CLAIMS

1. A performance isolation method for a send queue of an RDMA network interface card, comprising:

identifying, by an RDMA network interface card on a sending side, a tenant type based on a tenant type identifier comprised in a queue pair identifier, buffering a work queue element WQE of a send queue of an identified latency-sensitive tenant as a latency-sensitive WQE into a latency-sensitive group of a buffering module, and buffering a WQE of a send queue of an identified bandwidth-sensitive tenant as a bandwidth-sensitive WQE into a bandwidth-sensitive group of the buffering module, wherein at least some of to-be-processed WQEs buffered into the bandwidth-sensitive group of the buffering module are stored into a wait station, WQEs of different tenants in the latency-sensitive group and the bandwidth-sensitive group that are comprised in the buffering module are buffered based on different queues, and each node of a queue is a WQE;

determining, by using an inter-group scheduler, whether the WQE in the latency-sensitive group or the WQE in the bandwidth-sensitive group is scheduled, sending a scheduling request to a first scheduler when it is determined that the WQE in the latency-sensitive group is scheduled, and sending a scheduling request to a second scheduler when it is determined that the WQE in the bandwidth-sensitive group is scheduled;

scheduling, by the first scheduler, the latency-sensitive WQE in the latency-sensitive group based on the scheduling request of the inter-group scheduler by using a first scheduling algorithm, and returning a WQE scheduling result to the inter-group scheduler;

scheduling, by the second scheduler, the bandwidth-sensitive WQE in the wait station by using a second scheduling algorithm, and returning a WQE scheduling result to the inter-group scheduler;

transmitting, by the inter-group scheduler, the WQE scheduling results from the first scheduler and the second scheduler to a WQE processing module on the RDMA network interface card on the sending side; and

processing, by the WQE processing module, the latency-sensitive WQE based on the WQE scheduling result to output all data packets corresponding to the latency-sensitive WQE, and processing the bandwidth-sensitive WQE based on the WQE scheduling result to output, based on a path maximum transmission unit, one data packet corresponding to the bandwidth-sensitive WQE, and if there is a data packet that is not output, writing a WQE corresponding to the data

packet that is not output back to the wait station,

wherein the first scheduling algorithm is a fair round robin scheduling algorithm, and the step of scheduling, by the first scheduler, the latency-sensitive WQE in the latency-sensitive group based on the scheduling request of the inter-group scheduler by using a first scheduling algorithm, and returning a WQE scheduling result to the inter-group scheduler comprises:

receiving, by the first scheduler, the scheduling request of the inter-group scheduler, and traversing, based on the fair round robin scheduling algorithm, a queue of latency-sensitive WQEs of different tenants in an active linked list maintained by the first scheduler, to poll the queue of latency-sensitive WQEs of the different tenants in the latency-sensitive group, wherein the active linked list is configured to record a queue number of to-be-processed latency-sensitive WQEs of the different tenants in the latency-sensitive group, and one node of the active linked list corresponds to one non-empty queue of latency-sensitive WQEs; and

returning a WQE of the polled queue in a header of the active linked list as the WQE scheduling result to the inter-group scheduler, and deleting a corresponding node from the active linked list.

2. The method according to claim 1, wherein the method further comprises:

receiving an identification request of a tenant through a driver interface on the RDMA network interface card, and providing, based on the identification request, a tenant identifier that comprises a tenant type identifier to the tenant; and

creating a queue pair having the queue pair identifier for each tenant allocated with the tenant identifier, and allocating a queue pair resource for storing a WQE to the queue pair, wherein the queue pair identifier comprises the tenant type identifier.

3. The method according to claim 2, wherein the tenant type identifier is a binary value of a preset length, and the tenant type identifier is a last digit of the tenant identifier.

4. The method according to any one of claims 1 to 3, wherein the second scheduling algorithm is a token bucket scheduling algorithm, and the step of scheduling, by the second scheduler, the bandwidth-sensitive WQE in the bandwidth-sensitive group of the wait station by using a second scheduling algorithm, and returning a WQE scheduling result to the inter-group scheduler

comprises:

receiving, by the second scheduler, the scheduling request of the inter-group scheduler, and traversing, based on the token bucket scheduling algorithm, a token counter of a queue of bandwidth-sensitive WQEs of different tenants in an active linked list maintained by the second scheduler, to refresh a token and update a token counter count based on a predetermined refresh cycle and each token refresh quantity, wherein the active linked list is configured to record a queue number of to-be-processed bandwidth-sensitive WQEs of the different tenants in the bandwidth-sensitive group, and one node of the active inked list corresponds to one non-empty queue of bandwidth-sensitive WQEs.

6. The method according to any one of claims 1 to 5, wherein the method further comprises:

for the WQEs in the bandwidth-sensitive group in the wait station, maintaining states of all WQEs in the wait station by using a wait WQE state module, externally outputting all wait WQE states, and performing programming control on the WQE states based on set and reset interfaces provided by the wait WQE state module, wherein the wait WQE states comprise two states: an idle state and an occupied state.

6. The method according to any one of claims 1 to 5, wherein the step of determining, by using an inter-group scheduler, whether the WQE in the latency-sensitive group or the WQE in the bandwidth-sensitive group is scheduled comprises:

determining, by the inter-group scheduler based on dynamic weights configured for the latency-sensitive group and the bandwidth-sensitive group, whether to schedule the WQE in the latency-sensitive group or the WQE in the bandwidth-sensitive group by using a weighted round robin scheduling algorithm, and updating the dynamic weights based on each scheduling result.

7. The method according to claim 6, wherein the updating the dynamic weights based on each scheduling result comprises reducing, based on each scheduling result, a weight of the latency-sensitive group or the bandwidth-sensitive group corresponding to the scheduling result.

8. An RDMA network interface card, comprising:

a control module, configured to identify a tenant type based on a tenant type identifier comprised in a queue pair identifier, buffer a work queue element WQE of a send queue of an identified latency-sensitive tenant as a latency-sensitive WQE into a latency-sensitive group of a buffering module, and buffer a WQE of a send queue of an identified bandwidth-sensitive tenant as a bandwidth-sensitive WQE into a bandwidth-sensitive group of the buffering module, wherein at least some of to-be-processed WQEs buffered into the bandwidth-sensitive group of the buffering module are stored into a wait station, WQEs of different tenants in the latency-sensitive group and the bandwidth-sensitive group that are comprised in the buffering module are buffered

based on different queues, and each node of a queue is a WQE;

the wait station, configured to record at least some of to-be-processed WQEs in the bandwidth-sensitive group;

an inter-group scheduler, configured to determine whether the WQE in the latency-sensitive group or the WQE in the bandwidth-sensitive group is scheduled, send a scheduling request to a first scheduler when it is determined that the WQE in the latency-sensitive group is scheduled, send a scheduling request to a second scheduler when it is determined that the WQE in the bandwidth-sensitive group is scheduled, and transmit WQE scheduling results from the first scheduler and the second scheduler to a WQE processing module;

the first scheduler, configured to schedule the latency-sensitive WQE in the latency-sensitive group based on the scheduling request of the inter-group scheduler by using a first scheduling algorithm, and return a WQE scheduling result to the inter-group scheduler;

the second scheduler, configured to schedule the bandwidth-sensitive WQE in the bandwidth-sensitive group in the wait station by using a second scheduling algorithm, and return a WQE scheduling result to the inter-group scheduler; and

the WQE processing module, configured to process the latency-sensitive WQE based on the WQE scheduling result to output all data packets corresponding to the latency-sensitive WQE, and process the bandwidth-sensitive WQE based on the WQE scheduling result to output, based on a path maximum transmission unit, one data packet corresponding to the bandwidth-sensitive WQE, and if there is a data packet that is not output, write a WQE corresponding to the data packet that is not output back to the wait station,

wherein the first scheduling algorithm is a fair round robin scheduling algorithm;

the second scheduling algorithm is a token bucket scheduling algorithm; and

the inter-group scheduler determines, based on dynamic weights configured for the latency-sensitive group and the bandwidth-sensitive group, whether to schedule the WQE in the latency-sensitive group or the WQE in the bandwidth-sensitive group by using a weighted round robin scheduling algorithm, and updates the dynamic weights based on each scheduling result.

9. The RDMA network interface card according to claim 8, wherein the wait station comprises:

a true dual-port SRAM, comprising a first wait WQE read/write channel and a second wait WQE read/write channel, wherein the first wait WQE read/write channel is configured to transmit the bandwidth-sensitive WQE in the bandwidth-sensitive group in the buffering module to the wait station, and the second wait WQE read/write channel is configured to write the WQE corresponding to the data packet that is not output back to the wait station;

a wait WQE state module, configured to maintain states of all WQEs in the wait station and externally output all wait WQE states; and

set and reset interfaces, configured to perform programming control on the WQE states,

wherein the wait WQE states comprise two states: an idle state and an occupied state.