ms is said to be acceptable. Our measurements show the delays introduced by the MB86977 and MB86978 LSIs will be less than 1% of this 50 ms limit.

6.2 Jitter evaluation

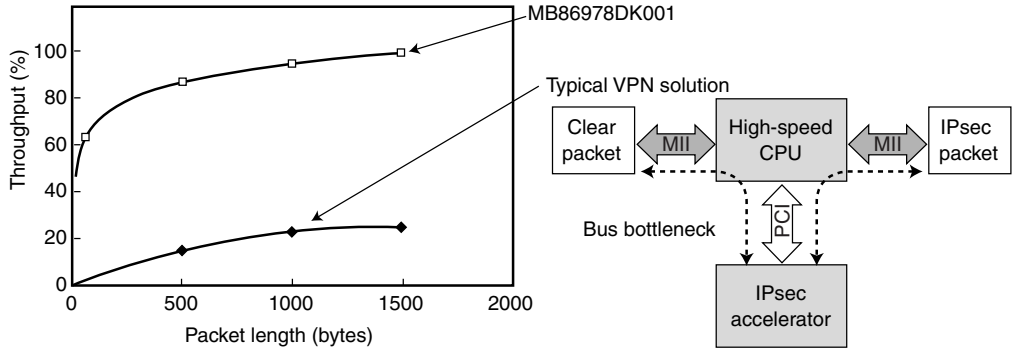
Jitter is a measure of the variability over time of latency across a network; in terms of voice quality, it is a more important parameter than throughput. We measured the jitter of VoIP packets that were received from the Internet by the MB86978DK001 and simultaneously received by a commodity broadband router. A comparison of the results is shown in **Figure 14**.

When the amount of downstream data increased, there was a significant increase in jitter in the commodity broadband router.

Table 4
Throughput evaluation.

Evaluation packet \ Encryption and authentication modes		Transport mode ESP with authentication AES HMAC-SHA-1		Tunnel mode ESP with authentication 3DES HMAC-SHA-1		Tunnel mode AH HMAC-SHA-1		Tunnel mode AH&ESP BIND AES HMAC-SHA-1	
		Encryption	Decryption	Encryption	Decryption	Encryption	Decryption	Encryption	Decryption
Short packet	Packet length (bytes)	64	106	64	114	64	108	80	150
	Theoretical throughput (%)	66.7	100	62.7	100	65.6	100	58.8	100
	Measured throughput (%)	66.7	100	62.7	100	62.7	100	58.8	100
Long packet	Packet length (bytes)	1474	1514	1464	1514	1474	1518	1450	1526
	Theoretical throughput (%)	97.4	100	96.7	100	97.1	100	95.1	100
	Measured throughput (%)	97.4	100	96.7	100	97.1	100	95.1	100

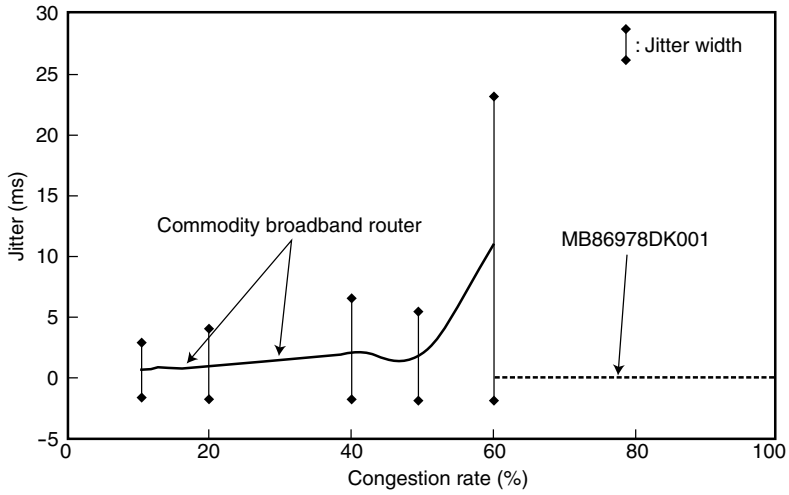
HMAC-SHA-1: Hash-based Message Authentication Code (HMAC) using the SHA-1 hash function.



(a) Throughput on tunnel mode DES encryption (with SHA-1 authentication)

(b) Configuration of typical VPN solution

Figure 13
Throughput comparison.



Measurement conditions: WAN 1504-byte-length packet → LAN1
WAN 94-byte-length packet (assuming voice transmission), 30 ms interval → LAN0

note: Commodity broadband router curve indicates average amount of jitter and jitter width.

Figure 14
Comparison of jitter at high congestion rates.

However, thanks to its Quality of Service (QoS) mechanism, the MB86978DK001's jitter remained in the microsecond region.

A VoIP terminal can generally compensate for jitter in the range of 10s of milliseconds, but jitter exceeding this amount results in lower voice quality.

Our evaluations show the MB86978DK001 can provide clear voice transmissions at much higher network loads than commodity routers.

As can be seen from these results, the MB86977 and MB86978 LSIs are highly superior solutions for VPN broadband routers used in home and SOHO environments.

7. Conclusion

In this paper, we introduced our MB86977 and MB86978 LSIs for high-speed network gateway functions. These LSIs are ideal for home gateway and broadband VPN router applications and have proved that bi-directional, full-wire IP packet processing — including IPsec and routing functions — can be performed over a 100 base-TX Ethernet link. It is expected that Gigabit Ethernet (GbE) connections will soon become mainstream in Internet connections and LAN environments. Accordingly, in the next development stage, we plan to apply the knowledge and design data we gained in developing the MB86977 and MB86978 LSIs so their architectures can be used in GbE applications.

Tomokazu Aoki, *Fujitsu LSI Solution Ltd.*
Mr. Aoki received the B.S. degree in Applied Physics and the M.S. degree in Energy Sciences from Tokyo Institute of Technology University, Tokyo, Japan in 1988 and 1990, respectively. He joined Fujitsu LSI Solution Ltd., Kawasaki, Japan in 2000, where he has been developing LSIs for network and security systems.

Teruhiko Nagatomo, *Fujitsu LSI Solution Ltd*.
Mr. Nagatomo received the B.S. and M.S. degrees in Welding Engineering from Osaka University, Japan in 1989 and 1991, respectively. He joined Fujitsu LSI Solution Ltd., Kawasaki, Japan in 2000, where he has been developing ICs for high-speed IP network systems.

Kazuya Asano, *Fujitsu LSI Solution Ltd*.
Mr. Asano received the B.S. and M.S. degrees in Applied Physics from the University of Tokyo, Tokyo, Japan in 1990 and 1992, respectively, and the M.S. degree in Electrical Engineering from UC Berkeley, California, USA in 1999. He joined Fujitsu LSI Solution Ltd., Kawasaki, Japan in 2000, where he has been designing LSI circuits for broadband communications.