

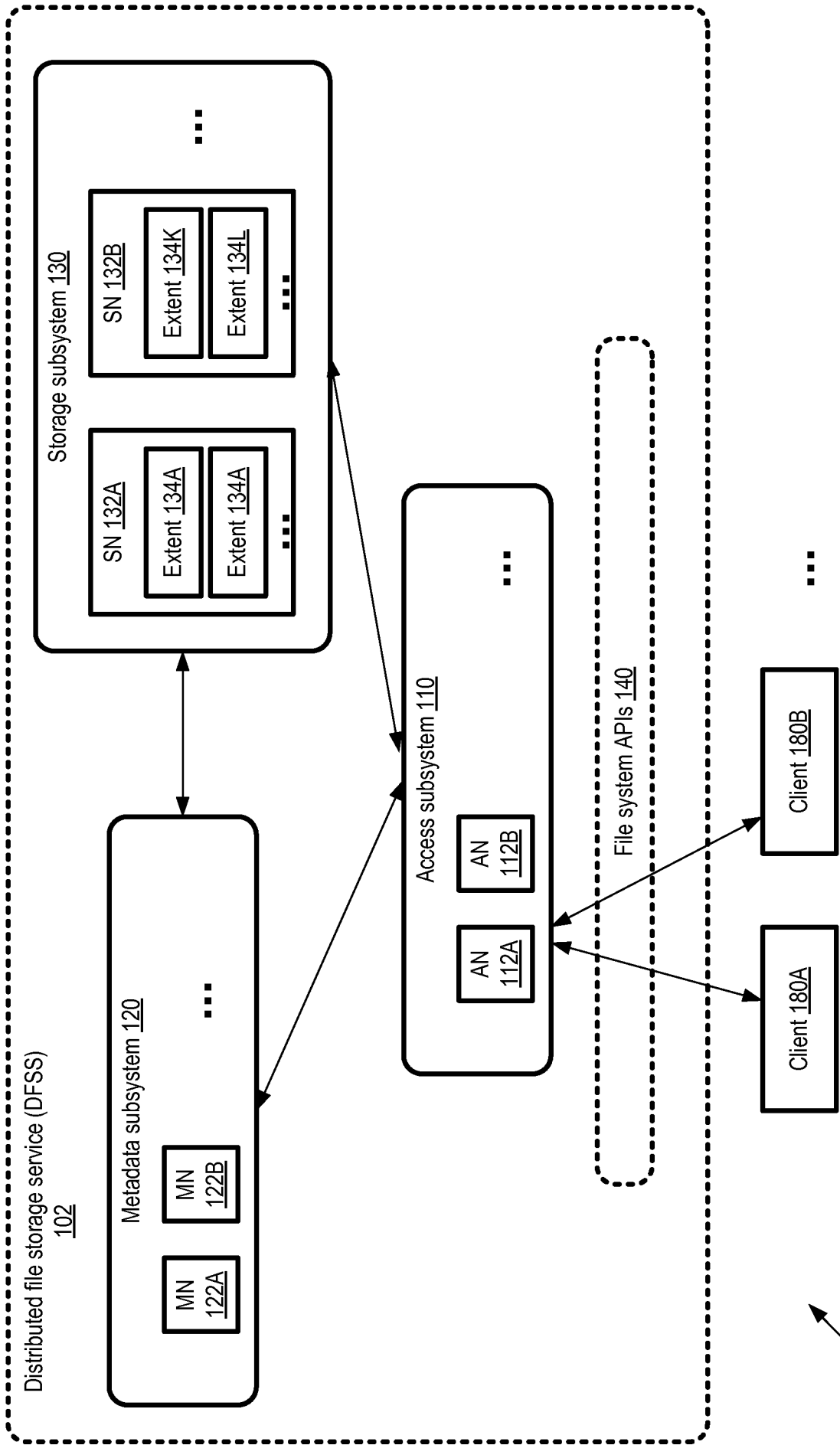


FIG. 1

System 100

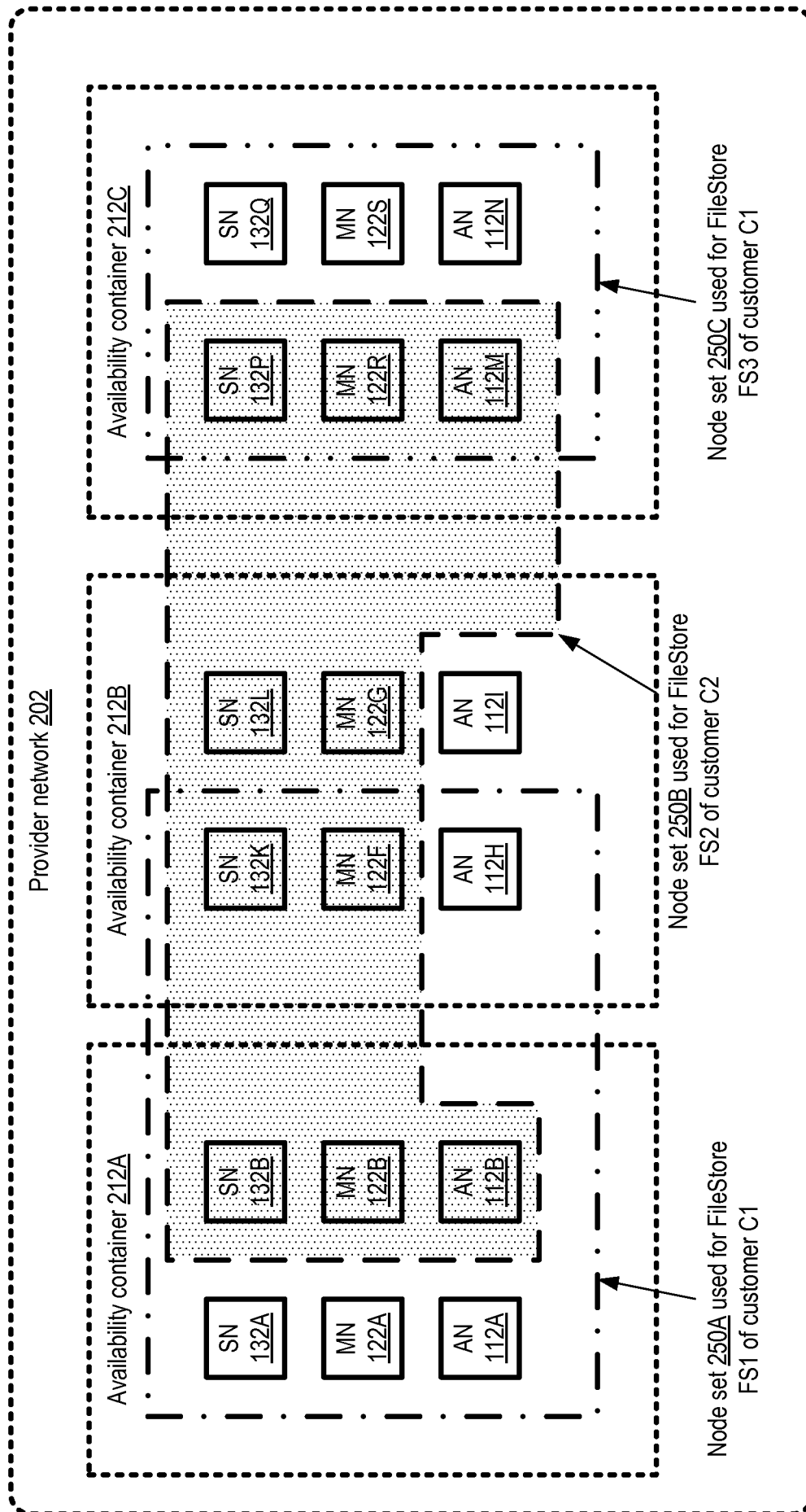


FIG. 2

3/69

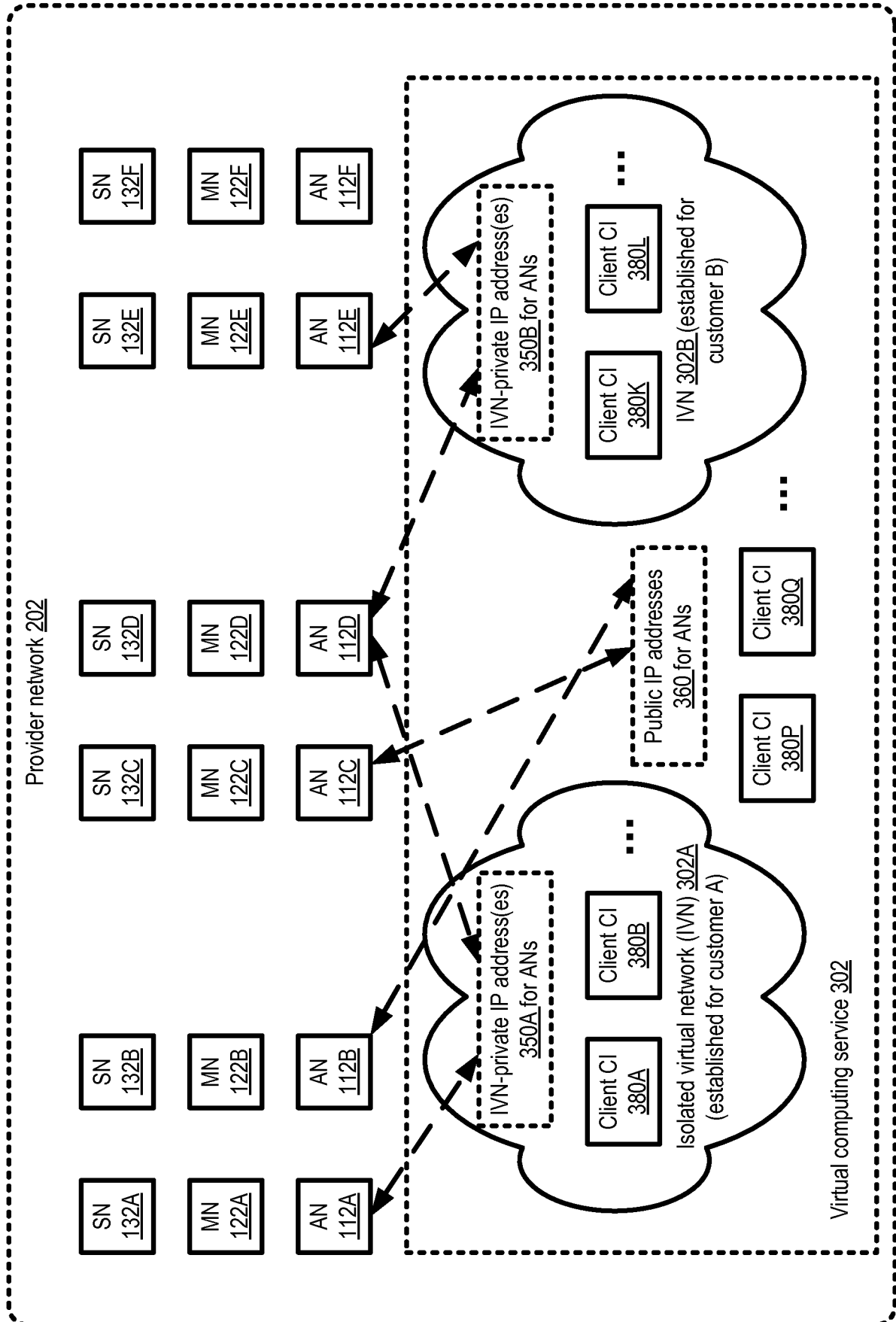


FIG. 3

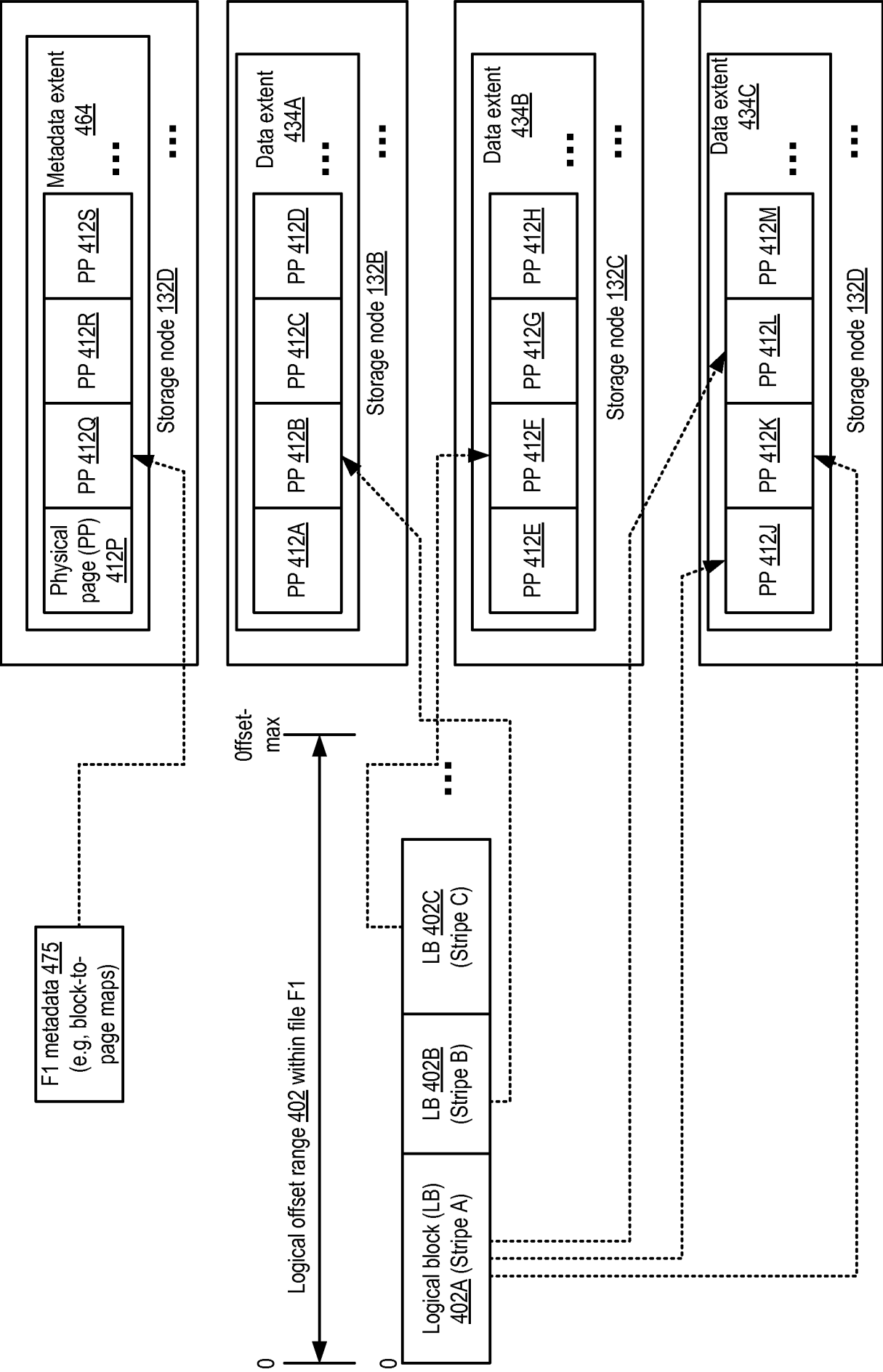


FIG. 4

...

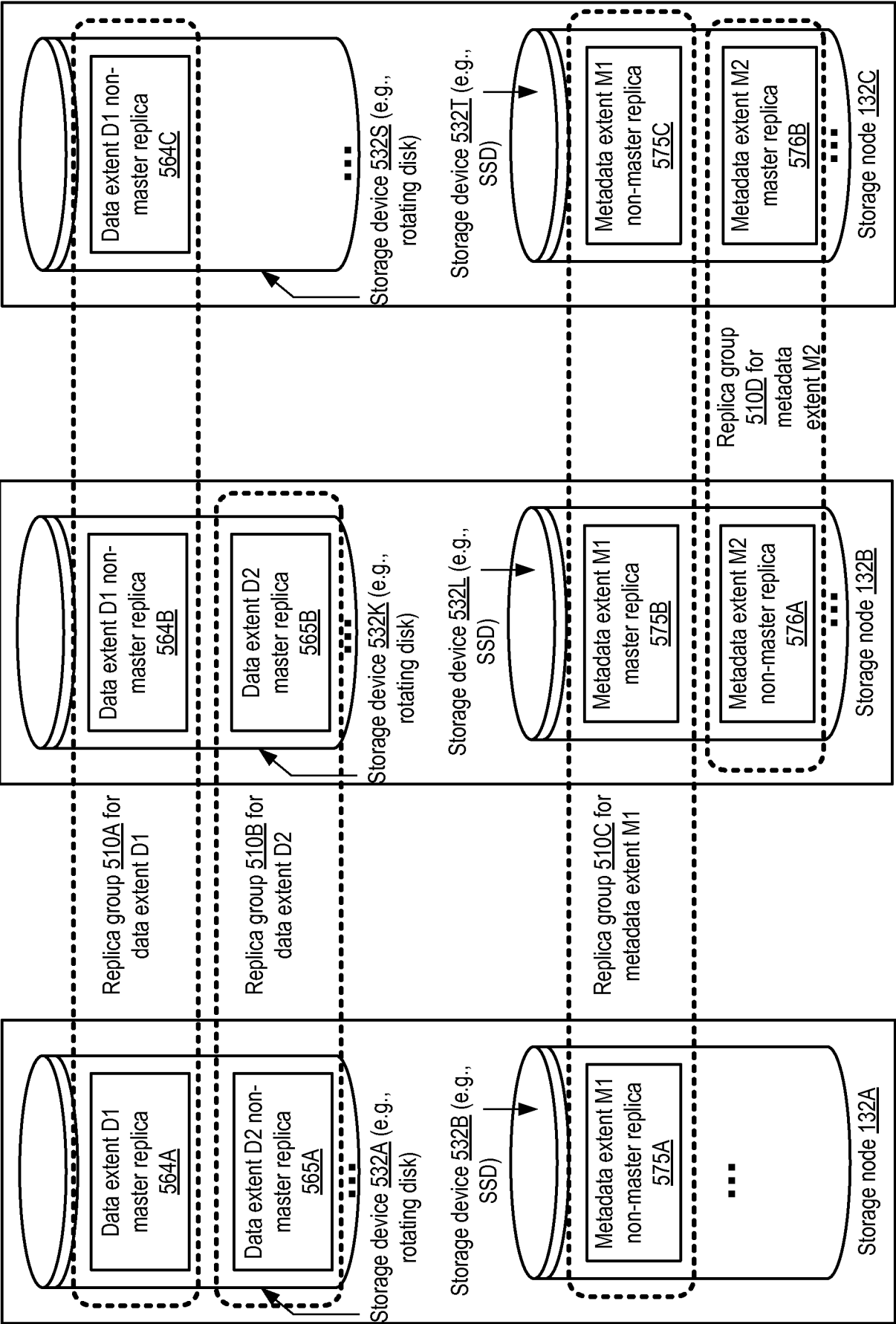


FIG. 5

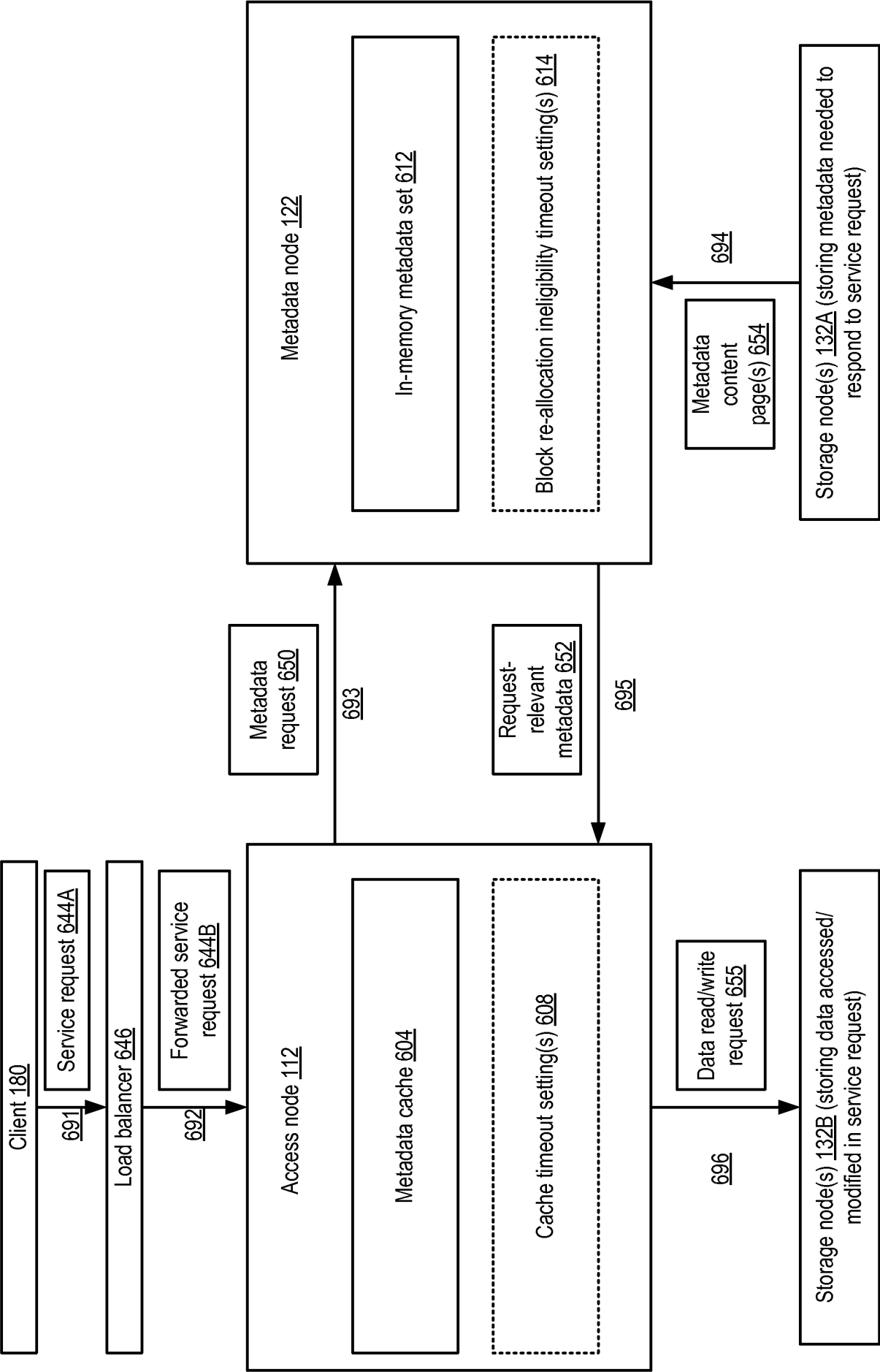


FIG. 6

7/69

File store name <u>702</u>	Metadata durability policy <u>704</u>	Metadata performance targets <u>706</u>	Metadata logical block policy <u>708</u>	Metadata physical page policy <u>710</u>	Metadata caching policy <u>712</u>	Data durability policy <u>714</u>	Data performance targets <u>716</u>	Data logical block policy <u>718</u>	Data physical page policy <u>720</u>
FS1	10-way full replication	W resp time < T1; R resp time < T2	Fixed LB size = 512KB, dynamic allocation	Fixed pp size = 8KB	Cache timeout = C1; block realloc ineligibility timeout = BR1	12/6 erasure coding	W resp time < T3; R resp time < T4; W tput > Y1; R tput > Z1	Dynamic LB sizes = 1K, 1K, 2K, 2K, 4K,...; static allocation	PP size = LB size
FS2	3-way full replication	R resp time < T7	Fixed LB size = 1MB, static allocation	Fixed pp size = 16KB	Cache timeout = C2; block realloc ineligibility timeout = BR2	6/4 erasure coding	W tput > Y2; R tput > Z2	Fixed LB size = 4MB, static allocation	PP size = 32KB
■ ■ ■ ■									

FIG. 7

8/69

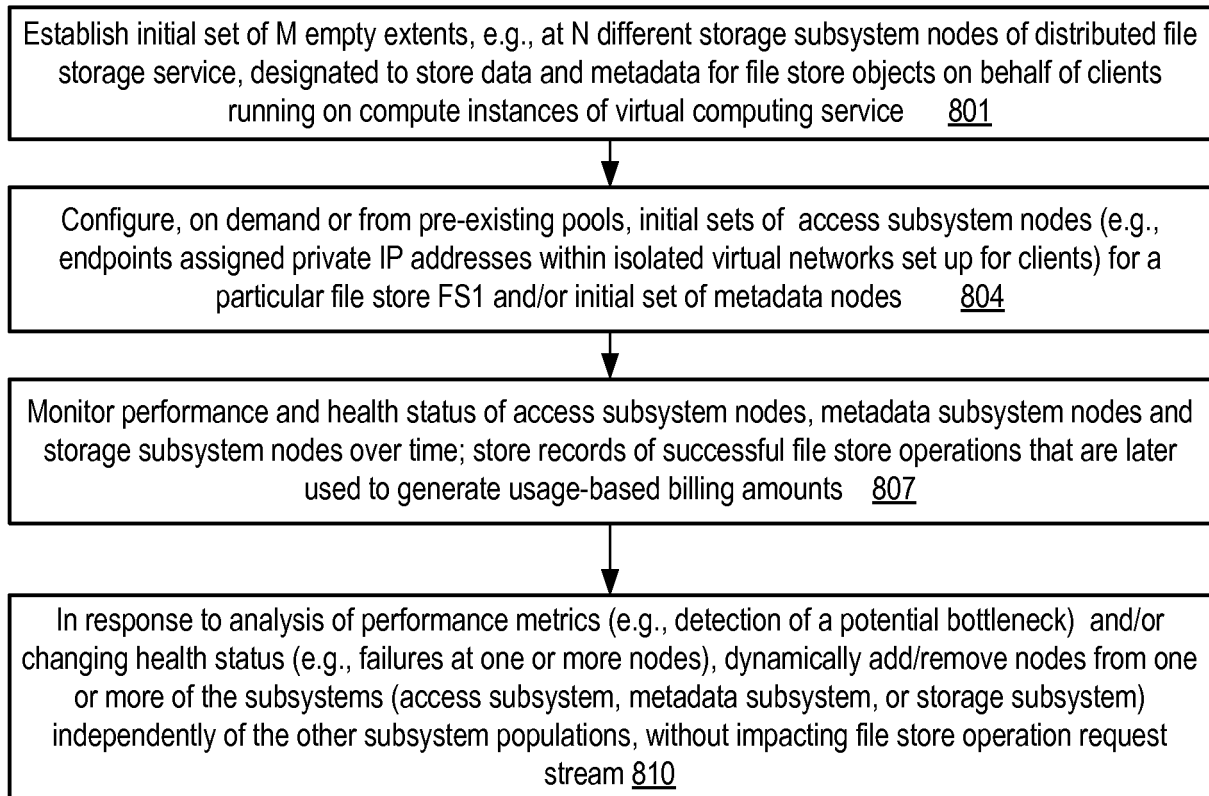


FIG. 8a

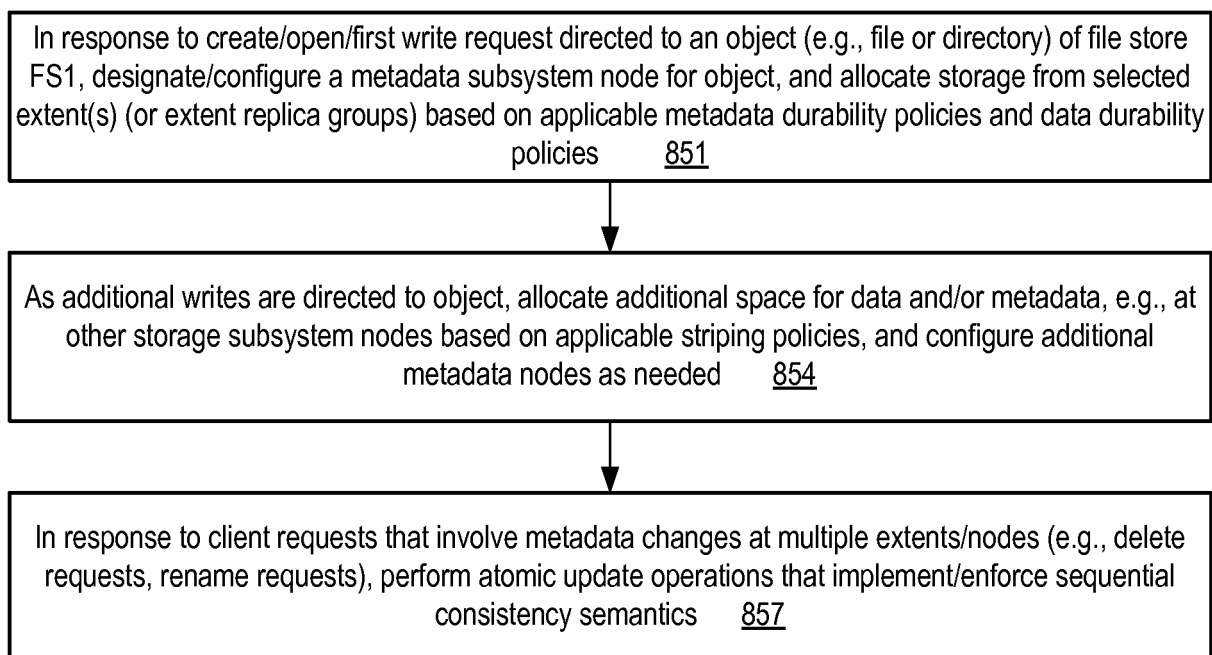


FIG. 8b

9/69

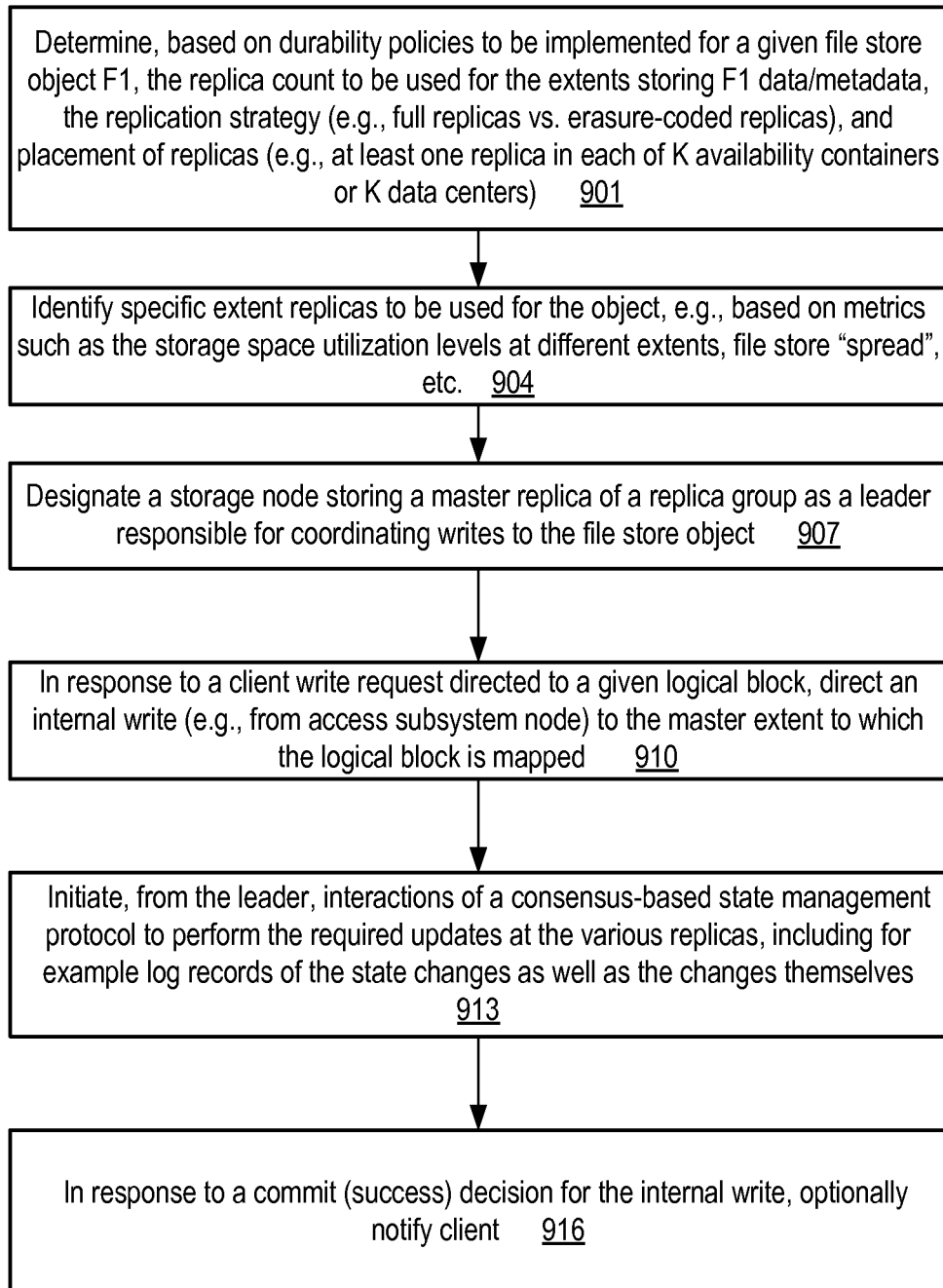


FIG. 9

10/69

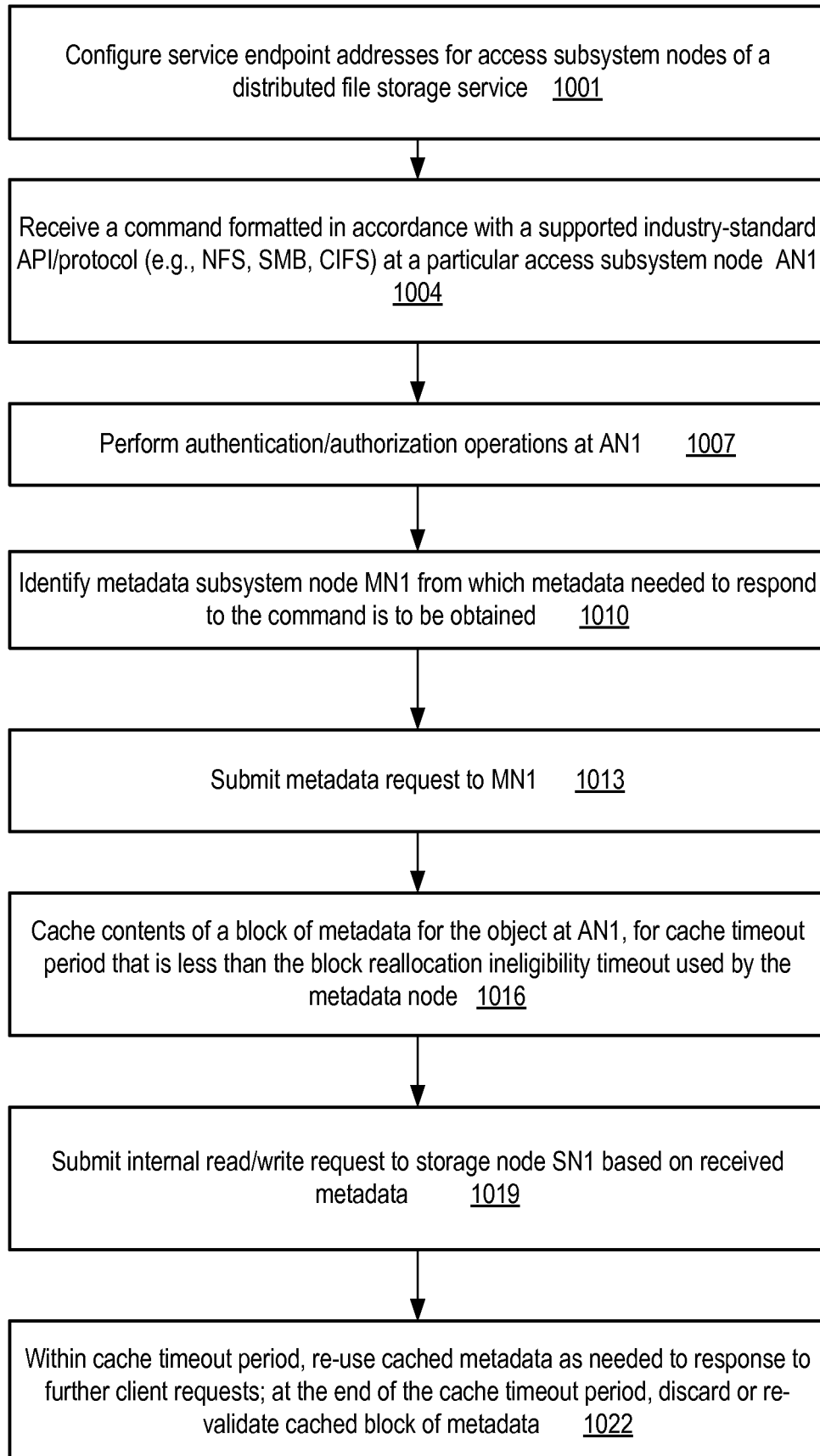


FIG. 10

11/69

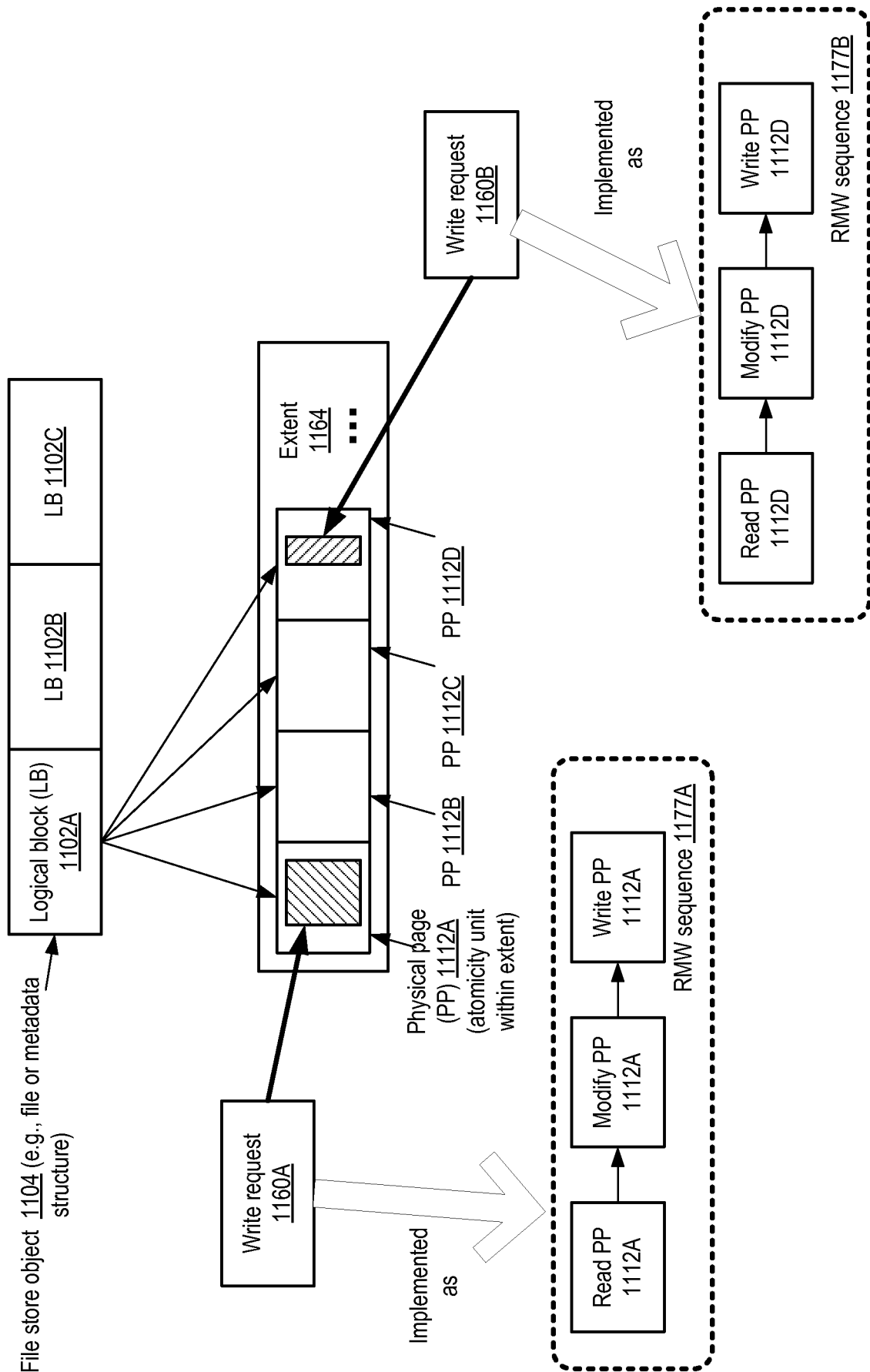


FIG. 11

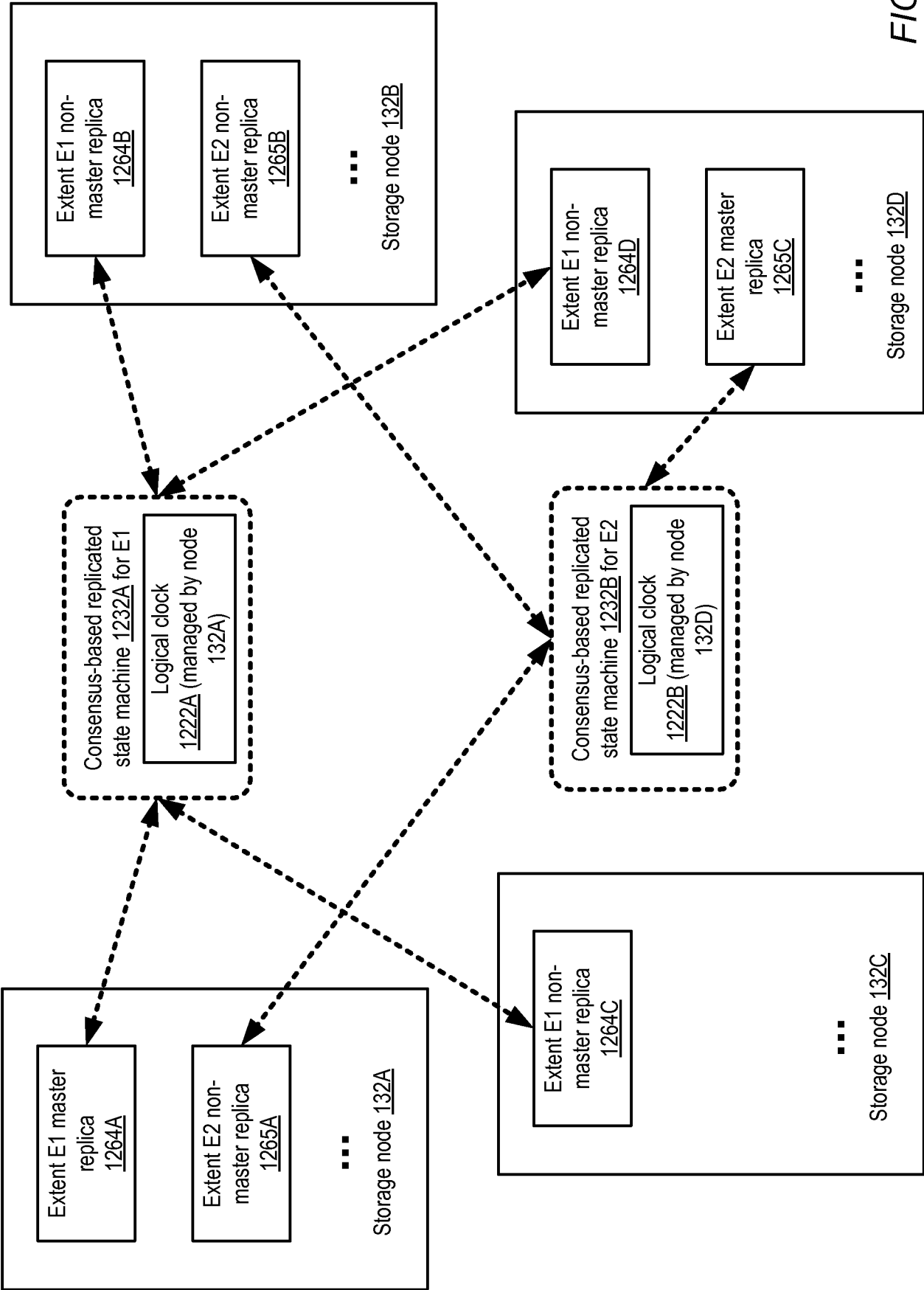


FIG. 12

13/69

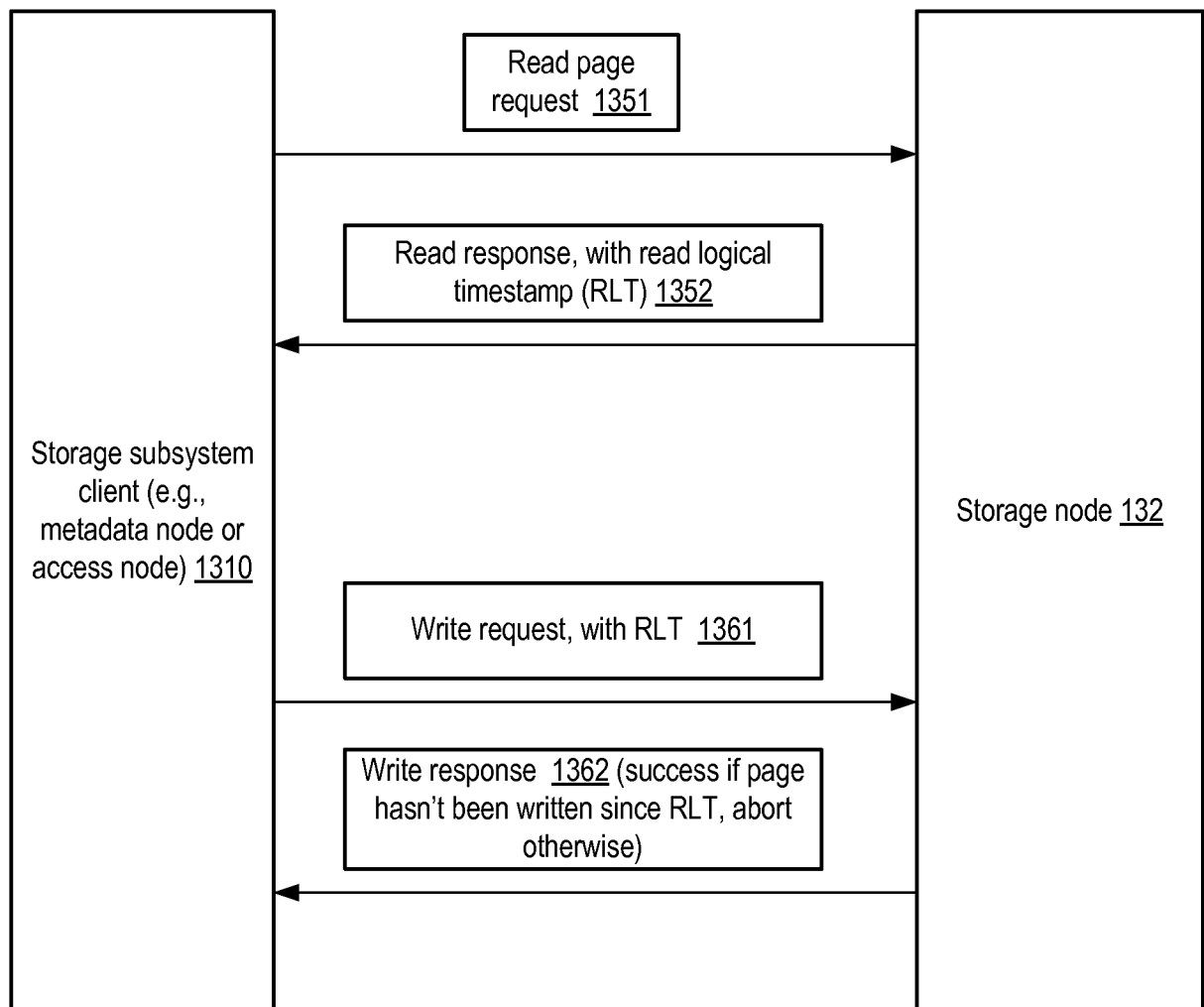


FIG. 13

14/69

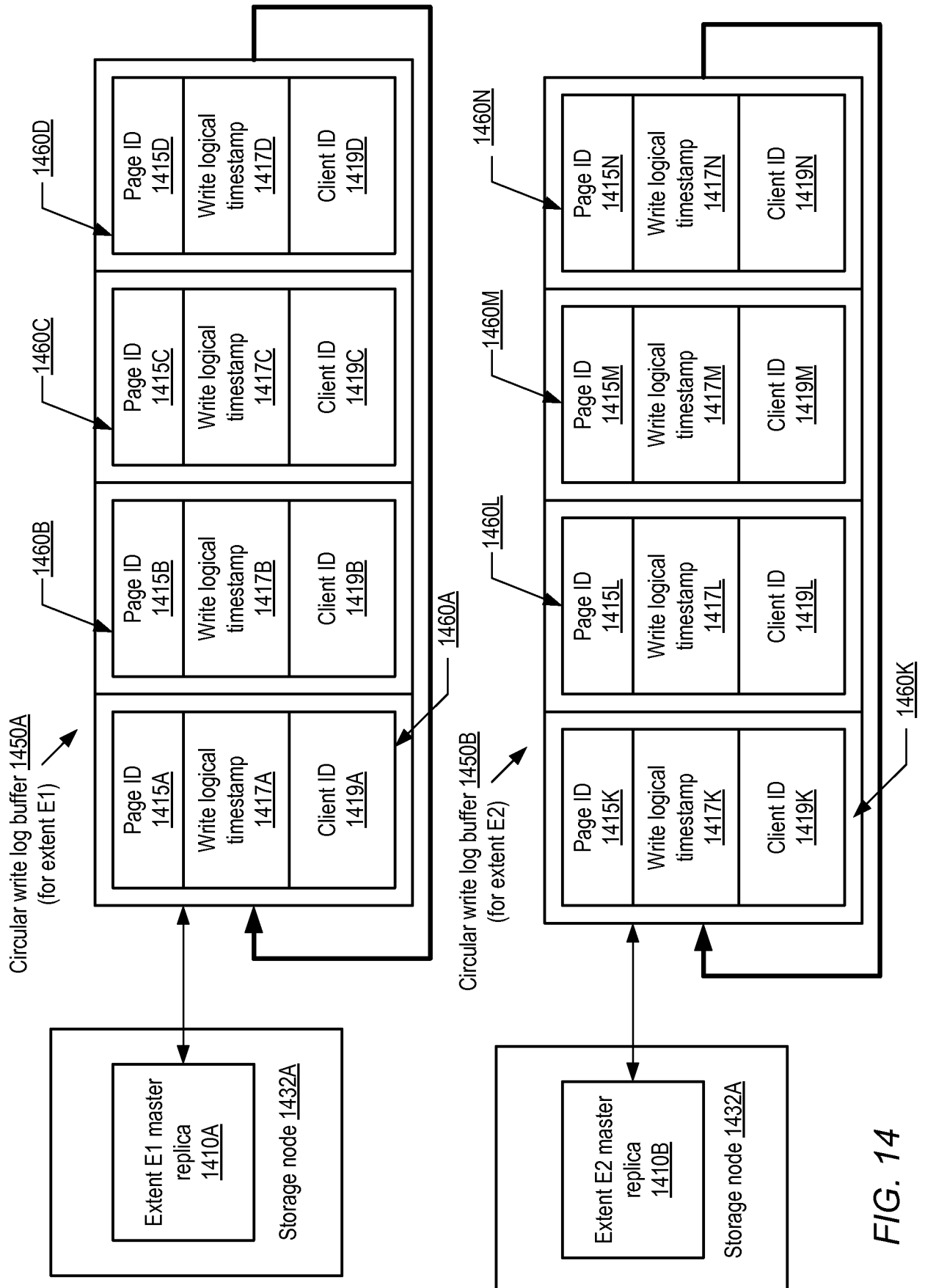


FIG. 14

15/69

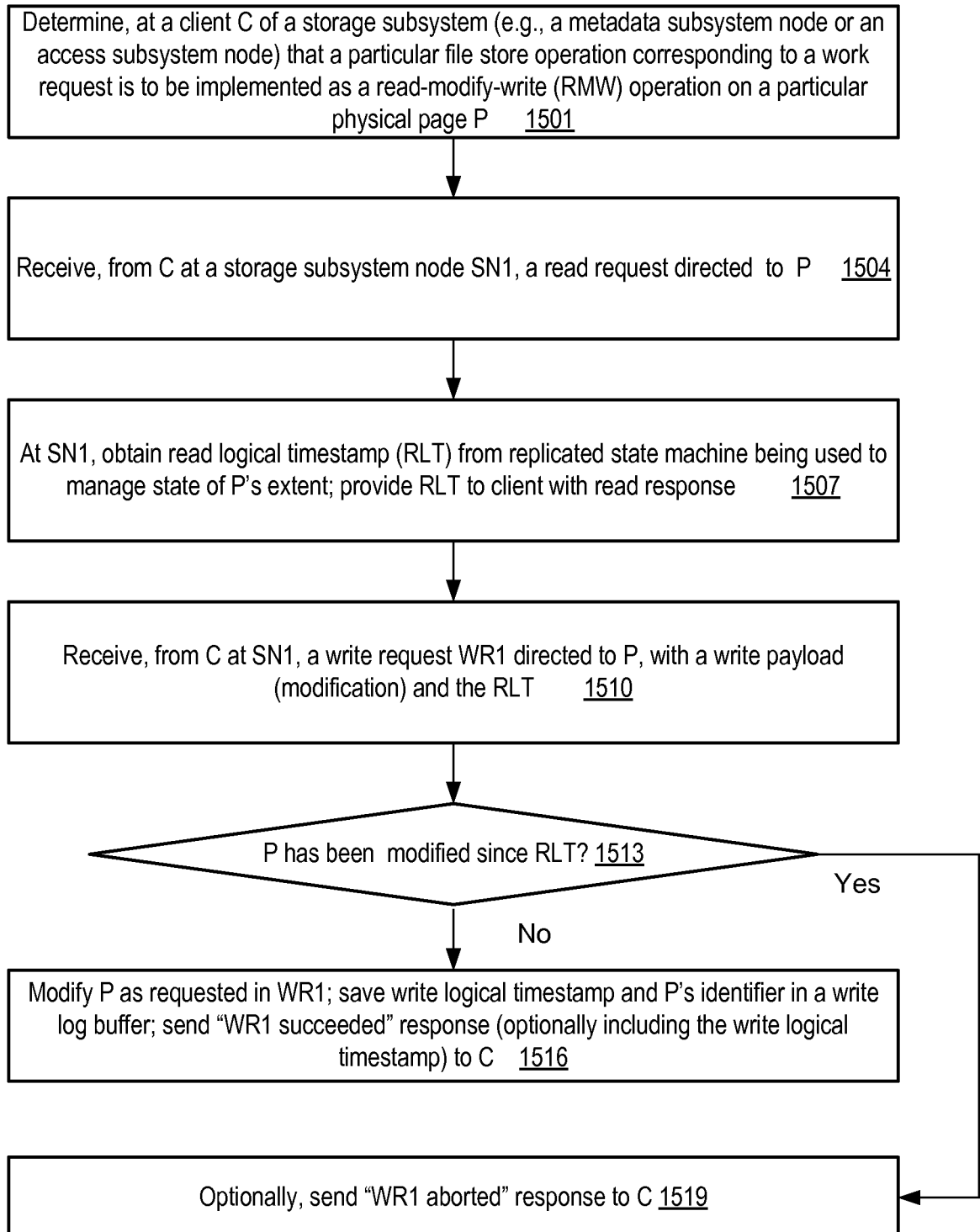


FIG. 15

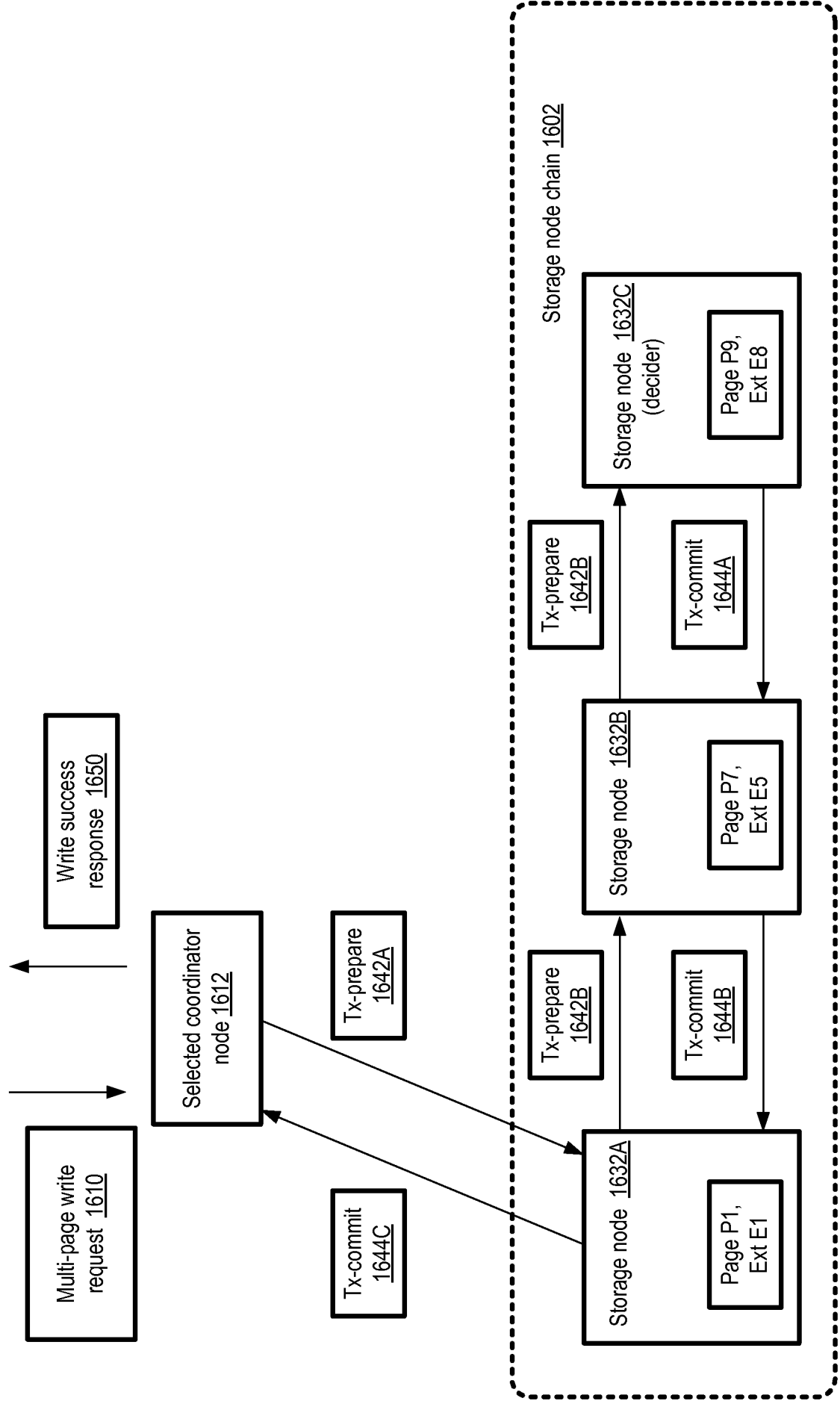


FIG. 16

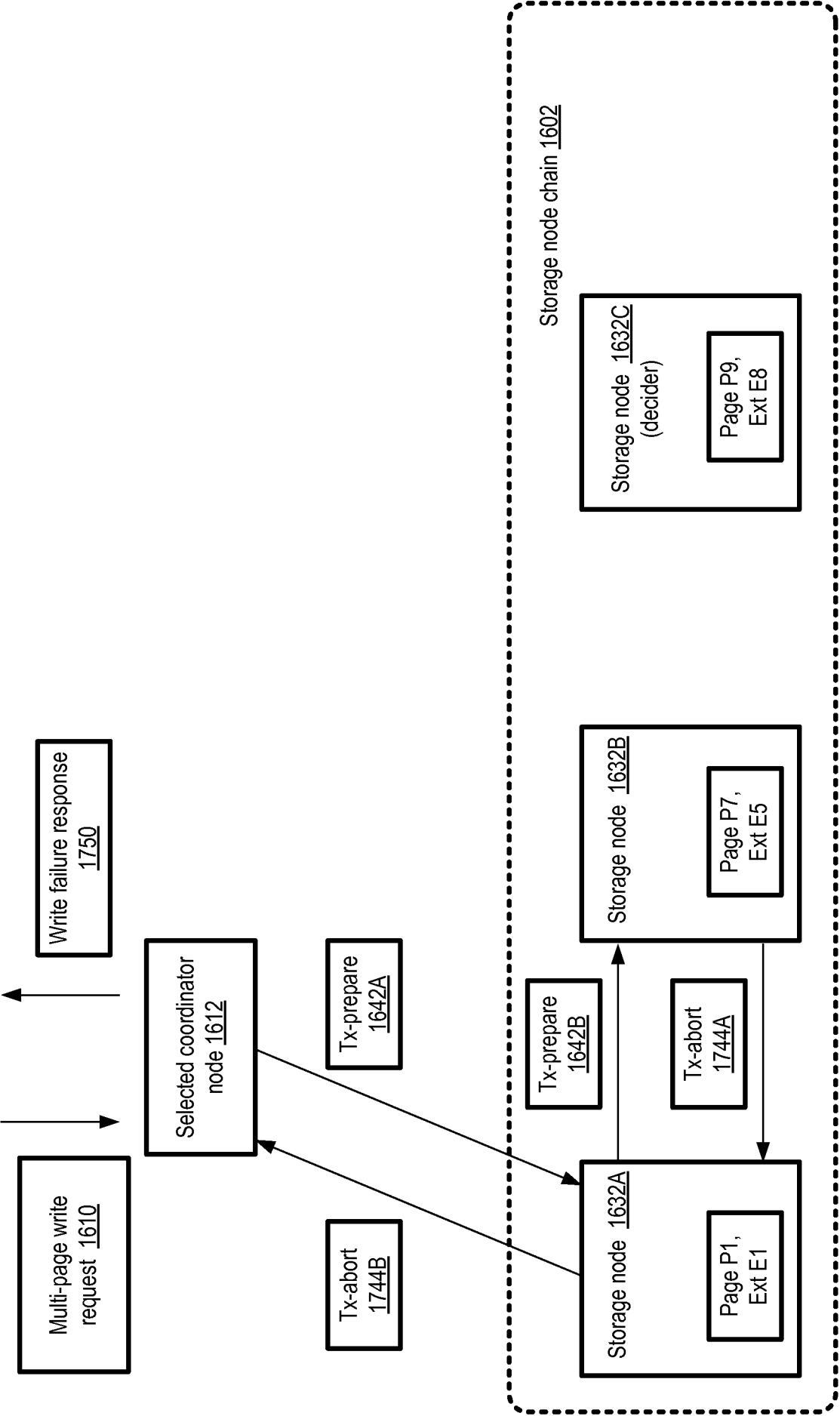


FIG. 17

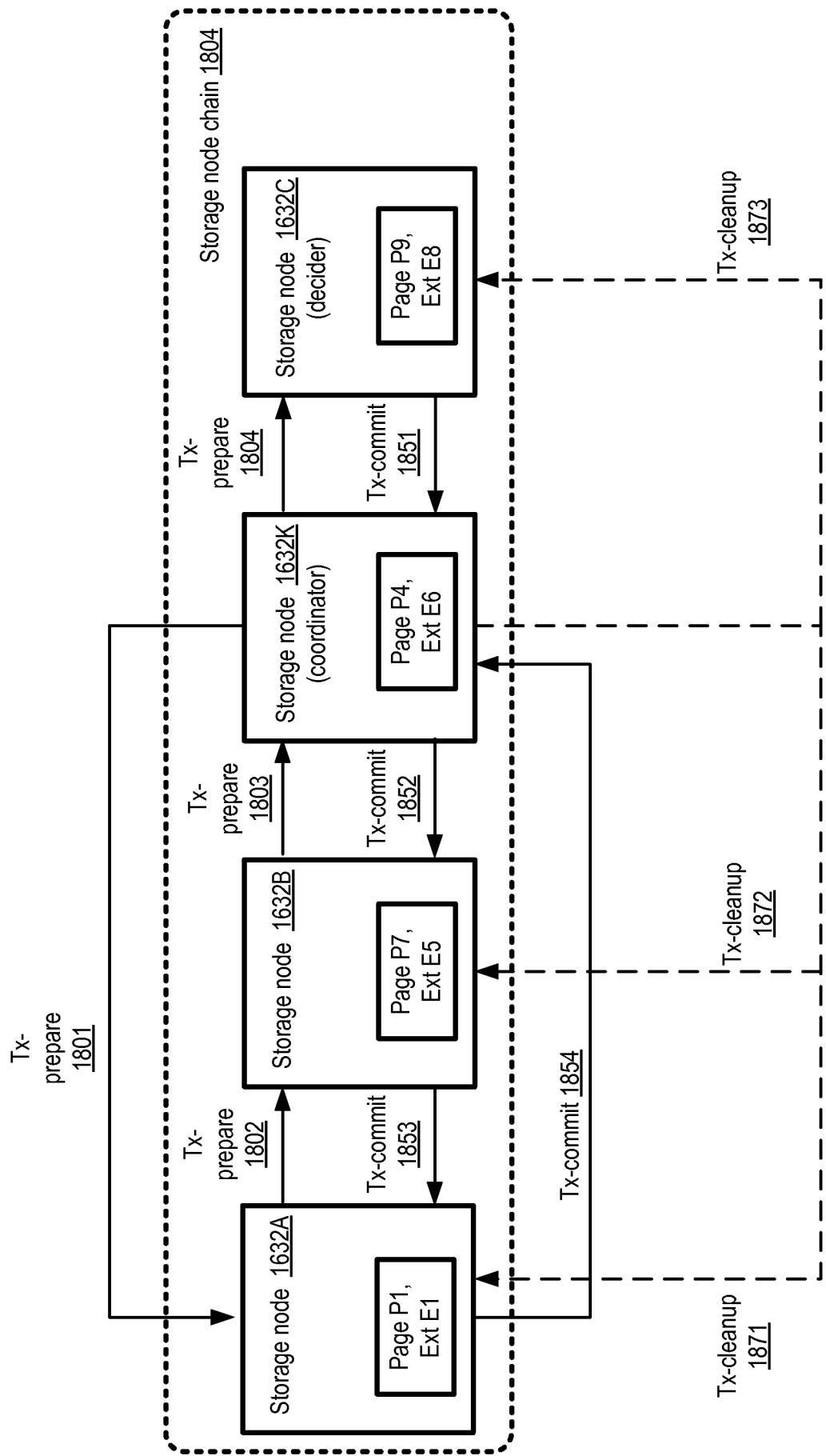


FIG. 18

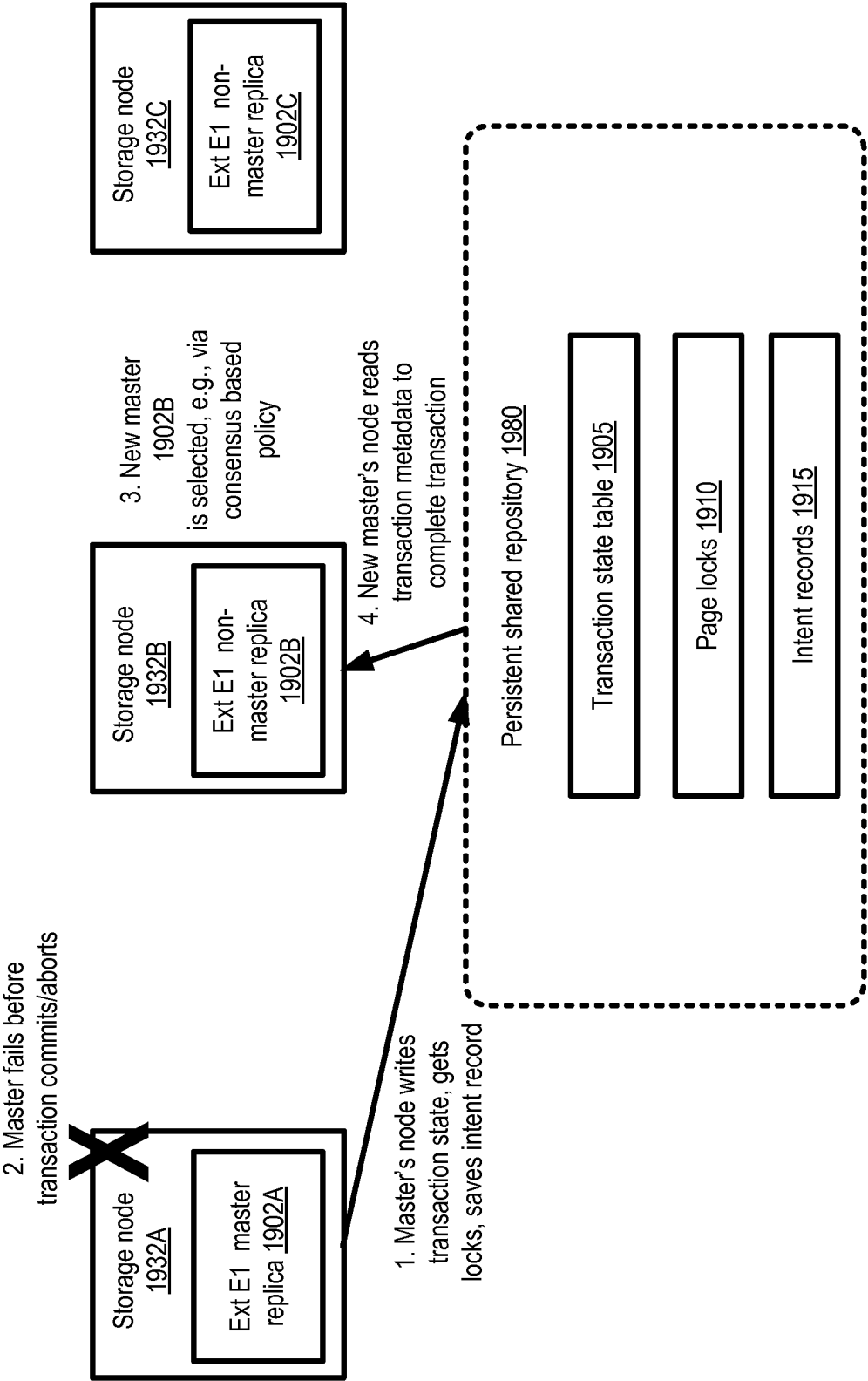


FIG. 19

20/69

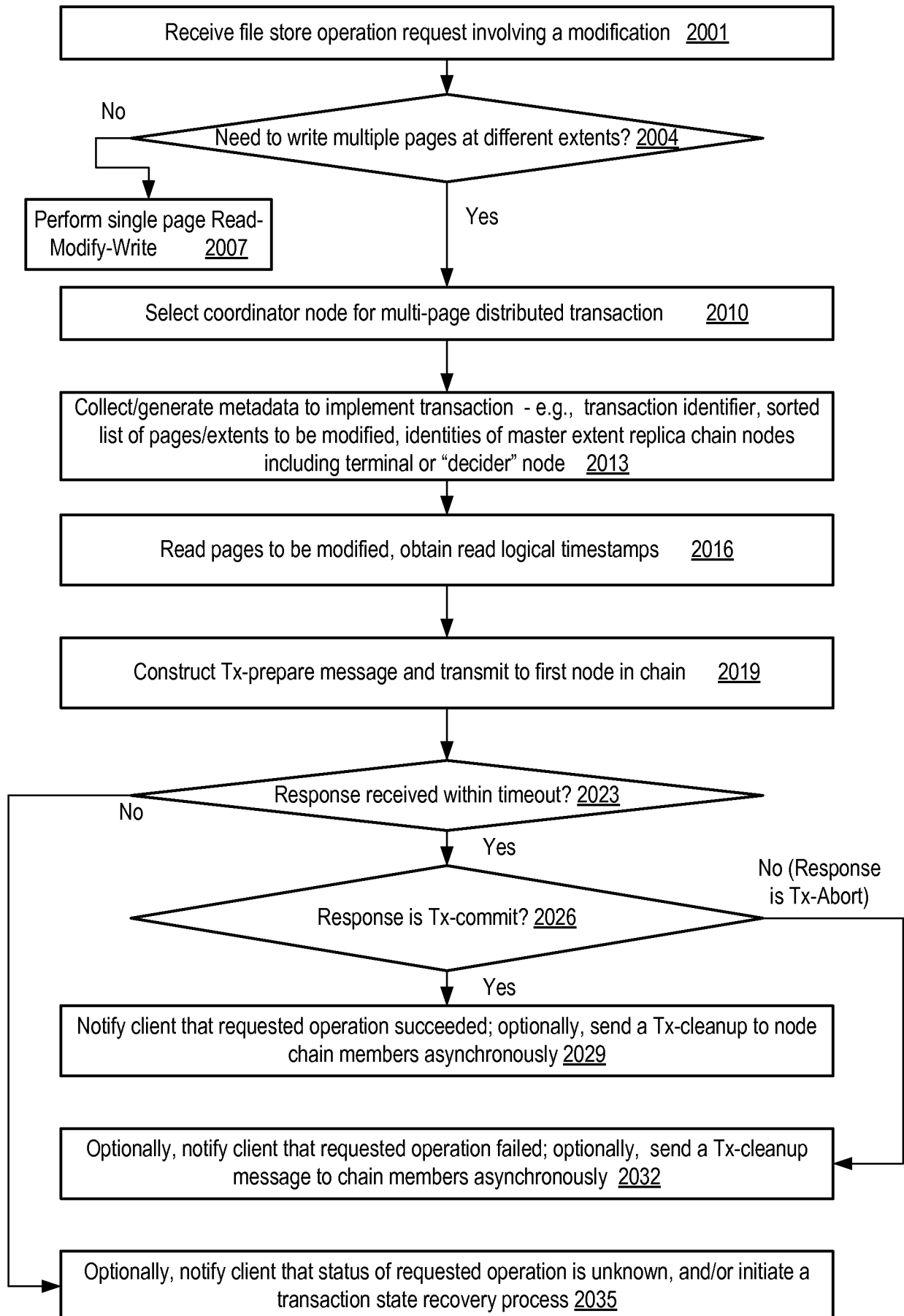


FIG. 20

21/69

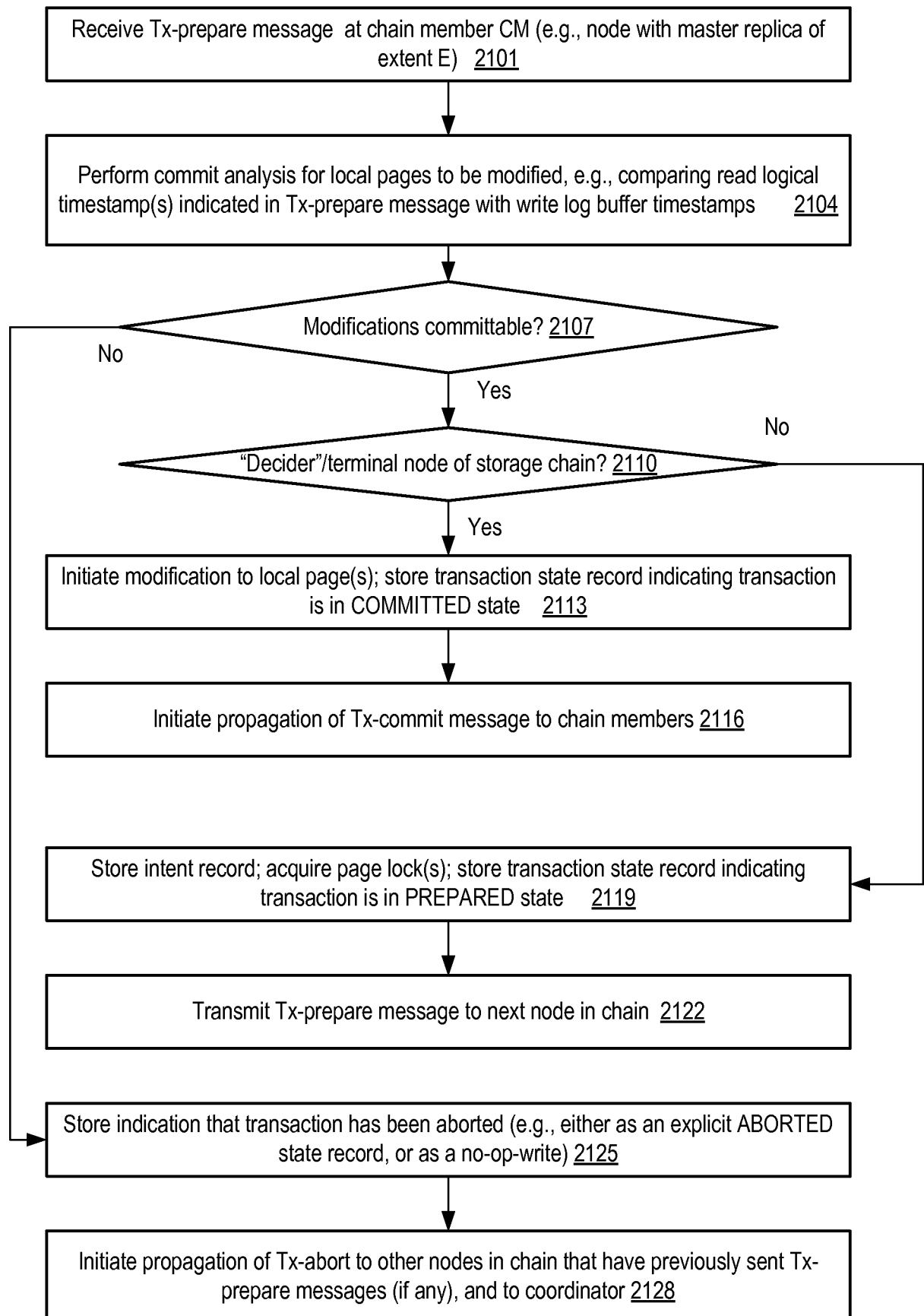


FIG. 21

22/69

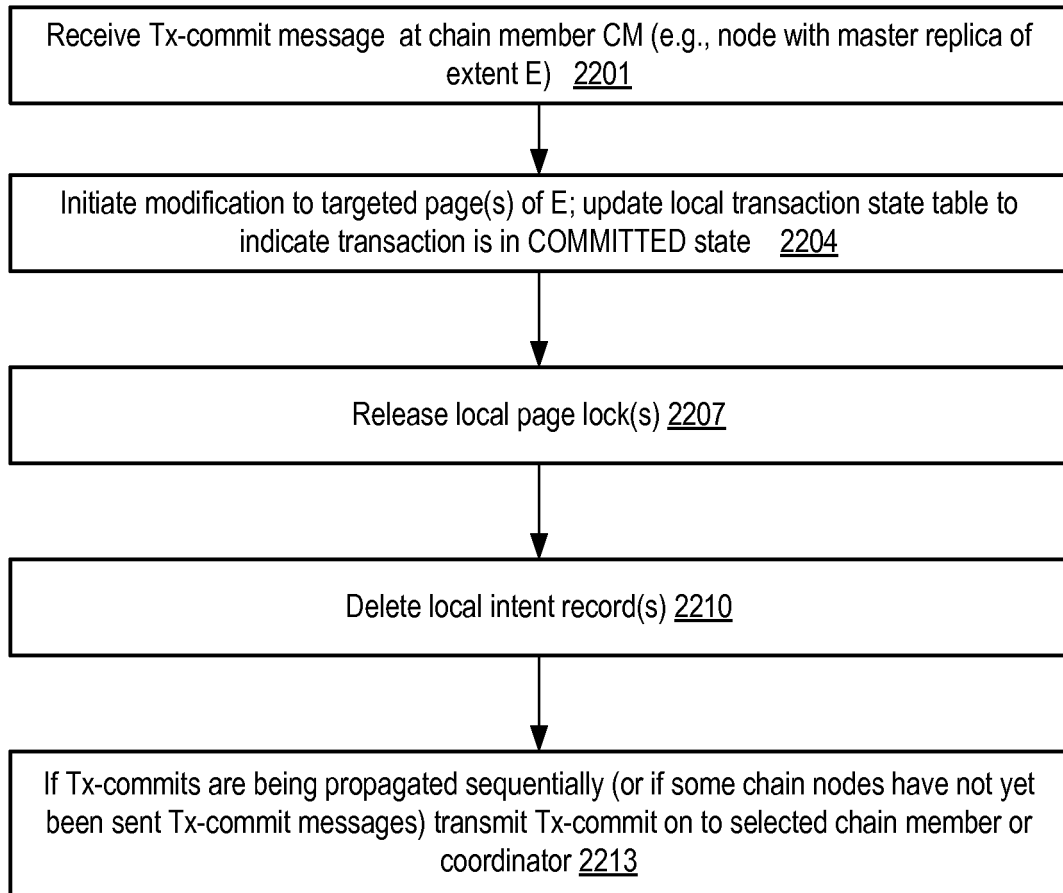


FIG. 22

23/69

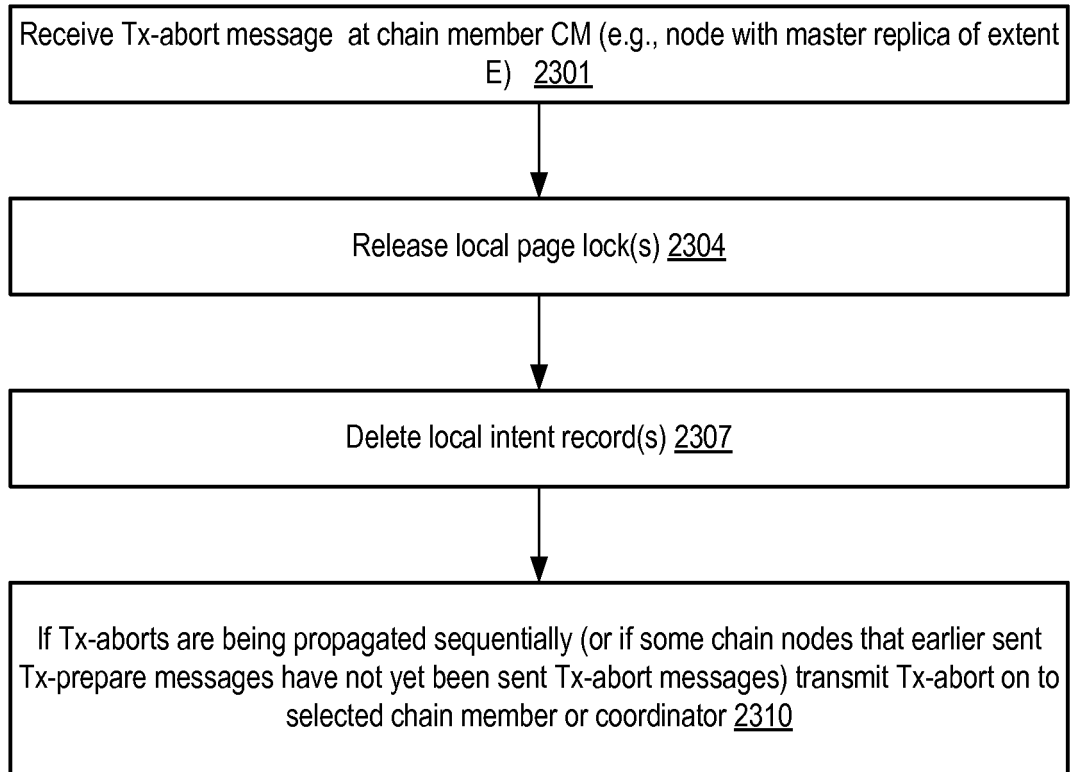


FIG. 23

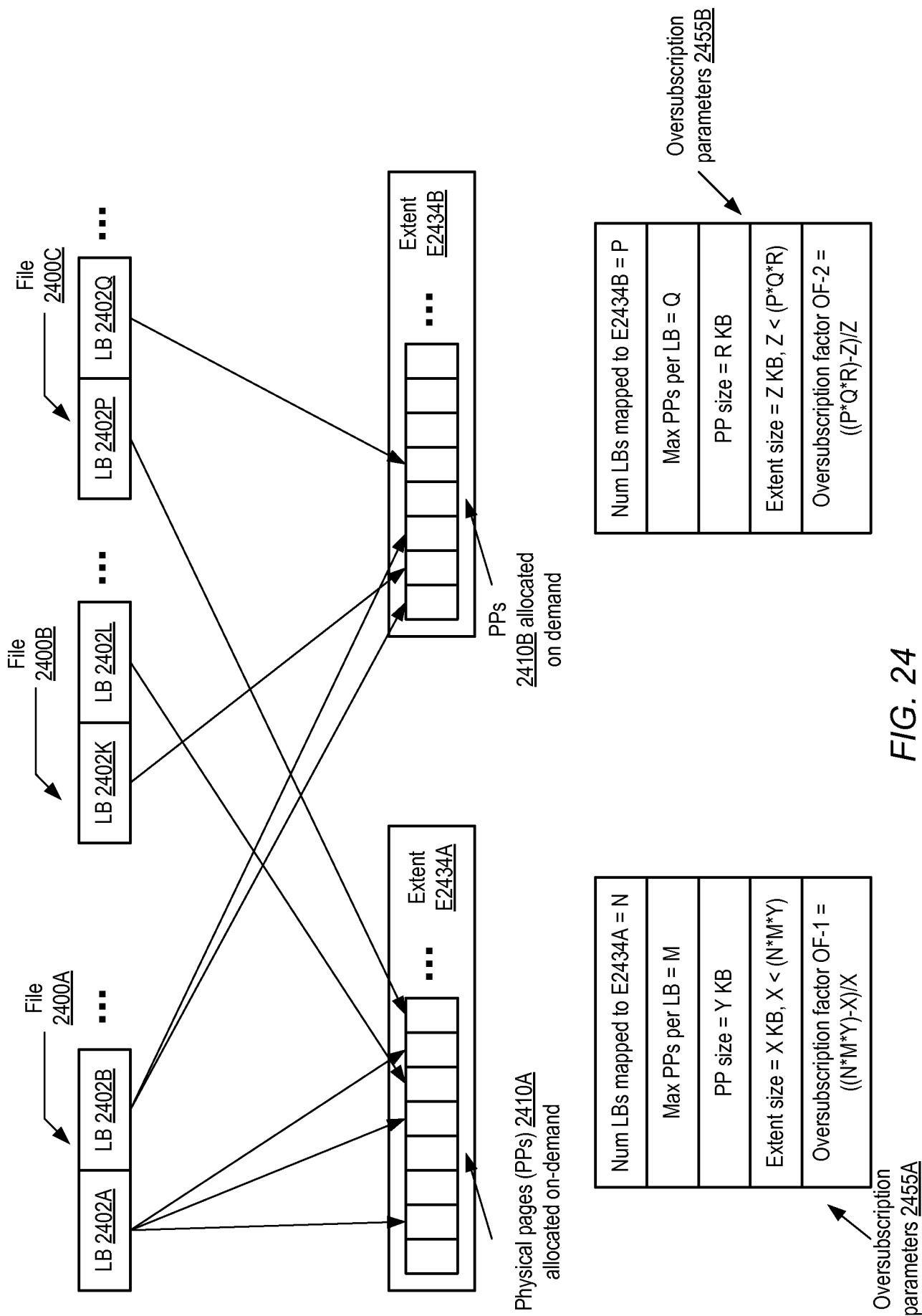


FIG. 24

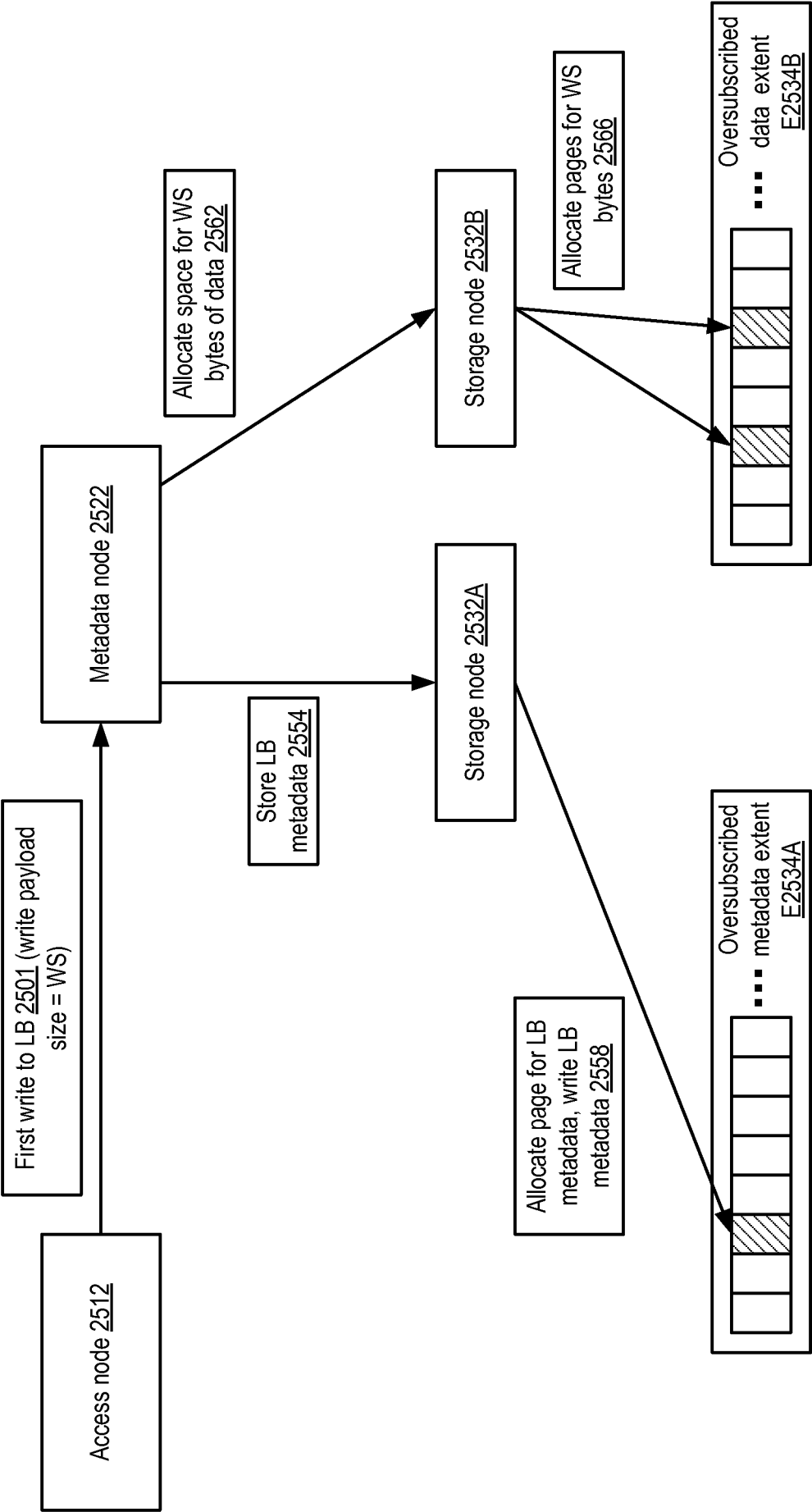


FIG. 25

26/69

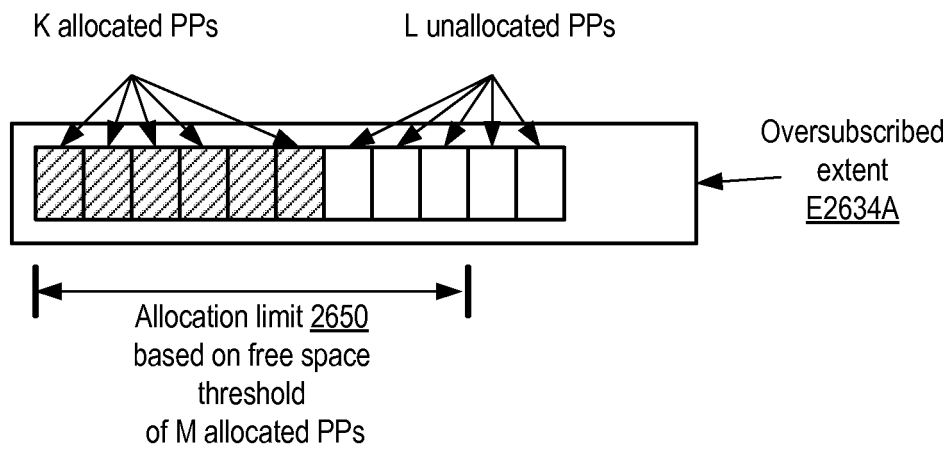


FIG. 26a

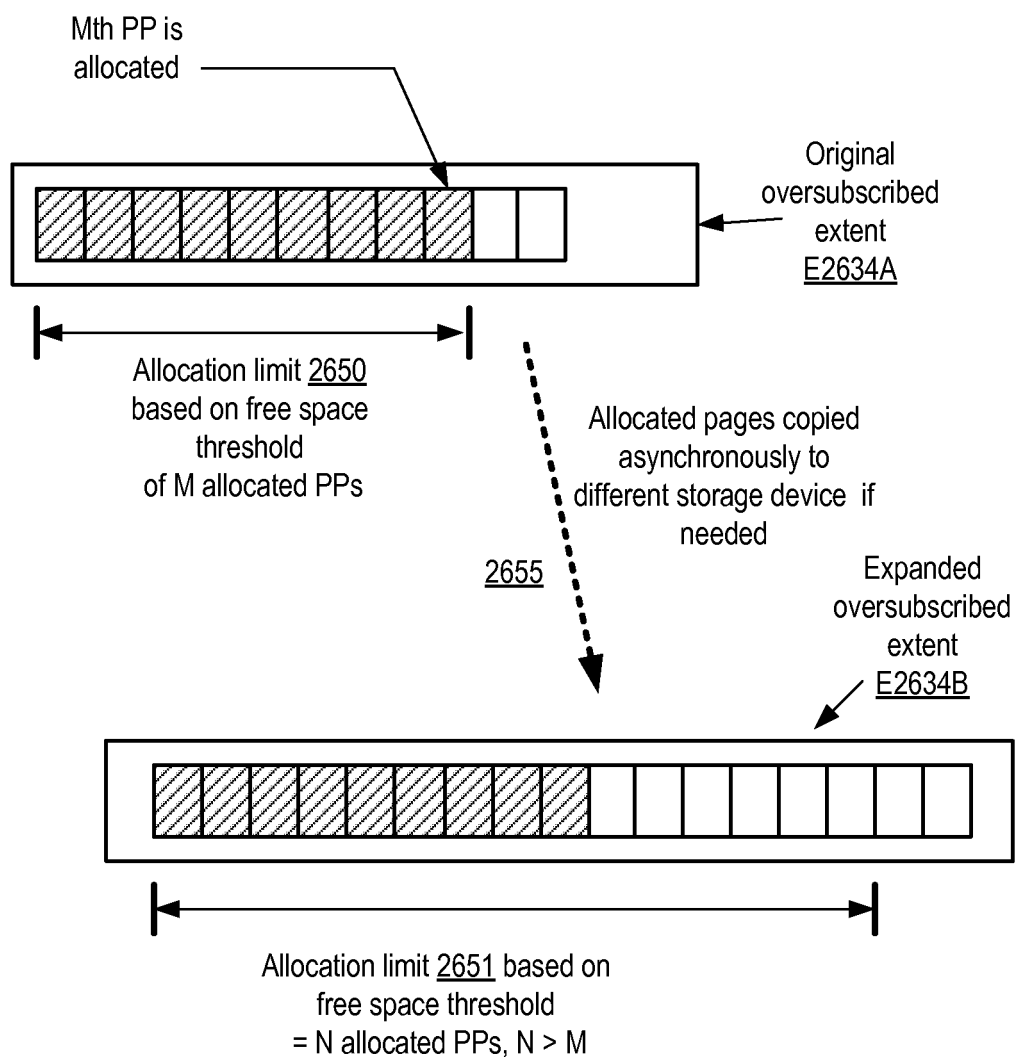


FIG. 26b

27/69

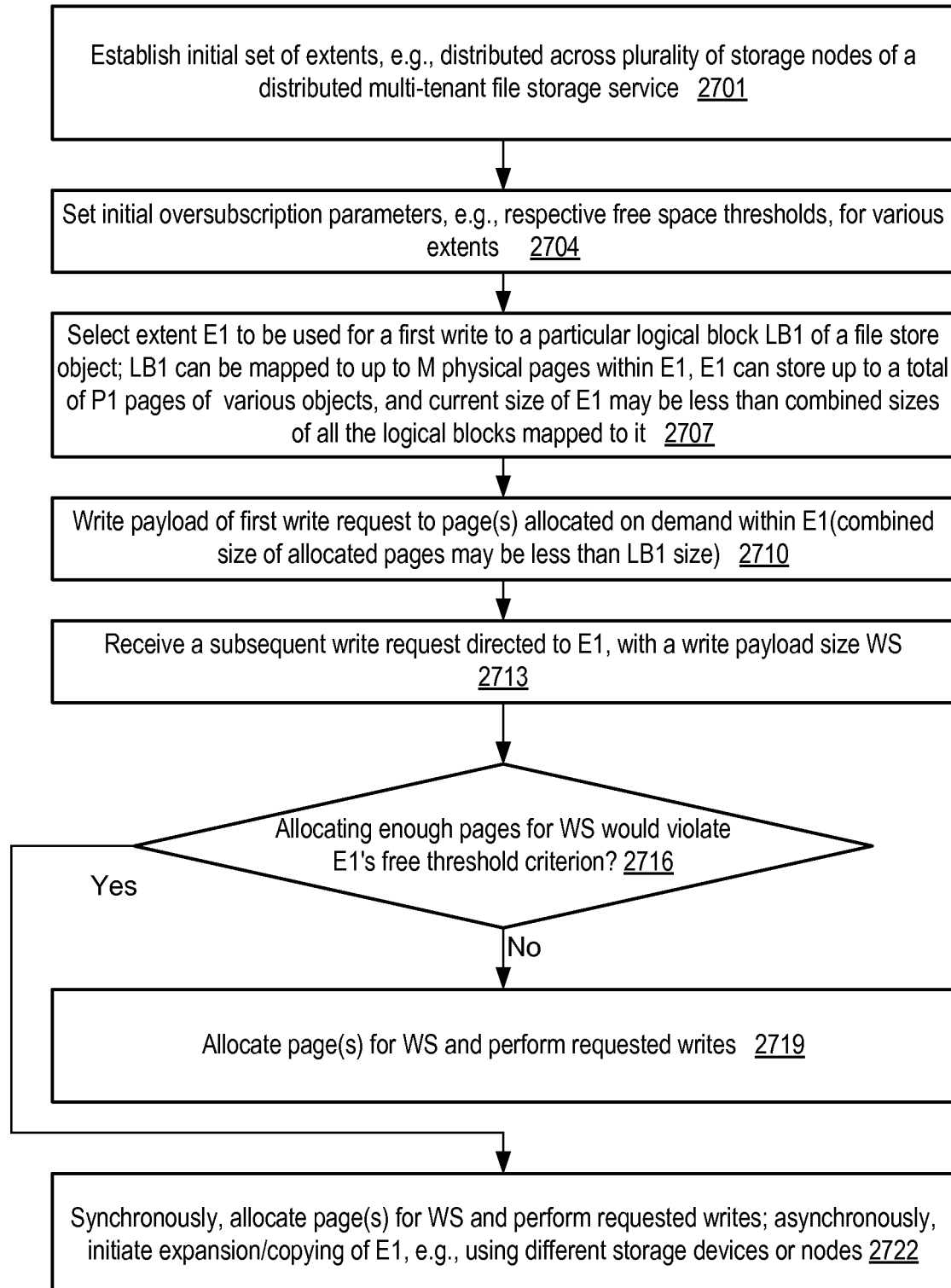


FIG. 27

28/69

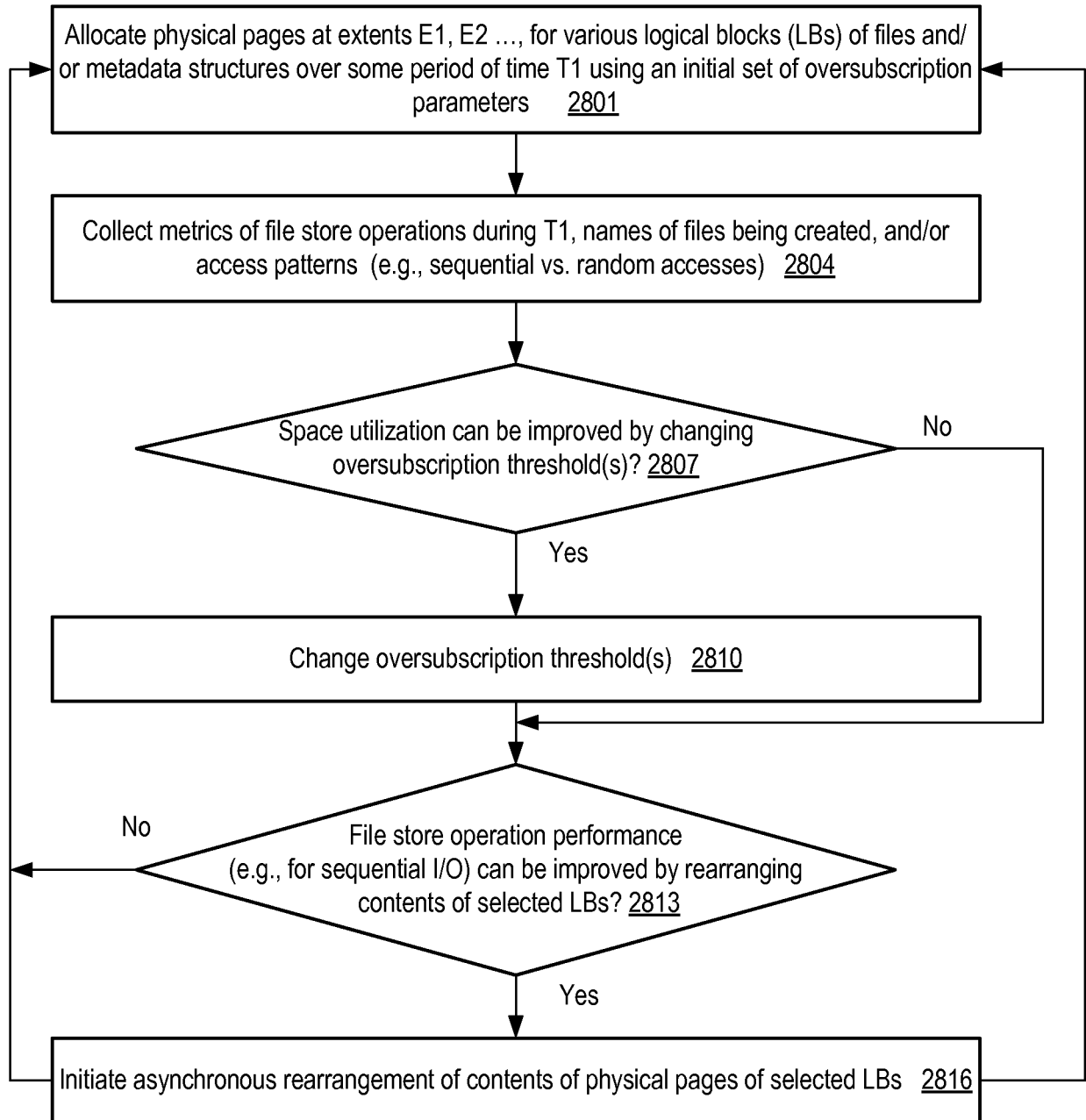


FIG. 28

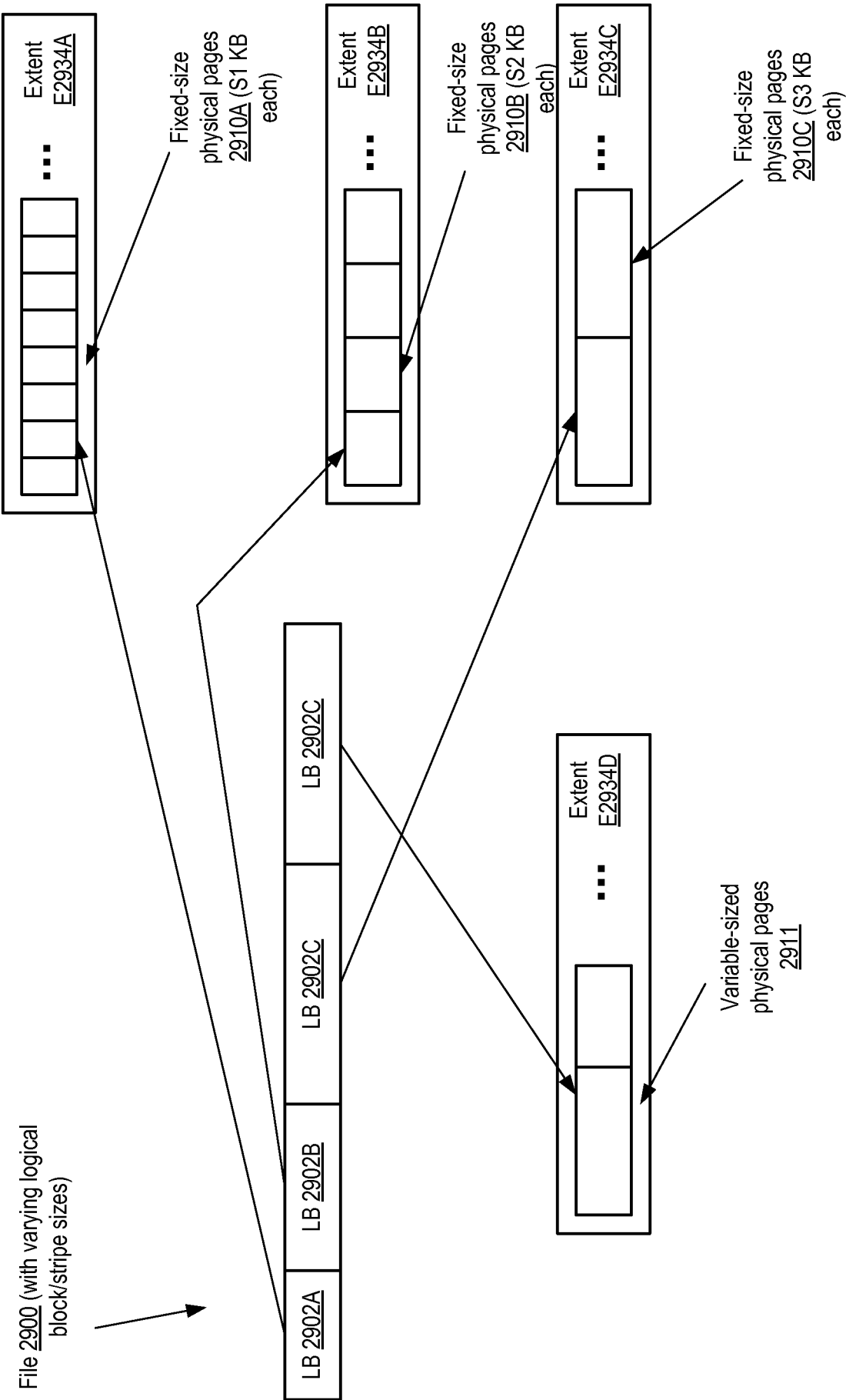


FIG. 29

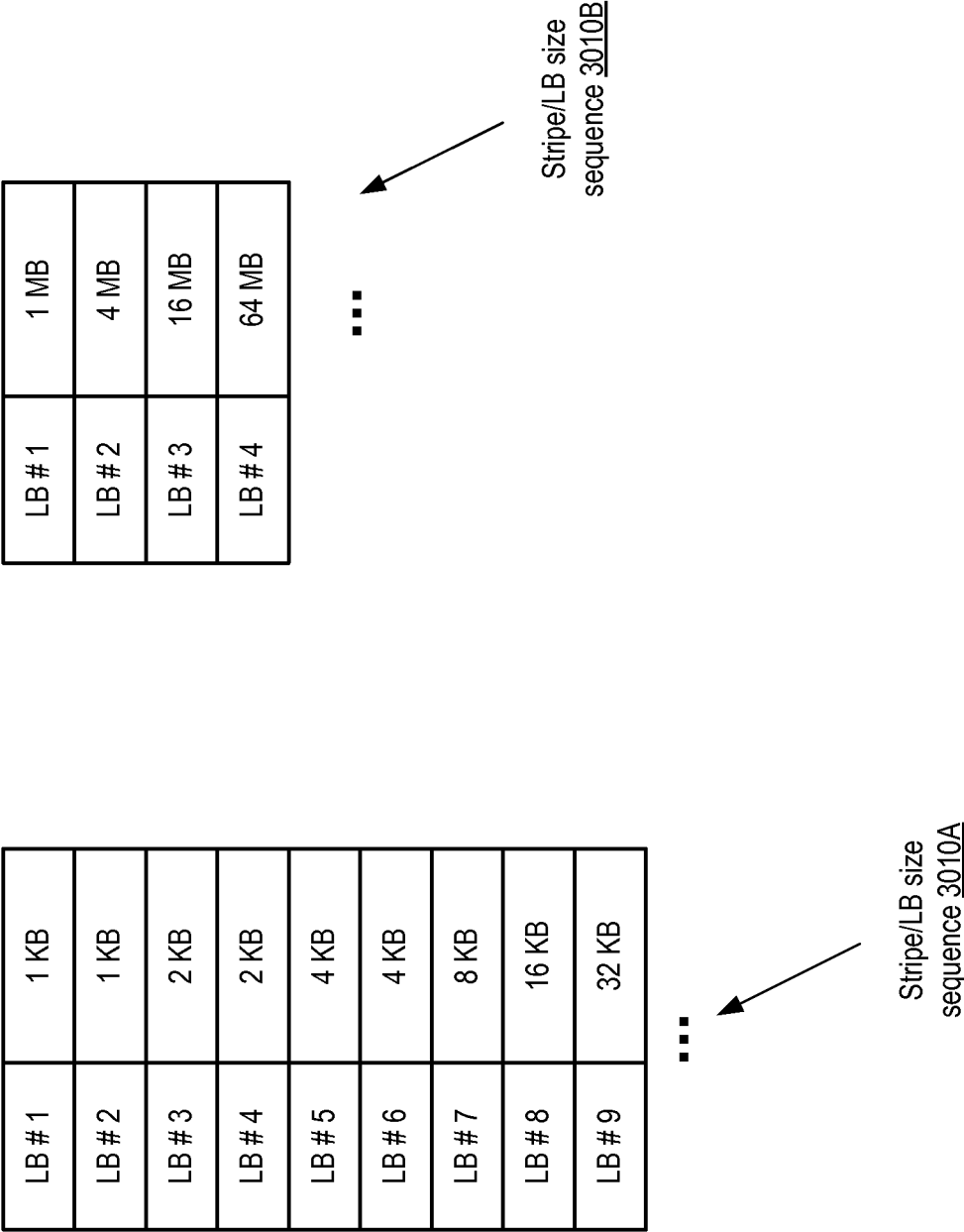


FIG. 30

31/69

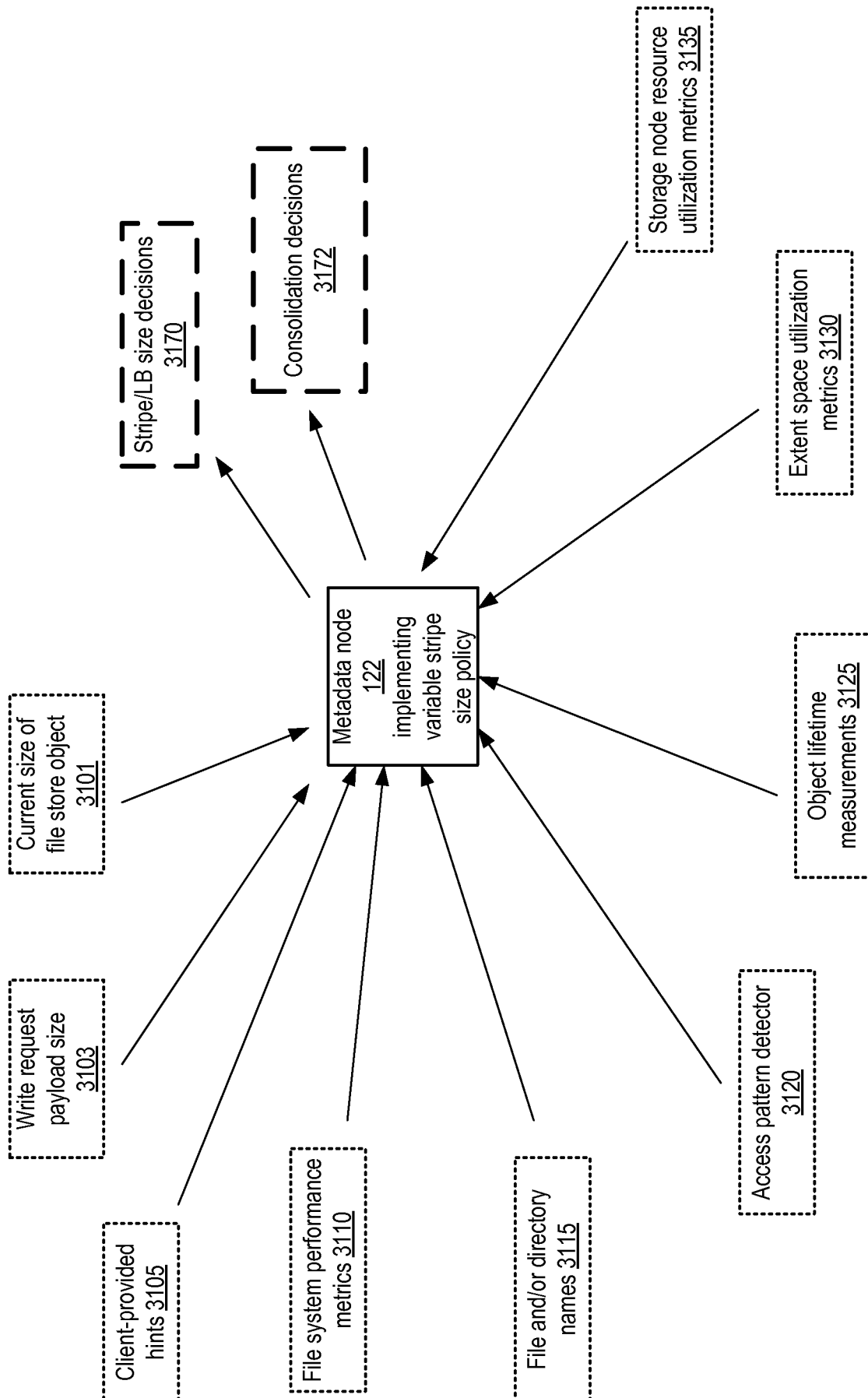


FIG. 31

32/69

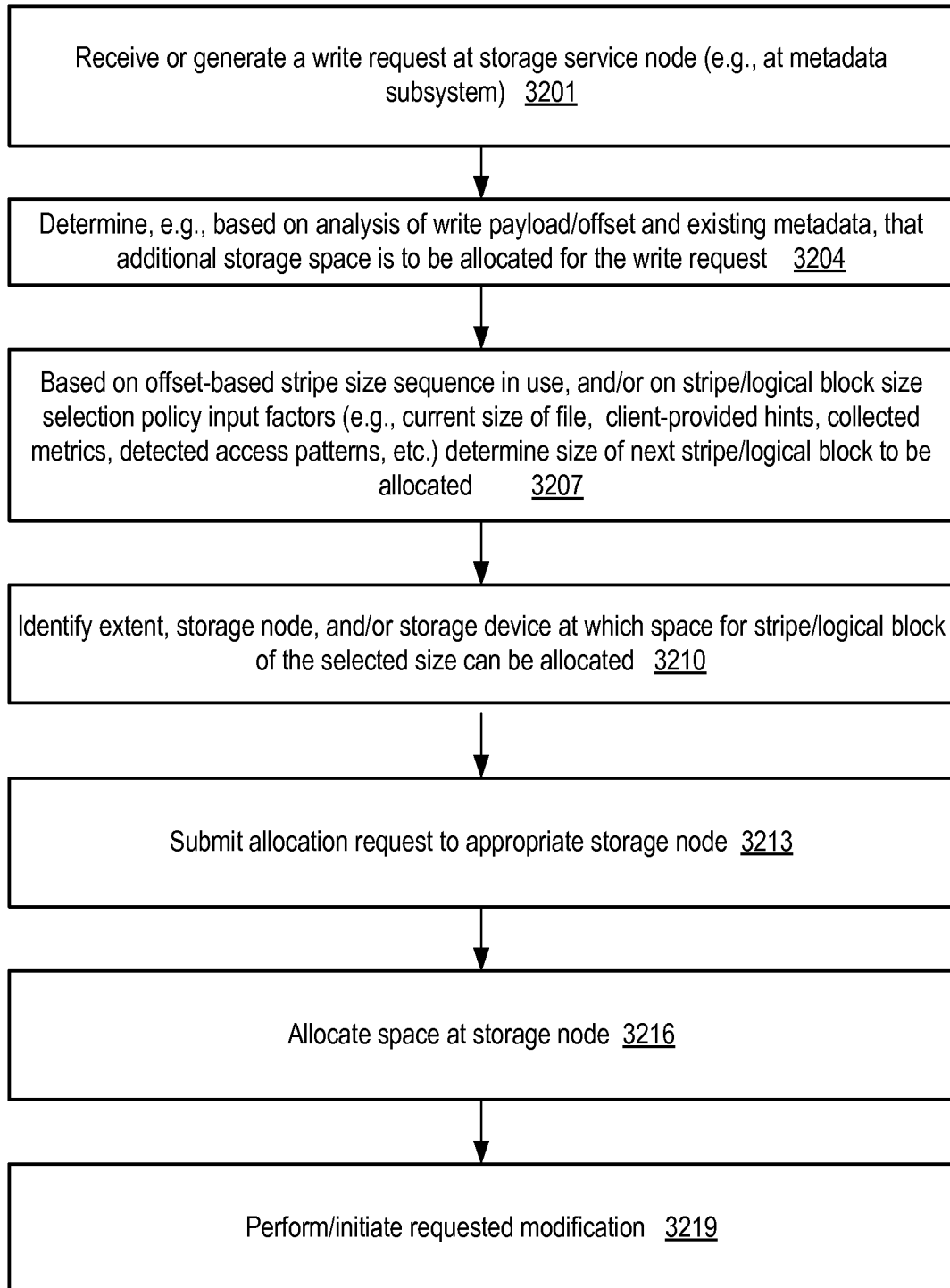


FIG. 32

33/69

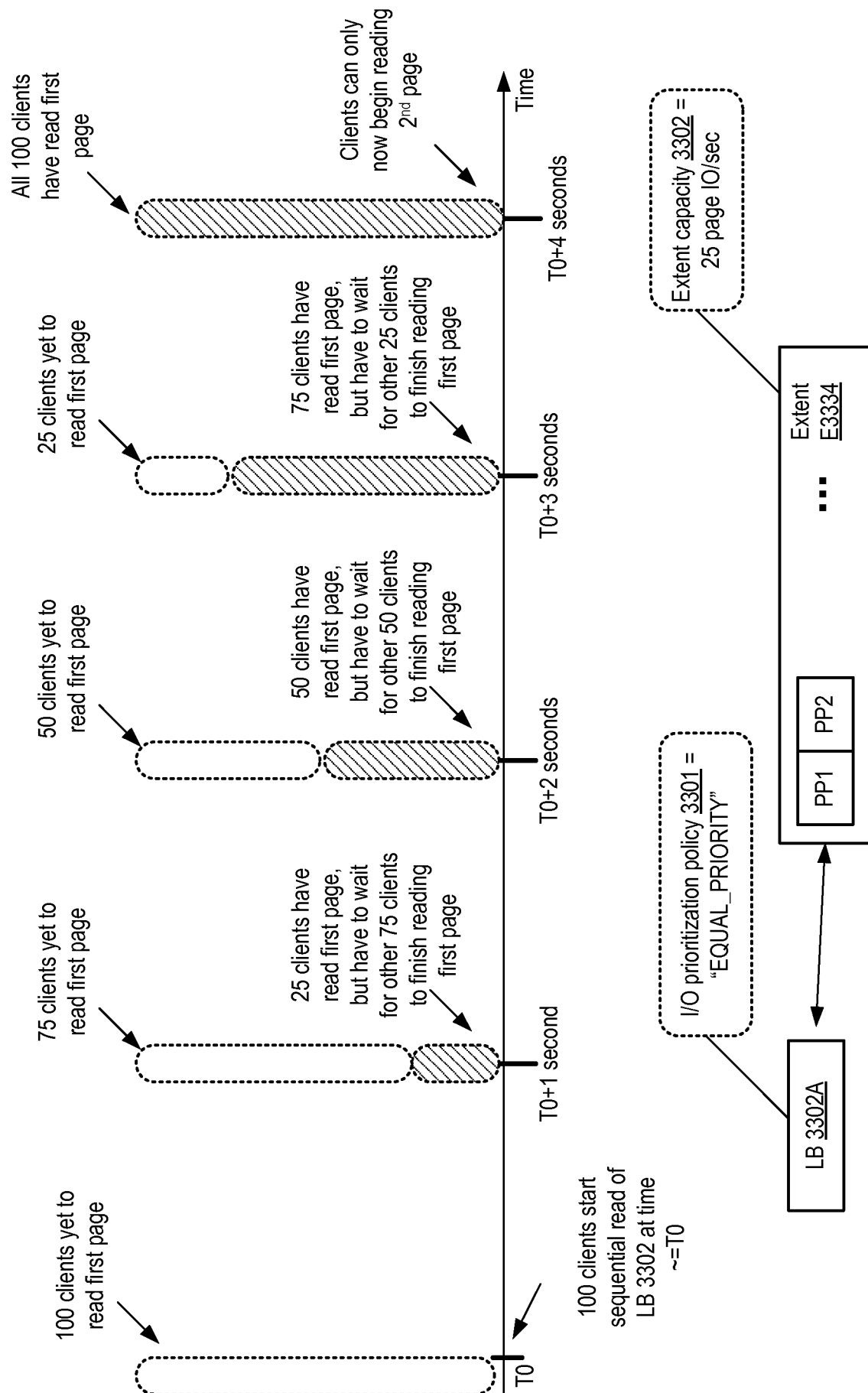


FIG. 33

34/69

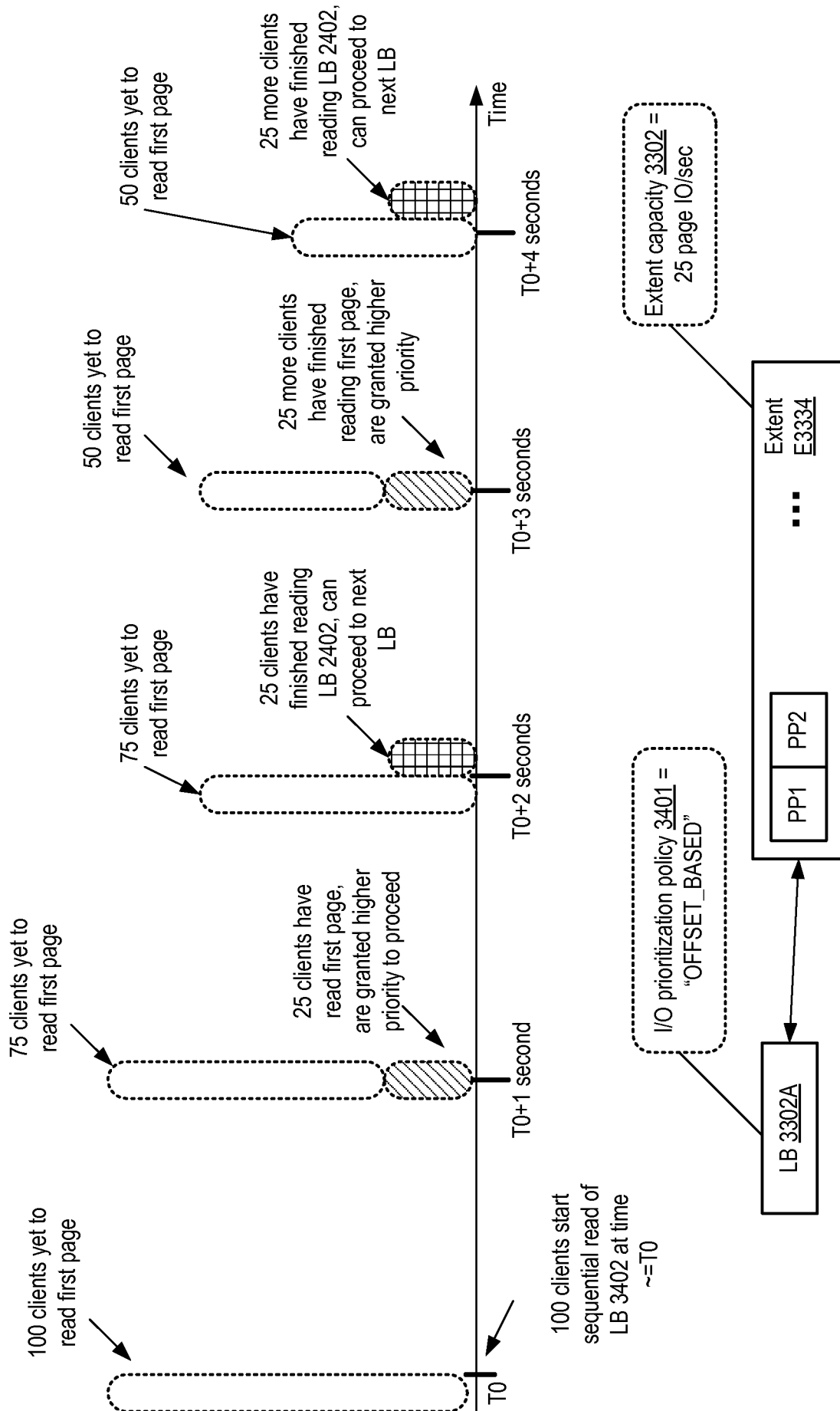


FIG. 34

35/69

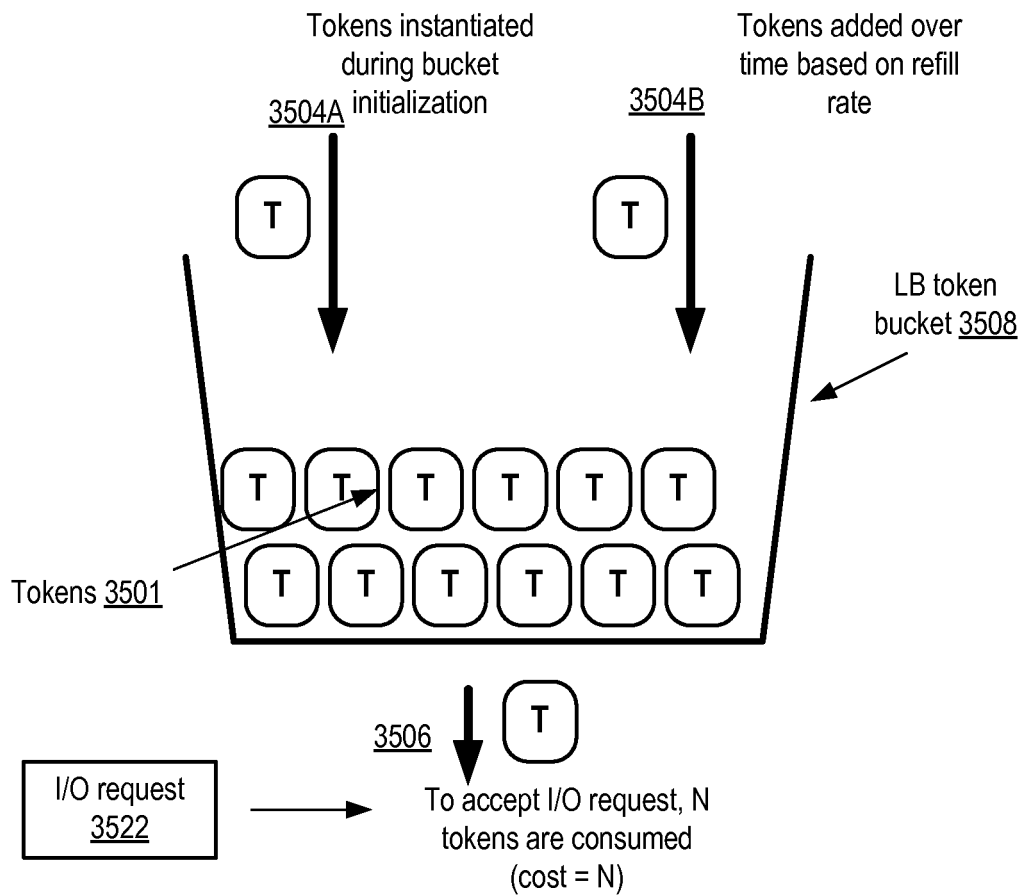


FIG. 35a

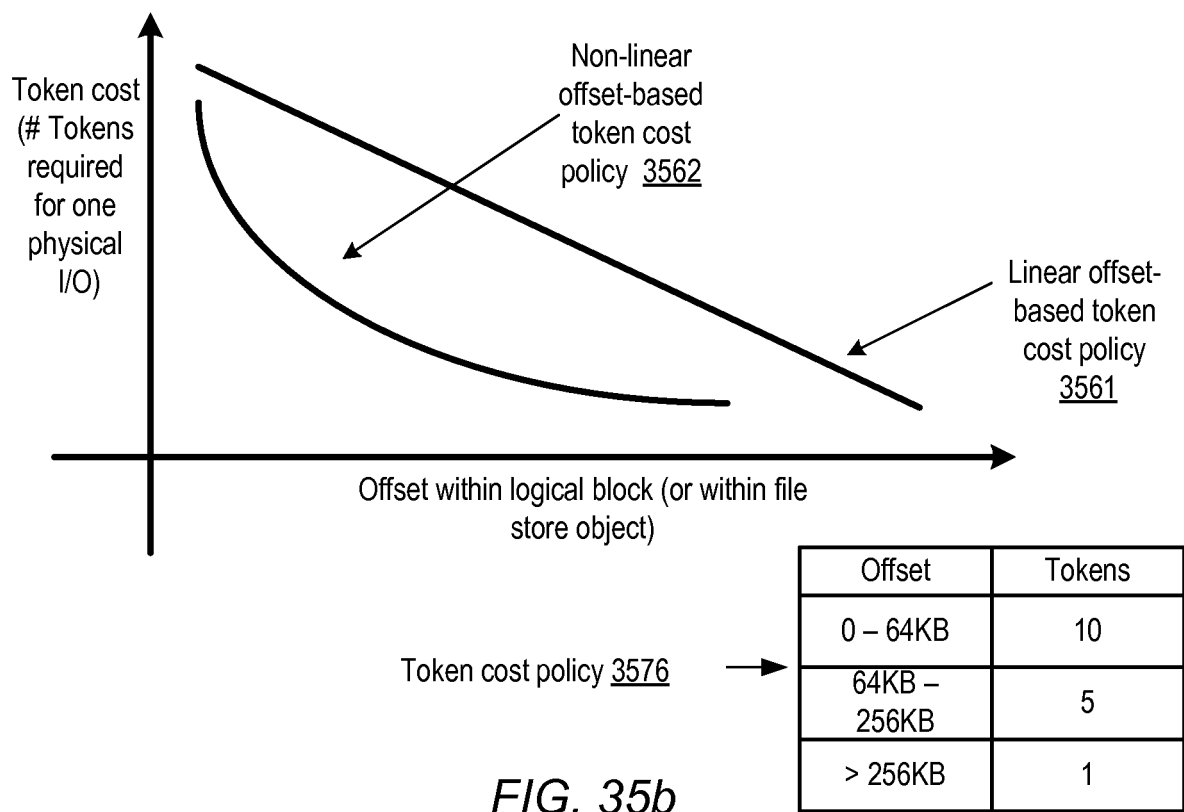


FIG. 35b

36/69

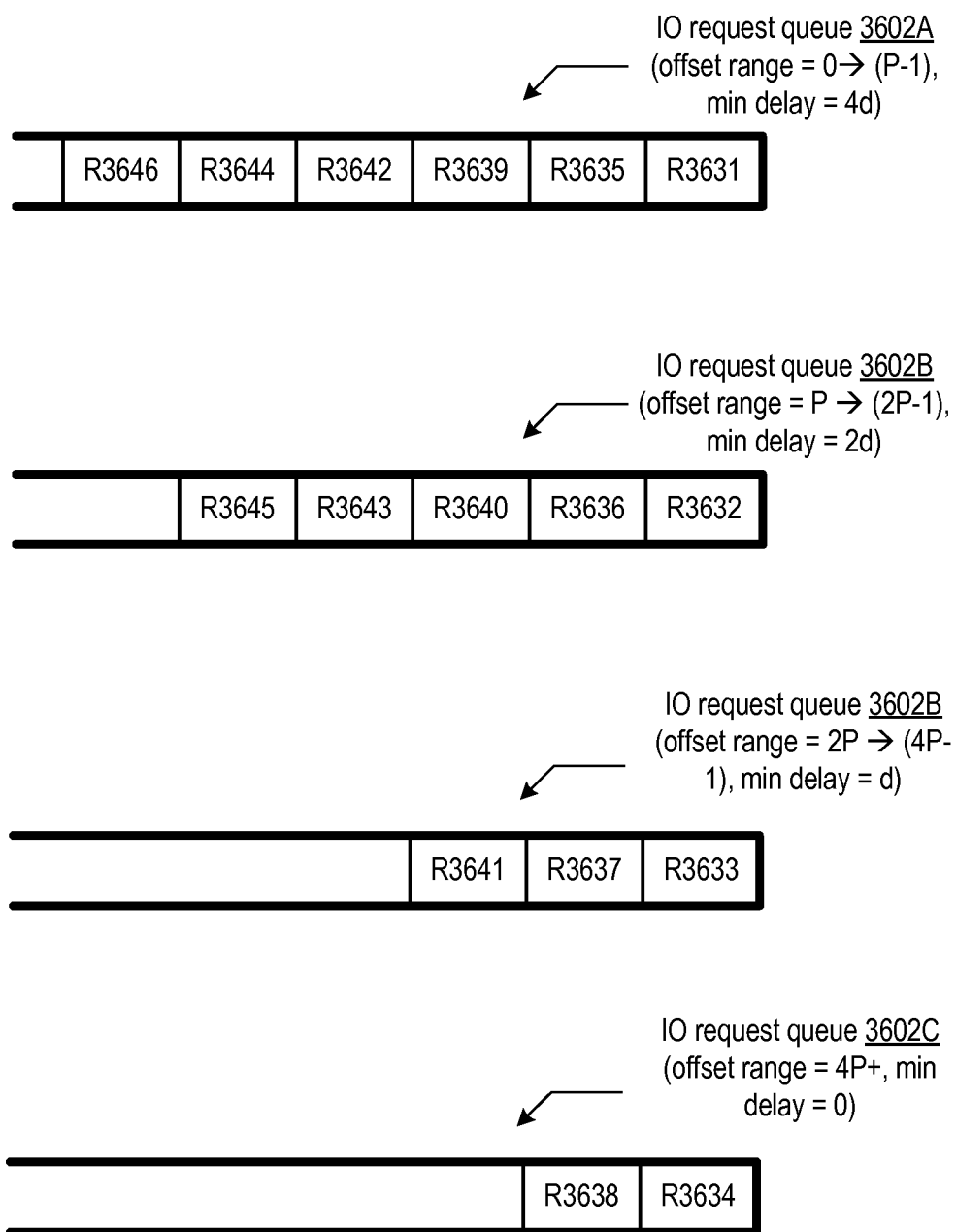


FIG. 36

3780

Content type <u>3702</u>	I/O type <u>3704</u>	Access pattern <u>3706</u>	I/O payload size <u>3708</u>	Object size <u>3710</u>	Congestion control parameter settings <u>3710</u>
Metadata	Read	Sequential	< 4KB	0 to S1 MB	Offset-dependent priority, linear
Metadata	Write	Random	Any	Any	Equal priority
Data	Read	Sequential	> 64KB	> 128MB	Offset-dependent priority, exponential decay

...

FIG. 37

38/69

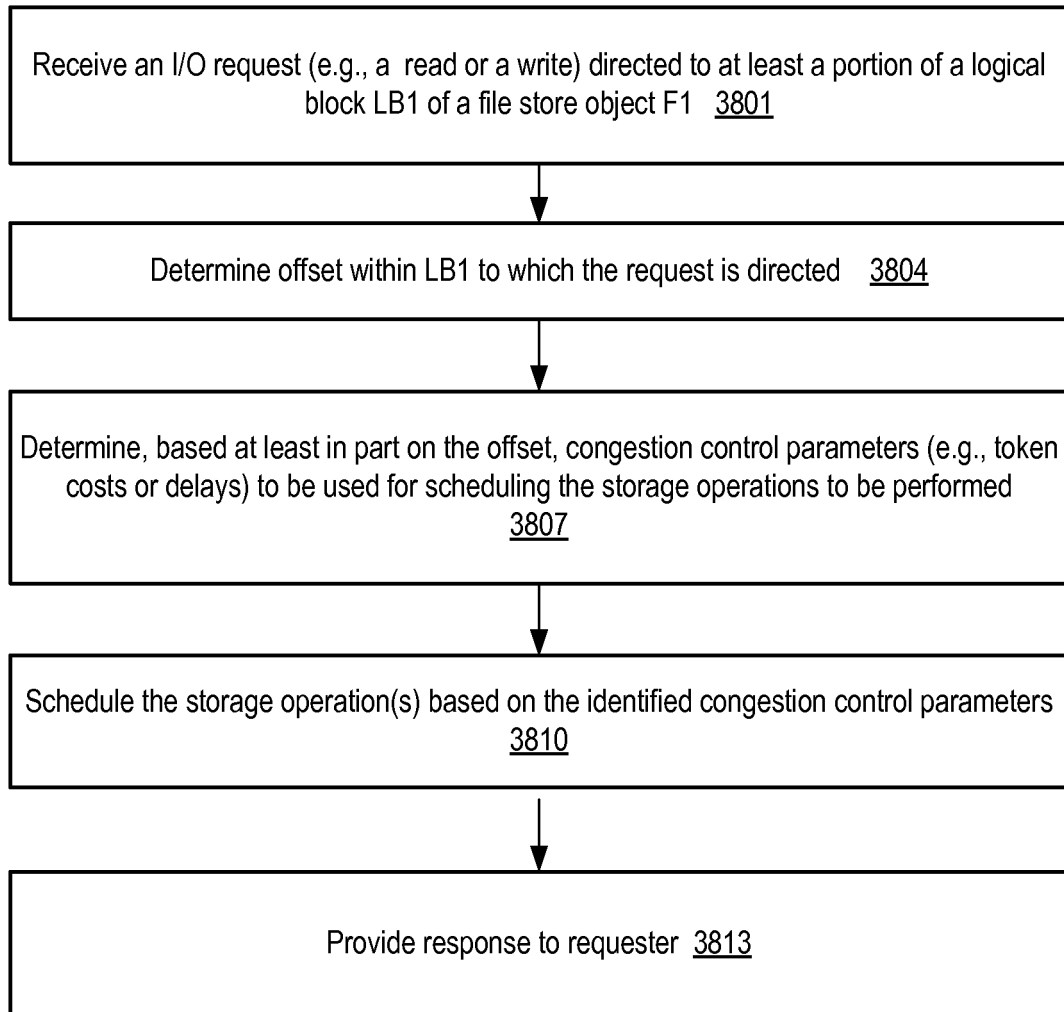


FIG. 38

39/69

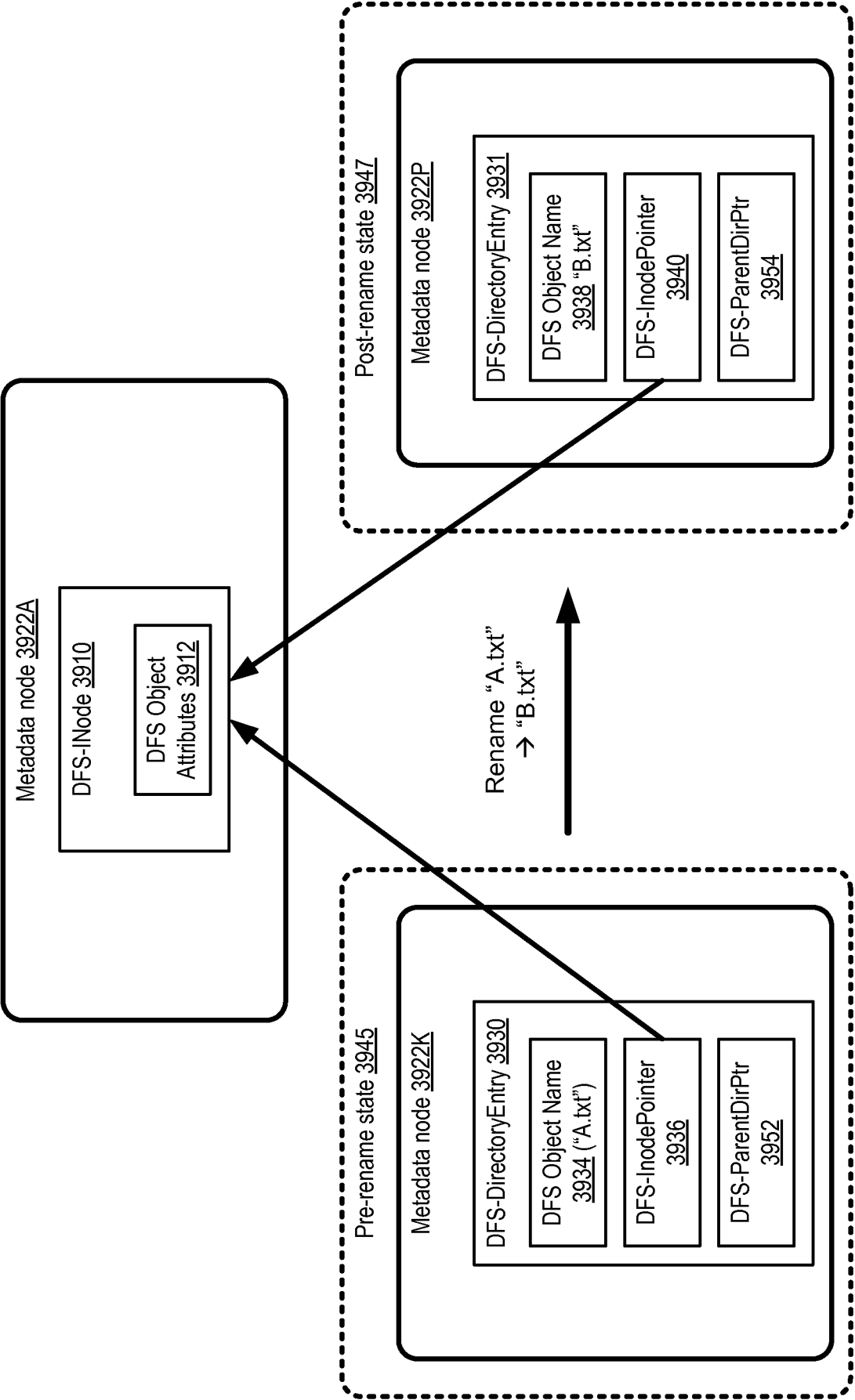


FIG. 39

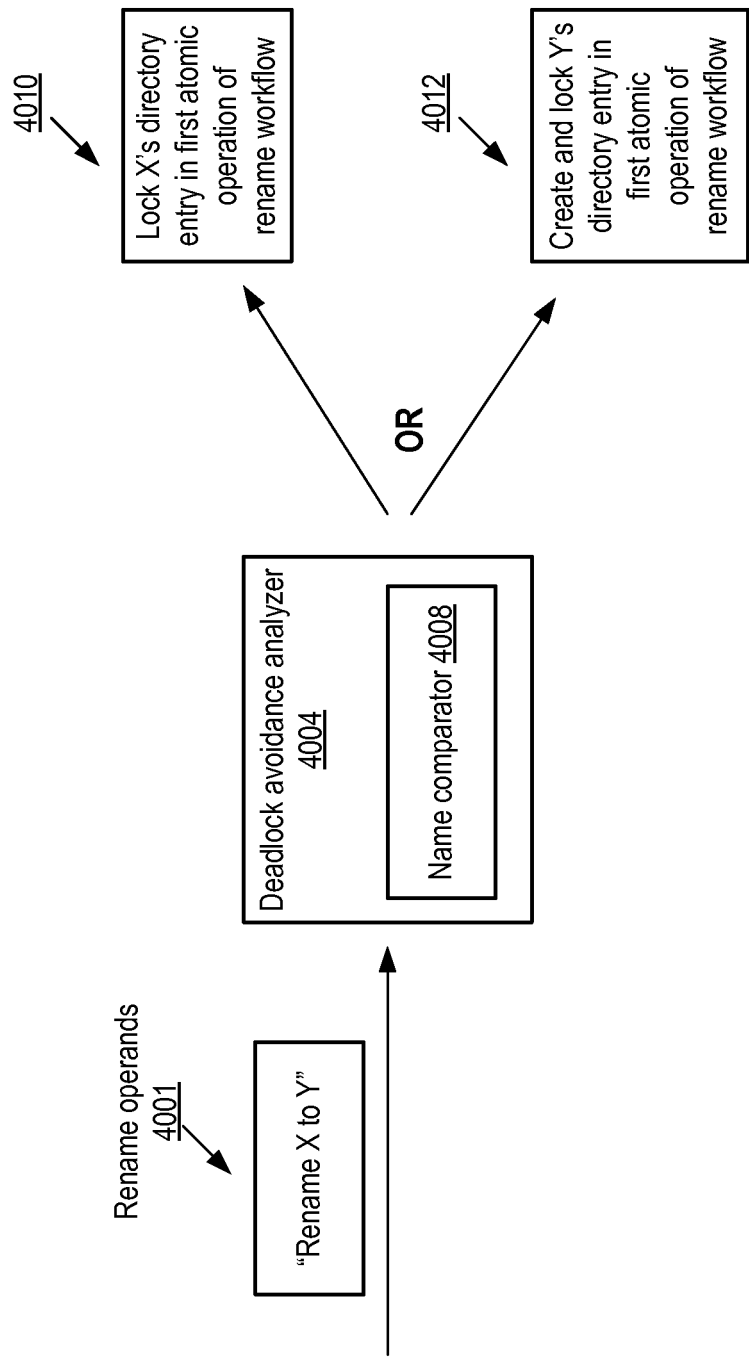


FIG. 40

41/69

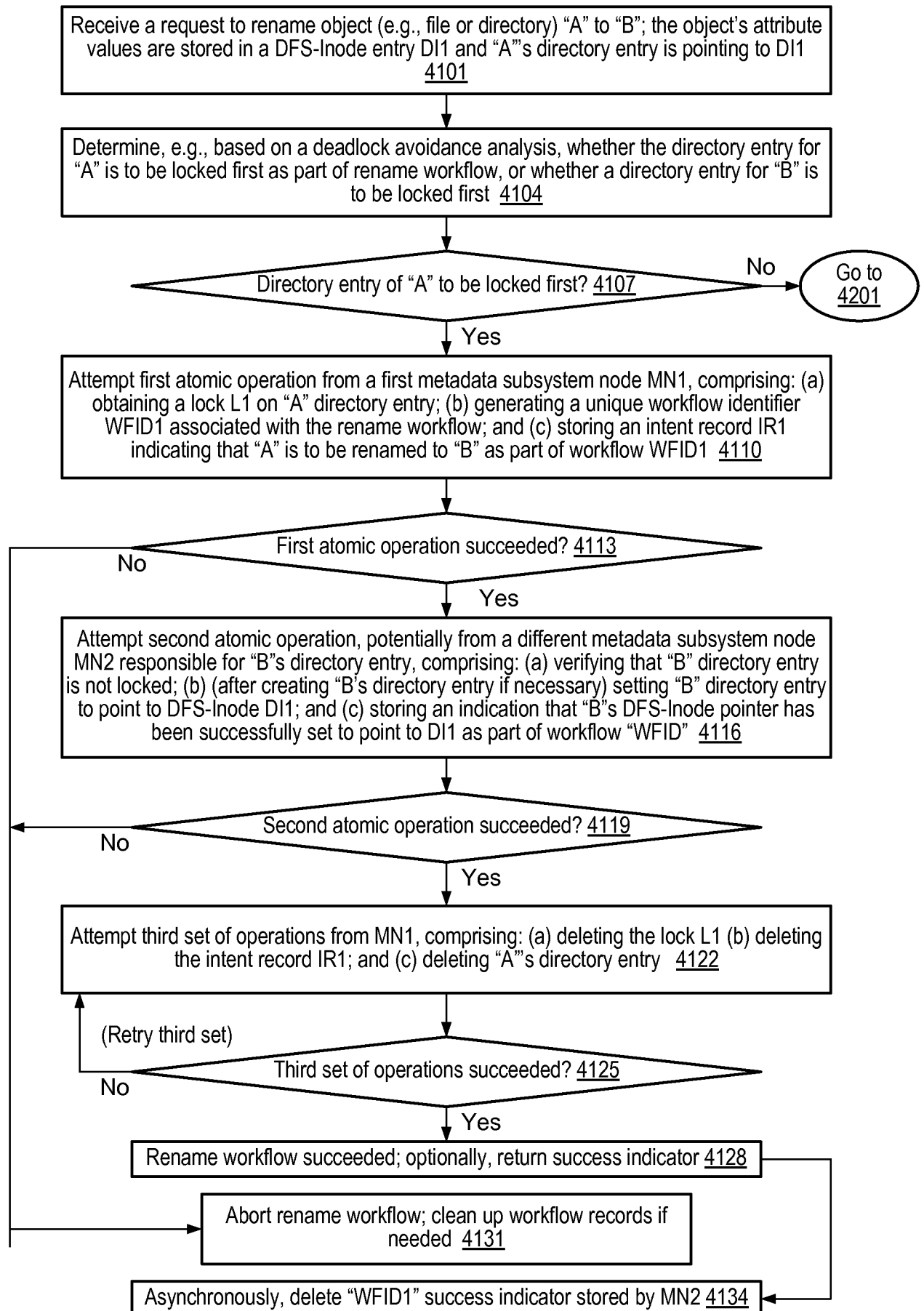


FIG. 41

42/69

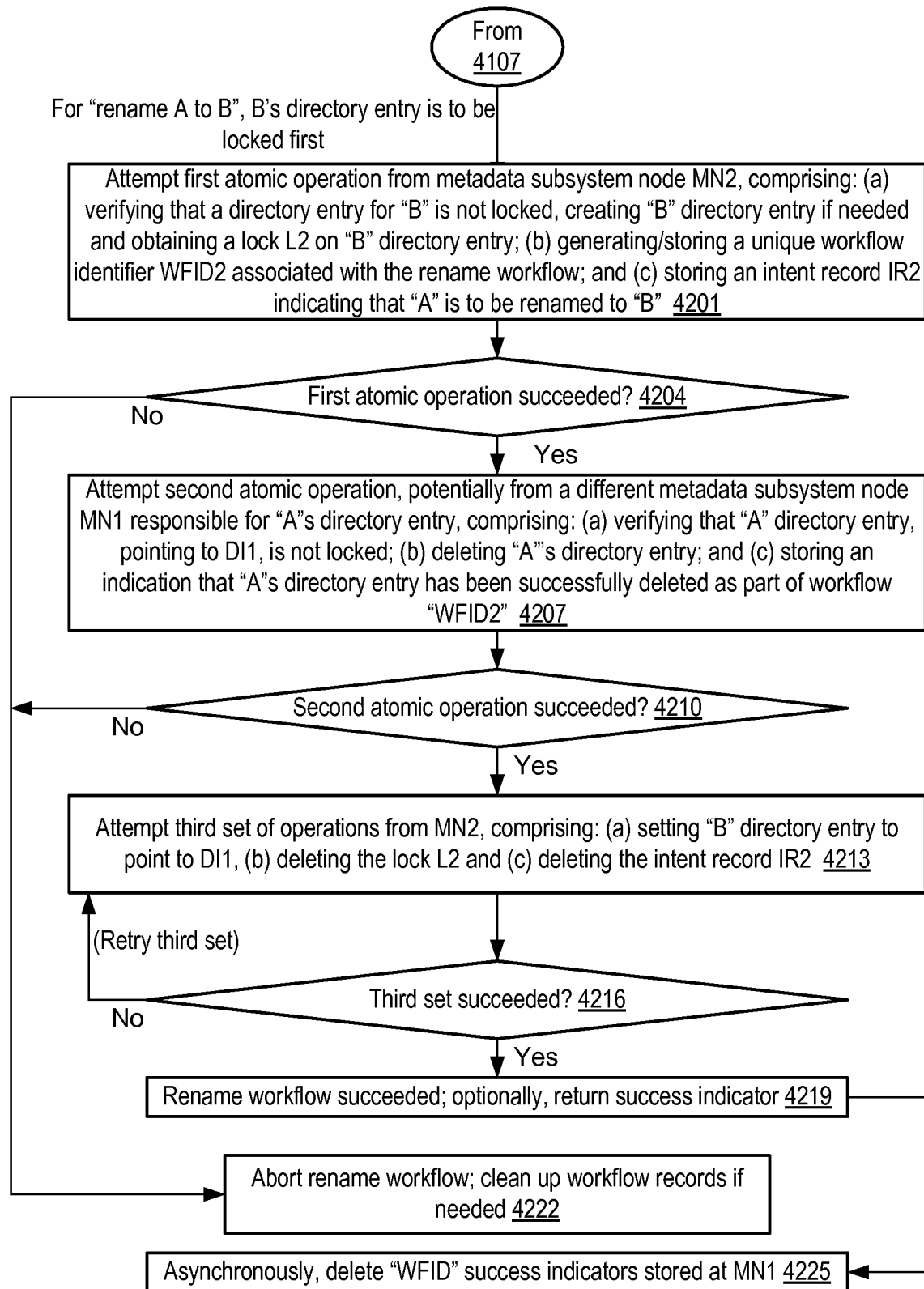


FIG. 42

43/69

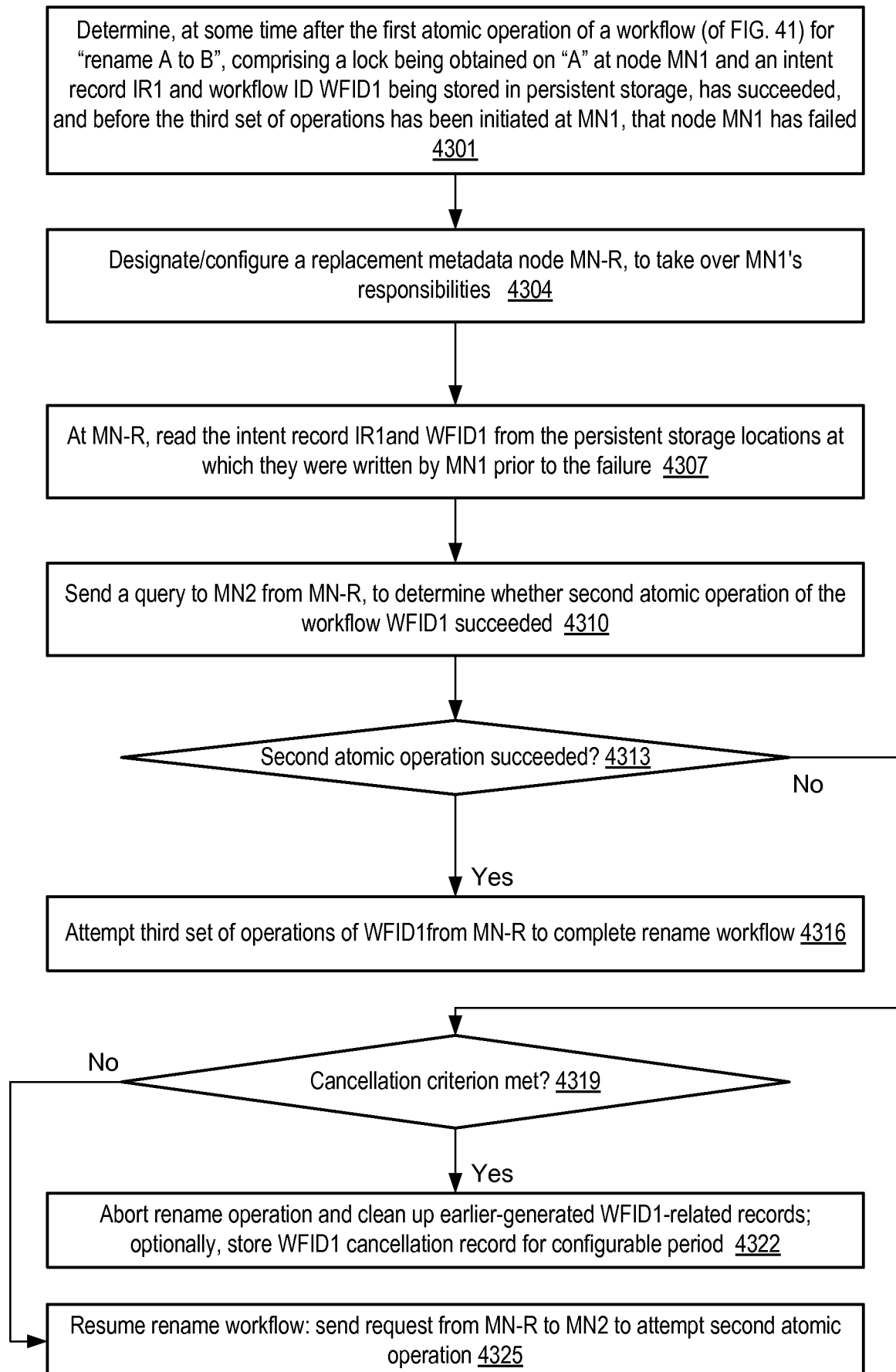


FIG. 43

44/69

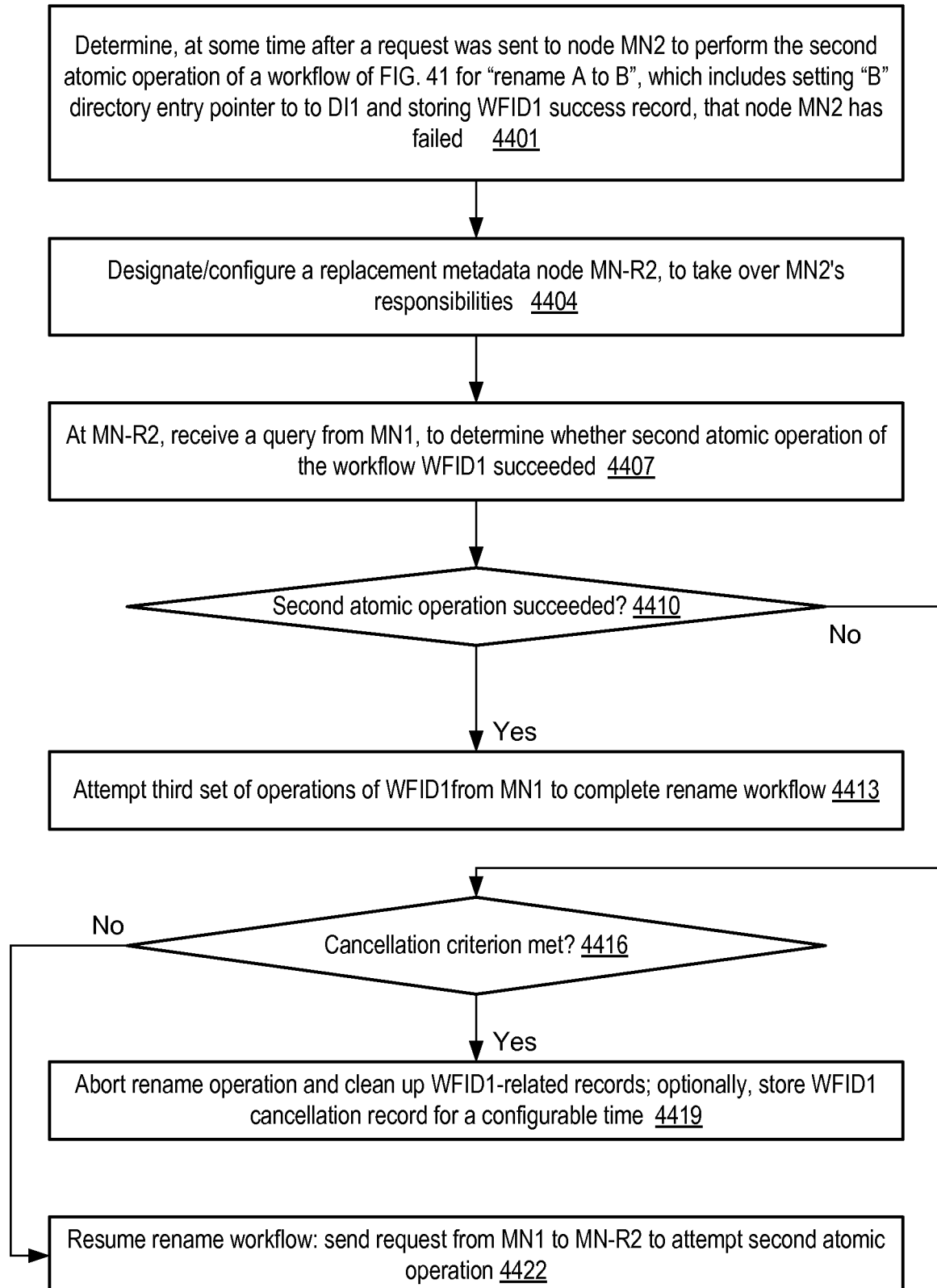
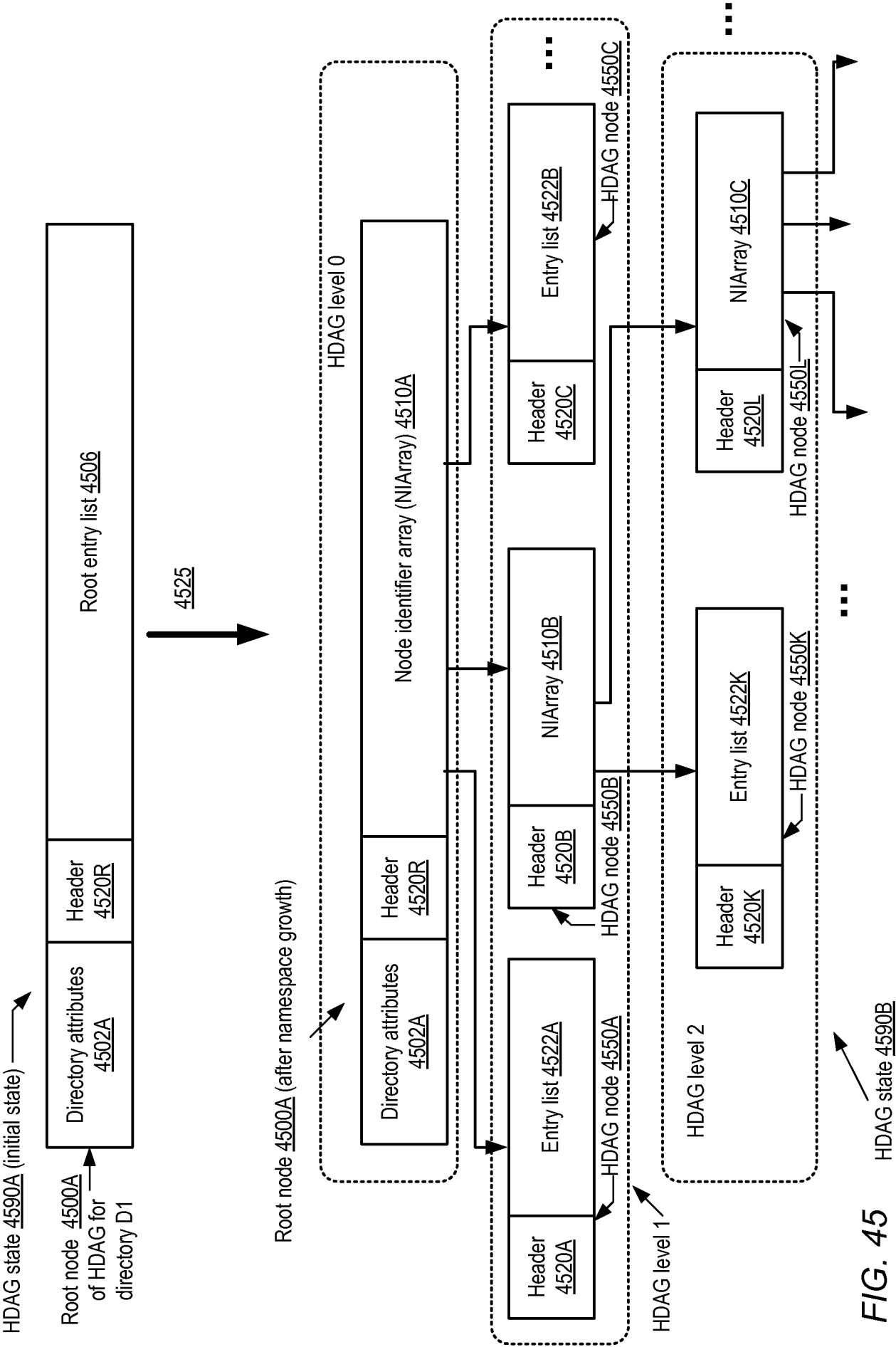


FIG. 44



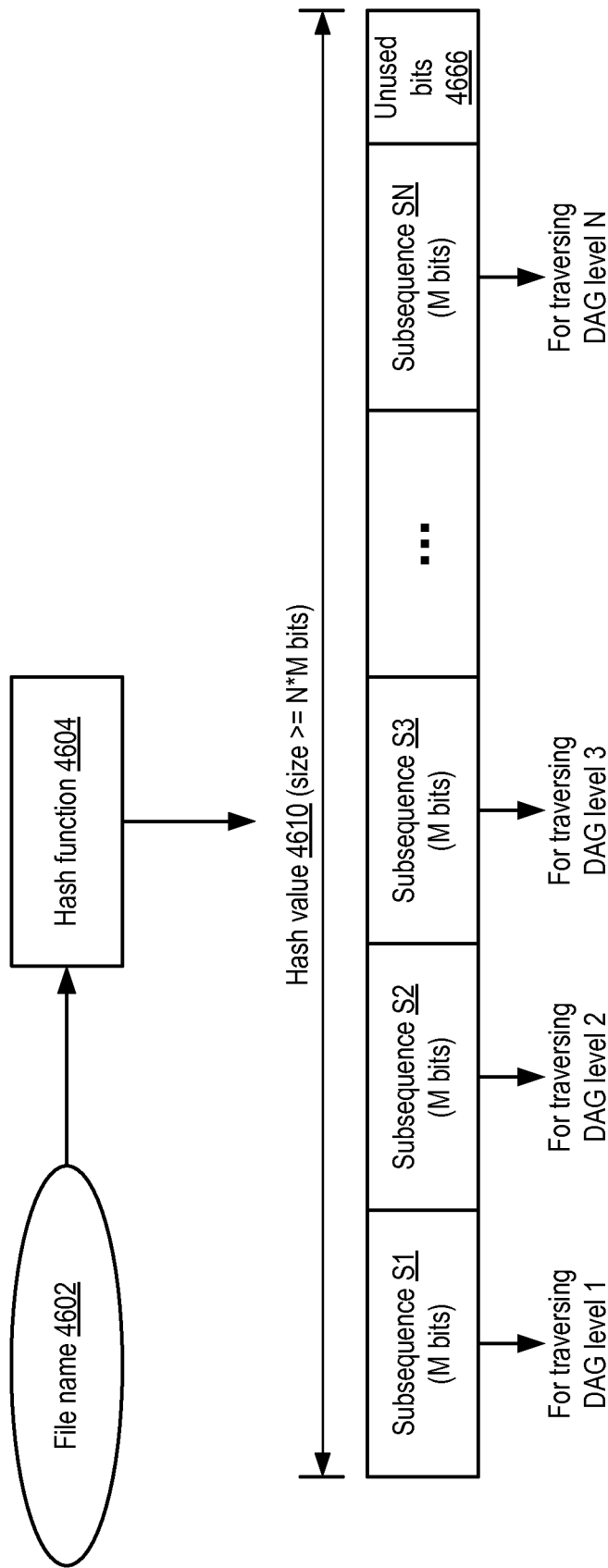


FIG. 46

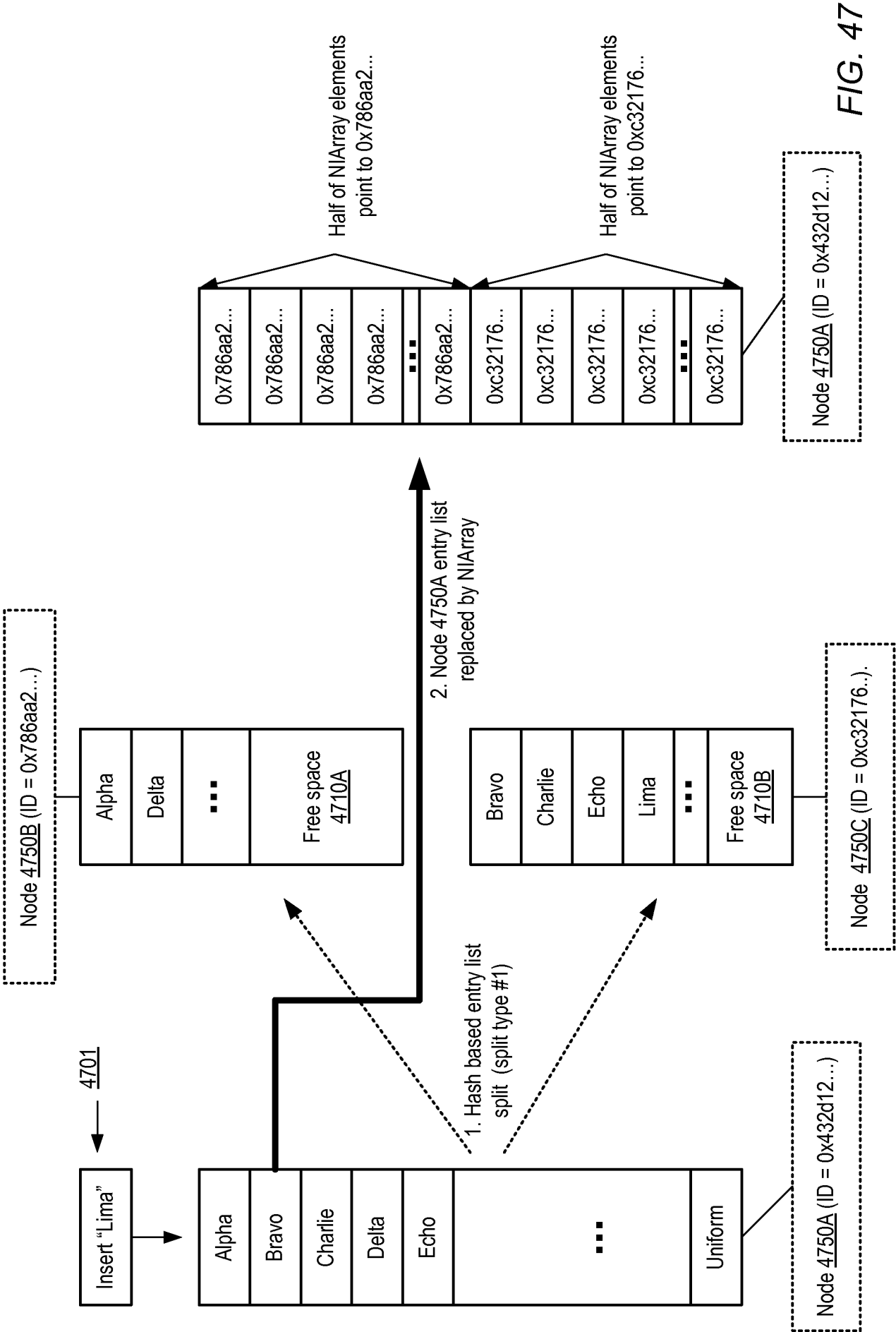
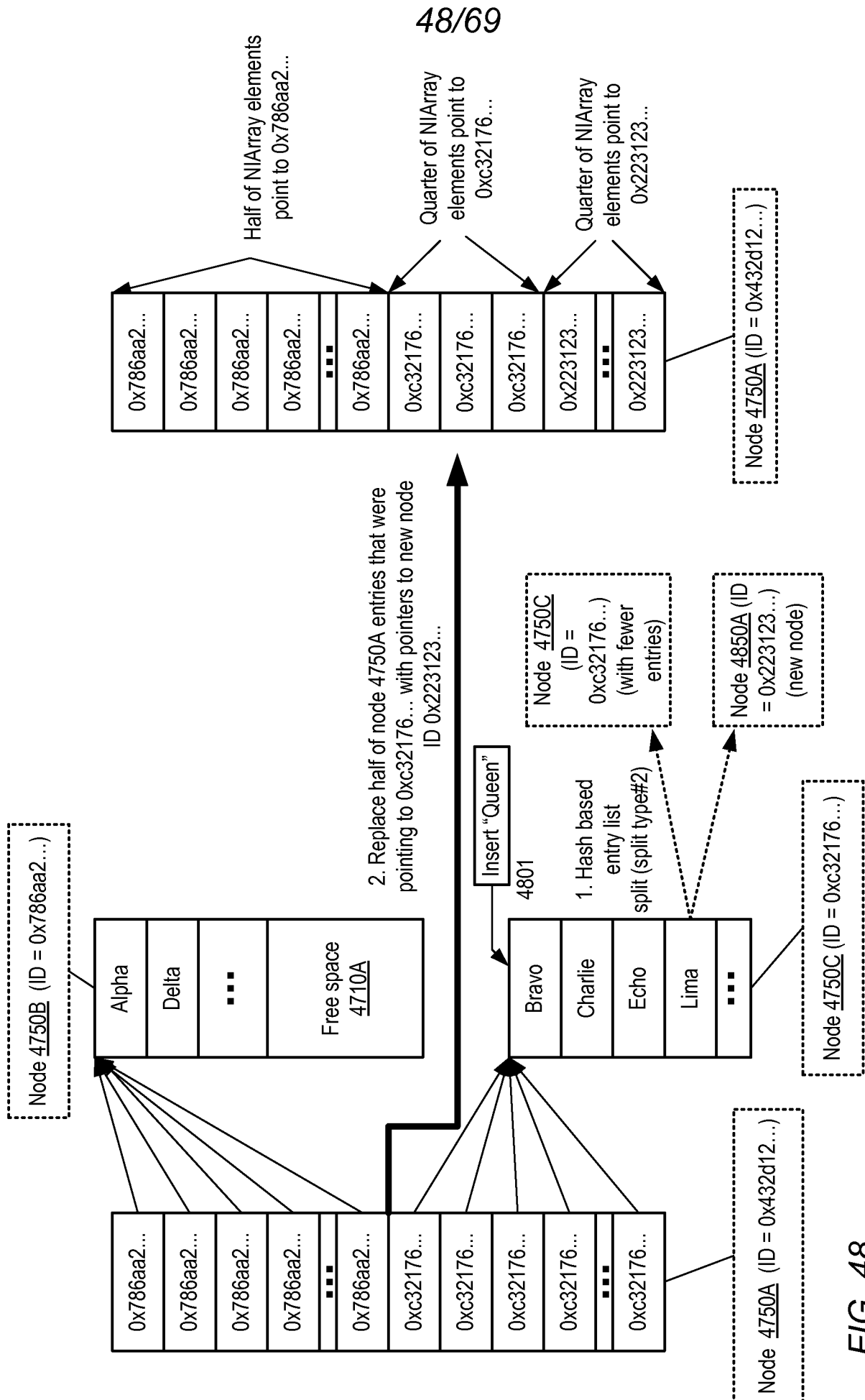


FIG. 47



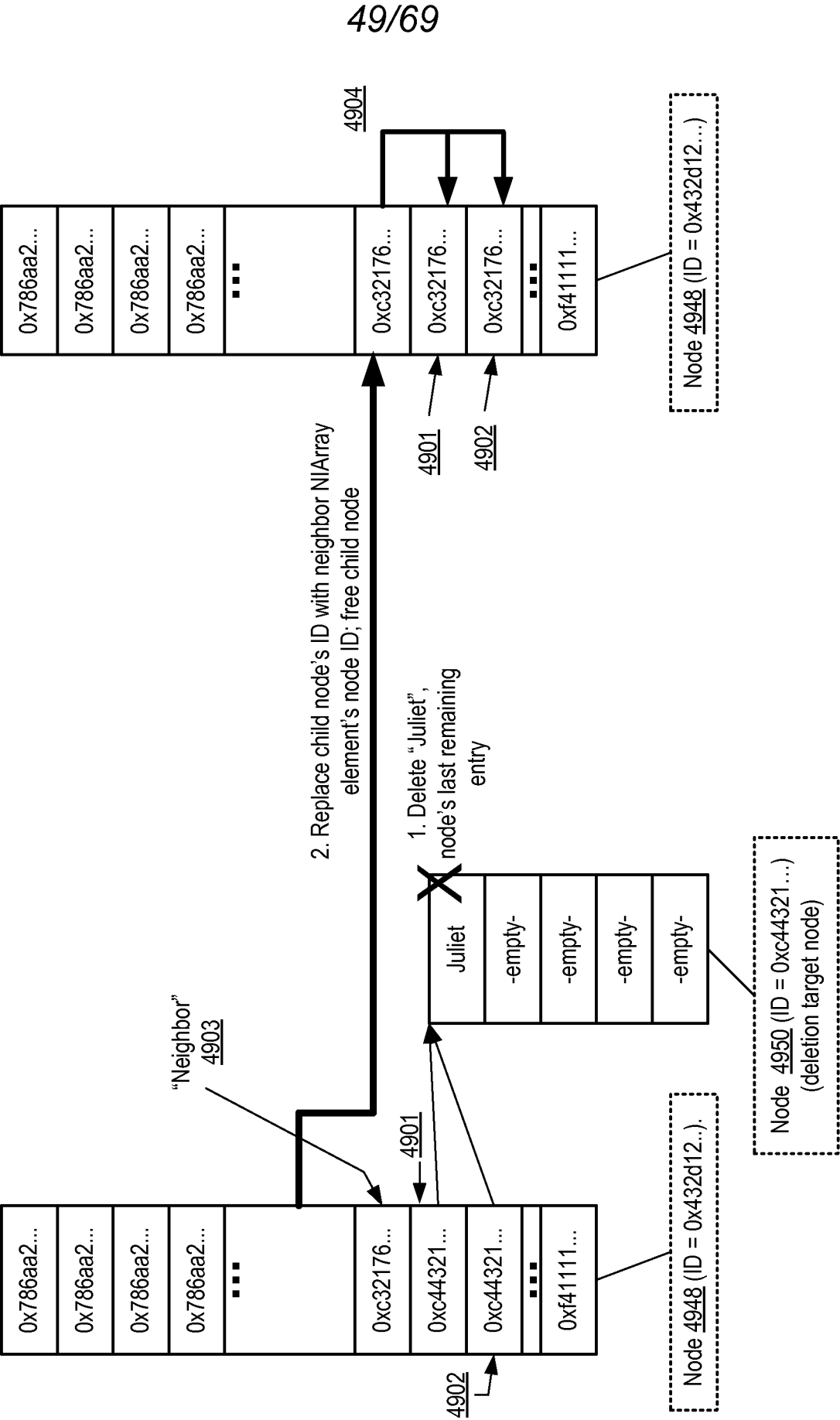


FIG. 49

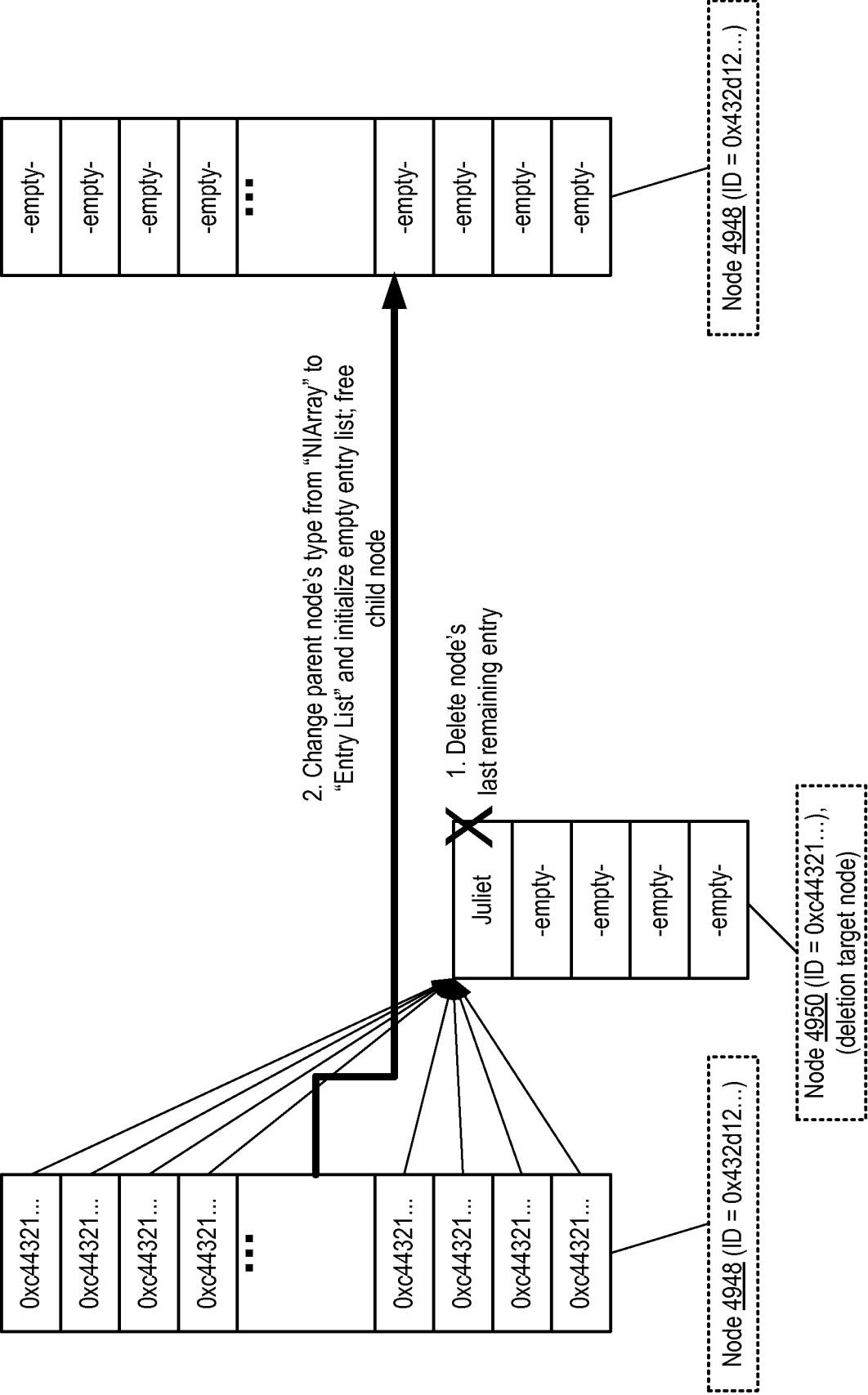


FIG. 50

51/69

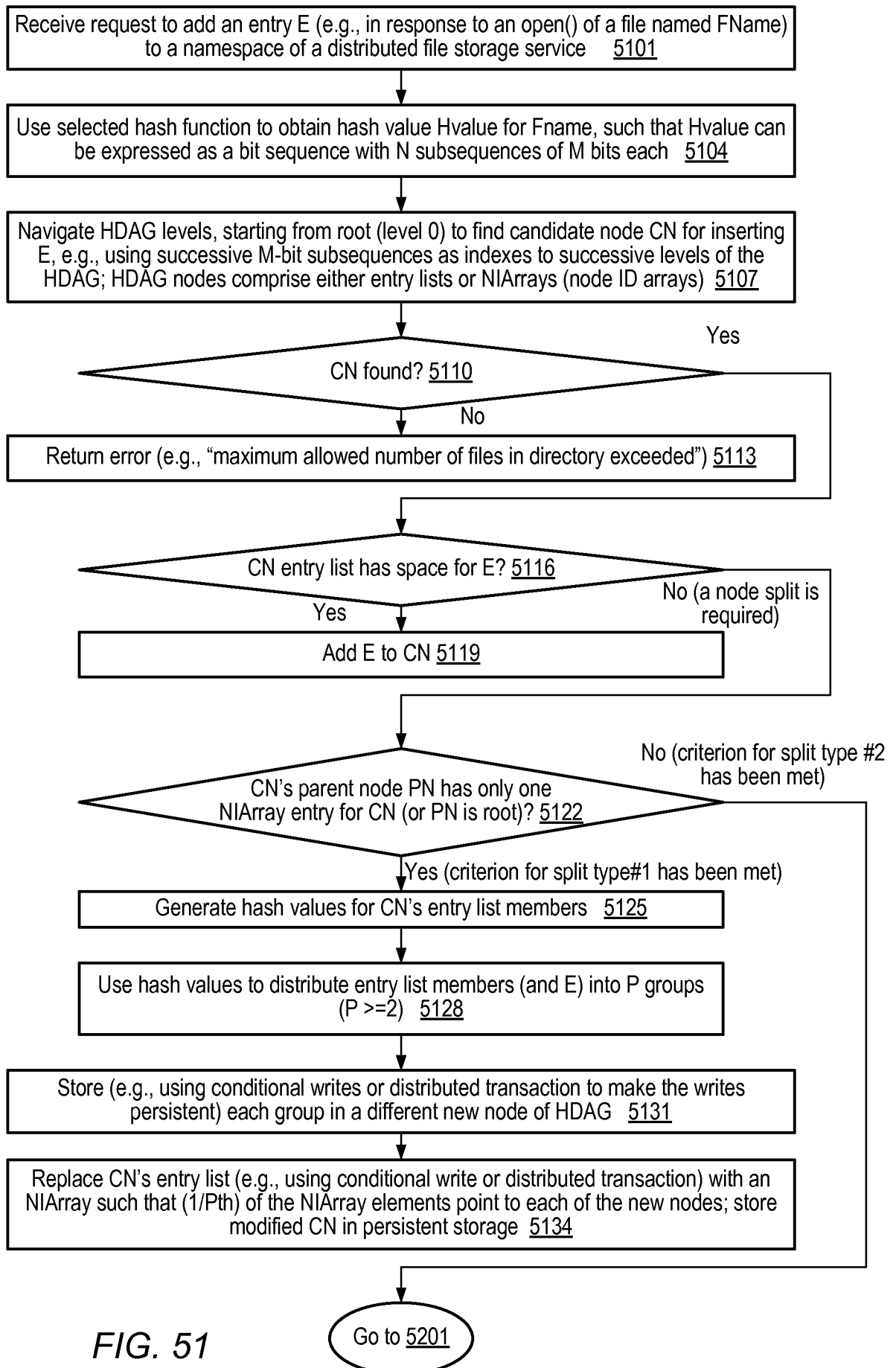


FIG. 51

52/69

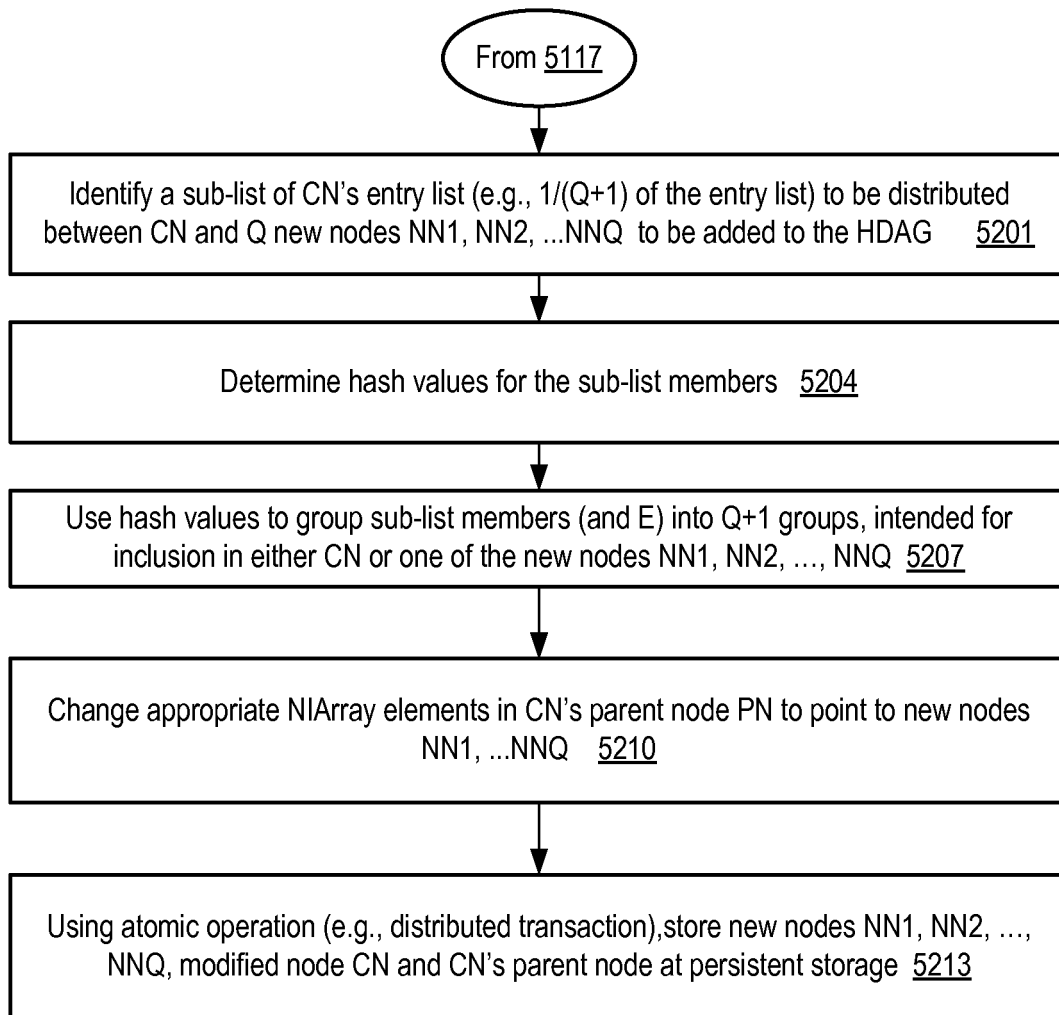


FIG. 52

53/69

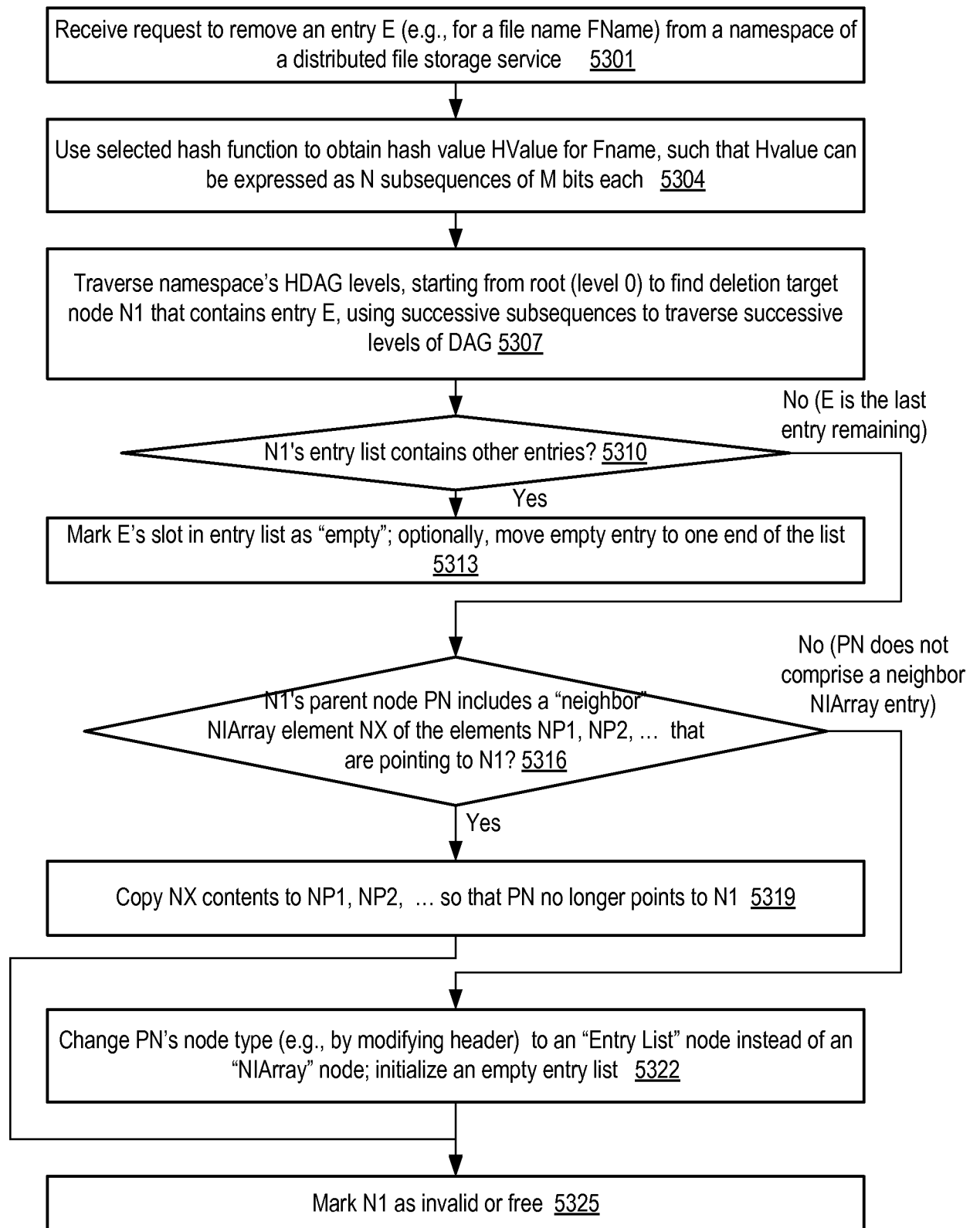


FIG. 53

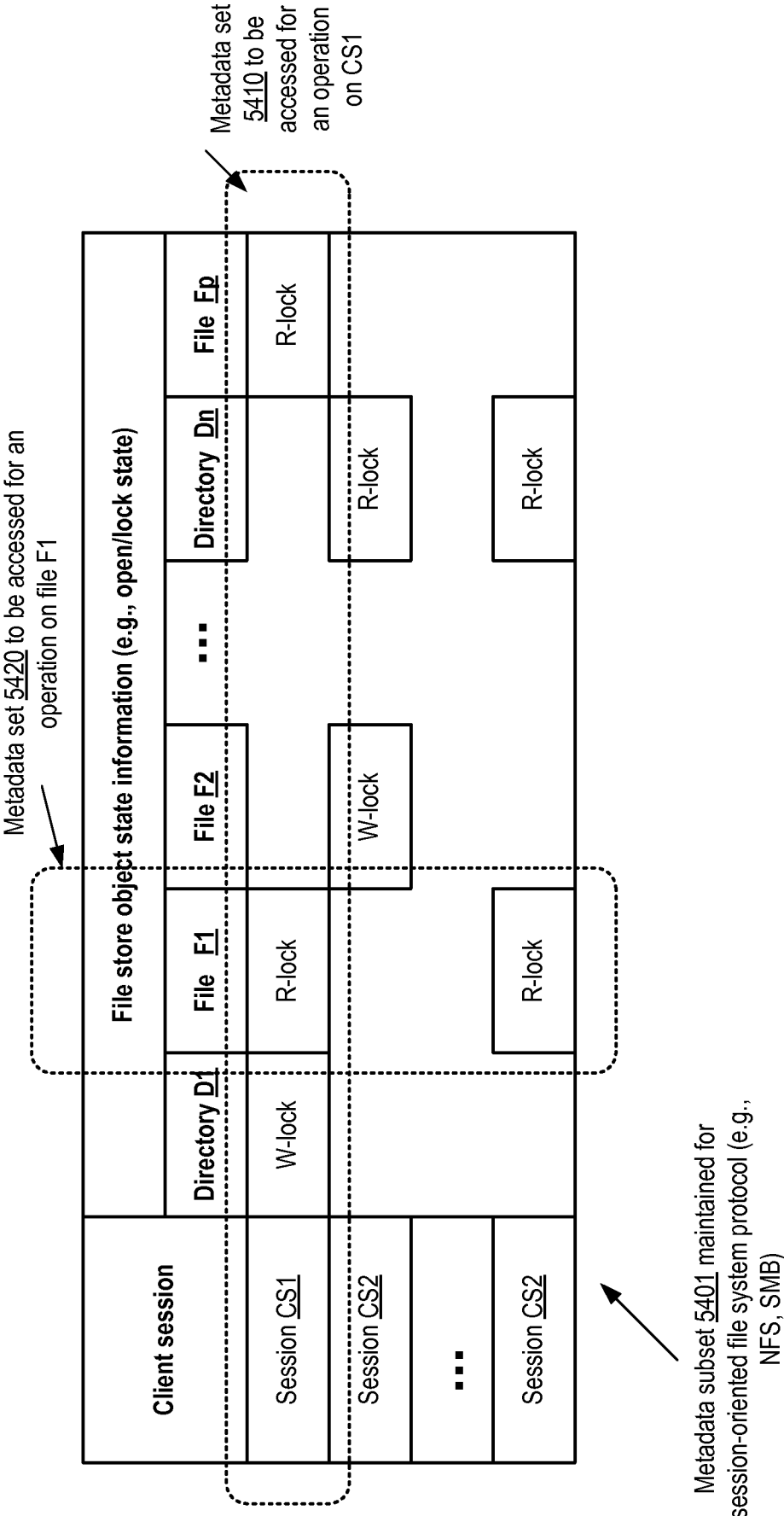


FIG. 54

55/69

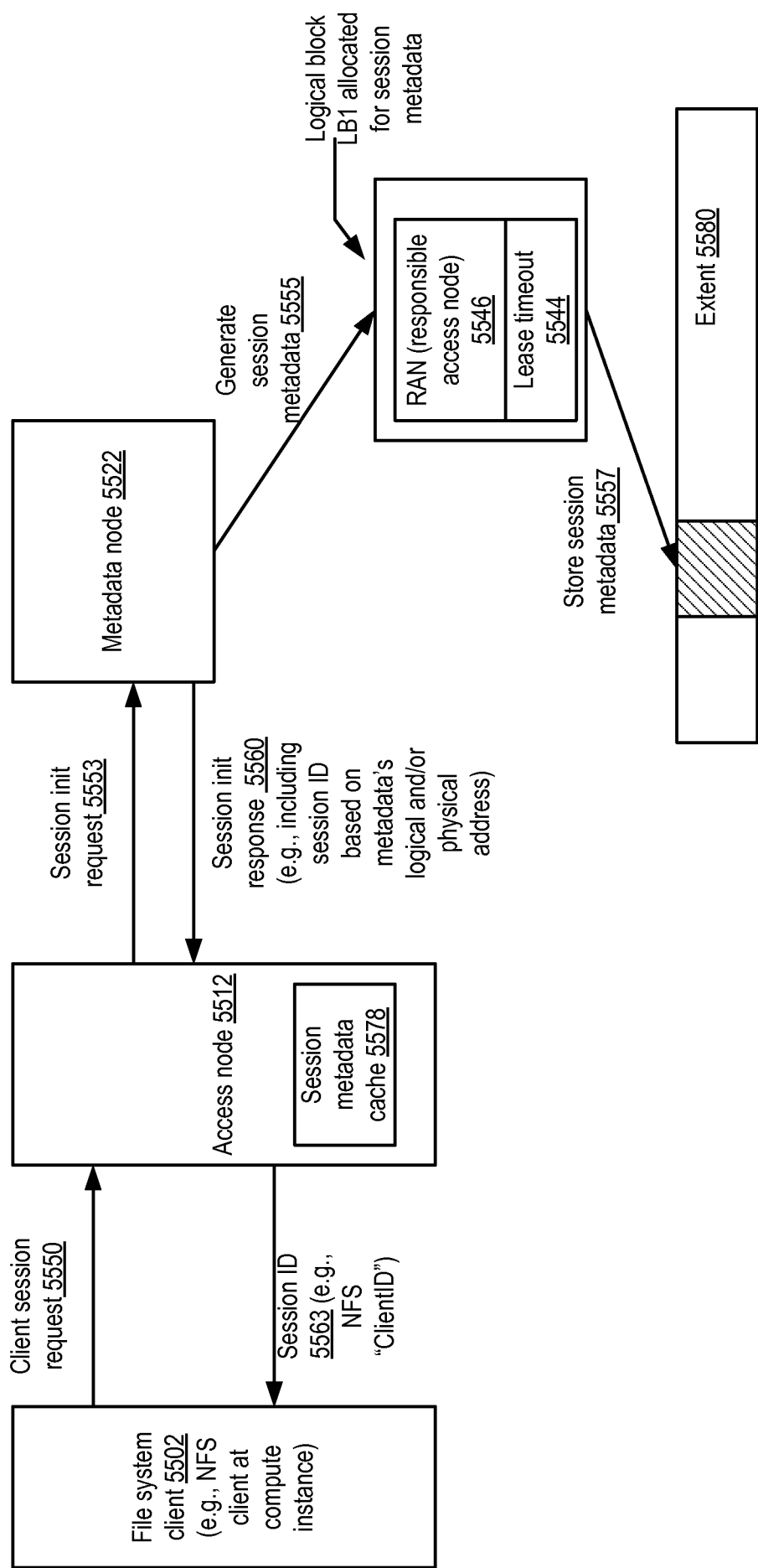


FIG. 55

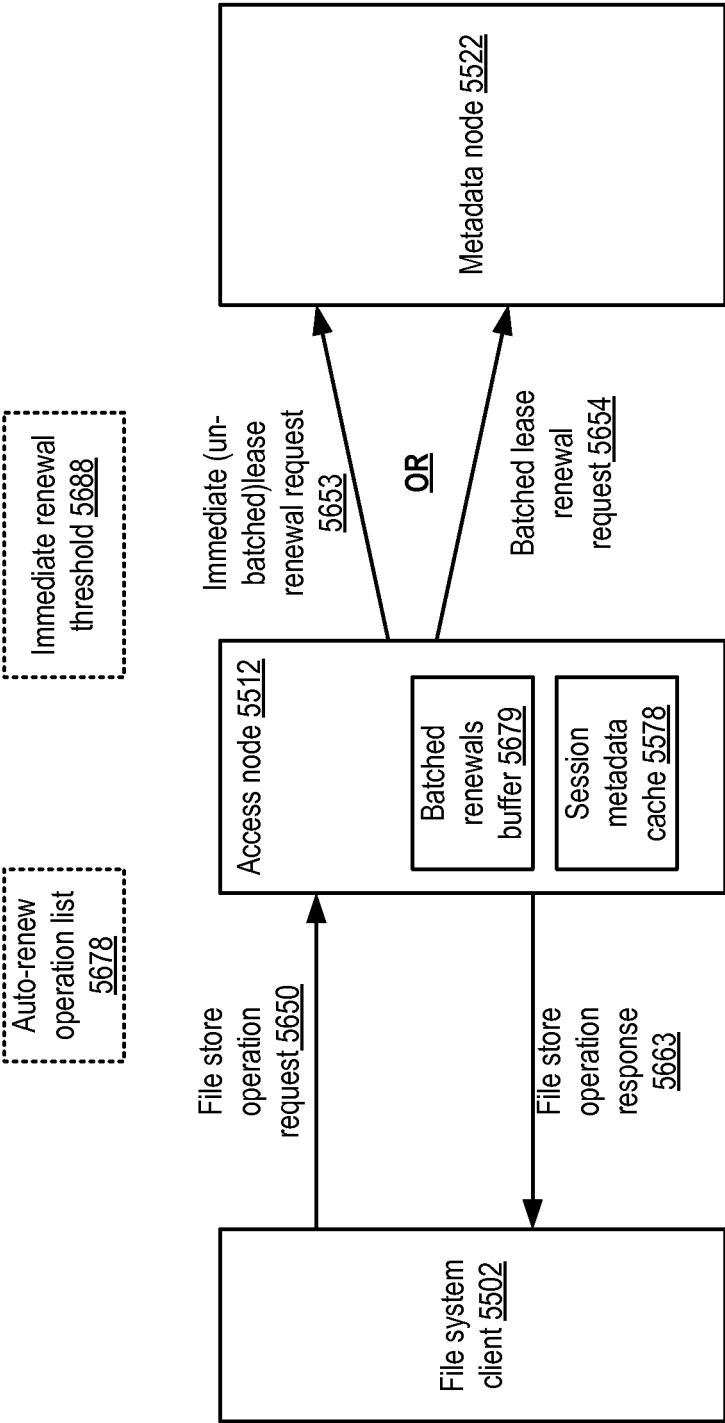


FIG. 56

57/69

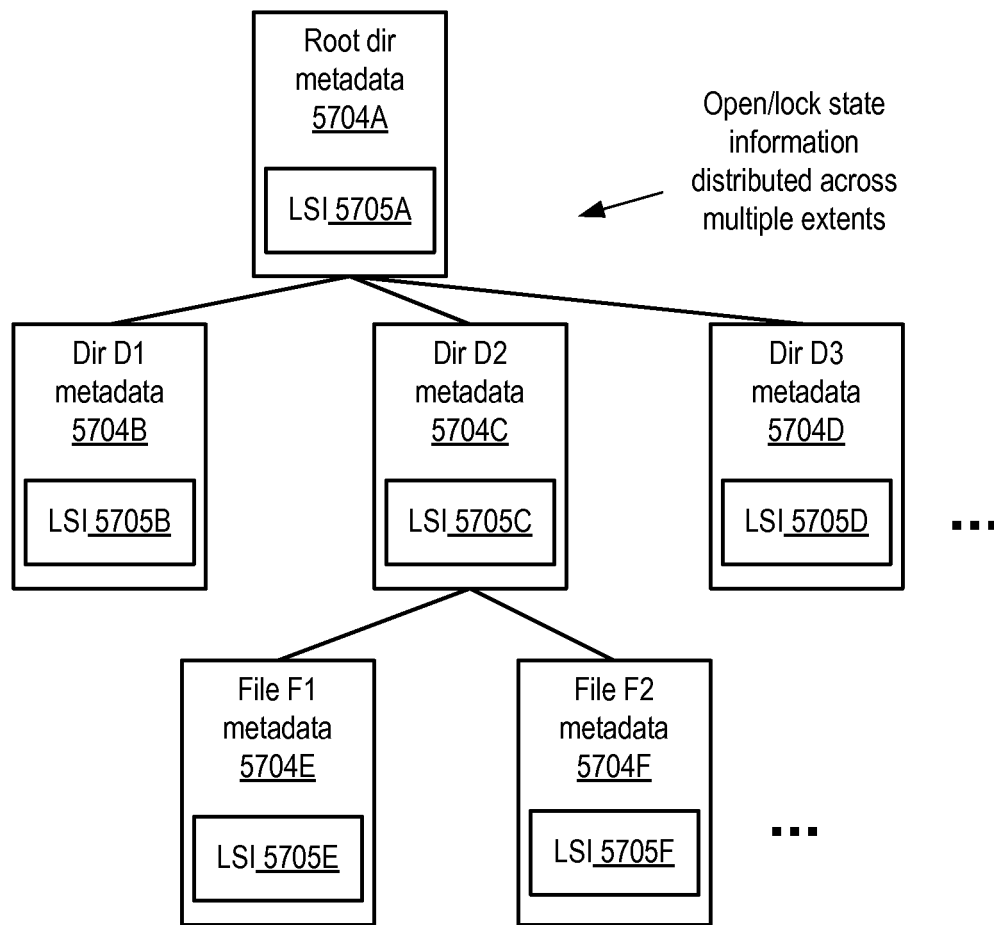


FIG. 57a

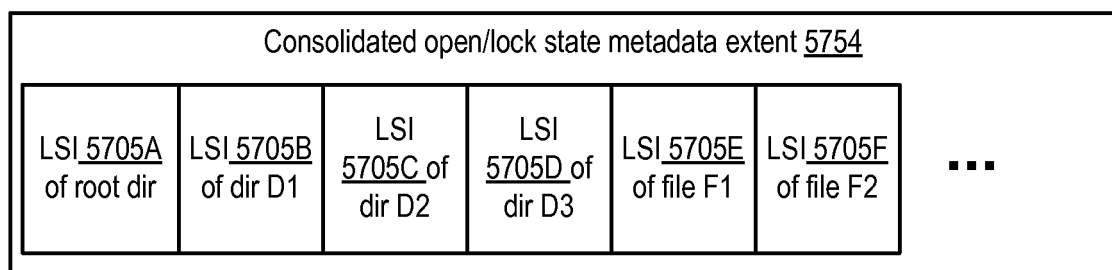


FIG. 57b

58/69

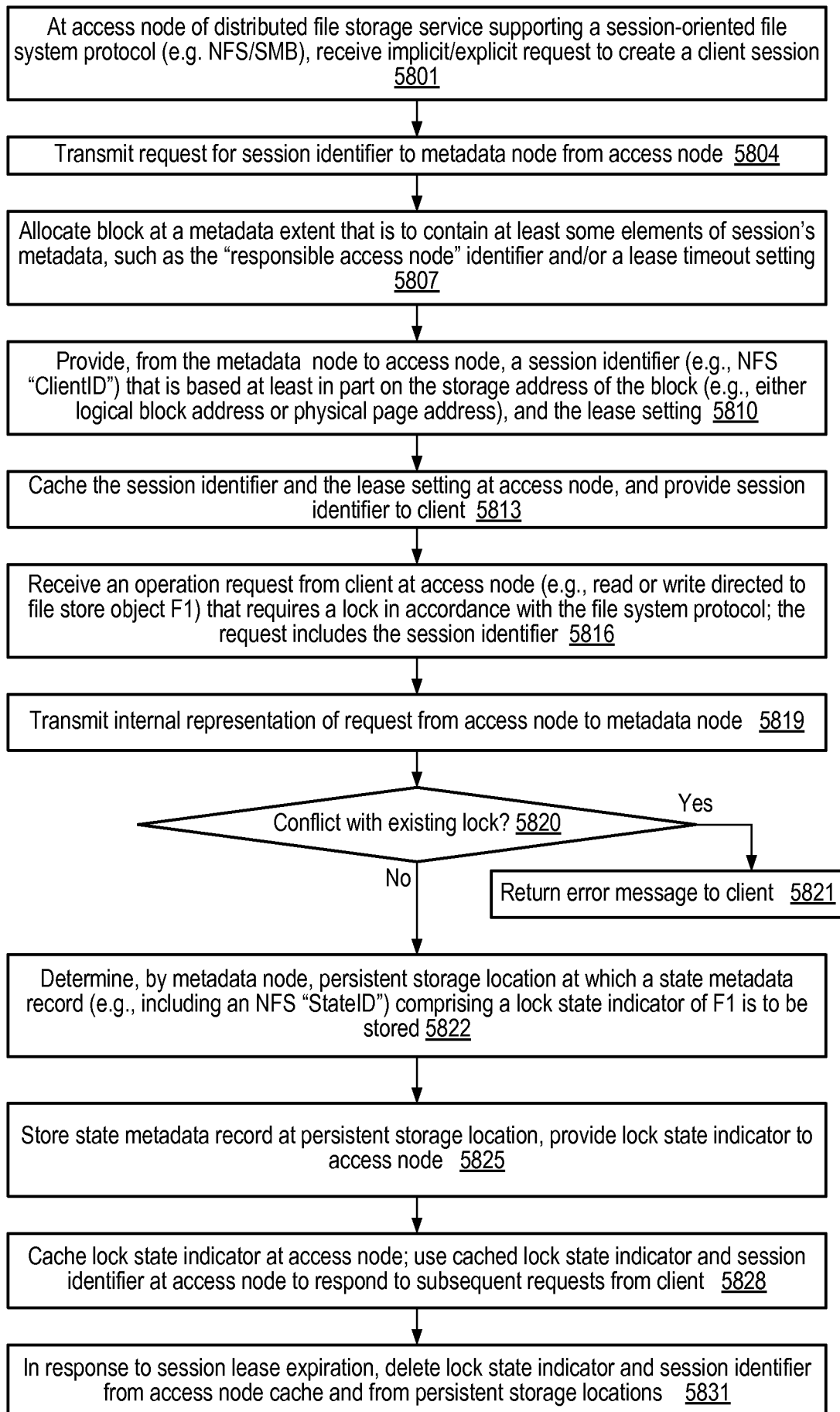


FIG. 58

59/69

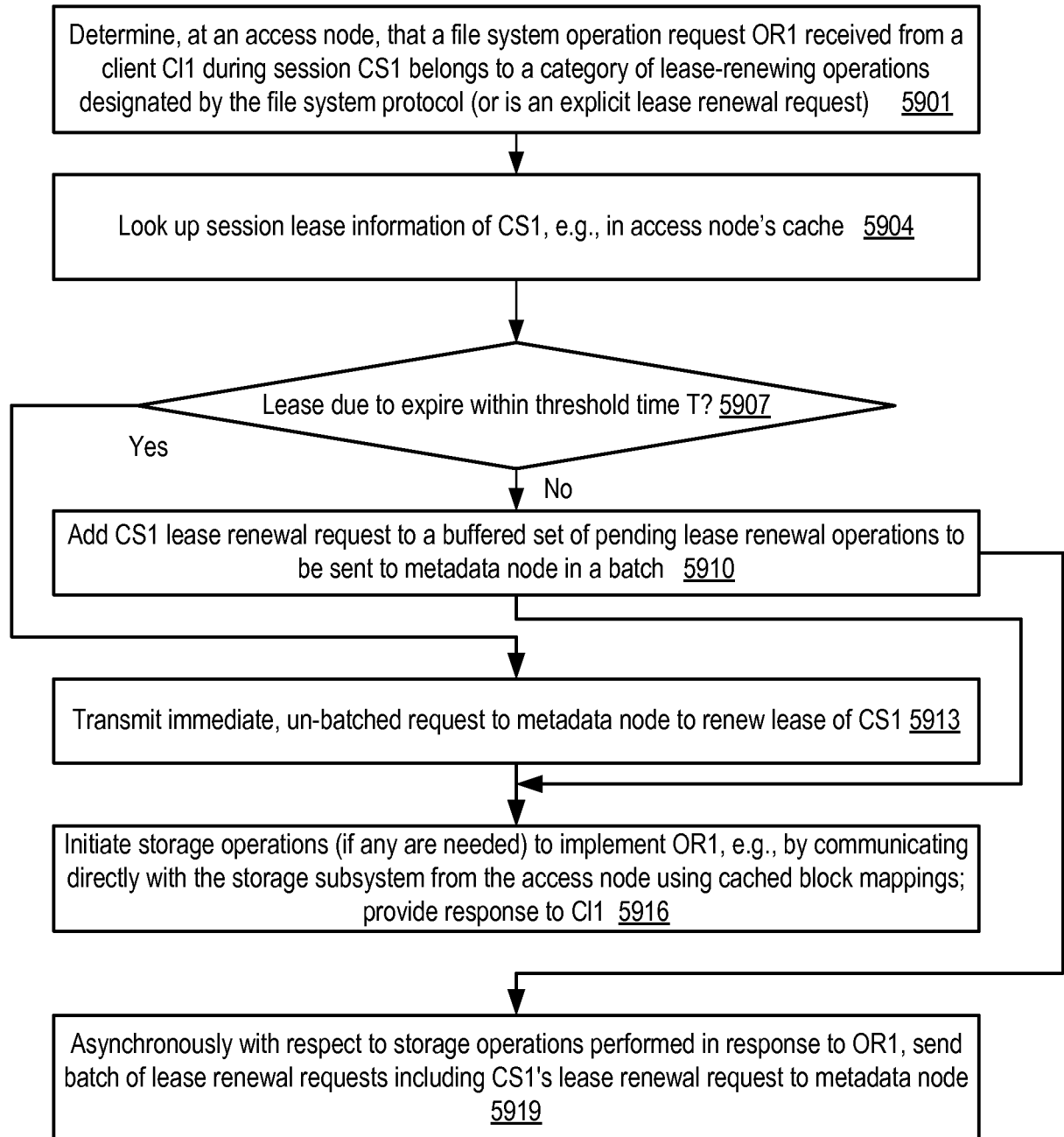


FIG. 59

60/69

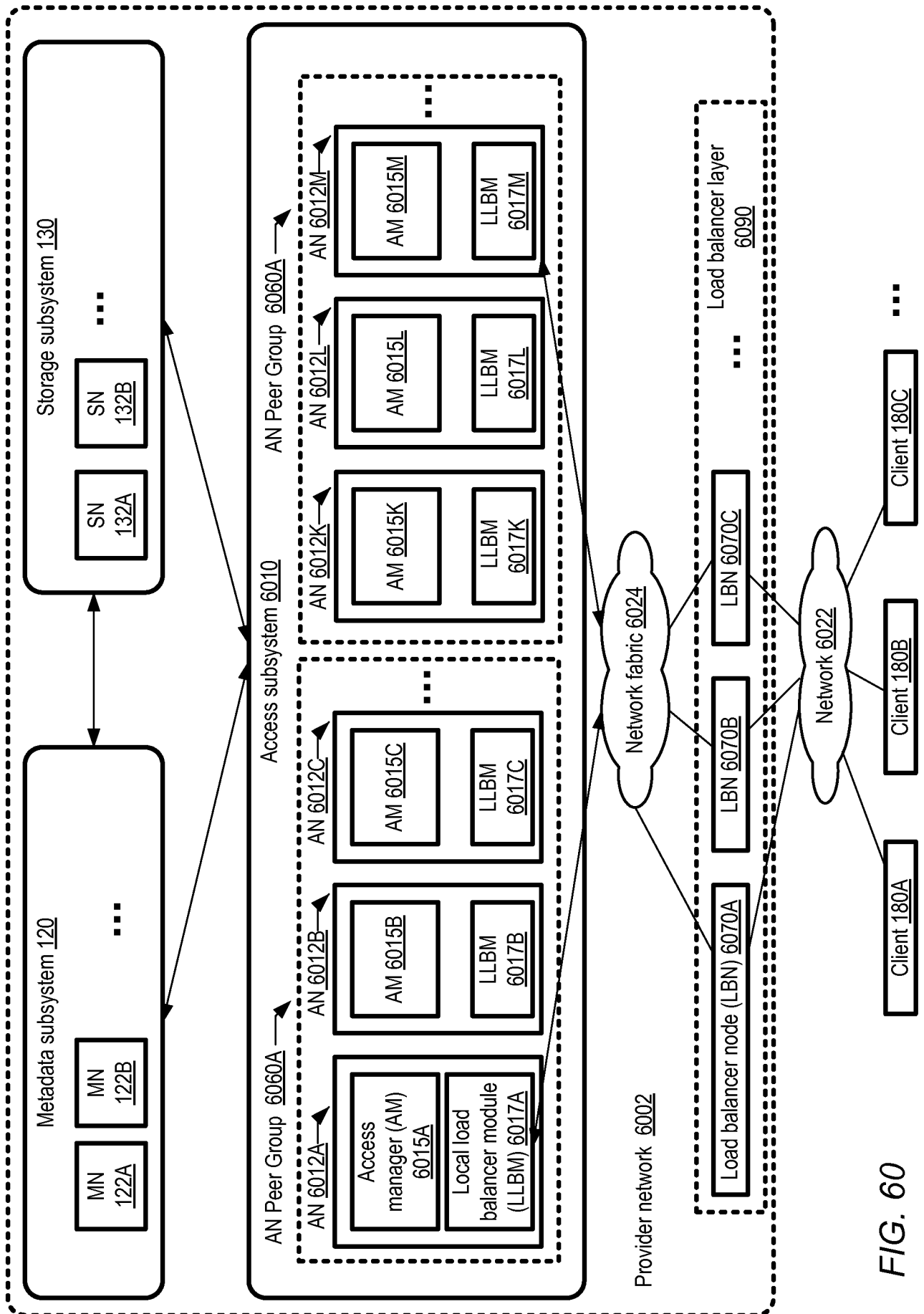


FIG. 60

61/69

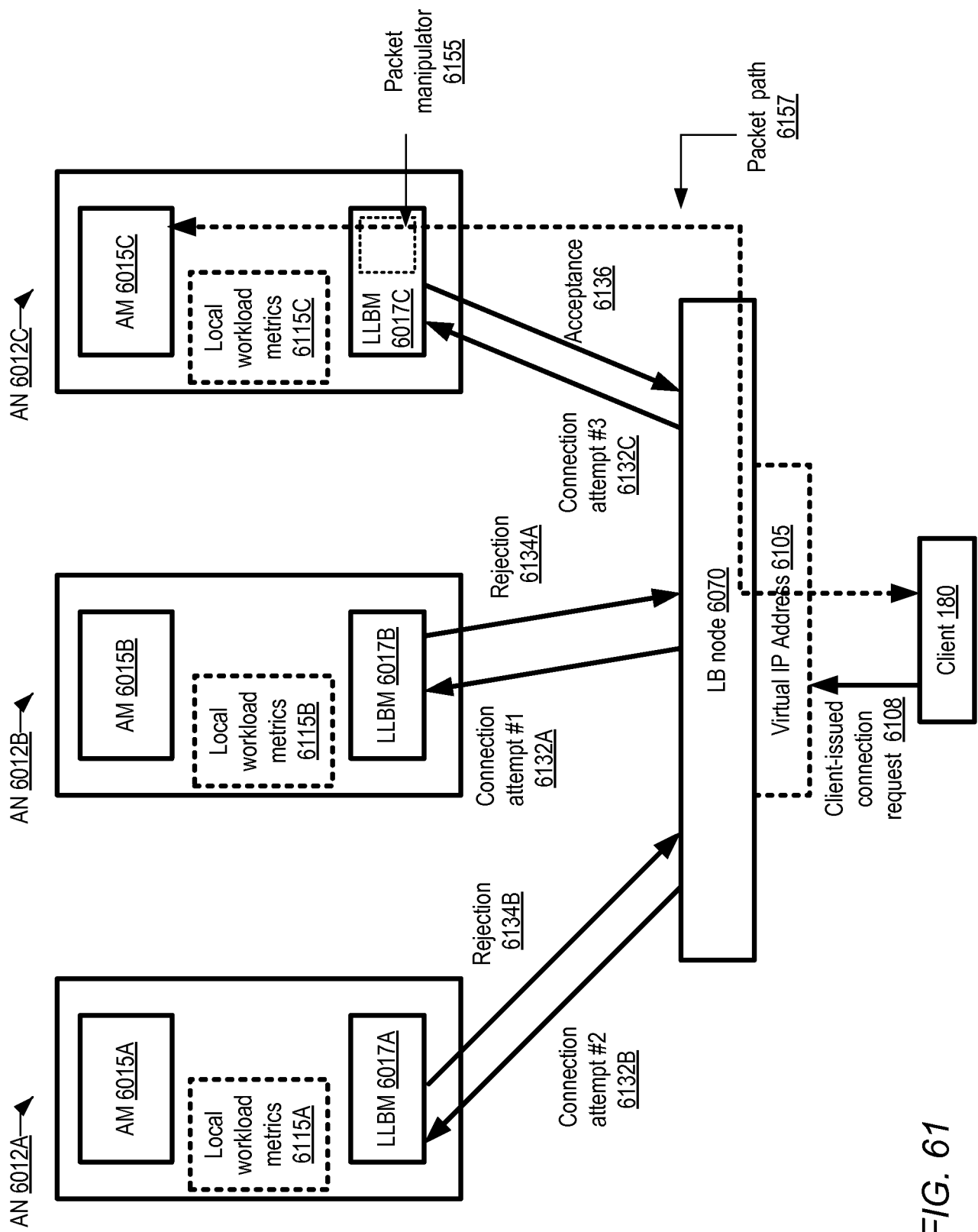


FIG. 61

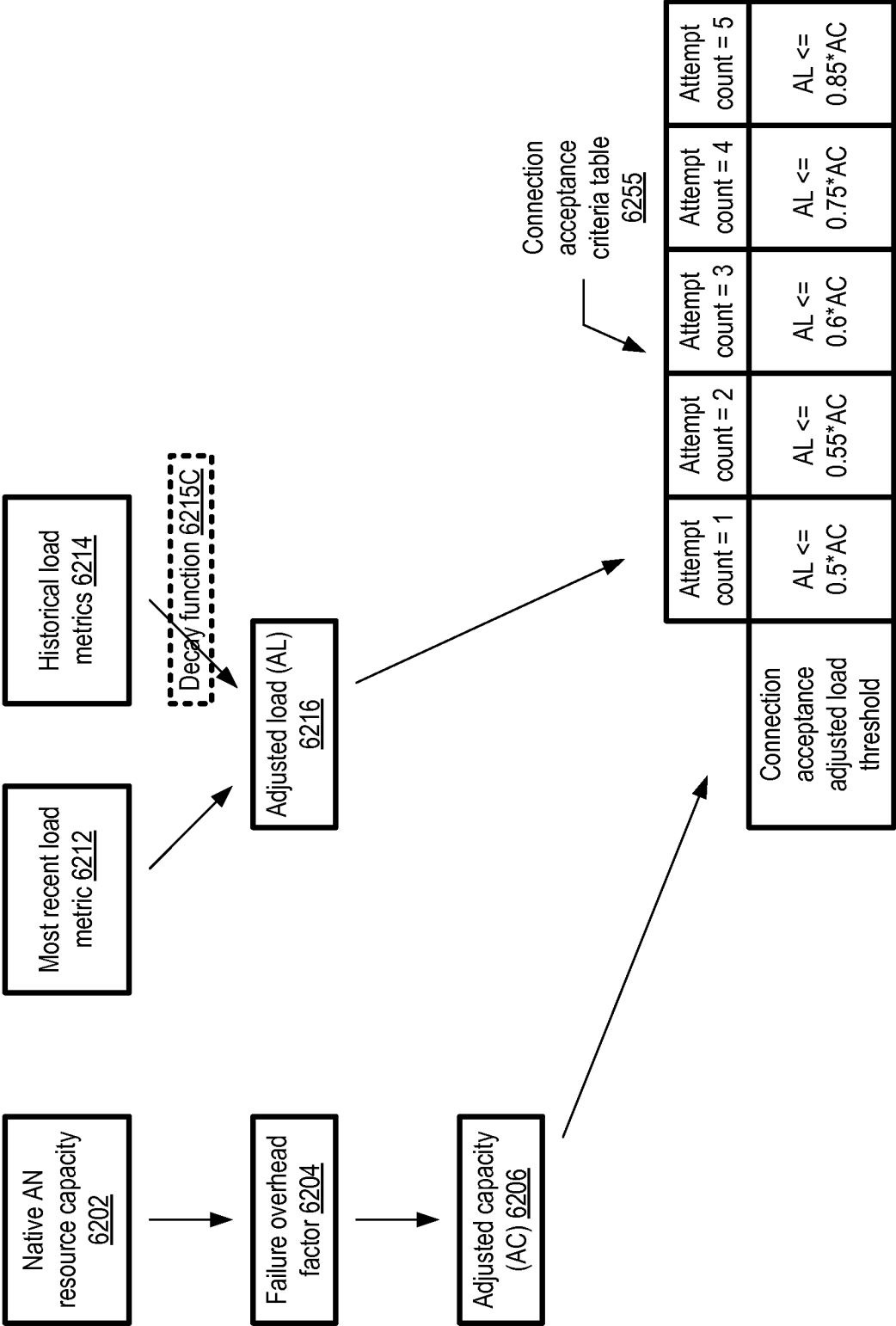


FIG. 62

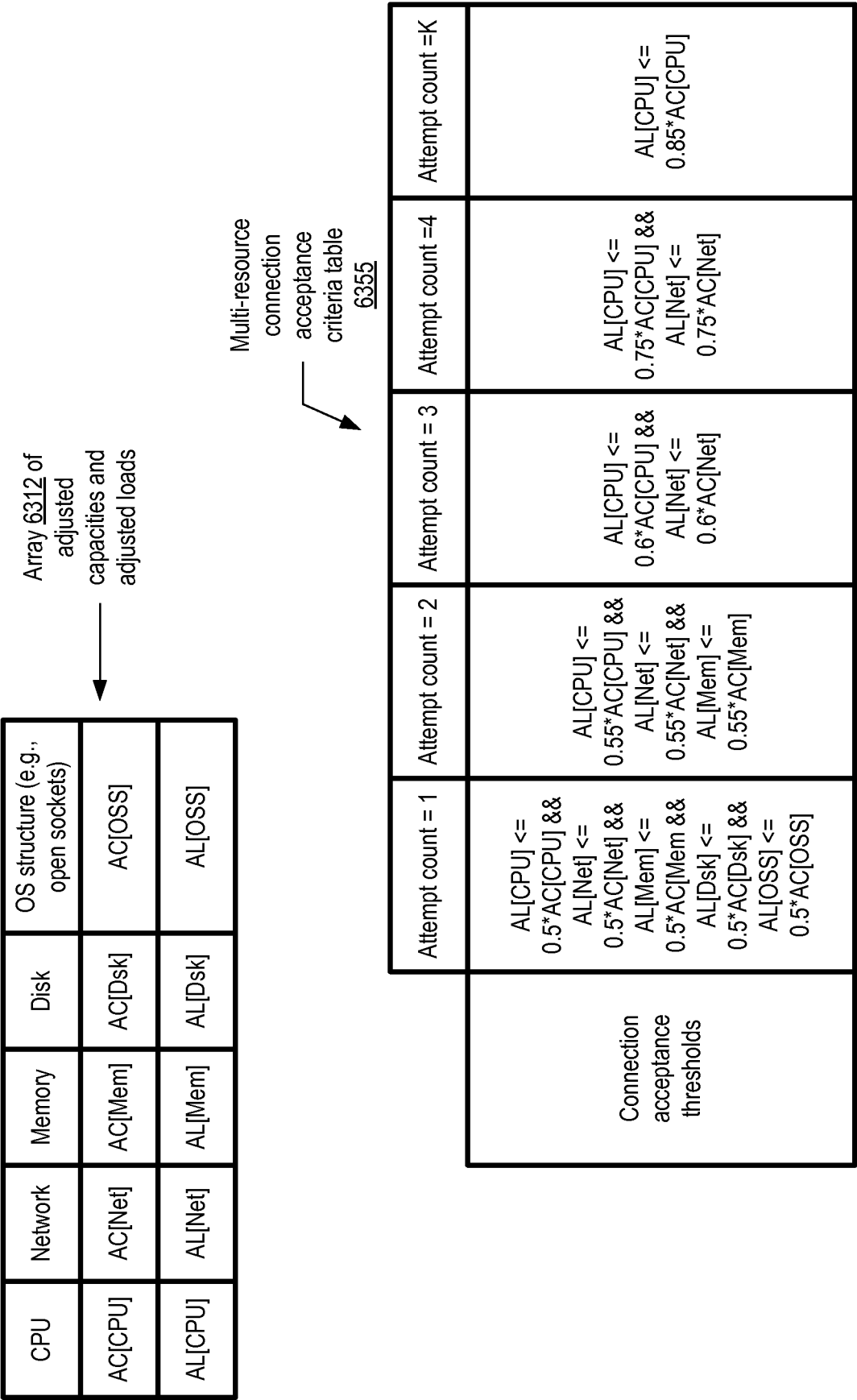


FIG. 63

64/69

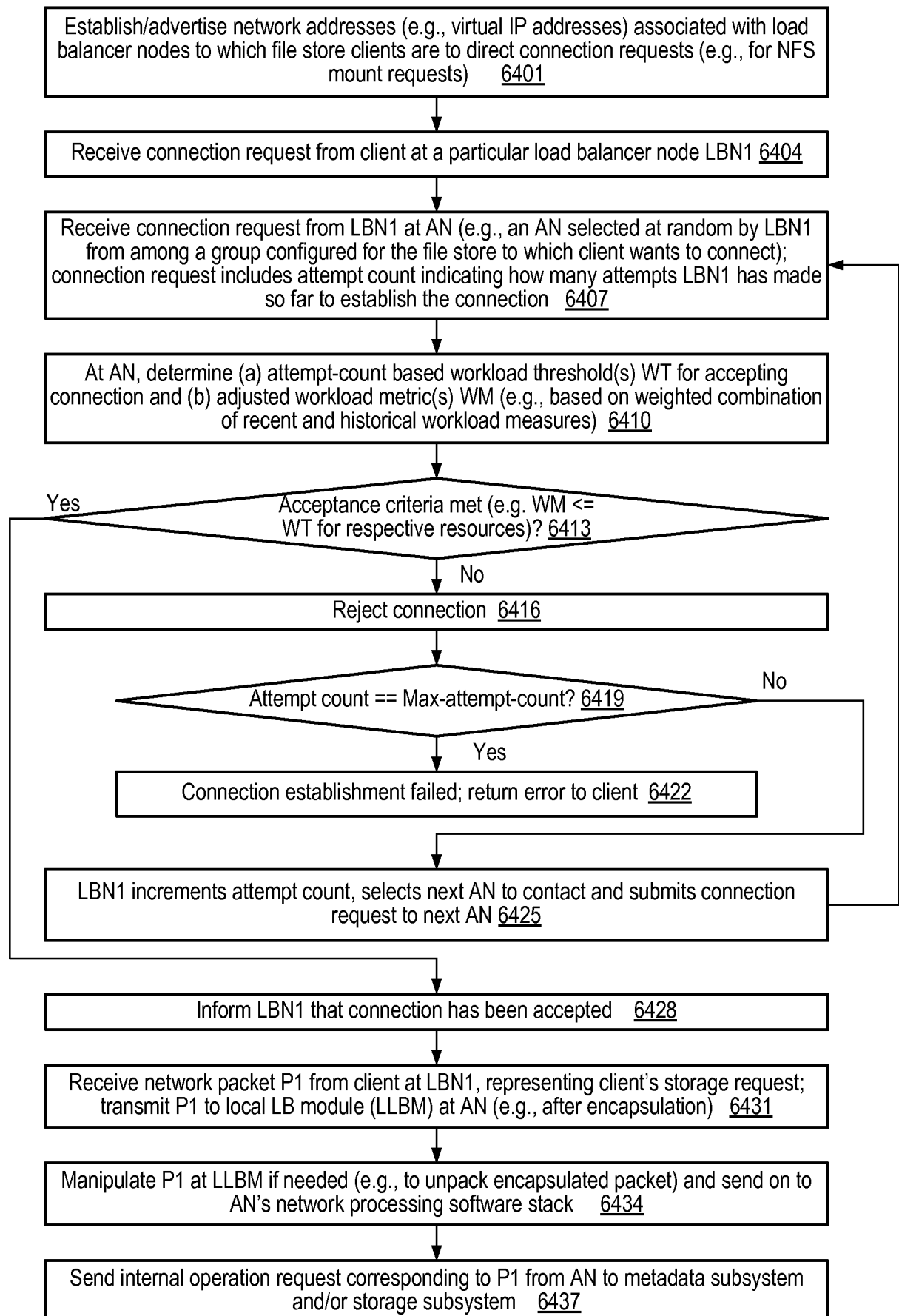
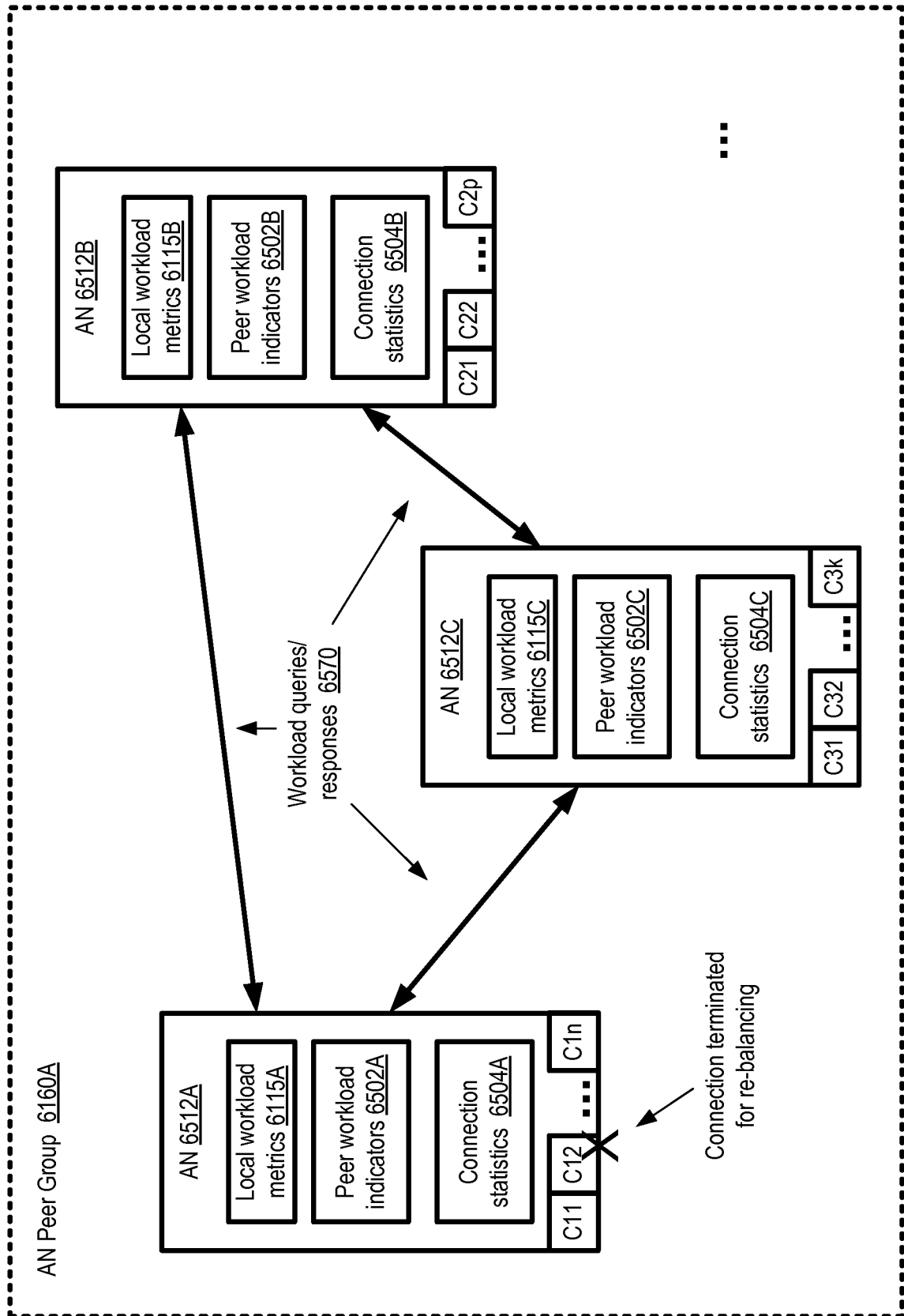


FIG. 64

65/69



Connection
acceptance and
rebalancing
criteria table
6655

Attempt count = 1	Attempt count = 2	Attempt count = 3	Attempt count = 4	Rebalancing
$\begin{aligned} &AL[CPU] \leq 0.5 * AC[CPU] \ \&\& \\ &AL[Net] \leq 0.5 * AC[Net] \ \&\& \\ &AL[Mem] \leq 0.5 * AC[Mem] \ \&\& \\ &AL[Dsk] \leq 0.5 * AC[Dsk] \ \&\& \\ &AL[OSS] \leq 0.5 * AC[OSS] \end{aligned}$	$\begin{aligned} &AL[CPU] \leq 0.55 * AC[CPU] \ \&\& \\ &AL[Net] \leq 0.55 * AC[Net] \ \&\& \\ &AL[Mem] \leq 0.55 * AC[Mem] \end{aligned}$	$\begin{aligned} &AL[CPU] \leq 0.6 * AC[CPU] \ \&\& \\ &AL[Net] \leq 0.6 * AC[Net] \end{aligned}$	$\begin{aligned} &AL[CPU] \leq 0.8 * AC[CPU] \ \&\& \\ &AL[Net] \leq 0.8 * AC[Net] \end{aligned}$	$\begin{aligned} &(AL[CPU] \geq 0.85 * AC[CPU] \\ &\quad \parallel \\ &AL[Net] \geq 0.85 * AC[Net]) \\ &\quad \&\& \\ &AL[Mem] \geq 0.85 * AC[Mem] \\ &\quad \&\& \\ &Seconds\text{-}since\text{-}last\text{-}rebalance > 600 \\ &\quad \&\& \\ &Estimated\text{-}peer\text{-}node\text{-}connection\text{-} \\ &acceptance\text{-}probability > 70\% \end{aligned}$

FIG. 66

67/69

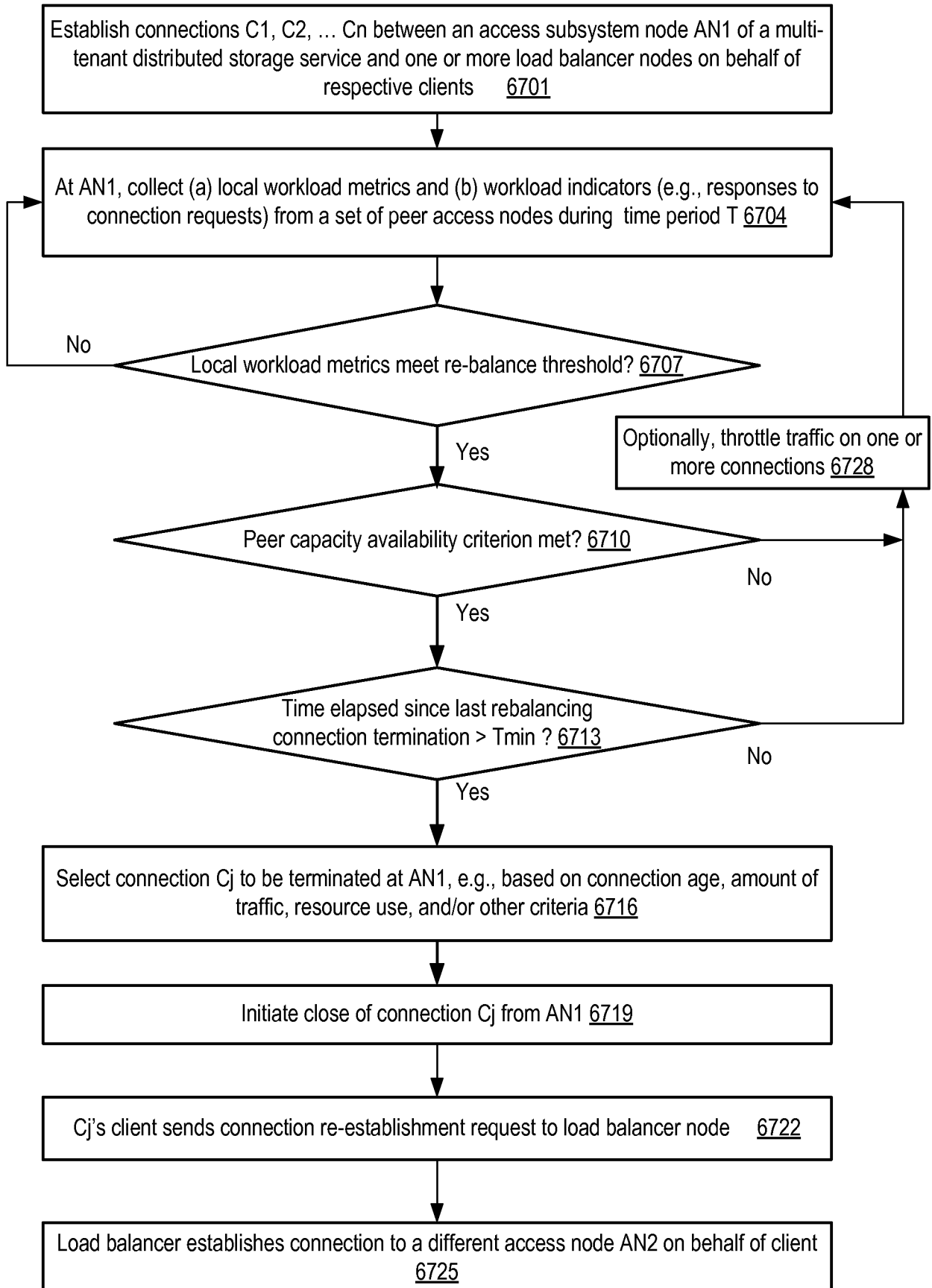


FIG. 67

68/69

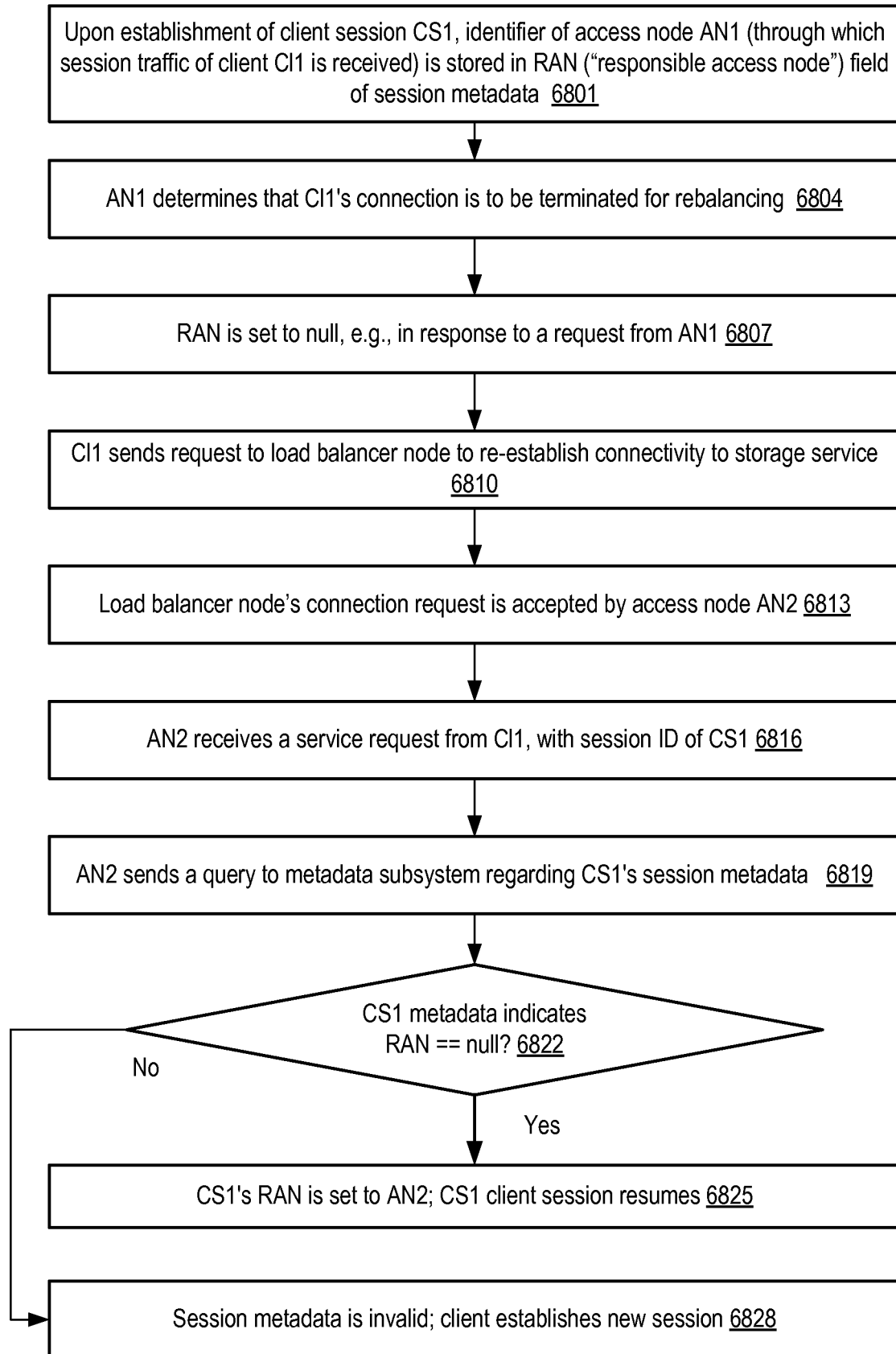


FIG. 68

69/69

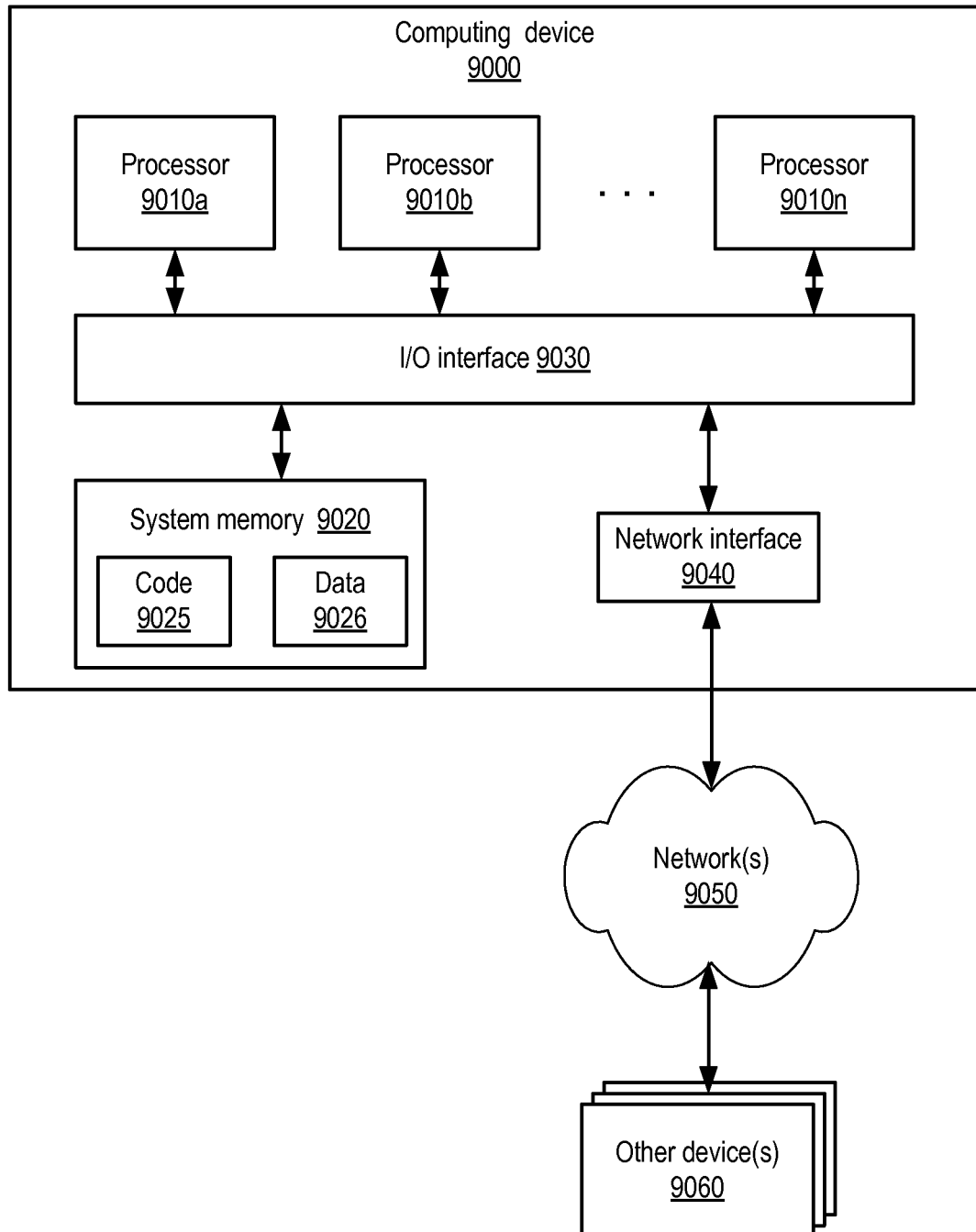


FIG. 69