



(12) **United States Patent**
Melvin

(10) **Patent No.:** **US 7,475,201 B1**
(45) **Date of Patent:** **Jan. 6, 2009**

(54) **PACKET PROCESSOR MEMORY**
INTERFACE WITH CONDITIONAL
DELAYED RESTART

(75) Inventor: **Stephen Waller Melvin**, San Francisco,
CA (US)

(73) Assignee: **Teplin Application Limited Liability**
Co., Wilmington, DE (US)

(*) Notice: Subject to any disclaimer, the term of this
patent is extended or adjusted under 35
U.S.C. 154(b) by 0 days.

(21) Appl. No.: **11/930,209**

(22) Filed: **Oct. 31, 2007**

Related U.S. Application Data

(63) Continuation of application No. 11/530,895, filed on
Sep. 11, 2006, which is a continuation of application
No. 11/306,375, filed on Dec. 26, 2005, now Pat. No.
7,107,402, which is a continuation of application No.
10/065,340, filed on Oct. 6, 2002, now Pat. No. 6,981,
110.

(60) Provisional application No. 60/347,781, filed on Oct.
23, 2001, now abandoned, provisional application No.
60/339,077, filed on Oct. 30, 2001, provisional appli-
cation No. 60/348,217, filed on Nov. 9, 2001, provi-
sional application No. 60/332,425, filed on Nov. 16,
2001.

(51) **Int. Cl.**
G06F 12/00 (2006.01)

(52) **U.S. Cl.** **711/154; 711/118; 711/158;**
711/221; 370/394; 709/216; 709/229; 712/218;
712/225

(58) **Field of Classification Search** **711/154,**
711/118, 158, 221; 370/394; 709/216, 229;
712/218, 225

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,193,188	A *	3/1993	Franaszek et al.	707/8
5,269,017	A	12/1993	Hayden et al.	
5,471,521	A	11/1995	Minakami	
5,655,096	A	8/1997	Branigin	
5,878,117	A	3/1999	Minakami	
5,905,998	A	5/1999	Ebrahim et al.	
5,931,957	A	8/1999	Konigsburg et al.	

(Continued)

OTHER PUBLICATIONS

S. Melvin, Y. Patt, "Handling of Packet Dependencies: A Critical
Issue for Highly Parallel Network Processors," International Confer-
ence on Compilers, Architectures, and Synthesis for Embedded Sys-
tems, Oct. 8-11, 2002, Grenoble, France.

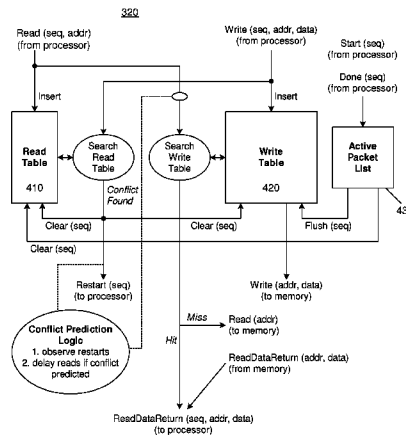
(Continued)

Primary Examiner—Stephen C Elmore
(74) *Attorney, Agent, or Firm*—McDonnell Boehnen Hulbert
& Berghoff LLP

(57) **ABSTRACT**

A mechanism receives memory reads and writes from a
packet processing engine, each memory access having an
associated packet identifier or sequence number. The mecha-
nism is placed between a processing element and a memory
system such that write data is buffered and information based
upon reads and writes is recorded. Upon receiving a memory
write, conflict detection logic determines if a conflict has
occurred and if packet processing for a packet needs to be
restarted. When such a conflict occurs, restart logic condi-
tionally delays the restart to the packet processing engine.
This can be accomplished using a stall signal or by delaying
the return of the first memory read after the restart. Such a
conditional delayed restart mechanism can optimize process-
ing based on the likelihood of multiple conflicts or a single
conflict.

19 Claims, 6 Drawing Sheets



U.S. PATENT DOCUMENTS

6,079,006	A	6/2000	Pickett	
6,389,016	B1	5/2002	Sabaa	
6,611,883	B1	8/2003	Avery	
6,665,708	B1*	12/2003	Tikekar et al.	709/215
6,665,755	B2	12/2003	Modelski	
6,738,379	B1	5/2004	Balazinski	
6,763,436	B2	7/2004	Gabber	
6,981,110	B1	12/2005	Melvin	
7,013,346	B1	3/2006	Tucker	
7,062,638	B2	6/2006	Yoaz	
7,089,404	B1	8/2006	Rozas	
7,107,402	B1	9/2006	Melvin	
2002/0073285	A1	6/2002	Butterworth	
2002/0112100	A1	8/2002	Zimmerman	
2002/0116587	A1	8/2002	Modelski	

OTHER PUBLICATIONS

M. Franklin, G. Sohi, "ARB: A hardware mechanism for dynamic reordering of memory references," IEEE Transactions on Computers, vol. 45, pp. 552-571, May 1996.

S. Gopal, T. N. Vijakumar, J. E. Smith, G. S. Sohi, "Speculative versioning cache," Proceedings of the Fourth International Symposium on High-Performance Computer Architecture (HPCA-4), Las Vegas, Feb. 1998.

G. Sohi, S. Breach, T. Vijaykumar, "Multiscalar processors," Proceedings of the 22nd Annual International Symposium on Computer Architecture, pp. 414-425, Ligure, Italy, Jun. 1995.

J. G. Steffan, T. Mowry, "The potential for using thread-level data speculation to facilitate automatic parallelization," Proceedings of

the Fourth International Symposium on High-Performance Computer Architecture (HPCA-4), Las Vegas, Feb. 1998.

L. Hammond, M. Willey, Kunle Olukotun, "Data Speculation Support for a Chip Multiprocessor," Proceedings of the Eighth International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS-VIII), San Jose, Oct. 1998.

J. Steffan, C. Colohan, Antonia Zhai, T. Mowry, "A Scalable Approach to Thread-Level Speculation," Proceedings of the 27th Annual International Symposium on Computer Architecture, Vancouver, Canada, Jun. 2000.

M. Cintra, J. Martinez, J. Torrellas, "Architectural Support for Scalable Speculative Parallelization in Shared-Memory Multiprocessors," Proceedings of the 27th Annual International Symposium on Computer Architecture, Vancouver, Canada, Jun. 2000.

J. Martinez, J. Torrellas, "Speculative Locks for Concurrent Execution of Critical Sections in Shared-Memory Multiprocessors," Workshop on Memory Performance Issues, International Symposium on Computer Architecture, Göteborg, Sweden, Jun. 2001.

R. Rajwar, J. Goodman, "Speculative Lock Elision: Enabling Highly Concurrent Multithreaded Execution," Proceedings of the 34th Annual International Symposium on Microarchitecture, Austin, Texas, Dec. 2001.

M. Herlihy, J.E.B. Moss, "Transactional Memory: Architectural Support for Lock-Free Data Structures," Proceedings of the International Conference on Computer Architecture, pp. 289-300, San Diego, California, May 1993.

R. S. Tomlinson, "Selecting Sequence Numbers," ACM Computer Communication Review, 25(1), pp. 45-53, Jan. 1995, originally published in Proceedings of ACM SIGCOMM/SIGOPS Interprocess Communications Workshop, Santa Monica, CA, Mar. 1975.

* cited by examiner

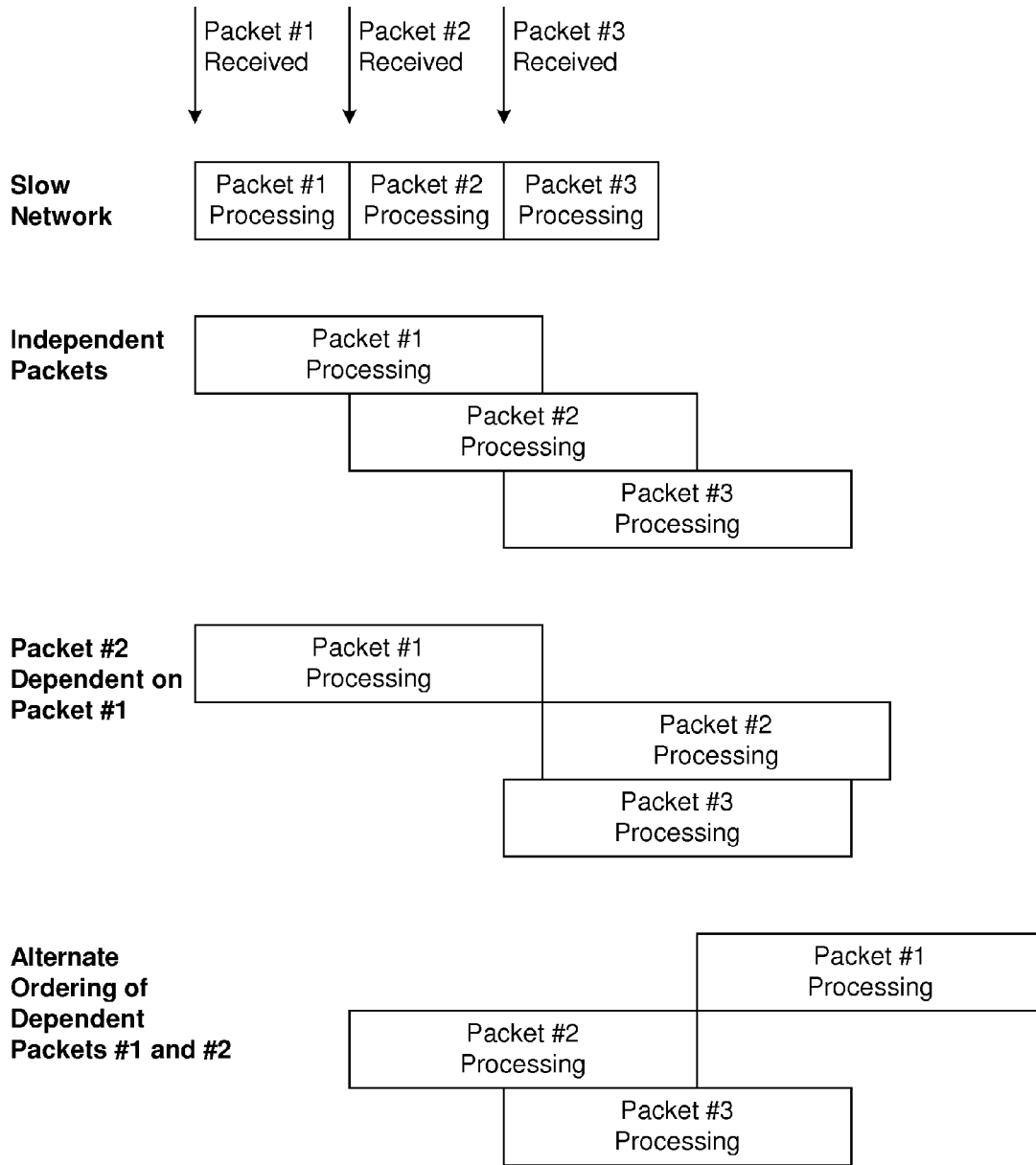


Fig. 1

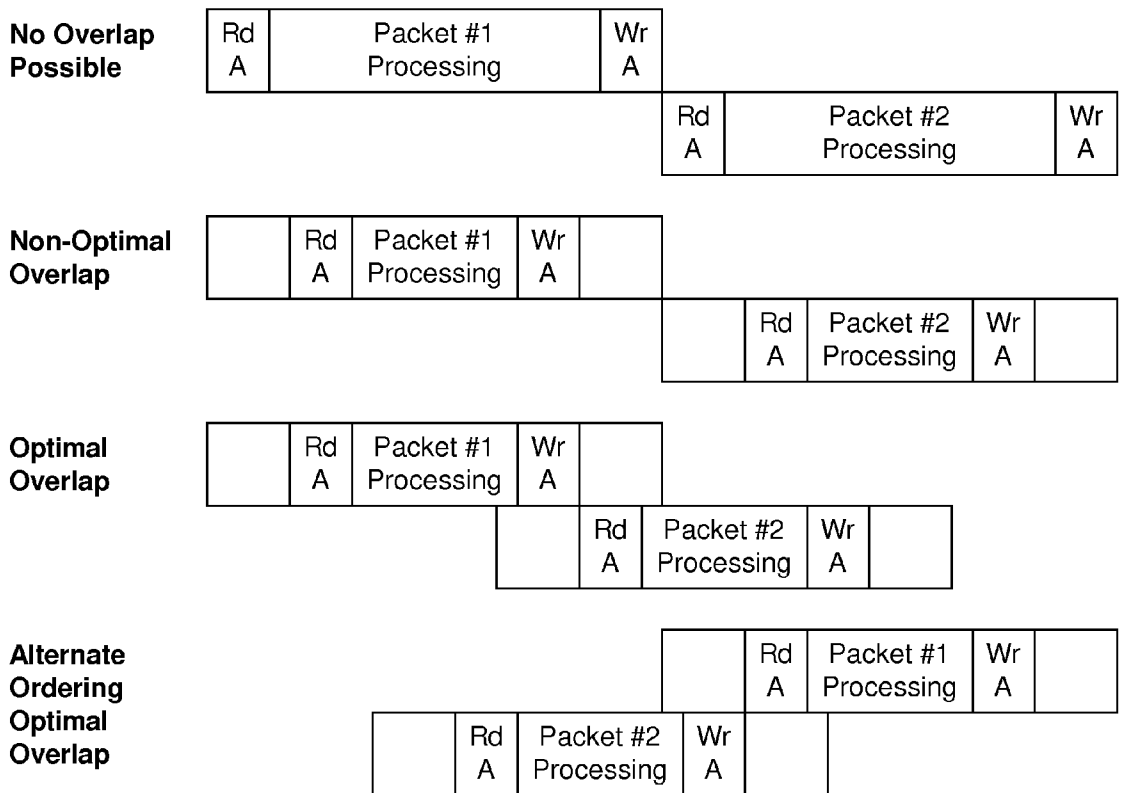


Fig. 2

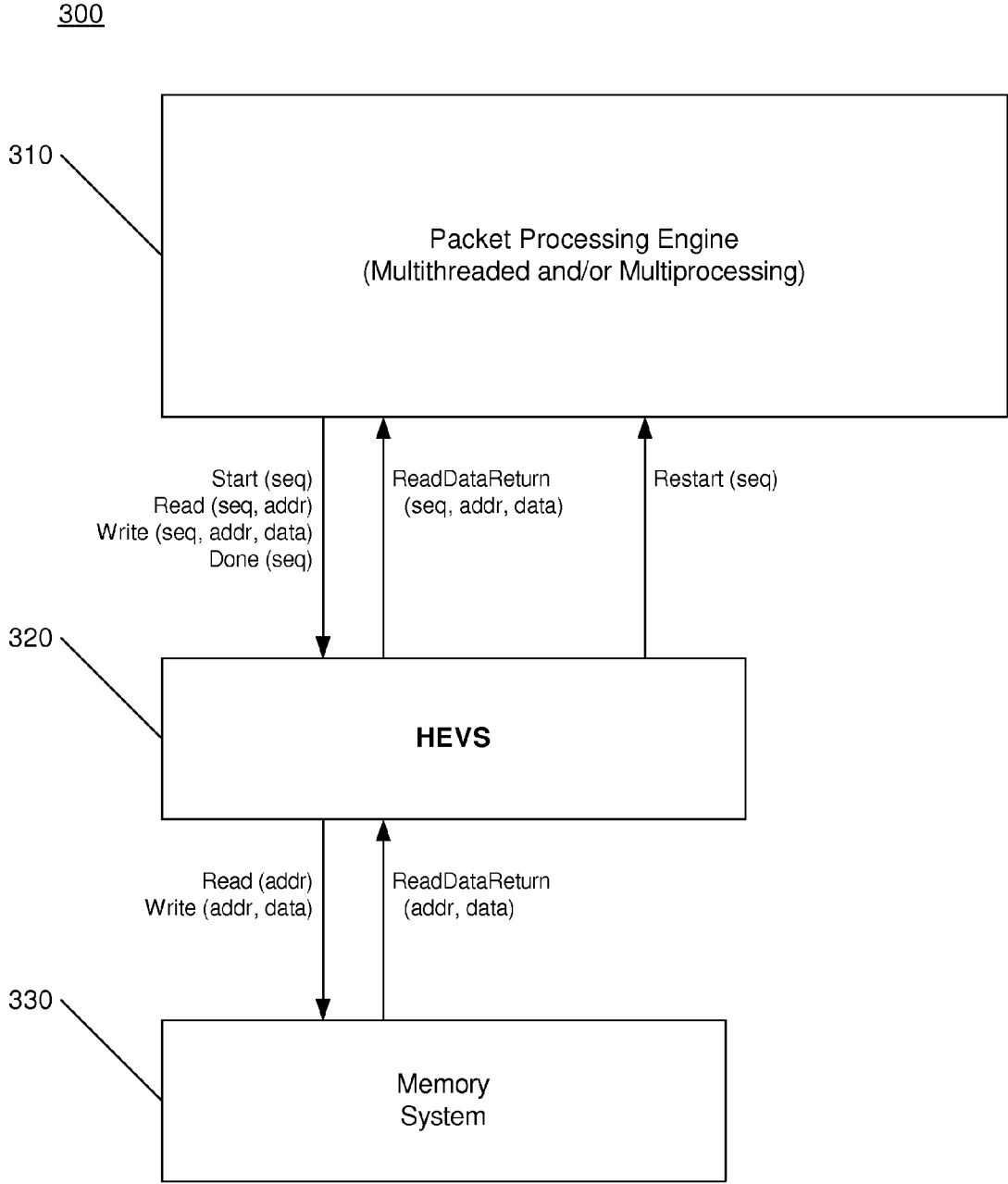


Fig. 3

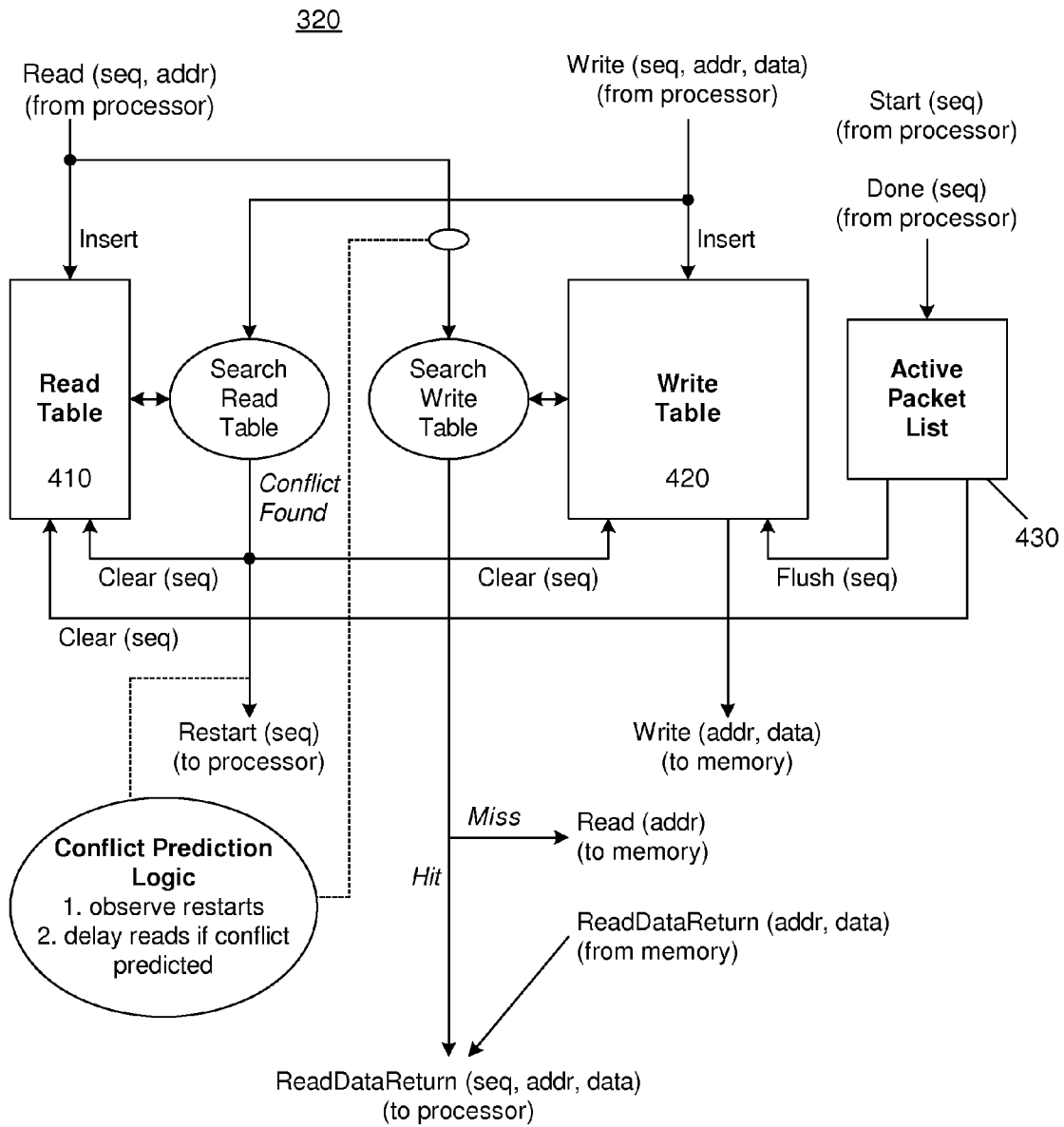
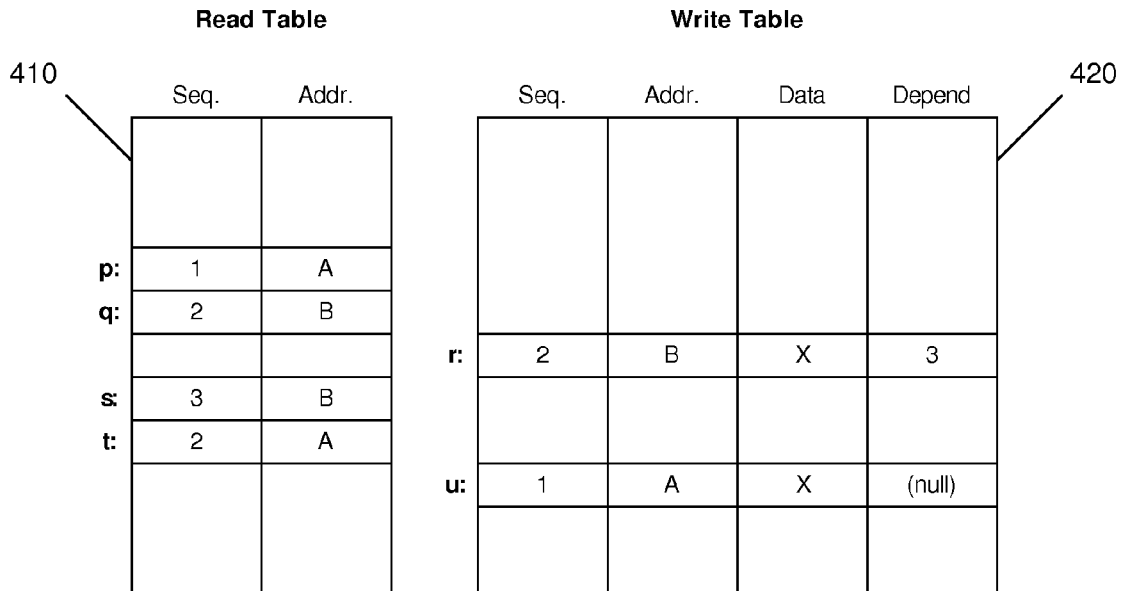


Fig. 4



Time Sequence:

1. Packet #1 reads location A
 Entry **p:** created in Read Table
 Write Table is searched, no matches found so memory read is performed
2. Packet #2 reads location B
 Entry **q:** created in Read Table
 Write Table is searched, no matches found so memory read is performed
3. Packet #2 writes location B
 Entry **r:** created in Write Table
 Read Table is searched, no conflicts found
4. Packet #3 reads location B
 Entry **s:** created in Read Table
 Write Table is searched, entry **r:** found, data X forwarded and dependency list updated
5. Packet #2 reads location A
 Entry **t:** created in Read Table
 Write Table is searched, no matches found so memory read is performed
6. Packet #1 writes location A
 Entry **u:** created in Write Table
 Read Table is searched for newer sequence read, entry **t:** is found
 Conflict is signaled to processor, Packet #2 is restarted
 Entry **q:** and all other sequence 2 entries are deleted
 Deletion of entry **r:** triggers Packet #3 restart signaled

Fig. 5

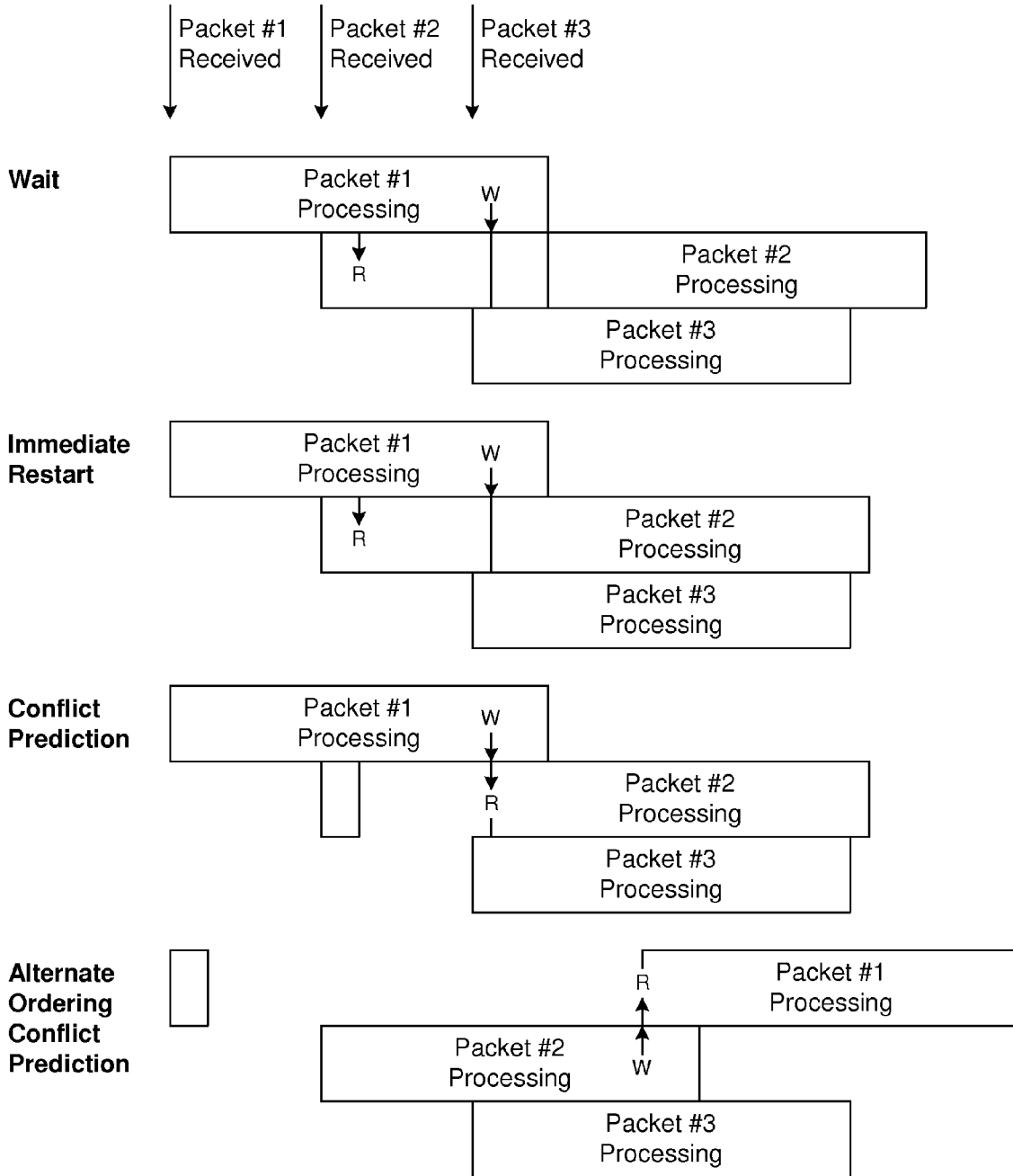


Fig. 6

1

**PACKET PROCESSOR MEMORY
INTERFACE WITH CONDITIONAL
DELAYED RESTART**

RELATED APPLICATIONS

This application is a continuation of application Ser. No. 11/530,895, filed on Sep. 11, 2006, which is a continuation of application Ser. No. 11/306,375, filed on Dec. 26, 2005, now U.S. Pat. No. 7,107,402, issued Sep. 12, 2006, which is a continuation of application Ser. No. 10/065,340, filed on Oct. 6, 2002, now U.S. Pat. No. 6,981,110, issued Dec. 27, 2005, which claims priority to application 60/347,781 filed on Oct. 23, 2001, application 60/339,077 filed on Oct. 30, 2001, application 60/348,217 filed on Nov. 9, 2001, and application 60/332,425 filed on Nov. 16, 2001. Each of these priority documents is incorporated herein by reference.

FIELD OF THE INVENTION

The present invention relates to the field of central processing unit (CPU) architecture and microarchitecture, and pertains more particularly to CPUs for networking applications that process multiple packets simultaneously.

BACKGROUND

Internet networking hardware involves processing of packets of information for many purposes and at many stages in a network. Routers, firewalls, gateways, load balancers and servers all process packets of information in some way. Where in the network the processing takes place (i.e. in the core or close to the edge) has a great deal to do with what types of processing needs to take place and how fast that processing must occur. In general, processing closer to the core takes place faster and involves less work. For example, many core routers perform only layer 2 packet forwarding (i.e. link layer header modification), which can be done with minimal processing overhead. Edge routers, however, typically perform more functions, such as traffic shaping, monitoring, billing and quality of service enforcement. In both situations, the need for processing is constantly evolving, and there is an increasing need to do more at faster rates.

Two key trends are the increase in network speed and the increase in the amount of processing that needs to take place at each stage in the network. Together these trends are forcing packet processing solutions into greater degrees of parallelism. FIG. 1 illustrates this point with four different scenarios for a packet processor. Here the term "packet processor" is used to generally refer to any processing engine that can perform programmable operations on packets of information.

In the first scenario of FIG. 1, the processing time of the packet is the same or smaller than the transmission time of the packet. In this scenario, the code need not be concerned with dependencies between packets, and ordinary single-threaded non-parallel processors can be used. In the other scenarios of FIG. 1, the processing time for a packet is substantially longer than the transmission time of one packet of information. The common trend is that the need for more complex operations (and thus larger workloads) and/or the increase in network speeds has lead to these situations.

In many cases the workload time is dominated by memory latency due to poor locality of data references and large working set sizes. This means that the limitation on packet throughput is driven by memory throughput, which has tended to increase at a rate even slower than single-threaded

2

processor performance, further driving packet processing solutions into parallel packet processing scenarios.

In the case that all packets can be operated on independently, as shown in the second scenario of FIG. 1, processing can be pipelined neatly and no conflict arises between code processing simultaneous packets. This would be the case in certain types of stateless firewalls and forwarding engines, where each packet is evaluated according to static rules and does not depend on any other packets. Thus, no state is changed by packet that affects a future packet. The forwarding tables and firewall rules might be dynamically modified, but this typically happens on a time scale orders of magnitude greater than the time to process a single packet. A parallel packet processing solution for this second scenario is relatively easy to implement. The code working on one packet need not be aware of other packets and there is no need to synchronize memory operations between packets.

In the more general case that dependencies can arise between packets, a more complicated situation exists. This is shown in the third and fourth scenarios of FIG. 1. This would be the case if both packets are from the same TCP connection and due to, for example, encryption or TCP state maintenance, there is a need to update state in memory between the processing of the two packets. One or more memory locations written by one packet will be read by the other packet. Note that packet #3 in these scenarios is independent from both packets and can be processed as soon as it arrives.

Other examples in which packet dependencies can arise would be the updating of traffic management counters and the updating of routing or address translation tables. In the latter case, two packets may be dependent even if they are from completely independent connections if they hash to the same table entry. One packet may want to modify a table entry while another packet is querying the same entry. The fourth scenario in FIG. 1 illustrates that in some, if not most cases it does not matter which order two dependent packets are processed, as long as they are serialized to prevent incorrect results.

In these cases where simultaneous processing of packets is required, and where dependencies can exist between packets, it can be complicated to enforce those dependencies. Currently, there are two common approaches to this problem. The first solution is a software solution, where software locks are included in the code to cause dependent packet processing to be delayed until an earlier packet has been completed. These software semaphores are used to lock out subsequent dependent packets from accessing state until the first packet has updated it. The second solution involves hardware, where packet classification hardware serializes all packets that can possibly be dependent. In a multiprocessor, this can involve generating a hash function that sends all packets of the same flow to the same processor, and distributes the load across multiple processors.

Unfortunately, packet processing code is often large and complex and modifying it to incorporate new locking mechanisms is not trivial. Even when such code is relatively simple, guaranteeing that software locks have been correctly programmed for all possible network traffic scenarios can be hard to verify. Furthermore, requiring hardware to enforce sequentiality when it is not needed lowers performance. This is because often the entire packet processing is not dependent such that a partial overlap is possible. The importance of a partial overlap of packet workload can be appreciated by referring to FIG. 2. In the case that a packet reads data as its first instruction and writes that same address as its last instruction, indeed there can be no overlap of processing. This is generally not the case however. The second scenario of FIG.

2 illustrates the case that the second packet can start before the first packet is completed, even though they are dependent. It is also the case that due to conditional branches, packets that are sometimes dependent may not always be dependent. Thus conservative locking and large grained locking can yield significantly sub-optimal solutions.

It is also the case that hardware solutions that group flows for multiprocessors suffer from the problem of guaranteeing that the grouping is relatively uniform over time in order to balance work across the multiple processing elements. The classification of packets to direct them to processing elements is constrained by having to preserve correctness and can't take advantage of a more dynamic load balancing approach.

What is needed is a hardware mechanism to preserve packet dependencies without requiring changes to software and allowing optimal enforcement of dependencies, such that packets are not serialized unless necessary by the overlying application.

BRIEF SUMMARY OF THE INVENTION

The present invention is a hardware mechanism for enforcing packet dependencies without any changes in software. This mechanism has the important advantage that packet dependencies are enforced by hardware only when required and potentially with no discarded work. The software can be written with the view that each packet is handled in order of arrival with no overlap with subsequent packets. This mechanism optimizes performance for the common case that no dependencies are present and doesn't require the hardware to be pre-configured with knowledge of what packet sequences will and will not be dependent.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 illustrates four different scenarios for how the processing of three packets can be overlapped.

FIG. 2 illustrates four different scenarios for the overlap of the processing of two packets in which a dependency exists.

FIG. 3 is a block diagram showing how the present invention fits within a packet processor.

FIG. 4 is a block diagram illustrating an embodiment of the present invention.

FIG. 5 illustrates detail of the read table and write table of an embodiment of the present invention and illustrates how the tables respond to a sequence of packet reads and writes.

FIG. 6 illustrates four different scenarios for the processing and restart of two packets in conjunction with an embodiment of the present invention.

DETAILED DESCRIPTION

The present invention is incorporated into a packet processor in which multiple packets are processed simultaneously. This can be implemented using a multiprocessor, a multi-threaded processor or a combination of both. The mechanism is not dependent on the type of processor used. It would be useful in any form of packet processor in which some form of simultaneity exists such that more than one packet at a time is being processed.

There are a few requirements placed on the packet processor. The first is that a time stamp or sequence number is associated with each packet as it arrives. The sequence number is used to enforce the sequential processing model. When a packet enters the processor, a new sequence number is generated and that sequence number is associated with the instructions that are processing that packet. The packet pro-

cessor is required to include this sequence number with all memory operations (reads and writes) performed. An additional requirement is that when a packet enters the processor and when a packet leaves the processor (or when processing ends for that packet), a notification of that event is made to the hardware. The information about when a packet enters and leaves is used to know when write data can be committed to memory.

Finally, the packet processor must have the ability to restart processing for each packet. To accomplish this a signal is generated that indicates that a conflict has been detected, which requires the packet processor to discard all work associated with the indicated sequence number and to restart processing that packet. The processor does not have to reconstruct any information, but only to start from the beginning of the packet again assuming that all memory writes have been discarded. It must be possible to restart a packet even after packet processing has ended for that packet. This is needed because an older packet that is still processing may do something that nullifies the processing that has already been completed by a younger dependent packet.

The present invention is responsible for guaranteeing that the results obtained while simultaneously processing multiple packets are the same as would be obtained if all packet processing occurred in the sequence of packet arrival. This hardware mechanism is placed between the packet processing engine and the memory system as shown in FIG. 3. Packet processing engine 310 is a parallel processing engine, incorporating multithreading and/or multiprocessing. The present invention (HEVS) 320 is situated between packet processing engine 310 and the memory system 330.

The packet processing engine 310 generates four types of commands to the HEVS mechanism 320: memory read, memory write, packet start and packet done. Each command includes a sequence number associated with the packet for which the operation applies. Returning to the packet processor are the read data returning, and a restart signal that is used to indicate to the packet processor that a particular packet needs to be restarted.

FIG. 4 illustrates a diagram of an embodiment of the present invention. There are two tables, a Read Table 410 and a Write Table 420 at the core of the design. The Read Table records all memory reads. The sequence number and the address are recorded, but not the data. Each time a read is accepted from the packet processor, an entry is inserted in the Read Table. To save space in the read table, it may be constructed to not save all address bits, but to discard some of the least significant address bits. If an entry being inserted matches an entry already in the table, it need not be inserted. The Read Table is searched whenever a write is accepted from the packet processor to determine if a conflict has arisen.

The Write Table 420 records all writes, including the sequence number, the address, the data and dependency information for forwarded reads. Each time a write is accepted from the packet processor, an entry is made in the Write Table. To save space in the Write Table, an entry can include a larger amount of storage than a single reference can generate. For example a Write Table entry could contain 64 bytes of data with valid bits associated with each byte. This would allow multiple writes to the same 64-byte area with the same sequence number to be combined into the same entry. Two different accesses with different sequence numbers cannot be combined. The Write Table is searched whenever a read is accepted to determine if data should be bypassed from the memory system. A record is kept of the fact that data has been

5

forwarded from one packet to another so that when a packet is restarted; all packets potentially corrupted by that packet are also restarted.

When a sequence number is completed, the active packet list **430** is updated and if the oldest packet active is no longer being processed, data from the Write Table is flushed to memory and all entries for that sequence number are cleared from the Read Table. Note that the packet start command is optional in that it could be inferred from the first read or write received with a new sequence number.

In summary, the operations that are performed for each type of command from the packet processor are as follows:

Read

1. Insert entry into Read Table including sequence number and address.
2. Search Write Table for the newest matching entry that has the same or an older sequence number. If one exists, forward that data back to the network processor and record dependency information. If none exists, send this request to the memory system.

Write

1. Insert entry into Write Table including sequence number, address and data
2. Search Read Table for any entry that is newer than this sequence number and matches the address. If one is found, then:
 - a. Signal restart to packet processor of the sequence number associated with the match
 - b. Delete all Read Table entries with that sequence number
 - c. Delete all Write Table entries with that sequence number

Repeat steps a, b and c. if there are multiple matching sequence numbers, and continue down all dependency chains

Start

1. Add entry to active packet list

Done

1. Mark packet as done in active packet list
2. If packet is the oldest done
 - a. Delete all Read Table entries for that sequence number
 - b. Flush all Write Table entries for that sequence number to memory
 - c. Delete all Write Table entries for that sequence number
 - d. Delete the packet from the active packet list

Repeat steps a. through d. if the next oldest packet is done

FIG. 5 illustrates a sequence of reads and writes further illustrating the operation of the read table **410** and the write table **420**. This sequence involves three packets performing a sequence of reads and writes. In this sequence, the final write to address A by packet #1 triggers a restart of both packets #2 and #3. This is due to the fact that data from packet #2 has been forwarded to packet #3. Since packet #2 has been restarted, packet #3 needs to be restarted as well.

However, it should be noted in this example that the read of location A by packet #2, which is the one that triggered the conflict, occurred after the write to location B by packet #2. This means that if packet #2 were able to be backed up and restarted between these two events, rather than at the beginning of processing, then packet #3 would not need to be restarted. This performance enhancement allows the amount of discarded work to be limited. It would involve a checkpoint-backup mechanism within the packet processor in conjunction with the use of sub-sequence numbers. The proces-

6

sor would then be able to restart processing from a safe checkpoint rather than having to back up to the beginning of the packet processing.

Another possible performance enhancement would be to save the data that is returned for each read in the Read Table. This would allow the conflict detection logic to determine if in fact wrong data had actually been returned, and only signal a restart if a data mismatch occurred. This would be advantageous if the packet processing code commonly overwrites data with the same value as is already stored, which may happen in certain applications. In addition, carefully written software could take advantage of a packet processor with this enhancement by purposefully writing an expected value into a variable early in its processing. In the common case, when the final value is written it will be the same and subsequent packets will not need to be restarted.

In many cases, it is not necessary to guarantee that processing yields the same results as if packets had been processed in the specific order of arrival, as long as the results are correct for any specific order. In other words, the packet processing code does not care which packet is processed first as long as there is a consistent order. Packets may in principle be reordered by the network in many cases and higher-level protocols cannot depend on any specific order. This principle has been illustrated in the last scenarios of FIGS. 1 and 2. (However, it is important to note that there are sometimes performance implications to reordering packets if code is written to optimize for cases where dependent packets are processed in order.)

An alternative mode of operation for the packet dependency mechanism described above is to not enforce a specific order, but to signal a conflict only when no consistent order exists. In this mode the packet sequence number is being used as merely a packet identifier rather than a numeric order. The mechanism is modified so that when a write occurs, a packet sequence is defined for any previous read. Then when two packets resolve to two conflicting sequences, one must be restarted.

The present invention can be used to enforce the correctness of a sequential packet processing model in a processor that processes packets simultaneously. Ideally, there should be a minimal amount of discarded work. (The importance of minimizing discarded work is particularly relevant in multi-threaded processors, where multiple threads all contend for the same execution resources.) In cases where conflicts are common, more careful optimization of the restart mechanism should be implemented. FIG. 6 illustrates four different scenarios for handling packet conflicts.

In this figure the arrow and the R represents a read in packet #2 for which a conflict arises, and the arrow and the W indicates a write in packet #1 that is the source of the conflict. In the first scenario, as soon as the conflict is detected, processing is stalled until the packet that generated the conflict has completed. This can be accomplished by having an additional stall signal to the packet processor, or by having the mechanism delay the return of the first read operation received after processing resumes for packet #2.

This first scenario is a conservative approach that essentially predicts that if a first conflict between two packets arises, then other conflicts between those two packets are likely to arise and therefore no more work should be expended on the second packet until the first packet is done.

In the second scenario of FIG. 6, processing is immediately restarted on the second packet after work is discarded and the Read Table and Write Table are flushed. This approach essentially predicts that there is likely to be only one conflict between two packets, so the second packet should not be

7

further delayed and processing should continue immediately. If a second conflict arises then the process is repeated. This may tie up processing resources within the packet processor that could be used for performing other tasks. The performance of the first two approaches shown in FIG. 6 is dependent on the traffic characteristics and on the characteristics of the packet workload.

Many more sophisticated mechanisms are possible. A conflict predictor could be implemented that would observe sequences of packet conflicts and develop a history to guess when memory operations would conflict. The information used in the predictor could be information from the packet header of each packet (e.g. source and destination IP number, and source and destination port number), as well as the memory addresses of the read and write involved. Ideally, a hardware conflict predictor would know when a read operation is received that it should not be allowed to complete since a write from an earlier packet is likely in the future. The third scenario of FIG. 6 illustrates the ideal situation when the hardware predictor is used to stall the memory read until the write from packet #1 occurs. The memory read is then completed and processing continues. In this case, there is no discarded work, and packet #2 is delayed the minimum time necessary to guarantee correct results.

Note that in the third scenario the second packet completes in the earliest time and the least machine resources are used. The advantage of a dynamic hardware conflict predictor is that it can adapt to changing application conditions and only apply stalls to read operations when there is a high likelihood of a conflict. The fourth scenario of FIG. 6 illustrates the opposite ordering of packets #1 and #2 in the case that the read and write occur at the same point. It may be acceptable for the hardware to guarantee either ordering rather than adhere to a strict order of arrival model.

The above description and teachings are related to various implementations of the present invention, and it should be clear to the skilled artisan that there are many alterations and amendments that might be made to the examples provided, without departing from the spirit and scope of the invention.

The invention claimed is:

1. An apparatus for interfacing a packet processing engine with a memory system comprising:

a first data storage element storing information based upon a memory read comprising a first packet identifier;

a second data storage element storing information based upon a memory write comprising a second packet identifier;

write processing and conflict detection logic coupled to said first data storage element and said second data storage element, wherein said second data storage element is modified based on said memory write and said first data storage element is searched based on said memory write, and wherein a determination is made whether a memory conflict exists between said memory read and said memory write; and

logic for signaling a restart coupled to said write processing and conflict detection logic and utilizing said first associated packet identifier, wherein the restart of packet processing is conditionally delayed after a restart is signaled.

2. The apparatus of claim 1 wherein said logic for signaling a restart asserts a stall signal for said first packet identifier if said restart of packet processing is to be delayed.

8

3. The apparatus of claim 1 wherein said logic for signaling a restart delays return of memory reads comprising said first packet identifier if said restart of packet processing is to be delayed.

4. The apparatus of claim 1 wherein the condition of delayed restart is based on traffic characteristics and packet workload.

5. The apparatus of claim 1 further comprising:
read processing logic coupled to said first data storage element and said second data storage element, wherein said first data storage element is modified based on said memory read and said second data storage element is searched based on said memory read.

6. The apparatus of claim 1 wherein:

said first data storage element comprises a plurality of entries each comprising a packet identifier and an address;

said second data storage element comprises a plurality of entries each comprising a packet identifier, write data and an address; and

wherein said memory write comprises an address and memory write data; and

wherein said write processing and conflict detection logic stores said address, said second packet identifier and said memory write data into said second data storage element and searches said first data storage element for an entry with an address matching said address and a packet identifier greater than said second packet identifier.

7. The apparatus of claim 1 wherein said first packet identifier and said second packet identifier are sequence numbers that indicate the order of arrival of packets at said packet processing engine.

8. The apparatus of claim 1 wherein said packet processing engine is a multithreaded processor that processes multiple packets in parallel.

9. The apparatus of claim 1 wherein said packet processing engine is a multiprocessor that processes multiple packets in parallel.

10. A method of interfacing a packet processor with a memory system comprising the steps of:

receiving a memory read comprising a first packet identifier from a packet processing engine;

storing information based upon said memory read into a first data storage element;

returning speculative read data to the packet processing engine responsive to said memory read;

receiving a memory write comprising a second packet identifier from said packet processing engine;

storing information based on said memory write into said second data storage element;

searching said first data storage element using information based on said memory write;

determining if said speculative read data is incorrect;

restarting processing of said packet associated with said speculative read data if said speculative read data is incorrect, wherein if said speculative data is incorrect, the restart of packet processing is conditionally delayed.

11. The method of claim 10 wherein the step of returning speculative read data comprises:

searching a second data storage element using information based on said memory read;

wherein if said step of searching detects a match, retrieving data from said second storage element; and if said step of searching does not detect a match, retrieving data from said memory system.

9

- 12. The method of claim 10 wherein:
 said memory read comprises a first memory address; and
 the step of storing information comprises storing at least a
 portion of said first memory address and said first packet
 identifier in a read table; and
 the step of searching comprises searching a write table for
 an entry with an address matching at least a portion of
 said first memory address and a packet identifier equal to
 or less than said first packet identifier.
- 13. The method of claim 10 wherein said step of restarting
 comprises asserting a stall signal for said first packet identifier
 if said packet restart is to be delayed.
- 14. The method of claim 10 wherein said step of restarting
 comprising delaying the return of memory reads comprising
 said first packet identifier if said packet restart is to be
 delayed.
- 15. The method of claim 10 wherein the condition of
 delayed restart is based on traffic characteristics and packet
 workload.
- 16. The method of claim 10 wherein
 said memory write comprises a second memory address
 and memory write data;

10

- said step of storing information based on said memory
 write comprises storing said second memory address,
 said second packet identifier and said memory write data
 in a write table; and
- 5 said step of searching said first data storage comprises
 searching a read table for an entry with an address
 matching at least a portion of said second memory
 address and a packet identifier greater than said second
 packet identifier.
- 10 17. The method of claim 10 wherein said first packet iden-
 tifier and said second packet identifier are sequence num-
 bers that indicate the order of arrival of packets at said packet
 processing engine.
- 15 18. The method of claim 10 wherein said packet processing
 engine is a multithreaded processor that processes multiple
 packets in parallel.
- 20 19. The method of claim 10 wherein said packet processing
 engine is a multiprocessor that processes multiple packets in
 parallel.

* * * * *