CLAIMS

What is claimed is:

1. An IO request scheduling method, comprising:

obtaining a target IO request to be scheduled, and determining whether the target IO request is a traffic-limited IO request;

when the target IO request is a traffic-unlimited first-type IO request, determining a current moment as a scheduling moment of the target IO request; or

when the target IO request is a traffic-limited second-type IO request, determining a scheduling moment of the target IO request based on a current resource traffic-limiting proportion, wherein the current resource traffic-limiting proportion is a proportion of resources used to process second-type IO requests in a current time period that the current moment belongs, the current resource traffic-limiting proportion is calculated based on a last resource traffic-limiting proportion and last load information of a last time period, the last load information is load information of first-type IO requests scheduled in the last time period, and the scheduling moment is not less than the current moment; and

adding the target IO request to an IO request set, wherein several IO requests in the IO request set are sequentially sent to a disk for processing in ascending order of scheduling moments,

wherein the determining a scheduling moment of the target IO request based on a current resource traffic-limiting proportion comprises:

calculating an expected scheduling moment of the target IO request based on the current resource traffic-limiting proportion and a scheduling moment of an obtained last second-type IO request, wherein the expected scheduling moment is negatively correlated with the current resource traffic-limiting proportion; and

determining the larger value of the current moment and the expected scheduling moment as the scheduling moment of the target IO request.

2. The method according to claim 1, wherein the first-type IO request comprises a foreground IO request, and the second-type IO request comprises a background IO request.

3. The method according to claim 1, wherein the adding the target IO request to an IO request set specifically comprises: adding the target IO request to an IO request set managed by using a minimum heap.

4. The method according to claim 1, wherein the determining a scheduling moment of the target IO request based on a current resource traffic-limiting proportion further comprises: determining an expected read/write frequency corresponding to request bandwidth of the target IO request based on performance of the disk; and

the calculating an expected scheduling moment of the target IO request based on the current resource traffic-limiting proportion and a scheduling moment of an obtained last second-type IO request specifically comprises: calculating the expected scheduling moment of the target IO request based on the expected read/write frequency, the current resource traffic-limiting proportion, and the scheduling moment of the obtained last second-type IO request, wherein the expected scheduling moment is negatively correlated with the expected read/write frequency.

5. The method according to any one of claims 1 to 4, wherein when the target IO request is a traffic-unlimited first-type IO request, the method further comprises:

determining a first moment of obtaining the target IO request;

at a second moment, selecting the target IO request from the IO request set, and sending the target IO request to the disk, wherein at the second moment, the scheduling moment of the target IO request is a smallest value of all IO requests comprised in the IO request set;

receiving a request result returned by the disk;

sending the request result to an application corresponding to the target IO request, and determining a third moment of sending the request result; and

determining a response time of the target IO request based on the third moment and the first moment.

6. The method according to claim 5, further comprising: when an end moment of the current time period is reached, determining a next resource traffic-limiting proportion of a next time period based on current load information of first-type IO requests scheduled in the current time period, wherein the current load information specifically comprises an average response time of the first-type IO requests scheduled in the current time period and average request bandwidth of the first-type IO requests scheduled in the current time period.

7. The method according to claim 6, wherein the determining a next resource traffic-limiting proportion of a next time period based on current load information of first-type IO requests scheduled in the current time period comprises:

determining an expected response time corresponding to the average request bandwidth based on the performance of the disk; and

when the average response time is not less than the expected response time, reducing the current resource traffic-limiting proportion based on a predetermined step to obtain the next resource traffic-limiting proportion of the next time period; or when the average response time is less than the expected response time, increasing the current resource traffic-limiting proportion based on a predetermined step to obtain the next resource traffic-limiting proportion of the next time period.

8. The method according to claim 6, wherein the next resource traffic-limiting proportion of the next time period is not greater than a first predetermined threshold and is not less than a second predetermined threshold.

9. An IO request scheduling method, comprising:

obtaining an IO request set, wherein the IO request set is obtained by using any one of claims 1 to 8, the IO request set comprises several IO requests, and each IO request has a corresponding scheduling moment;

selecting, from the IO request set, an IO request with a smallest scheduling moment as a current IO request; and

sending the current IO request to a disk, so that the disk processes the current IO request and returns a request result.

10. The method according to claim 9, wherein the sending the current IO request to a disk specifically comprises: sending the current IO request to the disk when a scheduling moment corresponding to the current IO request is less than a current moment.

11. An IO request scheduling apparatus, comprising:

a request acquisition unit, configured to obtain a target IO request to be scheduled;

a category determining unit, configured to determine whether the target IO request is a traffic-limited IO request;

a scheduling moment determining unit, configured to: when the target IO request is a traffic-unlimited first-type IO request, determine a current moment as a scheduling moment of the target IO request; or when the target IO request is a traffic-limited second-type IO request, determine a scheduling moment of the target IO request based on a current resource traffic-limiting proportion, wherein the current resource traffic-limiting proportion is a proportion of resources used to process second-type IO requests in a current time period that the current moment belongs, the current resource traffic-limiting proportion is calculated based on a last resource traffic-limiting proportion and last load information of a last time period, the last load information is load information of first-type IO requests scheduled in the last time period, and the scheduling moment is not less than the current moment; and

a request scheduling unit, configured to add the target IO request to an IO request set, wherein several IO requests in the IO request set are sequentially sent to a disk for processing in ascending order of scheduling moments.,

wherein when determining a scheduling moment of the target IO request based on a current resource traffic-limiting proportion, the scheduling moment determining unit is configured to:

calculate an expected scheduling moment of the target IO request based on the current resource traffic-limiting proportion and a scheduling moment of an obtained last second-type IO request,

wherein the expected scheduling moment is negatively correlated with the current resource traffic-limiting proportion; and

determine the larger value of the current moment and the expected scheduling moment as the scheduling moment of the target IO request

12. An IO request scheduling apparatus, comprising:

an acquisition unit, configured to obtain an IO request set, wherein the IO request set is obtained by using the apparatus according to claim 11, the IO request set comprises several IO requests, and each IO request has a corresponding scheduling moment;

a selection unit, configured to select, from the IO request set, an IO request with a smallest scheduling moment as a current IO request; and

a sending unit, configured to send the current IO request to a disk, so that the disk processes the current IO request and returns a request result.

13. A computing device, comprising a memory and a processor, wherein the memory stores a computer program, and when executing the computer program, the processor implements the method according to any one of claims 1 to 8 or claims 9 and 10.

14. A computer-readable storage medium, wherein a computer program is stored on the computer-readable storage medium, and when the computer program is executed in a computing device, the computing device performs the method according to any one of claims 1 to 8 or claims 9 and 10.