



(51) International Patent Classification:  
**H04L 12/861** (2013.01)

(21) International Application Number:  
PCT/CN2014/076020

(22) International Filing Date:  
23 April 2014 (23.04.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
61/816,261 26 April 2013 (26.04.2013) US  
14/252,722 14 April 2014 (14.04.2014) US

(71) Applicant: **MEDIATEK INC.** [CN/CN]; No. 1 Dusing Road 1st, Science-Based Industrial Park, Hsin-Chu, Taiwan (CN).

(72) Inventors: **YU, Yi-Hsin**; 12F., No.157, Sec. 1, Jiafong 2nd St., Jhubei City, Hsinchu County, Taiwan (CN). **CHEN, Yu-Hsun**; 6F., No. 223, Sec. 1, Liujia 1st Rd., Zhubei City, Hsinchu County, Taiwan (CN). **MA, Chang-Po**; 7F., No. 11, Ln. 42, Xinguang Rd., East Dist., Hsinchu City, Taiwan (CN). **LIN, Jui-Tse**; 4F., No. 2, Ln. 378, Sec. 2, Zhongxing Rd., Zhudong Township, Hsinchu County, Taiwan (CN). **LIU, Ta Hsing**; 1004 Chula Vista Terrace, Sunnyvale, California (US).

(74) Agent: **BEIJING SANYOU INTELLECTUAL PROPERTY AGENCY LTD.**; 16th Fl., Block A, Corporate Square, No. 35 Jinrong Street, Beijing 100033 (CN).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

**Declarations under Rule 4.17:**

— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))

**Published:**

— with international search report (Art. 21(3))

(54) Title: SCHEDULER FOR DECIDING FINAL OUTPUT QUEUE BY SELECTING ONE OF MULTIPLE CANDIDATE OUTPUT QUEUES AND RELATED METHOD

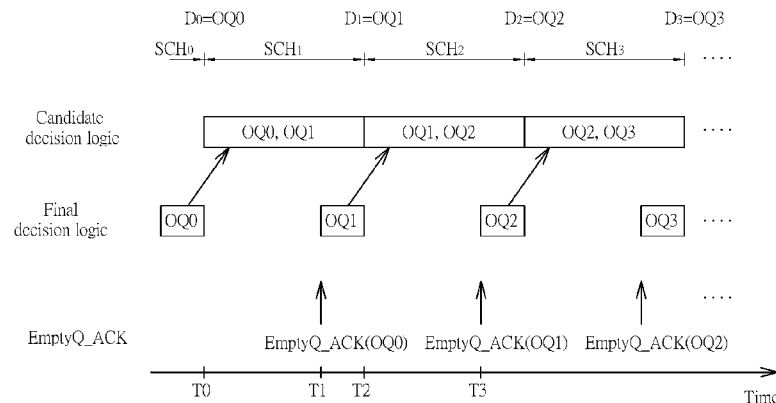
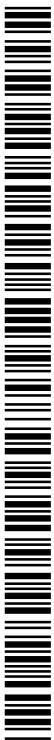


FIG. 2

(57) Abstract: A scheduler performs a plurality of scheduler operations each scheduling an output queue selected from a plurality of output queues associated with an egress port. The scheduler includes a candidate decision logic and a final decision logic. The candidate decision logic is arranged to decide a plurality of candidate output queues for a current scheduler operation, regardless of a resultant status of packet transmission of at least one scheduled output queue decided by at least one previous scheduler operation. The final decision logic is arranged to select one of the candidate output queues as a scheduled output queue decided by the current scheduler operation after obtaining the resultant status of packet transmission of the at least one scheduled output queue decided by the at least one previous scheduler operation.



**SCHEDULER FOR DECIDING FINAL OUTPUT QUEUE BY  
SELECTING ONE OF MULTIPLE CANDIDATE OUTPUT  
QUEUES AND RELATED METHOD**

**CROSS REFERENCE TO RELATED APPLICATIONS**

- 5 [0001] This application claims the benefit of U.S. provisional application No. 61/816,261, filed on 04/26/2013 and U.S. provisional application No. 14/252,722, filed on 04/14/2014, and incorporated herein by reference.

**BACKGROUND**

- 10 [0002] The disclosed embodiments of the present invention relate to forwarding packets, and more particularly, to a scheduler and related method for deciding a final output queue by selecting one of multiple candidate output queues.

- [0003] A network switch is a computer networking device that links different electronic devices. For example, the network switch receives an incoming packet  
15 generated from a source electronic device connected to it, and transmits an outgoing packet derived from the received packet to one or more destination electronic devices for which the received packet is meant to be received. In general, the network switch has a packet buffer for buffering packet data of packets received from ingress ports, and forwards the packets stored in the packet buffer through  
20 egress ports.

- [0004] Concerning ingress packets to be forwarded through the same egress port, the ingress packets come from source electronic device connected to different ingress ports. Hence, the network switch may create a plurality of output queues for ingress packets received by different ingress ports, respectively. For example, each  
25 of the output queues corresponding to the same egress port may be simply built by storing packet identifiers (packet IDs) of packets to thereby record a packet linked list indicative of an output sequence of the packets actually stored in the packet buffer. Since there are multiple output queues, a scheduler is therefore needed to perform a plurality of scheduler operations each used for making one output queue

decision that indicates which output queue is granted to output one packet to the egress port.

[0005] In general, the processing time required by a single scheduler operation should be shorter than a minimum packet transmission time (e.g., a transmission  
5 time of a 64-byte packet) to achieve the desired line rate. When the line rate becomes higher, the minimum packet transmission time is shorter accordingly. For example, when the network switch is used in a 10 Gigabit Ethernet (10GbE) environment, the minimum packet transmission time may be 67.2 ns (nanosecond);  
10 when the network switch is used in a 40 Gigabit Ethernet (40GbE) environment, the minimum packet transmission time may be 16.8 ns; and when the network switch is used in a 100 Gigabit Ethernet (100GbE) environment, the minimum packet transmission time may be 6.72 ns. Hence, regarding a high-speed network application, how to design a network switch meeting the strict scheduler timing constraint is a challenge to designers in the pertinent field.

15

## SUMMARY

[0006] In accordance with exemplary embodiments of the present invention, a scheduler for deciding a final output queue by selecting one of multiple candidate output queues and a related method are proposed to solve the above-mentioned  
20 problem.

[0007] According to a first aspect of the present invention, an exemplary scheduler for performing a plurality of scheduler operations each scheduling an output queue selected from a plurality of output queues associated with an egress port is disclosed. The exemplary scheduler includes a candidate decision logic and a final decision  
25 logic. The candidate decision logic is arranged to decide a plurality of candidate output queues for a current scheduler operation, regardless of a resultant status of packet transmission of at least one scheduled output queue decided by at least one previous scheduler operation. The final decision logic is arranged to select one of the candidate output queues as a scheduled output queue decided by the current  
30 scheduler operation after obtaining the resultant status of packet transmission of the at least one scheduled output queue decided by the at least one previous scheduler operation.

[0008] According to a second aspect of the present invention, an exemplary method for performing a plurality of scheduler operations each scheduling an output queue selected from a plurality of output queues associated with an egress port is disclosed.

5 The exemplary method includes: deciding a plurality of candidate output queues for a current scheduler operation, regardless of a resultant status of packet transmission of at least one scheduled output queue decided by at least one previous scheduler operation; and after obtaining the resultant status of packet transmission of the at least one scheduled output queue decided by the at least one previous scheduler operation, selecting one of the candidate output queues as a scheduled output queue  
10 decided by the current scheduler operation.

[0009] These and other objectives of the present invention will no doubt become obvious to those of ordinary skill in the art after reading the following detailed description of the preferred embodiment that is illustrated in the various figures and drawings.

15

## BRIEF DESCRIPTION OF DRAWINGS

[0010] FIG. 1 is a diagram illustrating a network switch according to an embodiment of the present invention.

20 [0011] FIG. 2 is a timing diagram illustrating a proposed solution for releasing the response time constraint according to an embodiment of the present invention.

[0012] FIG. 3 is a flowchart illustrating a method for scheduling an output queue according to an embodiment of the present invention.

## DETAILED DESCRIPTION

25 [0013] Certain terms are used throughout the description and following claims to refer to particular components. As one skilled in the art will appreciate, manufacturers may refer to a component by different names. This document does not intend to distinguish between components that differ in name but not function. In the following description and in the claims, the terms "include" and "comprise" are  
30 used in an open-ended fashion, and thus should be interpreted to mean "include, but

not limited to ...". Also, the term "couple" is intended to mean either an indirect or direct electrical connection. Accordingly, if one device is coupled to another device, that connection may be through a direct electrical connection, or through an indirect electrical connection via other devices and connections.

5 [0014] The concept of the present invention is to release the response time constraint of an acknowledgment message indicative of a resultant status of a packet transmission of a scheduled output queue by making multiple candidate output queue decisions before the acknowledgment message is ready. In this way, the packet bubble (i.e., idle egress port) probability in most cases can be reduced to  
10 thereby sustain the desired line rate. Generally speaking, the packet bubble can be completely avoided when the ingress packets are enough (i.e. the packets input to the ingress ports  $\cong$  the packets output from the egress pots). Further details of the proposed scheduling design are described hereinafter with reference to the accompanying drawings.

15 [0015] FIG. 1 is a diagram illustrating a network switch according to an embodiment of the present invention. The network switch 100 includes a buffer memory device 102, a dequeue device 104 and a scheduler 106. The buffer memory device 102 includes a packet buffer 112 and a queue manager 114. By way of example, the packet buffer 112 may be implemented using a dynamic random access memory  
20 (DRAM), and the queue manager 114 may include static random access memories (SRAMs) to store a plurality of packet linked lists acting as a plurality of output queues 115\_1-115\_N, respectively. When a packet is received from one of a plurality of ingress ports of the network switch 100, the packet is buffered in the packet buffer 112, and enqueued to one of the output queues 115\_1-115\_N. For  
25 example, packets received from an ingress port I (not shown) and required to be forwarded through an egress port J (not shown) are enqueued to the same output queue OQ<sub>I,J</sub>. For clarity and simplicity, only one egress port is shown in FIG. 1. Hence, a packet received from one of the ingress ports (e.g., port 1-port N) and decided to be forwarded through the egress port is processed by a packet processor  
30 (not shown) with using lots of lookup tables according to its packet attribute. Thus, the packet is enqueued to one of the output queues 115\_1-115\_N maintained by the queue manager 114. The scheduler 106 is arranged to properly schedule the output queues 115\_1-115\_N through respective priority configurations, such as rate

limitations of packets waiting to be forwarded through the same egress port. However, this is for illustrative purposes only, and is not meant to be a limitation of the present invention.

[0016] To relax the memory requirement of the output queues 115\_1-115\_N, a linked list structure is employed to serve as one output queue. For example, each of the output queues 115\_1-115\_N is a packet linked list composed of packet identifiers (packet IDs) of packets received from the same ingress port and required to be forwarded through the egress port. Hence, when a specific packet is dequeued from one output queue, a node which stores a packet ID of the specific packet is removed from the packet linked list, and the specific packet is read from the packet buffer 112 and output to the following stage (e.g., dequeue device 104) for packet forwarding through the egress port.

[0017] As the packet to be forwarded through the egress port may come from any of the ingress ports, the scheduler 106 is arranged to perform a plurality of scheduler operations one by one, where each scheduler operation makes one output queue decision for deciding which of the output queues 115\_1-115\_N is granted to output its packet data. That is, the scheduler 106 is arranged to properly schedule the output queues 115\_1-115\_N that have packets waiting to be forwarded through the same egress port.

[0018] In a conventional design, after a scheduler makes one output queue decision for granting one output queue, the scheduler does not make the next output queue decision until an acknowledgment message indicative of a resultant status of a packet transmission of the scheduled output queue is received. Thus, one scheduler operation performed by the conventional scheduler includes at least an operation of waiting for an acknowledgment message corresponding to a packet transmission of a current winner output queue and an operation of picking a next winner output queue from a plurality of output queues. In other words, the processing time required by one scheduler operation performed by the conventional scheduler includes at least a wait time and a decision-making time.

[0019] For example, an acknowledgement message EmptyQ\_ACK may indicate whether the scheduled output queue will become an empty queue after dequeued by the current packet transmission. If the acknowledgement message EmptyQ\_ACK indicates an empty queue, this scheduled output queue should not be selected as the next scheduled output queue; otherwise, the packet bubble (idle egress port) will

occur due to no packet transmission. Besides, the generation of the acknowledgement message EmptyQ\_ACK may need to search information recorded in a look-up table, and therefore may need several clock cycles. If the packet transmission of one packet dequeued from an output queue selected by one output queue decision is already complete and the conventional scheduler fails to make another output queue decision in time, the line rate of the egress port may decrease due to no packet transmission.

[0020] To release the response time constraint of an acknowledgment message indicative of a resultant status of a packet transmission of a scheduled output queue, the scheduler 106 of the present invention is configured to make multiple candidate output queue decisions, and then make a final output queue decision based on the multiple candidate output queue decisions. In this embodiment, the scheduler 106 includes a candidate decision logic 116 and a final decision logic 118. The candidate decision logic 116 is aware of queue statuses of the output queues 115\_1-115\_N. For example, the queue status of one output queue may indicate whether the output queue has packets to be serviced or not. The candidate decision logic 116 is arranged to decide a plurality of candidate output queues OQ\_1-OQ\_M for a current scheduler operation, regardless of a resultant status of packet transmission of at least one scheduled output queue OQ<sub>PRE</sub> decided by at least one previous scheduler operation. For example, the candidate decision logic 116 decides M candidate output queues OQ\_1-OQ\_M before acknowledgement messages EmptyQ\_ACK for (M-1) scheduled output queues OQ<sub>PRE</sub> decided by (M-1) previous scheduler operations are available to the scheduler 106, where M is a positive integer not smaller than 2.

[0021] In this embodiment, the candidate output queues OQ\_1-OQ\_M include a specific output queue that is different from the at least one scheduled output queue OQ<sub>PRE</sub> decided by the at least one previous scheduler operation. Under a condition where the acknowledgement messages EmptyQ\_ACK for (M-1) scheduled output queues OQ<sub>PRE</sub> decided by (M-1) previous scheduler operations are not ready yet, the candidate decision logic 116 first decides the best winner output queue (which has the highest decision priority) selected from the output queues 115\_1-115\_N associated with the egress port. When the best winner output queue is different from the (M-1) scheduled output queues OQ<sub>PRE</sub> decided by the (M-1) previous scheduler operations, the candidate decision logic 116 sets the best winner output queue as the specific output queue. However, when the best winner output queue is identical to

the (M-1) scheduled output queues  $OQ_{PRE}$  decided by the (M-1) previous scheduler operations, the candidate decision logic 116 further decides the second best winner output queue (which has the next decision priority) selected from the output queues 115\_1-115\_N associated with the egress port. When the second best winner output queue is different from the (M-1) scheduled output queues  $OQ_{PRE}$  decided by the (M-1) previous scheduler operations, the candidate decision logic 116 sets the second best winner output queue as the specific output queue. To put it simply, the candidate decision logic 116 will try to make output queue decisions according to decision priorities until the specific output queue different from the (M-1) scheduled output queues  $OQ_{PRE}$  decided by the (M-1) previous scheduler operations is found.

[0022] Since the candidate output queues  $OQ_1$ - $OQ_M$  are determined before the acknowledgement messages  $EmptyQ\_ACK$  for (M-1) scheduled output queues  $OQ_{PRE}$  decided by (M-1) previous scheduler operations are ready, the candidate output queues  $OQ_1$ - $OQ_M$  should include the (M-1) scheduled output queues  $OQ_{PRE}$  decided by the (M-1) previous scheduler operations due to the fact that one of the (M-1) scheduled output queues  $OQ_{PRE}$  should be allowed to be selected as the scheduled output queue  $OQ_{CUR}$  for the current scheduler operation. Hence, concerning the current scheduler operation, the candidate output queues  $OQ_1$ - $OQ_M$  determined by the candidate decision logic 116 are composed of (M-1) scheduled output queues  $OQ_{PRE}$  decided by the (M-1) previous scheduler operations and a different output queue (i.e., the aforementioned specific output queue) selected from the output queues 115\_1-115\_N associated with the egress port.

[0023] The final decision logic 118 is arranged to select one of the candidate output queues  $OQ_1$ - $OQ_M$  to serve as the scheduled output queue  $OQ_{CUR}$  decided by the current scheduler operation after obtaining the resultant status of packet transmission of the at least one scheduled output queue  $OQ_{PRE}$  decided by the at least one previous scheduler operation, such as acknowledgement messages  $EmptyQ\_ACK$  for (M-1) scheduled output queues  $OQ_{PRE}$  decided by (M-1) previous scheduler operations. More specifically, the final decision logic 118 checks the obtained resultant status (e.g.,  $EmptyQ\_ACK$ ) to decide which of the candidate output queues  $OQ_1$ - $OQ_M$  should be selected as the scheduled output queue  $OQ_{CUR}$  decided by the current scheduler operation. In one preferred embodiment, the value of M is equal to 2. Therefore, two candidate output queues  $OQ_1$ ,  $OQ_2$  include a single scheduled output queue  $OQ_{PRE}$  decided by a single previous scheduler operation



immediately preceding the current scheduler operation, and a specific output queue different from the scheduled output queue  $OQ_{PRE}$ . For example,  $OQ_1=OQ_{PRE}$  and  $OQ_2 \neq OQ_{PRE}$ .

[0024] When the obtained resultant status indicates that the scheduled output queue  $OQ_{PRE}$  is an empty queue, implying that the candidate output queue  $OQ_1$  is not a qualified candidate for the current scheduler operation, the final decision logic 118 selects the other candidate output queue  $OQ_2$  to serve as the scheduled output queue  $OQ_{CUR}$  decided by the current scheduler operation. When the obtained resultant status indicates that the scheduled output queue  $OQ_{PRE}$  is not an empty queue, implying that the candidate output queue  $OQ_1$  is a qualified candidate for the current scheduler operation, the final decision logic 118 refers to decision priorities of the candidate output queues  $OQ_1$ ,  $OQ_2$  to decide which of the candidate output queues  $OQ_1$ ,  $OQ_2$  should be selected to serve as the scheduled output queue  $OQ_{CUR}$  decided by the current scheduler operation. For example, when the candidate output queue  $OQ_1$  (i.e., previous scheduled output queue  $OQ_{PRE}$ ) is the best winner output queue and the other output queue  $OQ_2$  different from the previous scheduled output queue  $OQ_{PRE}$  is the second best winner output queue, the final decision logic 118 refers to decision priorities to select the candidate output queue  $OQ_1$  ( $OQ_1=OQ_{PRE}$ ) to serve as the scheduled output queue  $OQ_{CUR}$  decided by the current scheduler operation; and when the candidate output queue  $OQ_2$  different from the previous scheduled output queue  $OQ_{PRE}$  is the best winner output queue and the other candidate output queue  $OQ_1$  ( $OQ_1=OQ_{PRE}$ ) is the second best winner output queue, the final decision logic 118 refers to decision priorities to select the candidate output queue  $OQ_2$  ( $OQ_2 \neq OQ_{PRE}$ ) to serve as the scheduled output queue  $OQ_{CUR}$  decided by the current scheduler operation.

[0025] Please refer to FIG. 2, which is a timing diagram illustrating a proposed solution for releasing the response time constraint according to an embodiment of the present invention. At time  $T_0$ , the scheduler 106 finishes one scheduler operation  $SCH_0$  to output one output queue decision  $D_0=OQ_0$ , and starts the next scheduler operation  $SCH_1$ . After notified by the output queue decision  $D_0=OQ_0$ , the dequeue device 104 performs a packet dequeue operation upon the scheduled output queue  $OQ_0$  designated by the output queue decision  $D_0=OQ_0$ . In this embodiment, the candidate decision logic 116 decides two candidate output queues  $OQ_0$  and  $OQ_1$  before the acknowledgement message  $EmptyQ\_ACK(OQ_0)$  is ready. When the

acknowledgement message EmptyQ\_ACK(OQ0) is available at time  $T_1$ , the final decision logic 118 selects one of the candidate output queues OQ0 and OQ1 according to the acknowledgement message EmptyQ\_ACK(OQ0). In this example, the candidate output queue OQ1 is selected to serve as the scheduled output queue decided by the scheduler operation SCH<sub>1</sub>.

[0026] At time  $T_2$ , the scheduler 106 finishes one scheduler operation SCH<sub>1</sub> to output one output queue decision  $D_1=OQ1$ , and starts the next scheduler operation SCH<sub>2</sub>. After notified by the output queue decision  $D_1=OQ1$ , the dequeue device 104 performs a packet dequeue operation upon the scheduled output queue OQ1 designated by the output queue decision  $D_1=OQ1$ . Besides, the candidate decision logic 116 decides two candidate output queues OQ1 and OQ2 before the acknowledgement message EmptyQ\_ACK(OQ1) is ready. When the acknowledgement message EmptyQ\_ACK(OQ1) is available at time  $T_3$ , the final decision logic 118 selects one of the candidate output queues OQ1 and OQ2 according to the acknowledgement message EmptyQ\_ACK(OQ1). In this example, the candidate output queue OQ2 is selected to serve as the scheduled output queue decided by the scheduler operation SCH<sub>2</sub>.

[0027] It should be noted that the output queue decisions shown in FIG. 2 are for illustrative purposes only, and are not meant to be limitations of the present invention. For example, any scheduler that makes multiple candidate output queue decisions and then makes the final output queue decision for a current scheduler operation upon reception of a resultant status of packet transmission of a scheduled output queue decided by a previous scheduler operation falls within the scope of the present invention.

[0028] In this embodiment, the processing time of making candidate output queue decisions is able to cover the response time of the acknowledgement message EmptyQ\_ACK. The final output queue decision will be available at the end of the packet transmission of the current dequeued packet, such that no packet bubble (idle egress port) will happen in this case. More specifically, the candidate decision logic 116 of the scheduler 106 starts its designated operation every minimum packet transmission time to ensure the desired line rate without waiting for the acknowledgment message EmptyQ\_ACK (i.e. the scheduler 106 gets related information from the buffer memory device 102 to ensure the desired line rate), and the behavior of the final decision logic 118 of the scheduler 106 is equivalent to

having the final output queue right after the acknowledgement message EmptyQ\_ACK is ready. Besides, a network switch employing the proposed scheduler does not need a complicated status check mechanism implemented for speeding up generation of the acknowledgement message EmptyQ\_ACK. Thus, the high cost and difficulty to have a fast status check mechanism for the output queue can be avoided. Further, the architecture dependency of neighboring functional blocks in a network switch can be reduced/mitigated correspondingly.

[0029] Please refer to FIG. 3, which is a flowchart illustrating a method for scheduling an output queue according to an embodiment of the present invention.

10 Provided that the result is substantially the same, the steps are not required to be executed in the exact order shown in FIG. 3. The method may be employed by the scheduler 106 shown in FIG. 1, and may be briefly summarized as below.

[0030] Step 300: Start.

[0031] Step 302: Check if a minimum packet transmission time is expired. If yes, go to step 304 to start a new scheduler operation involving multiple candidate output queue decisions; otherwise, keep checking if a criterion of starting a new scheduler operation involving multiple candidate output queue decisions is met. In this embodiment, the proposed scheduler operation involving multiple candidate output queue decisions is performed every minimum packet transmission time, so as to ensure the desired line rate.

[0032] Step 304: Output an output queue decision decided by a previous scheduler operation, and start a current scheduler operation used for obtaining a next output queue decision.

[0033] Step 306: Make one candidate output queue decision to decide the best winner output queue.

[0034] Step 308: Check if the best winner output queue is identical to a scheduled output queue decided by the previous scheduler operation. If yes, go to step 310; otherwise, go to step 312.

[0035] Step 310: Make another candidate output queue decision to decide the second best winner output queue.

[0036] Step 312: Obtain multiple candidate output queues, including the scheduled output queue decided by the previous scheduler operation and a different output queue set by the best winner output queue/ second best winner output queue.

[0037] Step 314: Check if a resultant status of packet transmission of the scheduled

output queue decided by the previous scheduler operation is available. If yes, go to step 316; otherwise, go to step 314.

[0038] Step 316: Check the obtained resultant status to select one of the candidate output queues to serve as a scheduled output queue decided by the current scheduler operation.

[0039] Step 318: End.

[0040] As a person skilled in the art can readily understand details of each step shown in FIG. 3 after reading above paragraphs, further description is omitted here for brevity.

10 [0041] Those skilled in the art will readily observe that numerous modifications and alterations of the device and method may be made while retaining the teachings of the invention. Accordingly, the above disclosure should be construed as limited only by the metes and bounds of the appended claims.

## CLAIMS

1. A scheduler for performing a plurality of scheduler operations, each scheduling an output queue selected from a plurality of output queues associated with an egress port, the scheduler comprising:

5 a candidate decision logic, arranged to decide a plurality of candidate output queues for a current scheduler operation, regardless of a resultant status of packet transmission of at least one scheduled output queue decided by at least one previous scheduler operation; and

10 a final decision logic, arranged to select one of the candidate output queues as a scheduled output queue decided by the current scheduler operation after obtaining the resultant status of packet transmission of the at least one scheduled output queue decided by the at least one previous scheduler operation.

2. The scheduler of claim 1, wherein the candidate output queues include a specific output queue that is different from the at least one scheduled output queue decided by the at least one previous scheduler operation.

3. The scheduler of claim 2, wherein the candidate decision logic decides a best winner output queue selected from the output queues associated with the egress port; and the candidate decision logic sets the best winner output queue as the specific output queue when the best winner output queue is different from the at least one scheduled output queue decided by the at least one previous scheduler operation.

4. The scheduler of claim 3, wherein when the best winner output queue is identical to the at least one scheduled output queue decided by the at least one previous scheduler operation, the candidate decision logic further decides a second best winner output queue selected from the output queues associated with the egress port; and the candidate decision logic sets the second best winner output queue as the specific output queue when the second best winner output queue is different from the at least one scheduled output queue decided by the at least one previous scheduler operation.

5. The scheduler of claim 1, wherein the candidate output queues include the at least one scheduled output queue decided by the at least one previous scheduler operation and a different output queue selected from the output queues associated with the egress port.

6. The scheduler of claim 5, wherein the at least one scheduled output queue decided by the at least one previous scheduler operation only includes a single scheduled output queue decided by a single previous scheduler operation immediately preceding the current scheduler operation.

5 7. The scheduler of claim 5, wherein the final decision logic is further arranged to check the obtained resultant status to decide which of the candidate output queues is selected to serve as the scheduled output queue decided by the current scheduler operation.

10 8. The scheduler of claim 7, wherein when the obtained resultant status indicates that the at least one scheduled output queue is an empty queue, the final decision logic selects the different output queue to serve as the scheduled output queue decided by the current scheduler operation.

15 9. The scheduler of claim 7, wherein when the obtained resultant status indicates that the at least one scheduled output queue is not an empty queue, the final decision logic refers to decision priorities of the candidate output queues to decide which of the candidate output queues is selected to serve as the scheduled output queue decided by the current scheduler operation.

20 10. A method for performing a plurality of scheduler operations each scheduling an output queue selected from a plurality of output queues associated with an egress port, comprising:

deciding a plurality of candidate output queues for a current scheduler operation, regardless of a resultant status of packet transmission of at least one scheduled output queue decided by at least one previous scheduler operation; and

25 after obtaining the resultant status of packet transmission of the at least one scheduled output queue decided by the at least one previous scheduler operation, selecting one of the candidate output queues as a scheduled output queue decided by the current scheduler operation.

30 11. The method of claim 10, wherein the candidate output queues include a specific output queue that is different from the at least one scheduled output queue decided by the at least one previous scheduler operation.

12. The method of claim 11, wherein the step of deciding the candidate output queues comprises:

deciding a best winner output queue selected from the output queues associated with the egress port; and

when the best winner output queue is different from the at least one scheduled output queue decided by the at least one previous scheduler operation, setting the best winner output queue as the specific output queue.

13. The method of claim 12, wherein the step of deciding the candidate output queues further comprises:

when the best winner output queue is identical to the at least one scheduled output queue decided by the at least one previous scheduler operation, deciding a second best winner output queue selected from the output queues associated with the egress port; and

10 when the second best winner output queue is different from the at least one scheduled output queue decided by the at least one previous scheduler operation, setting the second best winner output queue as the specific output queue.

14. The method of claim 10, wherein the candidate output queues include the at least one scheduled output queue decided by the at least one previous scheduler operation and a different output queue selected from the output queues associated with the egress port.

15. The method of claim 14, wherein the at least one scheduled output queue decided by the at least one previous scheduler operation only includes a single scheduled output queue decided by a single previous scheduler operation immediately preceding the current scheduler operation.

16. The method of claim 14, wherein the step of selecting one of the candidate output queues as the scheduled output queue decided by the current scheduler operation comprises:

25 checking the obtained resultant status to decide which of the candidate output queues is selected to serve as the scheduled output queue decided by the current scheduler operation.

17. The method of claim 16, wherein the step of checking the obtained resultant status to decide which of the candidate output queues is selected to serve as the scheduled output queue decided by the current scheduler operation comprises:

30 when the obtained resultant status indicates that the at least one scheduled output queue is an empty queue, selecting the different output queue to serve as the scheduled output queue decided by the current scheduler operation.

18. The method of claim 16, wherein the step of checking the obtained resultant status to decide which of the candidate output queues is selected to serve as the

scheduled output queue decided by the current scheduler operation comprises:

when the obtained resultant status indicates that the at least one scheduled output queue is not an empty queue, referring to decision priorities of the candidate output queues to decide which of the candidate output queues is selected to serve as the

5 scheduled output queue decided by the current scheduler operation.



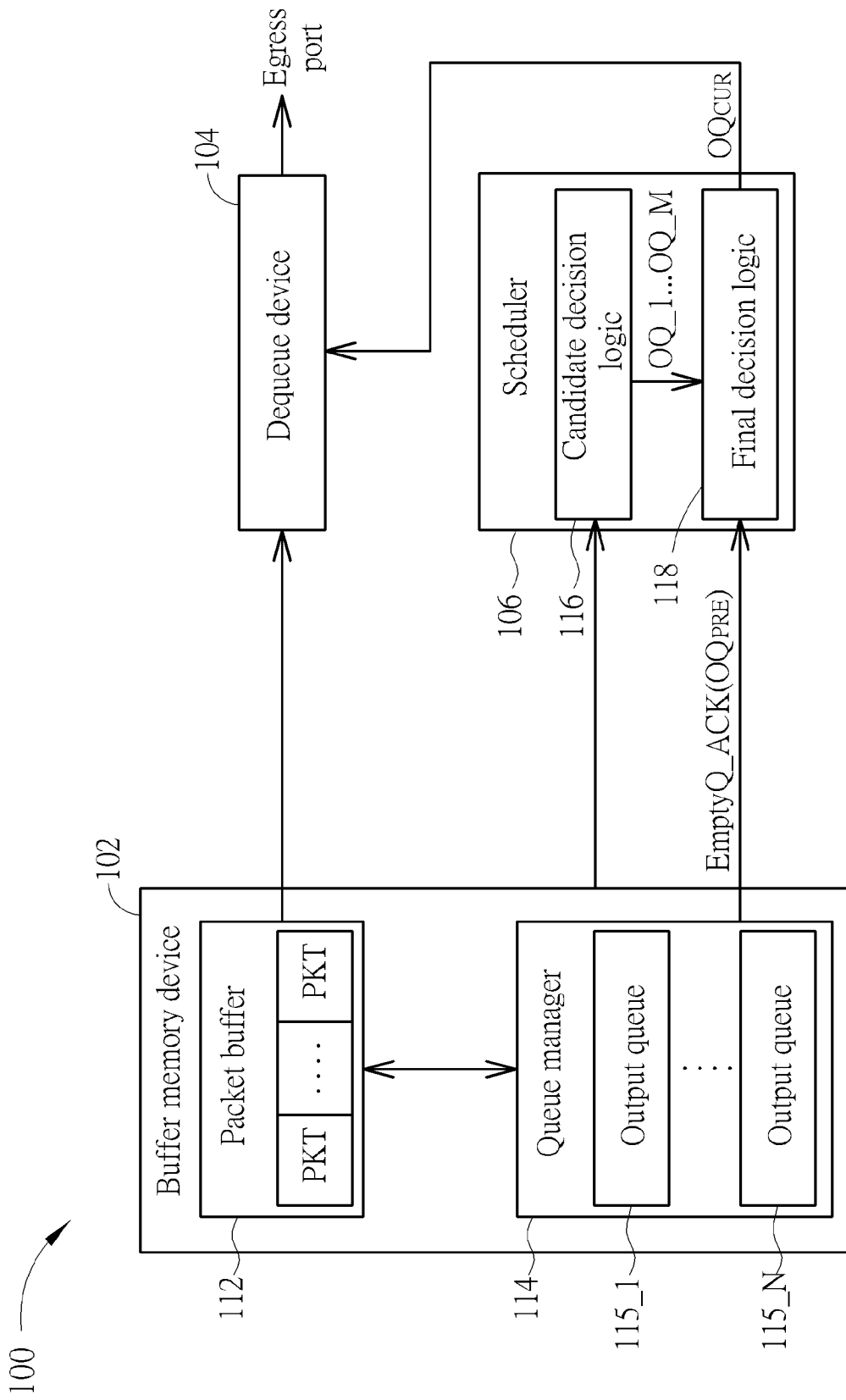


FIG. 1

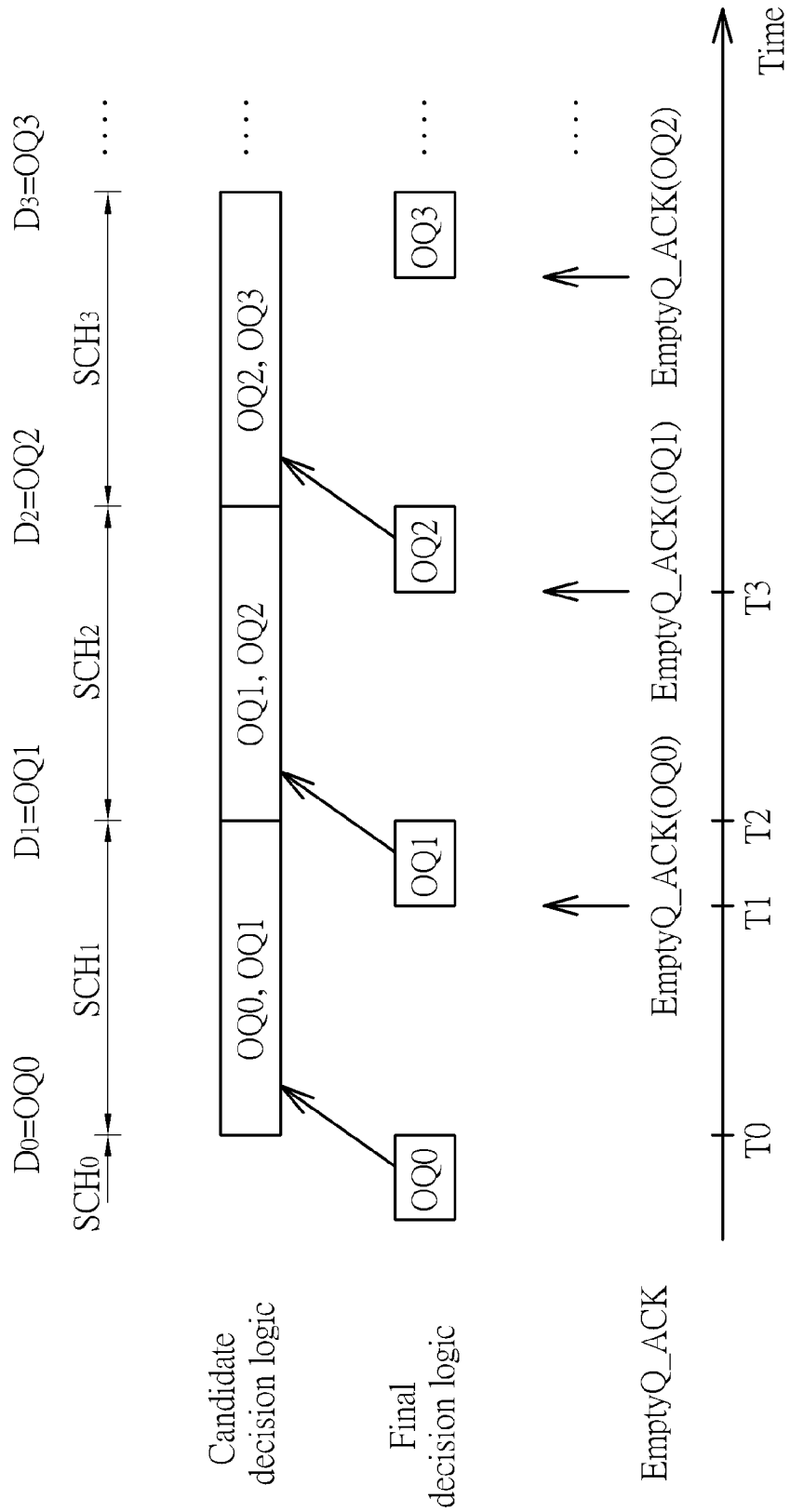


FIG. 2

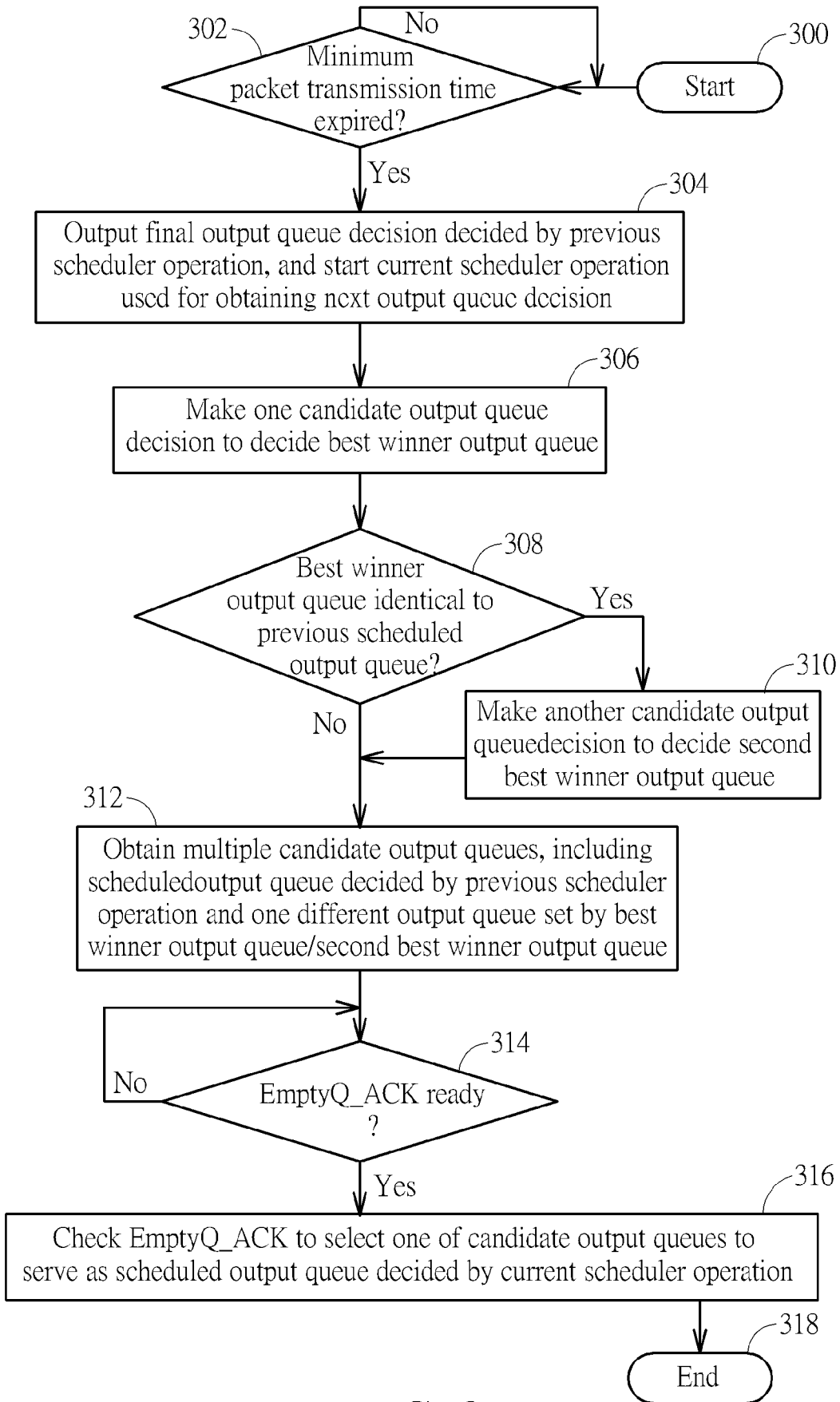


FIG. 3

## INTERNATIONAL SEARCH REPORT

International application No.

**PCT/CN2014/076020**

<b>A. CLASSIFICATION OF SUBJECT MATTER</b>		
H04L 12/861(2013.01)i		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b>		
Minimum documentation searched (classification system followed by classification symbols)		
H04L		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
CNPAT,CNKI,WPI,EPODOC,GOOGLE,IEEE: scheduler, schedul+, output w queue?, egress, port?, candidate, buffer, store, select, choose, logic, status, packet?, transmission		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 7391786B1 (CISCO TECHNOLOGY, INC.) 24 June 2008 (2008-06-24) the whole document	1-18
A	CN 101478483A (THE PLA INFORMATION ENGINEERING UNIVERSITY) 08 July 2009 (2009-07-08) the whole document	1-18
A	US 7373420B1 (NORTEL NETWORKS LIMITED) 13 May 2008 (2008-05-13) the whole document	1-18
A	US 2004/0246970A1 (ALCATEL) 09 December 2004 (2004-12-09) the whole document	1-18
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents:		
“A”	document defining the general state of the art which is not considered to be of particular relevance	“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
“E”	earlier application or patent but published on or after the international filing date	“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
“L”	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
“O”	document referring to an oral disclosure, use, exhibition or other means	“&” document member of the same patent family
“P”	document published prior to the international filing date but later than the priority date claimed	
Date of the actual completion of the international search		Date of mailing of the international search report
<b>03 July 2014</b>		<b>29 July 2014</b>
Name and mailing address of the ISA/ <b>STATE INTELLECTUAL PROPERTY OFFICE OF THE P.R.CHINA(ISA/CN) 6,Xitucheng Rd., Jimen Bridge, Haidian District, Beijing 100088 China</b>		Authorized officer  <b>YANG,Ying</b>
Facsimile No. <b>(86-10)62019451</b>		Telephone No. <b>(86-10)82245553</b>

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

International application No.

**PCT/CN2014/076020**

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
US 7391786B1	24 June 2008	None	
CN 101478483A	08 July 2009	None	
US 7373420B1	13 May 2008	None	
US 2004/0246970A1	09 December 2004	EP 1484939A1	08 December 2004
		DE 60305457T2	31 May 2007
		AT 327646T	15 June 2006