Hardware description:

Intel(R) Xeon(R) Gold 6150 CPU @ 2.70GHz
Intel Optane SSD 900P 260GB
ST4000NM000B-2TF100 HDD 4TB
260GB SSD and 4TB HDD are associated as bcache cache device

This test will be conducted using three servers of the same model. The three servers will conduct three sets of comparison tests: DSYNC random, DirectIO random, DSYNC random and sequential mixed tests. Each set of comparison data is performed using the same disk on the server with the same number, so that the other variables of each set of comparison data are maintained in a consistent state except for the differences in bcache.

DSYNC Testing

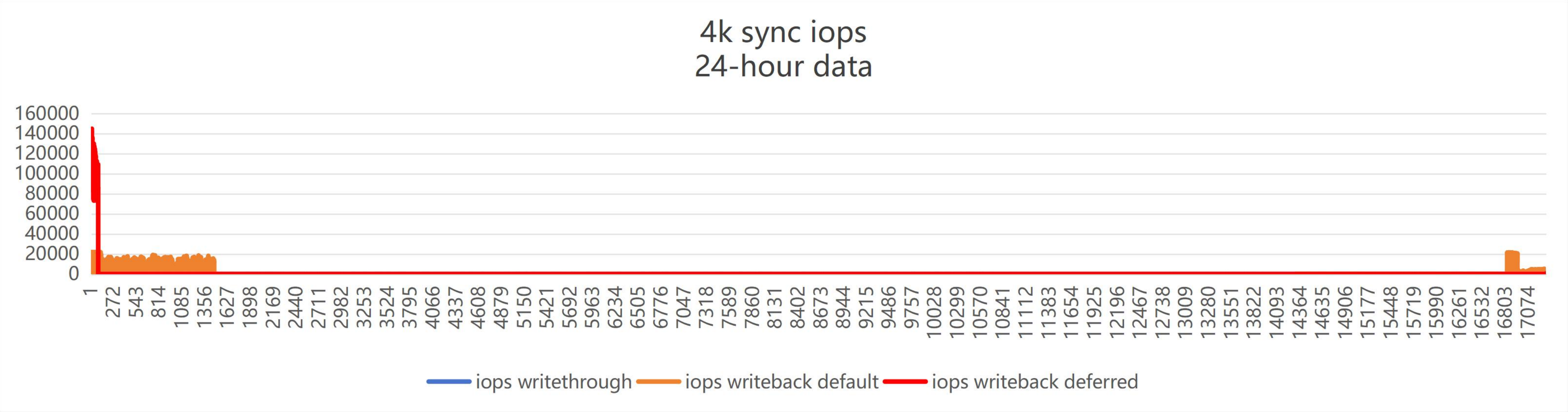
Test vdbench use case configuration:

```
messagescan=no
hd=default,vdbench=/root/vdbench,user=root,shell=ssh
hd=hd1,system=localhost
sd=sd1,hd=hd1,lun=/dev/bcache0,openflags=o_sync,threads=16
wd=wd1,sd=sd1,rdpct=0,xfersize=4k,seekpct=100,rhpct=0
rd=rd1,wd=wd1,iorate=max,fwd=yes,interval=5,elapsed=86400
```

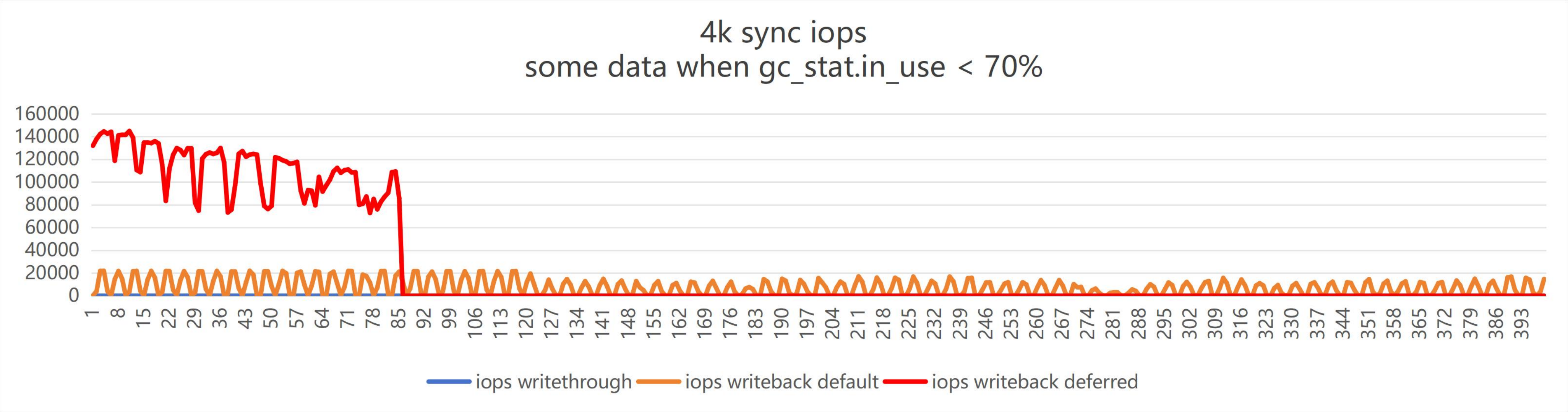
Test instructions:

1. Compare the random write performance of 4k dsync in the three cache working modes: writeback default, writeback deferred, and writethrough (both gc_after_writeback is enabled).
2. Compare the differences in efficiency of dirty data writeback in the two modes of writeback default and writeback deferred.

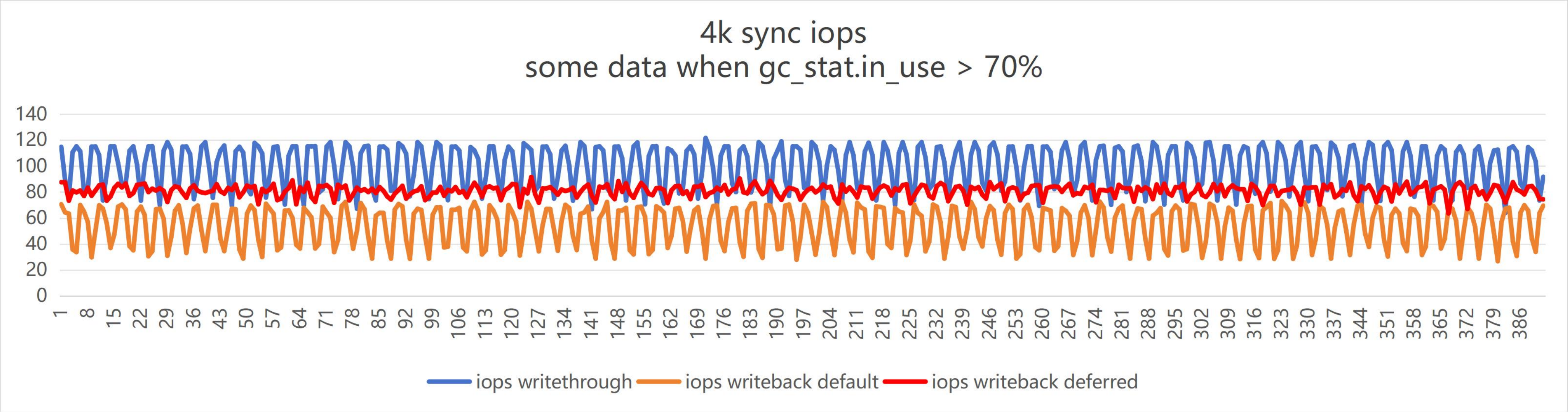
Data obtained after testing:



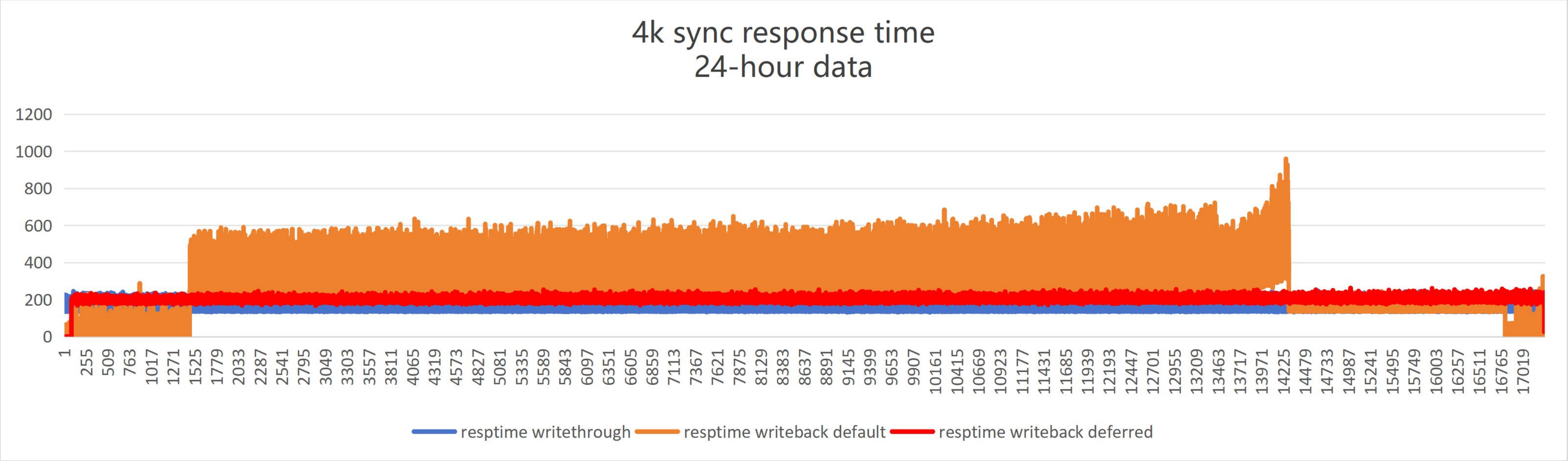
Note: The reason for the increase in yellow at the tail is that writeback default is the first to complete a round of dirty data writeback scans, which triggers GC (enble gc_after_writeback), which improves write efficiency. There are two reasons for taking the lead in completing dirty data writeback scans: Wirteback default writeback efficiency is about 6% higher than writebcak deferred writeback efficiency, followed by more than ten hours of long tests, writebcak default is about 30% lower than writeback deferred writeback speed. The blue part is covered with two other colors, and the picture will be enlarged and displayed later.



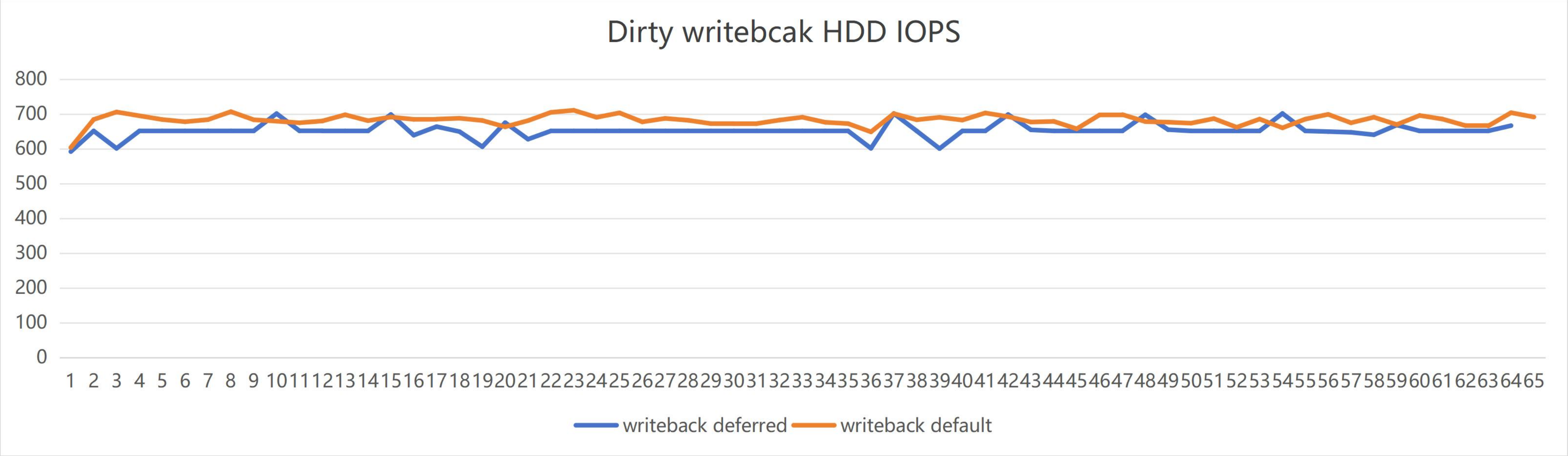
Note: The 24-hour comparison chart is compressed too much, and it is not easy to see the difference. This picture captures the starting part of the 24-hour data to enlarge. From the figure, see that the writeback deferred mode is extremely high at the beginning stage of writing iops. In a short period of time, gc_stat.in_use reaches 70%, and then iops starts to slow down. Wirteback default writes iops relatively lower, after taking a long time, gc_stat.in_use reaches 70%. Therefore, the reason why yellow is higher than red in the second half is seen in the figure. Blue keeps low on IOPS.



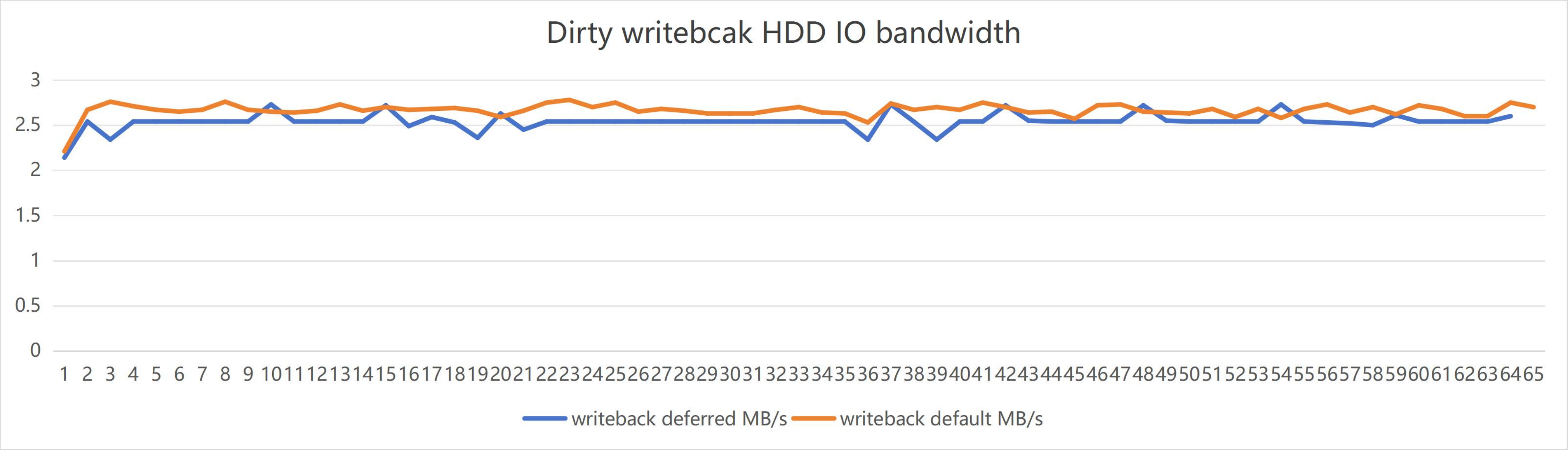
Note: In the 24-hour pressure test, nearly 90% of the time is in the state at gc_stat.in_use > 70%. It can be observed in the above figure that writeback deferred is about 30% higher than writeback default iops on average. The blue writethrough mode gc_stat.in_use has been maintained at a low level and has not reached the index of gc_stat.in_use > 70% of the data in the title.



Note: The red writeback deferred start stage, that is, before gc_stats.in_use < 70%, the delay is about 0.1 (the delay is low, and it reaches gc_stat.in_use>70% in a short time, so the comparison of the figure above is not obvious). The yellow writeback default at the tail is the first to complete the dirty data writeback scan and trigger the GC.



Note: This data is the HDD IOPS data captured when dirty data is written back after 4k random writing under the two modes of writeback deferred and writeback default separately.



Note: This data is the HDD IO bandwidth data captured when dirty data is written back after 4k random writing under the two modes of writeback deferred and writeback default separately.

DSYNC test summary:

1. When writeback deferred has a lot of cache space, 4k dsync performance has very good performance compared to writebcak default.
2. writeback deferred in gc_stat.in_use > 70% of cases, the performance of 4k dsync is about 30% higher than that of writebcak default.
3. Comparing writeback deferred with the writethrough mode that also has high data reliability, when gc_stat.in_use<70%, the performance of dsync is more than 1,000 times that of the latter. As the dirty data increases, affected by factors such as dirty data writeback, when gc_stat.in_use>70%, the performance of dsync is about 30% lower than the latter.
4. Compared with writeback default, in the static state, the write-back efficiency of pure dirty data in writeback deferred is about 6% lower.

DirectIO Testing

Test vdbench use case configuration:

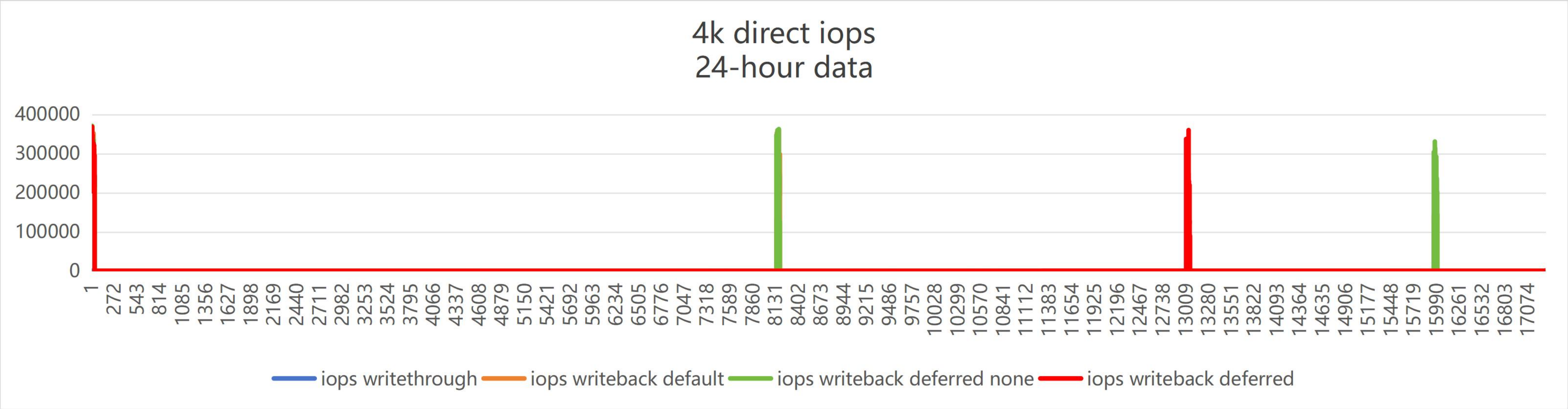
```
messagescan=no
hd=default,vdbench=/root/vdbench,user=root,shell=ssh
hd=hd1,system=localhost
```

```
sd=sd1,hd=hd1,lun=/dev/bcache0,openflags=o_direct,threads=16
wd=wd1,sd=sd1,rdpct=0,xfersize=4k,seekpct=100,rhpct=0
rd=rd1,wd=wd1,iorate=max,fwd=yes,interval=5,elapsed=86400
```

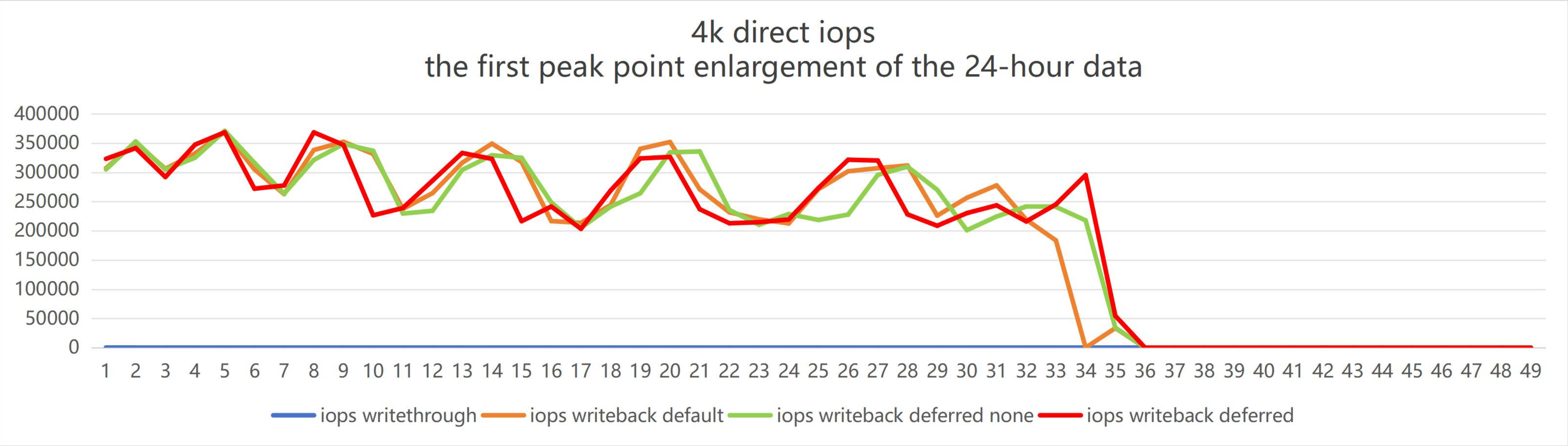
Test instructions:

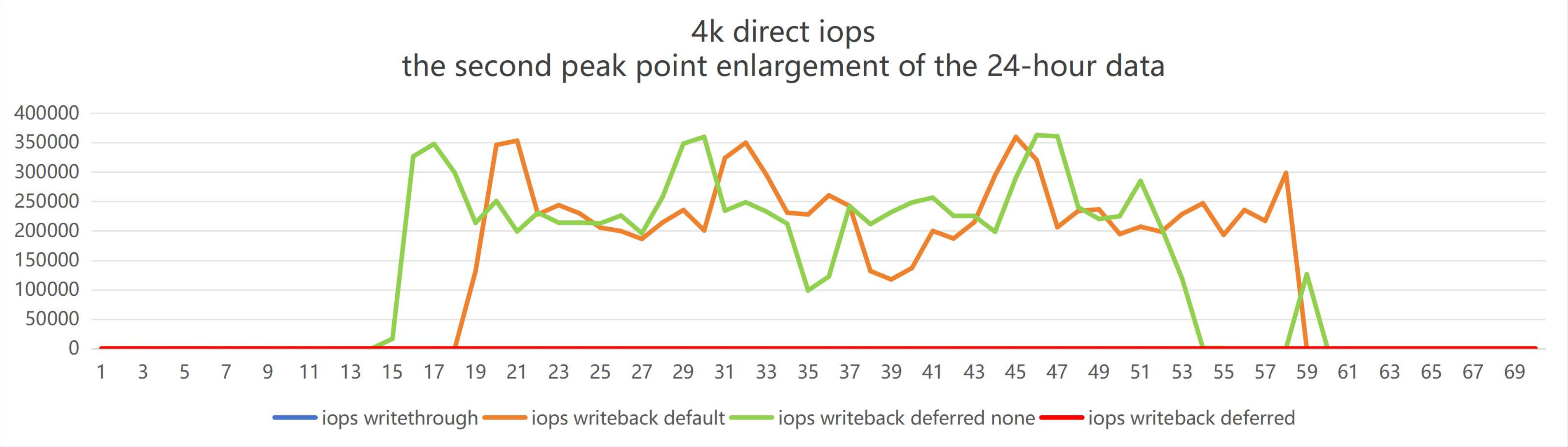
1. Compare the random write performance of 4k direct io in the three cache working modes of writeback default, writeback deferred, and writethrough (both gc_after_writeback is enabled). .
2. Compare the performance differences between writeback default (before patch) and writeback deferred none (after patch, but not deferred is enabled), with the purpose of verifying whether the patch will have an impact on the performance of the original code.

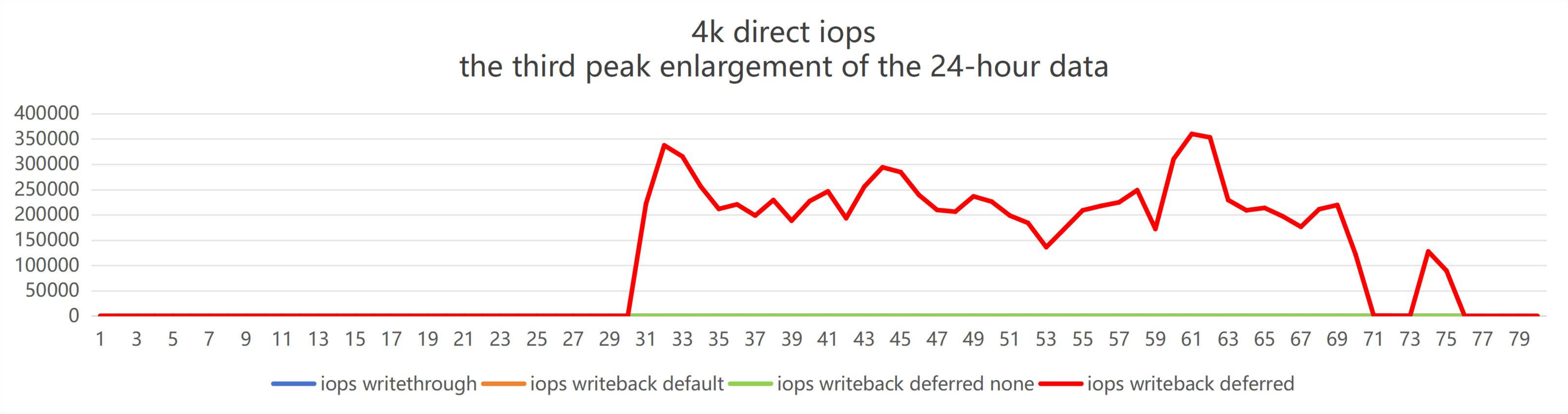
Data obtained after testing:

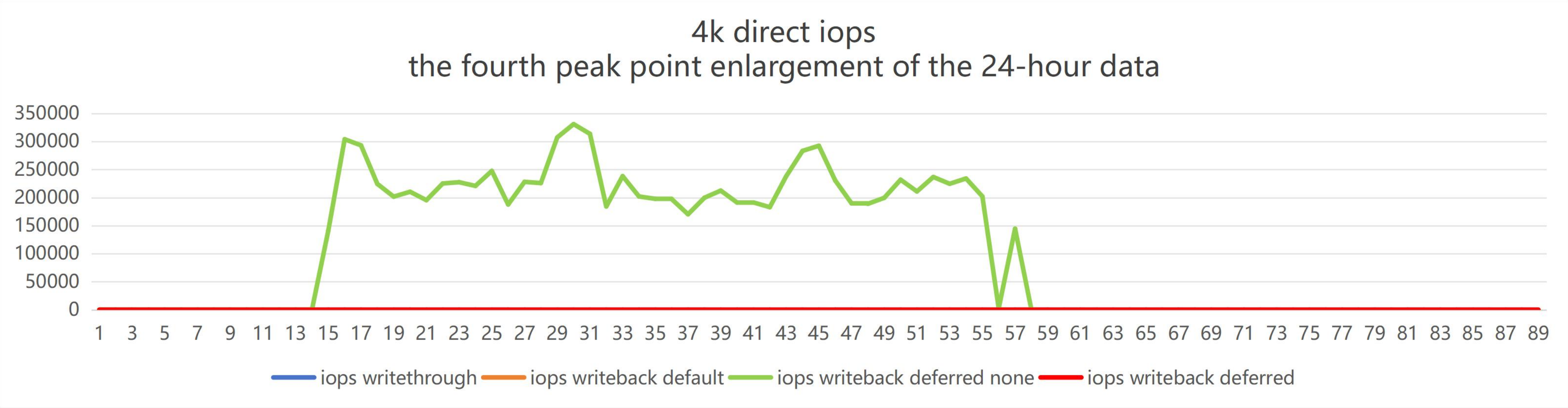


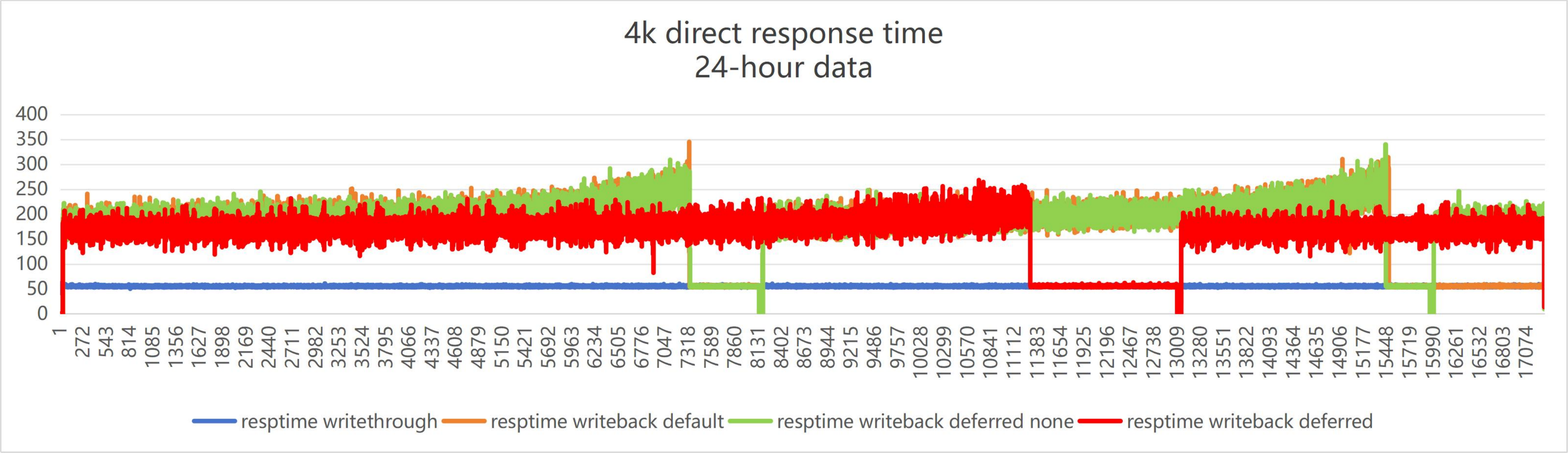
Note: There are 4 peak points in the figure, among which the first peak point has yellow, green, and red overlap, the second peak point has green and yellow overlap, the third peak point is only red, and the fourth peak point is only green. The four peak points are enlarged and displayed later.


4k direct iops
the first peak point enlargement of the 24-hour data
400000
350000
300000
250000
200000
150000
100000
50000
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
iops writethrough
iops writeback default
iops writeback deferred none
iops writeback deferred

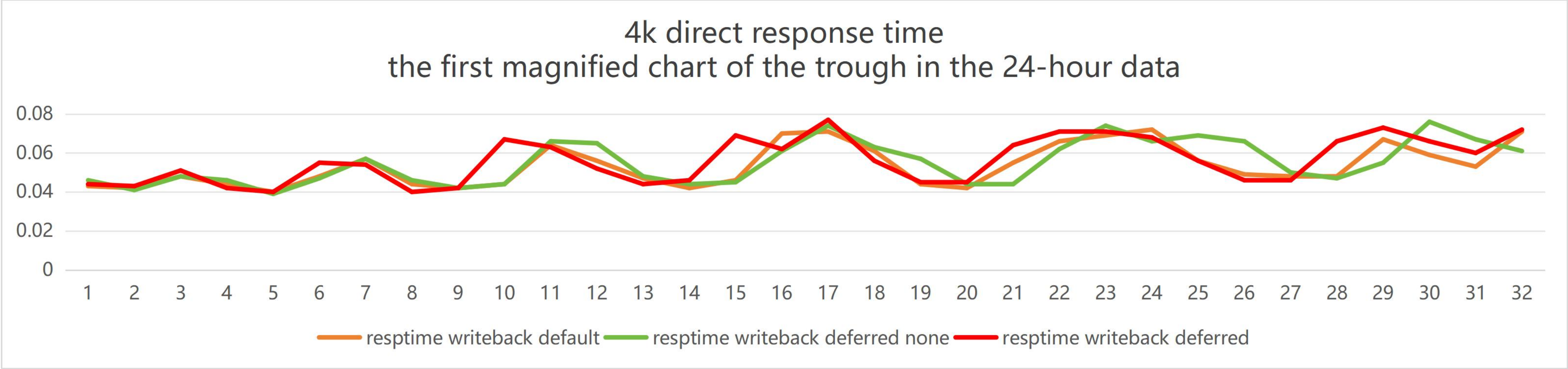

4k direct iops
the second peak point enlargement of the 24-hour data
400000
350000
300000
250000
200000
150000
100000
50000
0
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
31
33
35
37
39
41
43
45
47
49
51
53
55
57
59
61
63
65
67
69
iops writethrough
iops writeback default
iops writeback deferred none
iops writeback deferred


4k direct iops
the third peak enlargement of the 24-hour data
400000
350000
300000
250000
200000
150000
100000
50000
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69 71 73 75 77 79
iops writethrough
iops writeback default
iops writeback deferred none
iops writeback deferred


4k direct iops
the fourth peak point enlargement of the 24-hour data
350000
300000
250000
200000
150000
100000
50000
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69 71 73 75 77 79 81 83 85 87 89
iops writethrough
iops writeback default
iops writeback deferred none
iops writeback deferred



Note: The first low point in the chart is the overlap of yellow, green and red; the second low point is the overlap of yellow and green; the third low point is only red; and the fourth low point is only green



Note: Since the writethrough data value is around 50, while the data of the other three items in this graph are at the 0.0x level, if the writethrough data is retained, it will result in the comparison of the other three items not being displayed clearly. Therefore, the writethrough data is not shown here.

DirectIO test summary:

1. There is no difference in performance between writeback default (before patch) and writeback deferred none (after patch, but not deferred). The data of the two are highly coincidental, indicating that the patch will not have a significant negative impact on bcache performance.
2. In the random comparison of writeback deferred and writeback default, the performance data of IOPS and response time are basically the same, but the former is later than the latter GC.

DSYNC random and sequential mixed test

Test vdbench use case configuration:

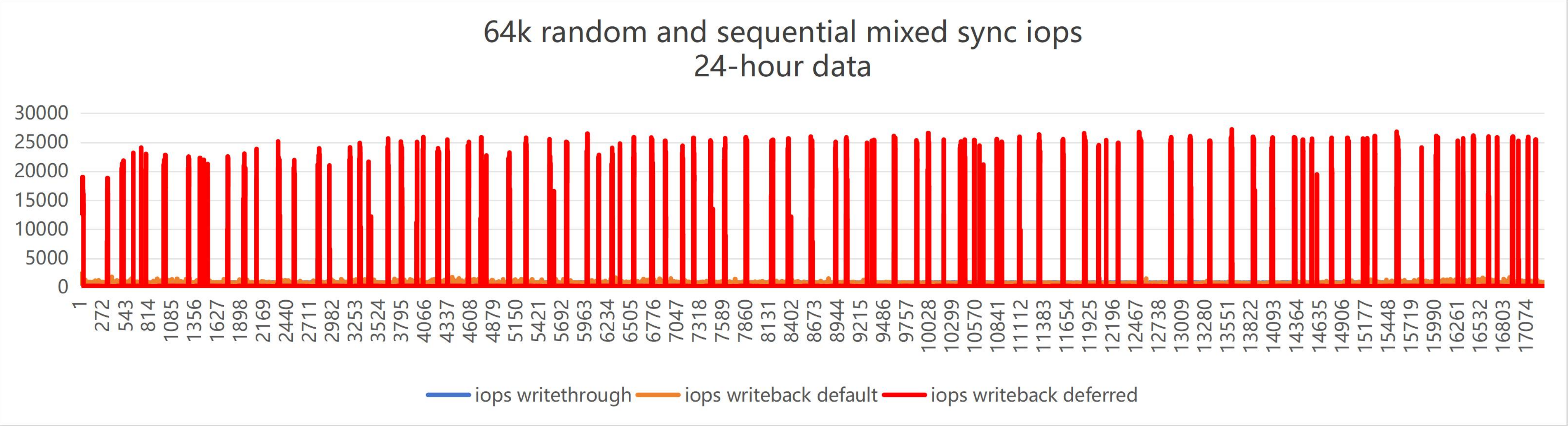
```
messagescan=no
hd=default,vdbench=/root/vdbench,user=root,shell=ssh
hd=hd1,system=localhost
sd=sd1,hd=hd1,lun=/dev/bcache0,openflags=o_sync,threads=15
wd=wd1,sd=sd1,rdpct=0,xfersize=64k,seekpct=1,rhpct=0
rd=rd1,wd=wd1,iorate=max,fwd=yes,interval=5,elapsed=86400
```

Note: This test case adopts the method of bs=64k and a random probability of 1% (seekpct=1) to simulate the scenario where sequential I/O larger than 4MB occurs during the test process. In such a scenario, partial bypass will occur to verify the performance of this patch in this situation.

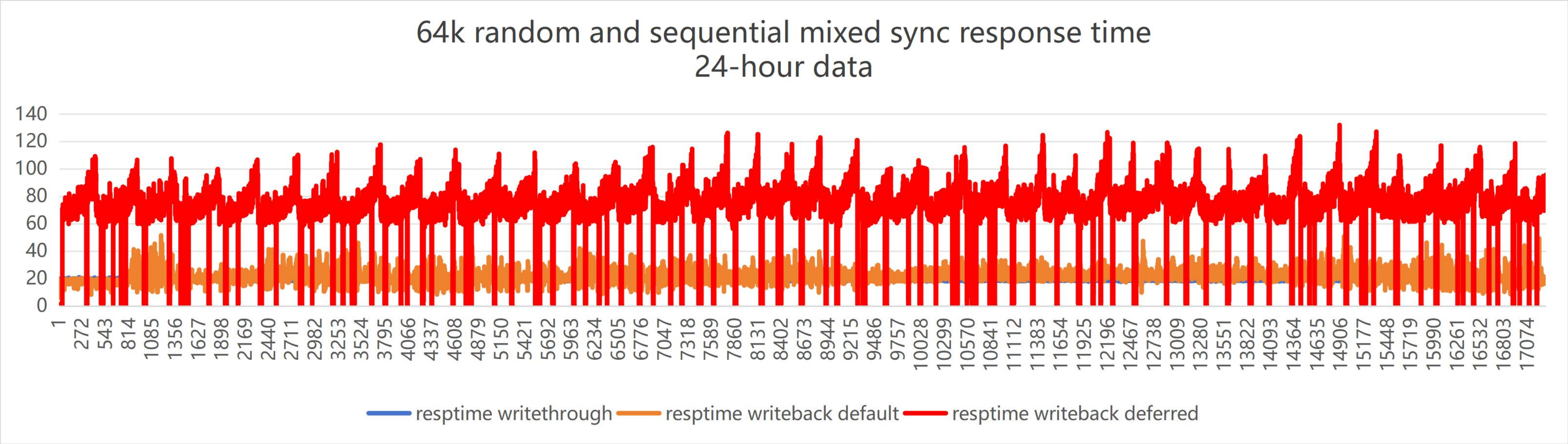
Test instructions:

1. Compare the write performance of 64K DSYNC random and sequential mixed write performance in three cache working modes: writeback default, writeback deferred, and writethrough (both gc_after_writeback is enabled).

Data obtained after testing:



Note: The average blue wirethrough iops maintains around 800, the average yellow writeback default maintains around 900, and the average red writeback deferred mode will quickly rise to 2w+ level after each triggering GC. Compared with the previous test of pure 4k sync random writing, this 64k random and sequential mixed read and write reading and writing can trigger GC frequently because of the large write IO, which is more in line with the state of the actual application scenario.



Note: As shown in the figure above, the writeback default write speed is low, and it is maintained at around 700 iops for a long time, which makes the cache disk maintain a low dirty data ratio and a low gc_stat.in_use for a long time. Therefore, when gc_stat.in_use > 70%, the writeback default has a lower response time than writeback deferred.

Summary of DSYNC random and sequential mixed test:

1. In the case of frequent bypass, writeback deferred still performs well. In contrast, writeback default IOPS has remained at a low level for a long time (gc_stat.in_use has not reached > 70%), which also caused this test case to fail to compare the data when both writeback deferred and writeback default are both gc_stat.in_use > 70%.

Comprehensive summary

In general, the writeback deferred mode has the same data reliability as Wirtethrough, and also has the write performance of writeback default that sacrifices a small part of the writeback efficiency, and has significantly improved dsync performance. In addition, from the DirectIO comparison test group, we can see that this patch did not have a bad impact on the performance of the original bcache driver.