BUFFER CONFIGURATION METHOD AND SWITCHING DEVICE

Abstract

A buffer configuration method and a switching device are disclosed, to ensure no packet loss. According to the method, a first switching device sends a first measurement frame to a second switching device; receives a second measurement frame sent by the second switching device, where the second measurement frame is generated based on the first measurement frame; determines a buffer configuration parameter based on the first measurement frame and the second measurement frame; and sets a local buffer based on the configuration parameter. The present disclosure is used to automatically configure a buffer of a switching device, thereby reducing required buffer space without packet loss.

Inventors: **Wu; Tao (Nanjing, CN), Yan; Jian (Nanjing, CN), Chen; Yongxian (Nanjing, CN), Liu; Heyang (Nanjing, CN), Wen; Huafeng (Nanjing, CN), Sun; Liyang (Leuven, BE)**

Applicant: **HUAWEI TECHNOLOGIES CO., LTD.** (Shenzhen, CN)

Family ID: 1000008586881

Appl. No.: 19/176271

Filed: **April 11, 2025**

Publication Classification

Int. Cl.: **H04L49/9005** (20220101); **H04L47/24** (20220101)

Background/Summary

CROSS-REFERENCE TO RELATED APPLICATIONS [0001] This application is a continuation of U.S. patent application Ser. No. 18/150,877, filed on Jan. 6, 2023, which is a continuation of International Application No. PCT/CN2021/102867, filed on Jun. 28, 2021, which claims priority to Chinese Patent Application No. 202010646219.4, filed on Jul. 7, 2020. All of the aforementioned patent applications are hereby incorporated by reference in their entireties.

TECHNICAL FIELD
[0002] The present disclosure relates to the field of network technologies, and in particular, to a buffer configuration method and a switching device.
BACKGROUND
[0003] A data packet forwarding process on the Internet is implemented based on a large quantity of switches. Each switch has a forwarding function. An upstream switch forwards a received data packet to a downstream switch, and the downstream switch continues to forward the data packet until the data packet arrives at a destination address. To ensure that the switch normally forwards the data packet, traffic of the switch needs to be controlled to avoid excessive load of the switch.
[0004] In the Institute of electrical and electronics engineers (IEEE) 802.1 and 802.3 standards, a priority-based flow control (priority-based flow control, PFC) mechanism is provided. In the PFC mechanism, when learning that congestion occurs in a priority queue of a downstream switch, an upstream switch suspends dequeuing a data packet from a priority queue. However, the configuration of a headroom (HDRM) of the switch is involved in the PFC mechanism. In the conventional technology, a user usually configures a value of an HDRM of a local buffer of a switch based on a default value that is of the HDRM and that is in a guide of a device manufacturer or that is set when the device is delivered from a factory.
[0005] However, the default value that is of the HDRM and that is in the guide of the device manufacturer or that is set when the device is delivered from the factory cannot be suitable for all networking scenarios. When the value of the HDRM is set to be excessively large, buffer space is wasted. If the value of the HDRM is set to be excessively small, a packet loss occurs.
SUMMARY
[0006] The present disclosure provides a buffer configuration method and a switching device, to automatically configure a local buffer to ensure no packet loss.
[0007] According to a first aspect, an embodiment of the present disclosure provides a buffer configuration method. The method includes:
[0008] A first switching device sends a first measurement frame to a second switching device.
[0009] The first switching device receives a second measurement frame sent by the second switching device. The second measurement frame is generated through triggering based on the first measurement frame.
[0010] The first switching device determines a buffer configuration parameter based on the first measurement frame and the second measurement frame.
[0011] The first switching device sets a local buffer based on the configuration parameter.
[0012] In this possible design, the method may be applied to a networking architecture including the first switching device and the second switching device, and the first switching device and the second switching device are communicatively connected through a port. The first switching device and the second switching device may be network devices or terminal devices. A common network device includes a switch, and a common terminal device includes a network adapter. A first measurement frame generator is disposed inside the first switching device, and the first measurement frame

generator may be any one of a processor, a field-programmable gate array (FPGA), and a PFC engine. The first switching device sends the first measurement frame to the second switching device. After receiving the first measurement frame, the second switching device triggers the second measurement frame to be generated. The first switching device determines the configuration parameter of the local buffer based on the first measurement frame and the second measurement frame, to implement an automatic configuration of the local buffer, thereby reducing use of the local buffer of the switching device, and improving a burst absorption capability and supporting more lossless queues while ensuring no packet loss.

[0013] In a possible design of the first aspect, that the first switching device determines a buffer configuration parameter based on the first measurement frame and the second measurement frame includes:

[0014] The first switching device obtains a first timestamp and a fourth timestamp. The first timestamp is a time at which the first switching device sends the first measurement frame, and the fourth timestamp is a time at which the first switching device receives the second measurement frame.

[0015] The first switching device determines the buffer configuration parameter based on the first timestamp and the fourth timestamp.

[0016] In this possible design, the time at which the first switching device sends the first measurement frame is recorded as the first timestamp, and the time at which the first switching device receives the second measurement frame is recorded as the fourth timestamp. The first timestamp and the fourth timestamp may be used to represent a data exchange speed between the first switching device and the second switching device. Therefore, the first switching device can determine the buffer configuration parameter based on the first timestamp and the fourth timestamp, so that the configuration parameter fits data transmission performance of the first switching device and the second switching device.

[0017] In a possible design of the first aspect, the first timestamp is a time at which the last bit of the first measurement frame enters a MAC layer of the first switching device, and the fourth timestamp is a time at which the last bit of the second measurement frame is sent from the MAC layer of the first switching device.

[0018] In a possible design of the first aspect, the first timestamp is a time at which the last bit of the first measurement frame is sent from a MAC layer of the first switching device, and the fourth timestamp is a time at which the last bit of the second measurement frame enters the MAC layer of the first switching device.

[0019] In a possible design of the first aspect, the first timestamp and the fourth timestamp are encapsulated in a time field following an Opcode field of the second measurement frame.

[0020] In this possible design, the first timestamp and the fourth timestamp may be encapsulated in the second measurement frame, so that a timestamp reading step can be omitted, and the first switching device can quickly obtain timestamp information. For example, a PFC frame format may be used as a format of the second measurement frame to store the first timestamp and the fourth timestamp in the time field following the Opcode field of the second measurement frame. Because a PFC frame is a MAC control frame, the MAC layer of the first switching device can directly identify the Opcode field, and then read the time field following the Opcode field to quickly obtain the timestamp information.

[0021] In a possible design of the first aspect, the first timestamp and the fourth timestamp are stored in the local buffer of the first switching device.

[0022] In this possible design, the first timestamp and the fourth timestamp are stored in the local buffer of the first switching device, and the first switching device obtains timestamp information from the local buffer.

[0023] In a possible design of the first aspect, the determining the buffer configuration parameter based on the first timestamp and the fourth timestamp includes:

[0024] The first switching device determines a timestamp difference based on the first timestamp and the fourth timestamp.

[0025] The first switching device determines the buffer configuration parameter based on the

timestamp difference, timestamp precision, a port rate between the first switching device and the second switching device, a configured maximum frame length, and a constant K. The constant K is provided by a device manufacturer.

[0026] In this possible design, the first switching device first calculates the timestamp difference based on the first timestamp and the fourth timestamp, and then determines the buffer configuration parameter based on the timestamp difference, the timestamp precision, the port rate between the first switching device and the second switching device, and the configured maximum frame length. Therefore, the first switching device can obtain the buffer configuration parameter associated with the exchange speed and the port rate between the first switching device and the second switching device and the configured maximum frame length, to implement an accurate configuration of the local buffer, thereby reducing use of the local buffer of the switching device, and improving a burst absorption capability and supporting more lossless queues while ensuring no packet loss.

[0027] In a possible design of the first aspect, the method further includes: The first switching device obtains a second timestamp and a third timestamp. The second timestamp is a time at which the second switching device receives the first measurement frame, and the third timestamp is a time at which the second switching device sends the second measurement frame.

[0028] In this possible design, the time at which the second switching device receives the first measurement frame is recorded as the second timestamp, and the time at which the second switching device sends the second measurement frame is recorded as the third timestamp. The second timestamp and the third timestamp may be used to represent delay time from receiving, by the second switching device, the first measurement frame to triggering the second measurement frame to be generated. Therefore, the first switching device can determine the buffer configuration parameter based on the first timestamp, the second timestamp, the third timestamp, and the fourth timestamp, so that the configuration parameter fits data transmission performance of the first switching device and the second switching device.

[0029] In a possible design of the first aspect, the second timestamp is a time at which the last bit of the first measurement frame is sent from a MAC layer of the second switching device, and the third timestamp is a time at which the last bit of the second measurement frame enters the MAC layer of the second switching device.

[0030] In a possible design of the first aspect, the second timestamp is a time at which the last bit of the first measurement frame enters a MAC layer of the second switching device, and the third timestamp is a time at which the last bit of the second measurement frame is sent from the MAC layer of the second switching device.

[0031] In a possible design of the first aspect, the second timestamp and the third timestamp are encapsulated in the time field following the Opcode field of the second measurement frame.

[0032] In this possible design, the second timestamp and the third timestamp may be encapsulated in the second measurement frame, so that a timestamp reading step can be omitted, and the first switching device can quickly obtain timestamp information. For example, the PFC frame format may be used as the format of the second measurement frame to store the second timestamp and the third timestamp in the time field following the Opcode field of the second measurement frame. Because the PFC frame is a MAC control frame, the MAC layer of the first switching device can directly identify the Opcode field, and then read the time field following the Opcode field to quickly obtain the timestamp information.

[0033] In a possible design of the first aspect, the second timestamp and the third timestamp are stored in the local buffer of the first switching device.

[0034] In this possible design, the second timestamp and the third timestamp are stored in the local buffer of the first switching device, and the first switching device obtains timestamp information from the local buffer.

[0035] In a possible design of the first aspect, the method further includes:

[0036] The first switching device determines a timestamp difference based on the first timestamp, the second timestamp, the third timestamp, and the fourth timestamp.

[0037] The first switching device determines the buffer configuration parameter based on the timestamp difference, timestamp precision, a port rate between the first switching device and the second switching device, a configured maximum frame length, and a constant K. The constant K is provided by a device manufacturer.

[0038] In this possible design, the first switching device first calculates the timestamp difference based on the first timestamp, the second timestamp, the third timestamp, and the fourth timestamp, and then determines the buffer configuration parameter based on the timestamp difference, the timestamp precision, the port rate between the first switching device and the second switching device, and the configured maximum frame length. Therefore, the first switching device can obtain the buffer configuration parameter associated with the exchange speed and the port rate between the first switching device and the second switching device and the configured maximum frame length, to implement an accurate configuration of the local buffer, thereby reducing use of the local buffer of the switching device, and improving a burst absorption capability and supporting more lossless queues while ensuring no packet loss.

[0039] In a possible design of the first aspect, the method further includes:

[0040] The first switching device sends the buffer configuration parameter to the second switching device, so that the second switching device sets a local buffer based on the configuration parameter.

[0041] In this possible design, the second switching device may further set the local buffer of the second switching device by using the buffer configuration parameter calculated by the first switching device, so that the second switching device can also implement an automatic configuration of the buffer.

[0042] In a possible design of the first aspect, before the first switching device sends the first measurement frame to the second switching device, the method further includes:

[0043] determining whether both the first switching device and the second switching device support an automatic buffer configuration function of priority-based flow control.

[0044] In this possible design, only when both the first switching device and the second switching device support the automatic buffer configuration function of priority-based flow control, the foregoing method can be applied to implement an automatic configuration of the local buffer. Therefore, before the method is applied, it is first determined whether both the first switching device and the second switching device support the automatic buffer configuration function of priority-based flow control.

[0045] In a possible design of the first aspect, the determining whether both the first switching device and the second switching device support an automatic buffer configuration function of priority-based flow control includes:

[0046] when the first switching device supports the automatic buffer configuration function of priority-based flow control, and received TLV-format data fed back by the second switching device includes a buffer configuration field, determining that both the first switching device and the second switching device support the automatic buffer configuration function of priority-based flow control.

[0047] In this possible design, a link layer discovery protocol (LLDP) may be used between the first switching device and the second switching device to determine whether both the first switching device and the second switching device support the automatic buffer configuration function of priority-based flow control.

[0048] In a possible design of the first aspect, when the automatic buffer configuration function is not supported by either the first or second switching device, the method further includes: [0049] giving a prompt that the local buffer needs to be manually set.

[0050] In a possible design of the first aspect, the buffer configuration parameter is used to represent a required buffer size, and a unit of the buffer configuration parameter includes a bit or a minimum storage unit of a chip buffer.

[0051] In a possible design of the first aspect, the buffer configuration parameter includes buffer space required for priority-based flow control.

[0052] In a possible design of the first aspect, the second measurement frame is generated by the

second switching device by performing loopback or acknowledgement processing on the first measurement frame.

[0053] According to a second aspect, the present disclosure provides a buffer configuration method. The method includes:

[0054] A second switching device receives a first measurement frame sent by a first switching device, and triggers, based on the first measurement frame, a second measurement frame to be generated.

[0055] The second switching device sends the second measurement frame to the first switching device.

[0056] The second switching device receives a buffer configuration parameter sent by the first switching device.

[0057] The second switching device sets a local buffer based on the configuration parameter.

[0058] In this possible design, the method may be applied to a networking architecture including the first switching device and the second switching device, and the first switching device and the second switching device are communicatively connected through a port. The first switching device and the second switching device may be network devices or terminal devices. A common network device includes a switch, and a common terminal device includes a network adapter. A second measurement frame generator is disposed inside the second switching device, and the second measurement frame generator may be any one of a processor, an FPGA, and a PFC engine. After receiving the first measurement frame sent by the first switching device, the second switching device triggers the second measurement frame to be generated, and sends the second measurement frame to the first switching device, so that the first switching device determines the configuration parameter of the local buffer based on the first measurement frame and the second measurement frame. Finally, the second switching device configures the local buffer based on the buffer configuration parameter determined by the first switching device, to implement an automatic configuration of the local buffer, thereby reducing use of the local buffer of the switching device, and improving an absorption burst capability and supporting more lossless queues while ensuring no packet loss.

[0059] In a possible design of the second aspect, before the second switching device receives the first measurement frame sent by the first switching device, the method further includes: [0060] determining whether both the first switching device and the second switching device support an automatic buffer configuration function of priority-based flow control.

[0061] In this possible design, only when both the first switching device and the second switching device support the automatic buffer configuration function of priority-based flow control, the foregoing method can be applied to implement an automatic configuration of the local buffer. Therefore, before the method is applied, it is first determined whether both the first switching device and the second switching device support the automatic buffer configuration function of priority-based flow control.

[0062] In a possible design of the second aspect, the determining whether both the first switching device and the second switching device support an automatic buffer configuration function of priority-based flow control includes: [0063] when the second switching device supports the automatic buffer configuration function of priority-based flow control, and received TLV-format data fed back by the first switching device includes a buffer configuration field, determining that both the first switching device and the second switching device support the automatic buffer configuration function of priority-based flow control.

[0064] In this possible design, a link layer discovery protocol (LLDP) may be used between the first switching device and the second switching device to determine whether both the first switching device and the second switching device support the automatic buffer configuration function of priority-based flow control.

[0065] In a possible design of the second aspect, the automatic buffer configuration function is not supported by either the first or second switching device, the method further includes: [0066] giving a prompt that the local buffer needs to be manually set.

[0067] In a possible design of the second aspect, the buffer configuration parameter is used to represent a required buffer size, and a unit of the buffer configuration parameter includes a bit or a

minimum storage unit of a chip buffer.

[0068] In a possible design of the second aspect, the buffer configuration parameter includes buffer space required for priority-based flow control.

[0069] In a possible design of the second aspect, the triggering, based on the first measurement frame, a second measurement frame to be generated includes:

[0070] The second switching device performs loopback or acknowledgement processing on the first measurement frame to generate the second measurement frame.

[0071] According to a third aspect, the present disclosure provides a first switching device, including a processor and a transceiver. The processor is configured to execute instructions in computer executable program code. When the processor executes the instructions, the instructions cause the first switching device to perform the method in any design of the first aspect.

[0072] According to a fourth aspect, the present disclosure provides a second switching device, including a processor and a transceiver. The processor is configured to execute instructions in computer executable program code. When the processor executes the instructions, the instructions cause the first switching device to perform the method in any design of the second aspect.

[0073] According to a fifth aspect, the present disclosure provides a networking system, including a first switching device and a second switching device. The first switching device and the second switching device are communicatively connected through a port. The first switching device performs the method in any design of the first aspect, and the second switching device performs the method in any design of the second aspect.

[0074] According to a sixth aspect, the present disclosure provides a chip, including a memory and a processor. The memory is configured to store program instructions, and the processor is configured to invoke the program instructions in the memory to perform the method of the first aspect in embodiments of the present disclosure.

[0075] According to a seventh aspect, the present disclosure provides a chip, including a memory and a processor. The memory is configured to store program instructions, and the processor is configured to invoke the program instructions in the memory to perform the method of the second aspect in embodiments of the present disclosure.

[0076] According to an eighth aspect, the present disclosure provides a readable storage medium. The readable storage medium stores a computer program, and when the computer program is executed, the method of the first aspect in embodiments of the present disclosure is implemented.

[0077] According to a ninth aspect, the present disclosure provides a readable storage medium. The readable storage medium stores a computer program, and when the computer program is executed, the method of the second aspect in embodiments of the present disclosure is implemented.

[0078] According to a tenth aspect, the present disclosure provides a program product. The program product includes a computer program, and the computer program is stored in a readable storage medium. At least one processor of a communication apparatus may read the computer program from the readable storage medium, and the at least one processor executes the computer program, so that the communication apparatus implements the method in any design of the first aspect in embodiments of the present disclosure.

[0079] According to an eleventh aspect, the present disclosure provides a program product. The program product includes a computer program, and the computer program is stored in a readable storage medium. At least one processor of a communication apparatus may read the computer program from the readable storage medium, and the at least one processor executes the computer program, so that the communication apparatus implements the method in any design of the second aspect in embodiments of the present disclosure.

[0080] In the present disclosure, the first switching device sends a measurement frame to the second switching device, and receives a measurement frame fed back by the second switching device, to determine a configuration parameter of a local buffer, so that the first switching device and the second switching device can implement an automatic configuration of the buffer. In this way, use of buffer

space is reduced while no packet loss is ensured, so that the switching device can increase buffer space used by a user to absorb a burst, thereby supporting more lossless queues.

Description

BRIEF DESCRIPTION OF DRAWINGS
[0081] FIG. **1** is a schematic diagram of a networking structure of a data center according to an embodiment of the present disclosure;
[0082] FIG. **2** is a schematic diagram **1** of a structure of communication between a first switching device and a second switching device according to an embodiment of the present disclosure;
[0083] FIG. **3** is a schematic diagram **2** of a structure of communication between a first switching device and a second switching device according to an embodiment of the present disclosure;
[0084] FIG. **4** is a flowchart **1** of a buffer configuration method according to an embodiment of the present disclosure;
[0085] FIG. **5** is a schematic diagram of a structure of a PFC frame according to an embodiment of the present disclosure;
[0086] FIG. **6** is a flowchart **2** of a buffer configuration method according to an embodiment of the present disclosure;
[0087] FIG. **7** is a flowchart **3** of a buffer configuration method according to an embodiment of the present disclosure;
[0088] FIG. **8** is a schematic diagram of an architecture of existing IEEE DCBX PFC TLV;
[0089] FIG. **9** is a schematic diagram of an architecture of extended and updated IEEE DCBX PFC TLV;
[0090] FIG. **10** is a schematic diagram of an architecture of existing first-version Intel DCBX PFC TLV;
[0091] FIG. **11** is a schematic diagram of an architecture of extended and updated first-version Intel DCBX PFC TLV;
[0092] FIG. **12** is a schematic diagram of an architecture of existing second-version Intel DCBX PFC TLV;
[0093] FIG. **13** is a schematic diagram of an architecture of extended and updated second-version Intel DCBX PFC TLV;
[0094] FIG. **14** is a flowchart **4** of a buffer configuration method according to an embodiment of the present disclosure;
[0095] FIG. **15** is a flowchart **5** of a buffer configuration method according to an embodiment of the present disclosure;
[0096] FIG. **16** is a schematic diagram **1** of a structure of a first switching device according to an embodiment of the present disclosure;
[0097] FIG. **17** is a schematic diagram **1** of a structure of a second switching device according to an embodiment of the present disclosure; and
[0098] FIG. **18** is a schematic diagram of a structure of a networking system according to an embodiment of the present disclosure.

DESCRIPTION OF EMBODIMENTS
[0099] FIG. **1** is a schematic diagram of a networking structure of a data center according to an embodiment of the present disclosure. As shown in FIG. **1**, the data center includes a plurality of network devices **11** and terminal devices **12**. At least two levels of data transmission links are formed between the network devices **11**, and an upstream network device **11** sends data to a downstream network device **11** until the downstream network device **11** sends the data to the terminal device **12**. To avoid a packet loss in a data transmission process, a priority-based flow control (PFC) mechanism is used for the network devices **11** and the terminal devices **12** in entire networking. Optionally, three parameters, Xoff (pause), Xon (resume), and an HDRM (headroom), corresponding to a port priority are configured on each network device **11** and terminal device **12**. Xoff and Xon may be configured by

a user based on a user requirement. A value of the HDRM is usually a default value that is of the HDRM and that is in a guide of a device manufacturer or that is set when the device is delivered from a factory. However, the default value that is of the HDRM and that is in the guide of the device manufacturer or that is set when the device is delivered from the factory is a fixed value, and cannot be suitable for all networking scenarios. When the value of the HDRM is set to be excessively large, buffer space is wasted. If the value of the HDRM is set to be excessively small, a packet loss occurs.

[0100] To resolve the problems in the conventional technology, embodiments of the present disclosure provide a buffer configuration method, so that a network device and a terminal device can automatically configure local buffers, thereby ensuring no packet loss.

[0101] FIG. **2** is a schematic diagram **1** of a structure of communication between a first switching device and a second switching device according to an embodiment of the present disclosure. As shown in FIG. **2**, a first switching device **20** and a second switching device **30** are communicatively connected through a port. The first switching device **20** includes a first measurement frame generator **21** and a first link layer **22**. The first measurement frame generator **21** may be any one of a processor, a field-programmable gate array (FPGA), and a PFC engine, and is used to generate a first measurement frame. The first link layer **22** is responsible for a MAC (media access control or medium access control) address of the first switching device **20**. The second switching device **30** includes a second measurement frame generator **31** and a second link layer **32**. The second measurement frame generator **31** may be any one of a processor, an FPGA, and a PFC engine, and is configured to perform loopback processing or acknowledgement processing on the received first measurement frame, to trigger a second measurement frame to be generated. The second link layer **32** is responsible for a MAC address of the second switching device **30**. Optionally, the first measurement frame generator **21** of the first switching device **20** generates the first measurement frame. After receiving the first measurement frame, the second link layer **32** of the second switching device **30** forwards the first measurement frame to the second measurement frame generator **31**. The second measurement frame generator **31** performs loopback processing or acknowledgement processing on the received first measurement frame to generate the second measurement frame, and sends the second measurement frame to the second link layer **32**. The second link layer **32** sends the second measurement frame to the first switching device **20** based on the MAC address of the first switching device **20**. A first timestamp t1, a second timestamp t2, a third timestamp t3, and a fourth timestamp t4 are encapsulated in the second measurement frame. The first timestamp t1 represents a time at which the first link layer **22** sends the first measurement frame to the second switching device **30**, the second timestamp t2 represents a time at which the second link layer **32** receives the first measurement frame, the third timestamp t3 represents a time at which the second link layer **32** sends the second measurement frame, and the fourth timestamp t4 represents a time at which the first link layer **22** receives the second measurement frame.

[0102] For example, the first switching device **20** first calculates a first timestamp difference t based on the first timestamp t1, the second timestamp t2, the third timestamp t3, and the fourth timestamp t4, and then calculates a value of an HDRM by using the first timestamp difference, timestamp precision, a port rate between the first switching device and the second switching device, a configured maximum frame length, a PFC measurement frame length, and a constant K. A calculation formula is as follows:

[00001]
$$HDRM = (t + 2 * \text{Timestampprecision}) * \text{SPEED} + 2 * \text{MAXFRAME} + \text{PFCMEASUREMENTFRAME} + K,$$
where

[0103] t=t4−t1−(t3−t2), Timestamp precision represents the timestamp precision, SPEED represents the port rate, MAX FRAME represents the configured maximum frame length, PFC MEASURE FRAME represents the PFC measurement frame length, and K is a constant. A value of K may be provided by a device manufacturer, and includes but is not limited to a receiving/transmission interface delay, a higher-layer delay, or the like defined in 802.1Qbb.

[0104] For example, the first switching device **20** sets an HDRM of a local buffer based on the calculated value of the HDRM. Optionally, the first switching device **20** may also send the calculated

value of the HDRM to the second switching device **30**, so that the second switching device **30** sets an HDRM of a local buffer based on the calculated value of the HDRM.

[0105] According to the structure of communication shown in FIG. **2**, a value of an HDRM of a local buffer can be flexibly independently configured based on a data transmission requirement between switching devices in actual networking, so that the value of the HDRM of the switching device is neither excessively large nor excessively small. In this way, use of buffer space is reduced while no packet loss is ensured, so that the switching device can increase buffer space used by a user to absorb a burst, thereby supporting more lossless queues.

[0106] FIG. **3** is a schematic diagram **2** of a structure of communication between a first switching device **20** and a second switching device **30** according to an embodiment of the present disclosure. As shown in FIG. **3**, the first switching device **20** and the second switching device **30** are communicatively connected through a port. The first switching device **20** includes a first measurement frame generator **21** and a first link layer **22**. The first measurement frame generator **21** may be any one of a processor, an FPGA, and a PFC engine, and is used to generate a first measurement frame. The first link layer **22** is responsible for a MAC address of the first switching device **20**. The second switching device **30** includes a second measurement frame generator **31** and a second link layer **32**. The second measurement frame generator **31** may be any one of a processor, an FPGA, and a PFC engine. The second link layer **32** is responsible for a MAC address of the second switching device **30**. Optionally, the first measurement frame generator **21** of the first switching device **20** generates the first measurement frame. After receiving the first measurement frame, the second link layer **32** of the second switching device **30** directly performs loopback processing or acknowledgement processing on the received first measurement frame to trigger a second measurement frame to be generated, and sends the second measurement frame to the first switching device **20** based on the MAC address of the first switching device **20**. A first timestamp t1 and a fourth timestamp t4 are encapsulated in the second measurement frame. The first timestamp t1 represents a time at which the first link layer **22** sends the first measurement frame to the second switching device **30**, and the fourth timestamp t4 represents a time at which the first link layer **22** receives the second measurement frame.

[0107] For example, the first switching device **20** first calculates a second timestamp difference t′ based on the first timestamp t1 and the fourth timestamp t4, and then calculates a value of an HDRM by using the second timestamp difference, timestamp precision, a port rate between the first switching device and the second switching device, a configured maximum frame length, a PFC measurement frame length, and a constant K. A calculation formula is as follows:

[00002]

$$HDRM = (t^{'} + 2 * Timestampprecision) * SPEED + 2 * MAXFRAME + PFCMEASUREFRAME + K,$$

where

[0108] t′=t4−t1, Timestamp precision represents the timestamp precision, SPEED represents the port rate, MAX FRAME represents the configured maximum frame length, PFC MEASURE FRAME represents the PFC measurement frame length, and K is a constant. A value of K may be provided by a device manufacturer, and includes but is not limited to a receiving/transmission interface delay, a higher-layer delay, or the like defined in 802.1Qbb.

[0109] For example, the first switching device **20** sets an HDRM of a local buffer based on the calculated value of the HDRM. Optionally, the first switching device **20** may also send the calculated value of the HDRM to the second switching device **30**, so that the second switching device **30** sets an HDRM of a local buffer based on the calculated value of the HDRM.

[0110] According to the structure of communication shown in FIG. **3**, a value of an HDRM of a local buffer can be flexibly independently configured based on a data transmission requirement between switching devices in actual networking, so that the value of the HDRM of the switching device is neither excessively large nor excessively small. In this way, use of buffer space is reduced while no packet loss is ensured, so that the switching device can increase buffer space used by a user to absorb a burst, thereby supporting more lossless queues.

[0111] The following describes the buffer configuration method in the present disclosure in detail by

using specific embodiments. It should be noted that the following specific embodiments may be combined with each other, and same or similar content is not repeatedly described in different embodiments.

[0112] FIG. **4** is a flowchart **1** of a buffer configuration method according to an embodiment of the present disclosure. Referring to FIG. **4**, the method in this embodiment includes the following steps.

[0113] S**101**: A first switching device sends a first measurement frame to a second switching device.

[0114] For example, in S**101**, the first switching device and the second switching device are communicatively connected through a port. The first switching device and the second switching device may be network devices or terminal devices. A common network device includes a switch, and a common terminal device includes a network adapter. A first measurement frame generator is disposed inside the first switching device, and the first measurement frame generator may be any one of a processor, an FPGA, and a PFC engine. The first switching device sends the first measurement frame generated by the first measurement frame generator to the second switching device.

[0115] S**102**: The first switching device receives a second measurement frame sent by the second switching device.

[0116] For example, in S**102**, a second measurement frame generator is disposed inside the second switching device, and the second measurement frame generator may be any one of a processor, an FPGA, and a PFC engine. After the second switching device receives the first measurement frame sent by the first switching device, the second measurement frame generator triggers the second measurement frame to be generated, and sends the second measurement frame to the first switching device. Optionally, the second switching device does not process the first measurement frame sent by the first switching device, but directly performs feedback to the first switching device after performing loopback or acknowledgement on the received first measurement frame.

[0117] S**103**: The first switching device determines a buffer configuration parameter based on the first measurement frame and the second measurement frame.

[0118] For example, in S**103**, the first switching device may determine the buffer configuration parameter based on a time of sending the first measurement frame and a time of receiving the second measurement frame. Optionally, the time of sending the first measurement frame and the time of receiving the second measurement frame may be stored in a local buffer as timestamp information, and the buffer configuration parameter may be calculated by reading the timestamp information. Alternatively, the timestamp information may be directly encapsulated in the second measurement frame, and when a MAC layer of the first switching device obtains the second measurement frame, the timestamp information included in the second measurement frame is directly read to calculate the buffer configuration parameter.

[0119] For example, the first switching device obtains a first timestamp and a fourth timestamp, and the first timestamp and the fourth timestamp may represent a data exchange speed between the first switching device and the second switching device. Therefore, the first switching device can determine the buffer configuration parameter based on the first timestamp and the fourth timestamp, so that the configuration parameter fits data transmission performance of the first switching device and the second switching device. The first timestamp and the fourth timestamp are encapsulated in the second measurement frame, so that the first switching device can omit a timestamp reading step, thereby increasing the speed of obtaining timestamp information.

[0120] In an optional implementation, the first timestamp is a time at which the last bit of the first measurement frame enters the MAC layer of the first switching device, and the fourth timestamp is a time at which the last bit of the second measurement frame is sent from the MAC layer of the first switching device.

[0121] In another optional implementation, the first timestamp is a time at which the last bit of the first measurement frame is sent from the MAC layer of the first switching device, and the fourth timestamp is a time at which the last bit of the second measurement frame enters the MAC layer of the first switching device.

[0122] FIG. **5** is a schematic diagram of a structure of a PFC frame according to an embodiment of the

present disclosure. As shown in FIG. **5**, the PFC frame sequentially includes a destination address, a source address, a length/type (Length/Type=0x8808), an operation code (Opcode), a priority enable vector, time fields (time[0], time[n], and time[7]), pad, and a frame check sequence (FCS) from top to bottom. An octet represents a unit.

[0123] Because the PFC frame is a MAC control frame, the MAC layer of the first switching device can directly identify the Opcode field, and then read the time field following the Opcode field. For example, a value assigned to the Opcode field in the existing PFC frame may be modified to obtain a modified frame structure. The modified frame structure may be used as a frame structure of the second measurement frame. For example, in the modified frame structure, a length of 16 octets is set in the time field following the Opcode field to store a timestamp.

[0124] For example, the first timestamp and the fourth timestamp may be stored in the time field following the Opcode field of the second measurement frame.

[0125] For example, the first switching device calculates a timestamp difference based on the first timestamp and the fourth timestamp that are included in the second measurement frame, and then determines the buffer configuration parameter based on the timestamp difference, timestamp precision, a port rate between the first switching device and the second switching device, and a configured maximum frame length. Therefore, the first switching device can obtain the buffer configuration parameter associated with the exchange speed and the port rate between the first switching device and the second switching device and the configured maximum frame length, to implement an accurate configuration of the local buffer, thereby reducing use of the local buffer of the switching device, and improving a burst absorption capability and supporting more lossless queues while ensuring no packet loss.

[0126] In an optional implementation, the first switching device may further obtain a second timestamp and a third timestamp. The second timestamp is a time at which the second switching device receives the first measurement frame, and the third timestamp is a time at which the second switching device sends the second measurement frame. For example, the time at which the second switching device receives the first measurement frame is recorded as the second timestamp, and the time at which the second switching device sends the second measurement frame is recorded as the third timestamp. The second timestamp and the third timestamp may be used to represent delay time from receiving, by the second switching device, the first measurement frame to triggering the second measurement frame to be generated. Therefore, the first switching device can determine the buffer configuration parameter based on the first timestamp, the second timestamp, the third timestamp, and the fourth timestamp, so that the configuration parameter fits data transmission performance of the first switching device and the second switching device.

[0127] For example, the second timestamp is a time at which the last bit of the first measurement frame is sent from a MAC layer of the second switching device, and the third timestamp is a time at which the last bit of the second measurement frame enters the MAC layer of the second switching device.

[0128] For example, the second timestamp is a time at which the last bit of the first measurement frame enters a MAC layer of the second switching device, and the third timestamp is a time at which the last bit of the second measurement frame is sent from the MAC layer of the second switching device.

[0129] In an optional implementation, the second timestamp and the third timestamp are encapsulated in the time field following the Opcode field of the second measurement frame. For example, the second timestamp and the third timestamp may be encapsulated in the second measurement frame, so that a timestamp reading step can be omitted, and the first switching device can quickly obtain timestamp information. For example, a PFC frame format may be used as a format of the second measurement frame to store the second timestamp and the third timestamp in the time field following the Opcode field of the second measurement frame. Because the PFC frame is a MAC control frame, the MAC layer of the first switching device can directly identify the Opcode field, and then read the time field following the Opcode field to quickly obtain the timestamp information.

[0130] In an optional implementation, the second timestamp and the third timestamp may be alternatively stored in the local buffer of the first switching device. It should be noted that when the second timestamp and the third timestamp are stored in the local buffer of the first switching device, the first switching device needs to obtain timestamp information from the local buffer.

[0131] In an optional implementation, the first switching device determines a timestamp difference based on the first timestamp, the second timestamp, the third timestamp, and the fourth timestamp. The first switching device determines the buffer configuration parameter based on the timestamp difference, timestamp precision, a port rate between the first switching device and the second switching device, a configured maximum frame length, and a constant K. The constant K is provided by a device manufacturer. For example, the first switching device first calculates the timestamp difference based on the first timestamp, the second timestamp, the third timestamp, and the fourth timestamp, and then determines the buffer configuration parameter based on the timestamp difference, the timestamp precision, the port rate between the first switching device and the second switching device, and the configured maximum frame length. Therefore, the first switching device can obtain the buffer configuration parameter associated with the exchange speed and the port rate between the first switching device and the second switching device and the configured maximum frame length, to implement an accurate configuration of the local buffer, thereby reducing use of the local buffer of the switching device, and improving a burst absorption capability and supporting more lossless queues while ensuring no packet loss.

[0132] S**104**: The first switching device sets the local buffer based on the configuration parameter.

[0133] For example, in S**104**, the first switching device sets a size of the local buffer to the value calculated in S**103**.

[0134] For example, the buffer configuration parameter is used to represent a required buffer size, and a unit of the buffer configuration parameter includes a bit or a minimum storage unit of a chip buffer. The buffer configuration parameters include a buffer space required for priority-based flow control.

[0135] For example, referring to FIG. **2** and FIG. **3**, when the configuration parameter is a value of an HDRM, the first switching device may calculate the timestamp difference T based on the first timestamp, the second timestamp, the third timestamp, and the fourth timestamp that are encapsulated in the second measurement frame, and then calculate the HDRM by using a formula. A specific calculation formula is as follows:

[00003]

$$HDRM = (T + 2 * \text{Timestampprecision}) * \text{SPEED} + 2 * \text{MAXFRAME} + \text{PFCMEASUREFRAME} + K,$$
where

[0136] T represents the timestamp difference, Timestamp precision represents the timestamp precision, SPEED represents the port rate, MAX FRAME represents the configured maximum frame length, PFC MEASURE FRAME represents a PFC measurement frame length, and K is a constant. A value of K may be provided by the device manufacturer, and includes but is not limited to a receiving/transmission interface delay, a higher-layer delay, or the like defined in 802.1Qbb.

[0137] In this embodiment, a timestamp difference of data frame transmission between the first switching device and the second switching device is obtained by constructing the first measurement frame and the second measurement frame and encapsulating a timestamp in the second measurement frame, and a buffer size required by the first switching device in a PFC mechanism is calculated based on the time difference, thereby implementing an automatic configuration of the local buffer on the side of the first switching device. More accurate calculation reduces use of the local buffer of the switching device, and can improve a burst absorption capability and support more lossless queues while ensuring no packet loss.

[0138] FIG. **6** is a flowchart **2** of a buffer configuration method according to an embodiment of the present disclosure. Referring to FIG. **6**, the method in this embodiment includes the following steps.

[0139] S**201**: A first switching device sends a first measurement frame to a second switching device.

[0140] S**202**: The first switching device receives a second measurement frame sent by the second switching device.

[0141] S**203**: The first switching device determines a buffer configuration parameter based on the first measurement frame and the second measurement frame.

[0142] S**204**: The first switching device sets a local buffer based on the configuration parameter.

[0143] For example, for a specific implementation process and implementation principle of S**201** to S**204** in this embodiment, refer to related descriptions of S**101** to S**104** in the method shown in FIG. **4**. Details are not described herein again.

[0144] S**205**: The first switching device sends the buffer configuration parameter to the second switching device, so that the second switching device sets a local buffer based on the configuration parameter.

[0145] For example, in S**205**, after calculating the buffer configuration parameter in S**203**, the first switching device may further send the configuration parameter to the second switching device, so that the second switching device also sets the local buffer based on the calculated configuration parameter.

[0146] In this embodiment, a time difference of data frame transmission between the first switching device and the second switching device is obtained by constructing the first measurement frame and the second measurement frame and encapsulating a timestamp in the second measurement frame, and a buffer size required by the first switching device in a PFC mechanism is calculated based on the time difference, thereby implementing an automatic configuration of the local buffer on the side of the second switching device. More accurate calculation reduces use of the local buffer of the switching device, and can improve a burst absorption capability and support more lossless queues while ensuring no packet loss.

[0147] FIG. **7** is a flowchart **3** of a buffer configuration method according to an embodiment of the present disclosure. Referring to FIG. **7**, the method in this embodiment includes the following steps.

[0148] S**301**: Determine whether both a first switching device and a second switching device support an automatic buffer configuration function of priority-based flow control; and if yes, perform S**302**; or if no, perform S**306**.

[0149] For example, in S**301**, the first switching device sends TLV (the TLV is a variable format, and the TLV means a type, a length, and a value)-format data including a buffer configuration field to the second switching device. If TLV-format data that is received by the first switching device and that is fed back by the second switching device also includes the buffer configuration field, it is determined that both the first switching device and the second switching device support the automatic buffer configuration function of priority-based flow control.

[0150] For example, in S**301**, the first switching device receives the TLV-format data that is sent by the second switching device and that includes the buffer configuration field. If the TLV-format data that is fed back by the first switching device to the second switching device also includes the buffer configuration field, it is determined that both the first switching device and the second switching device support the automatic buffer configuration function of priority-based flow control.

[0151] S**302**: The first switching device sends a first measurement frame to the second switching device.

[0152] S**303**: The first switching device receives a second measurement frame sent by the second switching device.

[0153] S**304**: The first switching device determines a buffer configuration parameter based on the first measurement frame and the second measurement frame.

[0154] S**305**: The first switching device sets a local buffer based on the configuration parameter.

[0155] For example, for a specific implementation process and implementation principle of S**302** to S**305** in this embodiment, refer to related descriptions of S**101** to S**104** in the method shown in FIG. **4**. Details are not described herein again.

[0156] S**306**: Give a prompt that needs to be manually set.

[0157] For example, in this embodiment, the first switching device and the second switching device first notify, by using an LLDP (link layer discovery protocol), whether each other has an automatic configuration function of a PFC HDRM. Optionally, this may be implemented by extending PFC configuration TLV. FIG. **8** is a schematic diagram of an architecture of existing IEEE DCBX PFC

TLV, and FIG. **9** is a schematic diagram of an architecture of extended and updated IEEE DCBX PFC TLV. 1 bit is removed from a reserved field of the existing IEEE DCBX PFC TLV in FIG. **8** to store an HDRM, to obtain the extended and updated IEEE DCBX PFC TLV shown in FIG. **9**. As shown in FIG. **8**, fields are sequentially a TLV type, a TLV information string length, an 802.1 organizationally unique identifier, an 802.1 subtype, willing, a MACsec bypass capability, the reserved field, a PFC capability, and PFC enable from left to right. As shown in FIG. **9**, fields are sequentially a TLV type, a TLV information string length, an 802.1 organizationally unique identifier, an 802.1 subtype, willing, a MACsec bypass capability, supporting automatic HDRM measurement, a reserved field, a PFC capability, and PFC enable from left to right.

[0158] For example, FIG. **10** is a schematic diagram of an architecture of existing first-version Intel DCBX PFC TLV, and FIG. **11** is a schematic diagram of an architecture of extended and updated first-version Intel DCBX PFC TLV. 1 bit is removed from a reserved field of the existing first-version Intel DCBX PFC TLV in FIG. **10** to store an HDRM, to obtain the extended and updated first-version Intel DCBX PFC TLV shown in FIG. **11**. As shown in FIG. **10**, fields are sequentially a TLV type, a TLV information string length, an operating version, a maximum version, enable, willing, an error, the reserved field, a subtype, and a bit mapping in an administration mode from left to right. As shown in FIG. **11**, fields are sequentially a TLV type, a TLV information string length, an operating version, a maximum version, enable, willing, an error, supporting automatic HDRM measurement, a reserved field, a subtype, and a bit mapping in an administration mode from left to right.

[0159] For example, FIG. **12** is a schematic diagram of an architecture of existing second-version Intel DCBX PFC TLV, and FIG. **13** is a schematic diagram of an architecture of extended and updated second-version Intel DCBX PFC TLV. 1 bit is removed from a reserved field of the existing second-version Intel DCBX PFC TLV in FIG. **12** to store an HDRM, to obtain the extended and updated second-version Intel DCBX PFC TLV shown in FIG. **13**. As shown in FIG. **12**, fields are sequentially a TLV type, a TLV information string length, an operating version, a maximum version, enable, willing, an error, the reserved field, a subtype, a PFC enable priority, and a maximum number of supported traffic classes from left to right. As shown in FIG. **13**, fields are sequentially a TLV type, a TLV information string length, an operating version, a maximum version, enable, willing, an error, supporting automatic HDRM measurement, a reserved field, a subtype, a PFC enable priority, and a maximum number of supported traffic classes from left to right.

[0160] A time difference of data frame transmission between the first switching device and the second switching device is obtained by constructing a PFC HDRM measurement frame and encapsulating a timestamp in the second measurement frame, and a size of an HDRM is calculated based on the time difference, thereby implementing an automatic HDRM configuration on the side of the first switching device and the side of the second switching device. More accurate calculation reduces use of the local buffer of the switching device, and can improve a burst absorption capability and support more lossless queues while ensuring no packet loss.

[0161] FIG. **14** is a flowchart **4** of a buffer configuration method according to an embodiment of the present disclosure. Referring to FIG. **14**, the method in this embodiment includes the following steps.

[0162] S**401**: A second switching device receives a first measurement frame sent by a first switching device, and triggers, based on the first measurement frame, a second measurement frame to be generated.

[0163] For example, in S**401**, the first switching device and the second switching device are communicatively connected through a port. The first switching device and the second switching device may be network devices or terminal devices. A common network device includes a switch, and a common terminal device includes a network adapter. A first measurement frame generator is disposed inside the first switching device, and the first measurement frame generator may be any one of a processor, an FPGA, and a PFC engine. The second switching device receives the first measurement frame sent by the first switching device. The first measurement frame is generated by the first measurement frame generator of the first switching device.

[0164] For example, a second measurement frame generator is disposed inside the second switching

device, and the second measurement frame generator may be any one of a processor, an FPGA, and a PFC engine. After receiving the first measurement frame sent by the first switching device, the second switching device sends the second measurement frame generated by the second measurement frame generator to the first switching device. Optionally, the second switching device does not process the first measurement frame sent by the first switching device, but directly performs feedback to the first switching device after performing loopback or acknowledgement on the received first measurement frame.

[0165] S**402**: The second switching device sends the second measurement frame to the first switching device.

[0166] For example, in S**402**, the second switching device sends the second measurement frame to the first switching device based on a MAC address of the first switching device.

[0167] S**403**: The second switching device receives a buffer configuration parameter sent by the first switching device.

[0168] For example, in S**403**, after calculating the buffer configuration parameter, the first switching device may further send the configuration parameter to the second switching device.

[0169] S**404**: The second switching device sets a local buffer based on the configuration parameter.

[0170] For example, in S**404**, the second switching device sets the local buffer based on the configuration parameter calculated by the first switching device. A time difference of data frame transmission between the first switching device and the second switching device is obtained by constructing the first measurement frame and the second measurement frame and encapsulating a timestamp in the second measurement frame, and a size of an HDRM is calculated based on the time difference, thereby implementing an automatic HDRM configuration on the side of the second switching device. More accurate calculation reduces use of the local buffer of the switching device, and can improve a burst absorption capability and support more lossless queues while ensuring no packet loss.

[0171] FIG. **15** is a flowchart **5** of a buffer configuration method according to an embodiment of the present disclosure. Referring to FIG. **15**, the method in this embodiment includes the following steps.

[0172] S**501**: Determine whether both a first switching device and a second switching device support an automatic buffer configuration function of priority-based flow control; and if yes, perform S**502**; or if no, perform S**507**.

[0173] For example, in S**501**, the second switching device receives TLV-format data that is sent by the first switching device and that includes a buffer configuration field. If TLV-format data that is fed back by the second switching device to the first switching device also includes the buffer configuration field, it is determined that both the first switching device and the second switching device support the automatic buffer configuration function of priority-based flow control.

[0174] For example, in S**501**, the second switching device sends the TLV-format data including the buffer configuration field to the first switching device. If the TLV-format data that is received by the second switching device and that is fed back by the first switching device also includes the buffer configuration field, it is determined that both the first switching device and the second switching device support the automatic buffer configuration function of priority-based flow control.

[0175] S**502**: The second switching device receives a first measurement frame sent by the first switching device.

[0176] S**503**: The second switching device performs loopback processing on the first measurement frame to obtain a second measurement frame.

[0177] S**504**: The second switching device sends the second measurement frame to the first switching device.

[0178] S**505**: The second switching device receives a buffer configuration parameter sent by the first switching device.

[0179] S**506**: The second switching device sets a local buffer based on the configuration parameter.

[0180] For example, for a specific implementation process and implementation principle of S**502** to S**506** in this embodiment, refer to related descriptions of S**401** to S**405** in the method shown in FIG.

14. Details are not described herein again.

[0181] S**507**: Give a prompt that needs to be manually set.

[0182] In this embodiment, a time difference of data frame transmission between the first switching device and the second switching device is obtained by constructing the first measurement frame and the second measurement frame and encapsulating a timestamp in the second measurement frame, and a buffer size required by the first switching device in a PFC mechanism is calculated based on the time difference, thereby implementing an automatic configuration of local buffers on the side of the first switching device and the side of the second switching device. More accurate calculation reduces use of the local buffer of the switching device, and can improve a burst absorption capability and support more lossless queues while ensuring no packet loss.

[0183] FIG. **16** is a schematic diagram **1** of a structure of a first switching device according to an embodiment of the present disclosure. Referring to FIG. **16**, the first switching device in this embodiment includes a processor **61** and a transceiver **62**.

[0184] The transceiver **62** is configured to send a first measurement frame to a second switching device.

[0185] The transceiver **62** is configured to receive a second measurement frame sent by the second switching device. The second measurement frame is generated through triggering based on the first measurement frame.

[0186] The processor **61** is configured to determine a buffer configuration parameter based on the first measurement frame and the second measurement frame.

[0187] The processor **61** is configured to set a local buffer based on the configuration parameter.

[0188] In this possible design, the first switching device may be applied to a networking architecture including the first switching device and the second switching device, and the first switching device and the second switching device are communicatively connected through a port. The first switching device and the second switching device may be network devices or terminal devices. A common network device includes a switch, and a common terminal device includes a network adapter. A first measurement frame generator is disposed inside the first switching device, and the first measurement frame generator may be any one of a processor, an FPGA, and a PFC engine. The first switching device sends the first measurement frame to the second switching device. After receiving the first measurement frame, the second switching device triggers the second measurement frame to be generated. The first switching device determines the configuration parameter of the local buffer based on the first measurement frame and the second measurement frame, to implement an automatic configuration of the local buffer, thereby reducing use of the local buffer of the switching device, and improving a burst absorption capability and supporting more lossless queues while ensuring no packet loss.

[0189] In a possible design, the processor **61** is specifically configured to: [0190] obtain a first timestamp and a fourth timestamp, where the first timestamp is a time at which the first switching device sends the first measurement frame, and the fourth timestamp is a time at which the first switching device receives the second measurement frame; and [0191] determine the buffer configuration parameter based on the first timestamp and the fourth timestamp.

[0192] In this possible design, the time at which the first switching device sends the first measurement frame is recorded as the first timestamp, and the time at which the first switching device receives the second measurement frame is recorded as the fourth timestamp. The first timestamp and the fourth timestamp may be used to represent a data exchange speed between the first switching device and the second switching device. Therefore, the first switching device can determine the buffer configuration parameter based on the first timestamp and the fourth timestamp, so that the configuration parameter fits data transmission performance of the first switching device and the second switching device.

[0193] In a possible design, the first timestamp is a time at which the last bit of the first measurement frame enters a MAC layer of the first switching device, and the fourth timestamp is a time at which the last bit of the second measurement frame is sent from the MAC layer of the first switching device.

[0194] In a possible design of the first aspect, the first timestamp is a time at which the last bit of the

first measurement frame is sent from a MAC layer of the first switching device, and the fourth timestamp is a time at which the last bit of the second measurement frame enters the MAC layer of the first switching device.

[0195] In a possible design of the first aspect, the first timestamp and the fourth timestamp are encapsulated in a time field following an Opcode field of the second measurement frame.

[0196] In this possible design, the first timestamp and the fourth timestamp may be encapsulated in the second measurement frame, so that a timestamp reading step can be omitted, and the first switching device can quickly obtain timestamp information. For example, a PFC frame format may be used as a format of the second measurement frame to store the first timestamp and the fourth timestamp in the time field following the Opcode field of the second measurement frame. Because a PFC frame is a MAC control frame, the MAC layer of the first switching device can directly identify the Opcode field, and then read the time field following the Opcode field to quickly obtain the timestamp information.

[0197] In a possible design of the first aspect, the first timestamp and the fourth timestamp are stored in the local buffer of the first switching device.

[0198] In this possible design, the first timestamp and the fourth timestamp are stored in the local buffer of the first switching device, and the first switching device obtains timestamp information from the local buffer.

[0199] In a possible design, the processor **61** is specifically configured to: [0200] determine a timestamp difference based on the first timestamp and the fourth timestamp; and [0201] determine the buffer configuration parameter based on the timestamp difference, timestamp precision, a port rate between the first switching device and the second switching device, a configured maximum frame length, and a constant K provided by a device manufacturer.

[0202] In this possible design, the first switching device first calculates the timestamp difference based on the first timestamp and the fourth timestamp, and then determines the buffer configuration parameter based on the timestamp difference, the timestamp precision, the port rate between the first switching device and the second switching device, and the configured maximum frame length. Therefore, the first switching device can obtain the buffer configuration parameter associated with the exchange speed and the port rate between the first switching device and the second switching device and the configured maximum frame length, to implement an accurate configuration of the local buffer, thereby reducing use of the local buffer of the switching device, and improving a burst absorption capability and supporting more lossless queues while ensuring no packet loss.

[0203] In a possible design, the processor **61** is further configured to: [0204] obtain a second timestamp and a third timestamp, where the second timestamp is a time at which the second switching device receives the first measurement frame, and the third timestamp is a time at which the second switching device sends the second measurement frame.

[0205] In this possible design, the time at which the second switching device receives the first measurement frame is recorded as the second timestamp, and the time at which the second switching device sends the second measurement frame is recorded as the third timestamp. The second timestamp and the third timestamp may be used to represent delay time from receiving, by the second switching device, the first measurement frame to triggering the second measurement frame to be generated. Therefore, the first switching device can determine the buffer configuration parameter based on the first timestamp, the second timestamp, the third timestamp, and the fourth timestamp, so that the configuration parameter fits data transmission performance of the first switching device and the second switching device.

[0206] In a possible design of the first aspect, the second timestamp is a time at which the last bit of the first measurement frame is sent from a MAC layer of the second switching device, and the third timestamp is a time at which the last bit of the second measurement frame enters the MAC layer of the second switching device.

[0207] In a possible design of the first aspect, the second timestamp is a time at which the last bit of the first measurement frame enters a MAC layer of the second switching device, and the third

timestamp is a time at which the last bit of the second measurement frame is sent from the MAC layer of the second switching device.

[0208] In a possible design of the first aspect, the second timestamp and the third timestamp are encapsulated in the time field following the Opcode field of the second measurement frame.

[0209] In this possible design, the second timestamp and the third timestamp may be encapsulated in the second measurement frame, so that a timestamp reading step can be omitted, and the first switching device can quickly obtain timestamp information. For example, the PFC frame format may be used as the format of the second measurement frame to store the second timestamp and the third timestamp in the time field following the Opcode field of the second measurement frame. Because the PFC frame is a MAC control frame, the MAC layer of the first switching device can directly identify the Opcode field, and then read the time field following the Opcode field to quickly obtain the timestamp information.

[0210] In a possible design of the first aspect, the second timestamp and the third timestamp are stored in the local buffer of the first switching device.

[0211] In this possible design, the second timestamp and the third timestamp are stored in the local buffer of the first switching device, and the first switching device obtains timestamp information from the local buffer.

[0212] In a possible design, the processor **61** is specifically configured to: [0213] determine a timestamp difference based on the first timestamp, the second timestamp, the third timestamp, and the fourth timestamp; and [0214] determine the buffer configuration parameter based on the timestamp difference, timestamp precision, a port rate between the first switching device and the second switching device, a configured maximum frame length, and a constant K provided by a device manufacturer.

[0215] In this possible design, the first switching device first calculates the timestamp difference based on the first timestamp, the second timestamp, the third timestamp, and the fourth timestamp, and then determines the buffer configuration parameter based on the timestamp difference, the timestamp precision, the port rate between the first switching device and the second switching device, and the configured maximum frame length. Therefore, the first switching device can obtain the buffer configuration parameter associated with the exchange speed and the port rate between the first switching device and the second switching device and the configured maximum frame length, to implement an accurate configuration of the local buffer, thereby reducing use of the local buffer of the switching device, and improving a burst absorption capability and supporting more lossless queues while ensuring no packet loss.

[0216] In a possible design, the processor **61** is specifically configured to: [0217] send the buffer configuration parameter to the second switching device, so that the second switching device sets a local buffer based on the configuration parameter.

[0218] In this possible design, the second switching device may further set the local buffer of the second switching device by using the buffer configuration parameter calculated by the first switching device, so that the second switching device can also implement an automatic configuration of the buffer.

[0219] In a possible design, the processor **61** is further configured to: [0220] determine whether both the first switching device and the second switching device support an automatic buffer configuration function of priority-based flow control.

[0221] In this possible design, only when both the first switching device and the second switching device support the automatic buffer configuration function of priority-based flow control, the foregoing method can be applied to implement an automatic configuration of the local buffer. Therefore, before the method is applied, it is first determined whether both the first switching device and the second switching device support the automatic buffer configuration function of priority-based flow control.

[0222] In a possible design, the determining whether both the first switching device and the second switching device support an automatic buffer configuration function of priority-based flow control

includes: [0223] when the first switching device supports the automatic buffer configuration function of priority-based flow control, and received TLV-format data fed back by the second switching device includes a buffer configuration field, determining that both the first switching device and the second switching device support the automatic buffer configuration function of priority-based flow control.

[0224] In this possible design, a link layer discovery protocol (LLDP) may be used between the first switching device and the second switching device to determine whether both the first switching device and the second switching device support the automatic buffer configuration function of priority-based flow control.

[0225] In a possible design, the automatic buffer configuration function is not supported by either the first or second switching device, the processor **61** is further configured to: [0226] give a prompt that the local buffer needs to be manually set.

[0227] In a possible design, the buffer configuration parameter is used to represent a required buffer size, and a unit of the buffer configuration parameter includes a bit or a minimum storage unit of a chip buffer.

[0228] In a possible design, the buffer configuration parameter includes buffer space required for priority-based flow control.

[0229] In a possible design, the second measurement frame is generated by the second switching device by performing loopback or acknowledgement processing on the first measurement frame.

[0230] In this embodiment, a time difference of data frame transmission between the first switching device and the second switching device is obtained by constructing the first measurement frame and the second measurement frame and encapsulating a timestamp in the second measurement frame, and a buffer size required by the first switching device in a PFC mechanism is calculated based on the time difference, thereby implementing an automatic configuration of local buffers on the side of the first switching device and the side of the second switching device. More accurate calculation reduces use of the local buffer of the switching device, and can improve a burst absorption capability and support more lossless queues while ensuring no packet loss.

[0231] FIG. **17** is a schematic diagram **1** of a structure of a second switching device according to an embodiment of the present disclosure. Referring to FIG. **17**, the second switching device in this embodiment includes a processor **71** and a transceiver **72**.

[0232] The transceiver **72** is configured to: receive a first measurement frame sent by a first switching device, and trigger, based on the first measurement frame, a second measurement frame to be generated.

[0233] The transceiver **72** is configured to send the second measurement frame to the first switching device.

[0234] The transceiver **72** is configured to receive a buffer configuration parameter sent by the first switching device.

[0235] The processor **71** is configured to set a local buffer based on the configuration parameter.

[0236] In this possible design, the second switching device may be applied to a networking architecture including the first switching device and the second switching device, and the first switching device and the second switching device are communicatively connected through a port. The first switching device and the second switching device may be network devices or terminal devices. A common network device includes a switch, and a common terminal device includes a network adapter. A second measurement frame generator is disposed inside the second switching device, and the second measurement frame generator may be any one of a processor, an FPGA, and a PFC engine. After receiving the first measurement frame sent by the first switching device, the second switching device triggers the second measurement frame to be generated, and sends the second measurement frame to the first switching device, so that the first switching device determines the configuration parameter of the local buffer based on the first measurement frame and the second measurement frame. Finally, the second switching device configures the local buffer based on the buffer configuration parameter determined by the first switching device, to implement an automatic configuration of the local buffer, thereby reducing use of the local buffer of the switching device, and improving an absorption burst

capability and supporting more lossless queues while ensuring no packet loss.

[0237] In a possible design, the processor **71** is further configured to: [0238] determine whether both the first switching device and the second switching device support an automatic buffer configuration function of priority-based flow control.

[0239] In this possible design, only when both the first switching device and the second switching device support the automatic buffer configuration function of priority-based flow control, the foregoing method can be applied to implement an automatic configuration of the local buffer. Therefore, before the method is applied, it is first determined whether both the first switching device and the second switching device support the automatic buffer configuration function of priority-based flow control.

[0240] In a possible design, the processor **71** is specifically configured to: [0241] when the second switching device supports the automatic buffer configuration function of priority-based flow control, and received TLV-format data fed back by the first switching device includes a buffer configuration field, determining that both the first switching device and the second switching device support the automatic buffer configuration function of priority-based flow control.

[0242] In this possible design, a link layer discovery protocol (LLDP) may be used between the first switching device and the second switching device to determine whether both the first switching device and the second switching device support the automatic buffer configuration function of priority-based flow control.

[0243] In a possible design, the processor **71** is further configured to: [0244] give a prompt that the local buffer needs to be manually set.

[0245] In a possible design, the buffer configuration parameter is used to represent a required buffer size, and a unit of the buffer configuration parameter includes a bit or a minimum storage unit of a chip buffer.

[0246] In a possible design, the buffer configuration parameter includes buffer space required for priority-based flow control.

[0247] In a possible design, the triggering, based on the first measurement frame, a second measurement frame to be generated includes:

[0248] The second switching device performs loopback or acknowledgement processing on the first measurement frame to generate the second measurement frame.

[0249] In this embodiment, a time difference of data frame transmission between the first switching device and the second switching device is obtained by constructing the first measurement frame and the second measurement frame and encapsulating a timestamp in the second measurement frame, and a buffer size required by the first switching device in a PFC mechanism is calculated based on the time difference, thereby implementing an automatic configuration of local buffers on the side of the first switching device and the side of the second switching device. More accurate calculation reduces use of the local buffer of the switching device, and can improve a burst absorption capability and support more lossless queues while ensuring no packet loss.

[0250] FIG. **18** is a schematic diagram of a structure of a networking system according to an embodiment of the present disclosure. Referring to FIG. **18**, the networking system in this embodiment includes a first switching device **81** and a second switching device **82**, and the first switching device **81** and the second switching device **82** are communicatively connected through a communication port **83**.

[0251] For example, the first switching device **81** performs the methods shown in FIG. **4**, FIG. **6**, and FIG. **7**, and the second switching device performs the methods shown in FIG. **14** and FIG. **15**. For a specific implementation process and implementation principle of the first switching device and the second switching device, refer to related descriptions in the embodiments shown in FIG. **3** to FIG. **7**, FIG. **14**, and FIG. **15**. Details are not described herein again.

[0252] An embodiment of the present disclosure provides a computer-readable storage medium. The computer-readable storage medium stores instructions, and when the instructions are executed, a computer performs the method performed by the terminal device in the foregoing embodiment of the

present disclosure.

[0253] An embodiment of the present disclosure provides a computer-readable storage medium. The computer-readable storage medium stores instructions, and when the instructions are executed, a computer performs the method performed by the network device in the foregoing embodiment of the present disclosure.

[0254] In the several embodiments provided in the present disclosure, it should be understood that the disclosed apparatuses and methods may be implemented in other manners. For example, the foregoing apparatus embodiments are merely examples. For example, division into the units is merely logical function division, and may be other division during actual implementation. For example, a plurality of units or components may be combined or integrated into another system, or some features may be ignored or not performed. In addition, the displayed or discussed mutual couplings or direct couplings or communication connections may be implemented through some interfaces. The indirect couplings or communication connections between the apparatuses or units may be implemented in electrical, mechanical, or another form.

[0255] The units described as separate parts may or may not be physically separate, and parts displayed as units may or may not be physical units, in other words, may be located in one position, or may be distributed on a plurality of network units. Some or all of the units may be selected based on actual requirements to achieve the objectives of the solutions of embodiments.

[0256] In addition, functional units in embodiments of the present disclosure may be integrated into one processing unit, each of the units may exist alone physically, or two or more units may be integrated into one unit. The integrated unit may be implemented in a form of hardware, or may be implemented in a form of hardware in addition to a software functional unit.

[0257] It should be noted that, in embodiments of the present disclosure, division into the modules is an example and is merely logical function division, and may be other division during actual implementation. Functional modules in embodiments of the present disclosure may be integrated into one processing module, or each of the modules may exist alone physically, or two or more modules are integrated into one module. The integrated module may be implemented in a form of hardware, or may be implemented in a form of a software functional module.

[0258] When the integrated module is implemented in the form of a software functional module and sold or used as an independent product, the integrated module may be stored in a computer-readable storage medium. Based on such an understanding, the technical solutions of the present disclosure essentially, or the part contributing to the conventional technology, or all or a part of the technical solutions may be implemented in a form of a computer software product. The computer software product is stored in a storage medium and includes several instructions for instructing a computer device (which may be a personal computer, a server, or a network device) or a processor to perform all or a part of the steps of the methods in embodiments of the present disclosure. The foregoing storage medium includes any medium that can store program code, such as a USB flash drive, a removable hard disk, a read-only memory (ROM), a random access memory (RAM), a magnetic disk, or an optical disc.

[0259] All or some of the foregoing embodiments may be implemented by using software, hardware, firmware, or any combination thereof. When software is used to implement the embodiments, all or a part of the embodiments may be implemented in a form of a computer program product. The computer program product includes one or more computer instructions. When the computer instructions are loaded and executed on a computer, all or some of the procedures or functions according to embodiments of the present disclosure are generated. The computer may be a general-purpose computer, a dedicated computer, a computer network, or another programmable apparatus. The computer instructions may be stored in a computer-readable storage medium or may be transmitted from a computer-readable storage medium to another computer-readable storage medium. For example, the computer instructions may be transmitted from a website, computer, server, or data center to another website, computer, server, or data center in a wired (for example, a coaxial cable, an optical fiber, or a digital subscriber line (DSL)) or wireless (for example, infrared, radio, or microwave)

manner. The computer-readable storage medium may be any usable medium accessible by a computer, or a data storage device, such as a server or a data center, integrating one or more usable media. The usable medium may be a magnetic medium (for example, a floppy disk, a hard disk, or a magnetic tape), an optical medium (for example, a digital versatile disc (DVD)), a semiconductor medium (for example, a solid-state disk (SSD)), or the like.

Claims

1. A buffer configuration method, comprising: sending, by a first switching device, a first measurement frame to a second switching device; receiving, by the first switching device, a second measurement frame sent by the second switching device, wherein the second measurement frame is generated through triggering based on the first measurement frame; determining, by the first switching device, a value of a priority-based flow control (PFC) headroom based on the first measurement frame and the second measurement frame; and setting, by the first switching device, a local buffer based on the value of the PFC headroom.

2. The method according to claim 1, wherein the determining, by the first switching device, a value of a PFC headroom based on the first measurement frame and the second measurement frame comprises: obtaining, by the first switching device, a first timestamp and a fourth timestamp, wherein the first timestamp is a time at which the first switching device sends the first measurement frame, and the fourth timestamp is a time at which the first switching device receives the second measurement frame; and determining, by the first switching device, the value of the PFC headroom based on the first timestamp and the fourth timestamp.

3. The method according to claim 2, wherein the first timestamp is a time at which the last bit of the first measurement frame enters a MAC layer of the first switching device, and the fourth timestamp is a time at which the last bit of the second measurement frame is sent from the MAC layer of the first switching device.

4. The method according to claim 2, wherein the first timestamp is a time at which the last bit of the first measurement frame is sent from a MAC layer of the first switching device, and the fourth timestamp is a time at which the last bit of the second measurement frame enters the MAC layer of the first switching device.

5. The method according to claim 2, wherein the first timestamp and the fourth timestamp are stored in the local buffer of the first switching device.

6. The method according to claim 2, wherein the determining, by the first switching device, the value of the PFC headroom based on the first timestamp and the fourth timestamp comprises: determining, by the first switching device, a timestamp difference based on the first timestamp and the fourth timestamp; and determining, by the first switching device, the value of the PFC headroom based on the timestamp difference, timestamp precision, a port rate between the first switching device and the second switching device, a configured maximum frame length, and a constant K, wherein the constant K is provided by a device manufacturer.

7. The method according to claim 1, further comprising: sending, by the first switching device, the value of the PFC headroom to the second switching device, so that the second switching device sets a local buffer based on the value of the PFC headroom.

8. The method according to claim 1, further comprising: determining, by the first switching vice, whether both the first switching device and the second switching device support an automatic buffer configuration function of priority-based flow control.

9. The method according to claim 8, wherein the determining, by the first switching device, whether both the first switching device and the second switching device support an automatic buffer configuration function of priority-based flow control comprises: when the first switching device supports the automatic buffer configuration function of priority-based flow control, and received TLV-format data fed back by the second switching device comprises a buffer configuration field, determining, by the first switching device, that both the first switching device and the second

switching device support the automatic buffer configuration function of priority-based flow control.

10. The method according to claim 8, wherein, when the first switching device and the second switching device comprise a switching device that does not support the automatic buffer configuration function of priority-based flow control, the method further comprises: giving, by the first switching device, a prompt that the local buffer needs to be manually set.

11. The method according to claim 1, wherein the second measurement frame is generated by the second switching device by acknowledgement processing on the first measurement frame.

12. A buffer configuration method, comprising: receiving, by a second switching device, a first measurement frame sent by a first switching device, and triggering, by the second switching device, a second measurement frame to be generated based on the first measurement frame; sending, by the second switching device, the second measurement frame to the first switching device; receiving, by the second switching device, a value of a priority-based flow control; (PFC) headroom sent by the first switching device; and setting, by the second switching device, a local buffer based on the value of the priority-based flow control; (PFC) headroom.

13. The method according to claim 12, further comprising: determining, by the second switching device, whether both the first switching device and the second switching device support an automatic buffer configuration function of priority-based flow control.

14. The method according to claim 13, wherein the determining, by the second switching device, whether both the first switching device and the second switching device support an automatic buffer configuration function of priority-based flow control comprises: when the second switching device supports the automatic buffer configuration function of priority-based flow control, and received TLV-format data fed back by the first switching device comprises a buffer configuration field, determining, by the second switching device, that both the first switching device and the second switching device support the automatic buffer configuration function of priority-based flow control.

15. The method according to claim 13, wherein, when the first switching device and the second switching device comprise a switching device that does not support the automatic buffer configuration function of priority-based flow control, the method further comprises: giving, by the second switching device, a prompt that the local buffer needs to be manually set.

16. A first switching device, comprising a processor; a memory storing program instructions that, when executed by the processor, cause the first switching device to: send a first measurement frame to a second switching device; receive a second measurement frame sent by the second switching device, wherein the second measurement frame is generated through triggering based on the first measurement frame; determine a value of a priority-based flow control (PFC) headroom based on the first measurement frame and the second measurement frame; and set, a local buffer based on the value of the PFC headroom.

17. The first switching device according to claim 16, wherein the determining of a value of a PFC headroom based on the first measurement frame and the second measurement frame comprises: obtaining a first timestamp and a fourth timestamp, wherein the first timestamp is a time at which the first switching device sends the first measurement frame, and the fourth timestamp is a time at which the first switching device receives the second measurement frame; and determining the value of the PFC headroom based on the first timestamp and the fourth timestamp.

18. The first switching device according to claim 17, wherein the first timestamp is a time at which the last bit of the first measurement frame enters a MAC layer of the first switching device, and the fourth timestamp is a time at which the last bit of the second measurement frame is sent from the MAC layer of the first switching device.

19. The first switching device according to claim 17, wherein the first timestamp is a time at which the last bit of the first measurement frame is sent from a MAC layer of the first switching device, and the fourth timestamp is a time at which the last bit of the second measurement frame enters the MAC layer of the first switching device.

20. The first switching device according to claim 17, wherein the first timestamp and the fourth timestamp are stored in the local buffer of the first switching device.

21. The first switching device according to claim 17, wherein the determining of the value of the PFC headroom based on the first timestamp and the fourth timestamp comprises: determining a timestamp difference based on the first timestamp and the fourth timestamp; and determining the value of the PFC headroom based on the timestamp difference, timestamp precision, a port rate between the first switching device and the second switching device, a configured maximum frame length, and a constant K, wherein the constant K is provided by a device manufacturer.

22. The first switching device according to claim 16, wherein the program instructions, when executed by the processor, further cause the first switching device to: send the value of the PFC headroom to the second switching device, so that the second switching device sets a local buffer based on the value of the PFC headroom.

23. The first switching device according to claim 16, wherein the program instructions, when executed by the processor, further cause the first switching device to: determine whether both the first switching device and the second switching device support an automatic buffer configuration function of priority-based flow control.

24. The first switching device according to claim 23, wherein the determining of whether both the first switching device and the second switching device support an automatic buffer configuration function of priority-based flow control comprises: when the first switching device supports the automatic buffer configuration function of priority-based flow control, and received TLV-format data fed back by the second switching device comprises a buffer configuration field, determining that both the first switching device and the second switching device support the automatic buffer configuration function of priority-based flow control.

25. The first switching device according to claim 23, wherein, when the first switching device and the second switching device comprise a switching device that does not support the automatic buffer configuration function of priority-based flow control, the program instructions, when executed by the processor, further cause the first switching device to: give a prompt that the local buffer needs to be manually set.

26. The first switching device according to claim 16, wherein the second measurement frame is generated by the second switching device by acknowledgement processing on the first measurement frame.

27. A second switching device, comprising: a processor; a memory storing program instructions that, when executed by the processor, cause the second switching device to: receive, a first measurement frame sent by a first switching device, and trigger based on the first measurement frame, a second measurement frame to be generated; send, the second measurement frame to the first switching device; receive, a value of a priority-based flow control; (PFC) headroom sent by the first switching device; and set, a local buffer based on the value of the priority-based flow control: (PFC) headroom.

28. The second switching device according to claim 27, wherein the program instructions, when executed by the processor, further cause the second switching device to: determine whether both the first switching device and the second switching device support an automatic buffer configuration function of priority-based flow control.

29. The second switching device according to claim 28, wherein the determining of whether both the first switching device and the second switching device support an automatic buffer configuration function of priority-based flow control comprises: when the second switching device supports the automatic buffer configuration function of priority-based flow control, and received TLV-format data fed back by the first switching device comprises a buffer configuration field, determining that both the first switching device and the second switching device support the automatic buffer configuration function of priority-based flow control.

30. The second switching device according to claim 29, wherein when the first switching device and the second switching device comprise a switching device that does not support the automatic buffer configuration function of priority-based flow control, wherein the program instructions, when executed by the processor, further cause the second switching device to: give a prompt that the local buffer needs to be manually set.