



[12] 发明专利说明书

[21] ZL 专利号 98109322.1

[45] 授权公告日 2005 年 4 月 13 日

[11] 授权公告号 CN 1197009C

[22] 申请日 1998.5.27 [21] 申请号 98109322.1

[30] 优先权

[32] 1997.5.27 [33] JP [31] 136847/1997

[71] 专利权人 株式会社东芝

地址 日本神奈川县

[72] 发明人 平山秀昭 白木原敏雄

审查员 陈汝岩

[74] 专利代理机构 中国专利代理(香港)有限公司

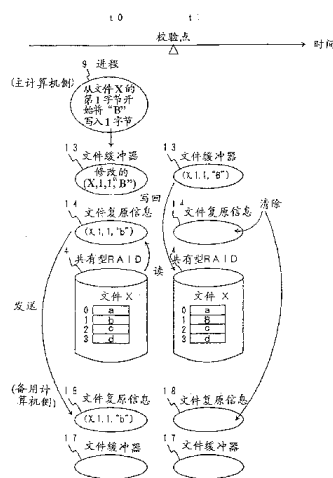
代理人 程天正 叶恺东

权利要求书 5 页 说明书 9 页 附图 13 页

[54] 发明名称 文件系统和文件管理方法

[57] 摘要

一种可以极大地提高共有型 RAID 的盘利用率的文件系统。在更新配置在共有型 RAID 上的文件时，读出更新前的数据，作为文件复原信息预先保持到运用系统计算机和待机系统计算机中；在进程中断(失败)时，使用该文件复原信息将文件复原为校验点采集时的状态。在指示了下一个校验点的采集时，就丢弃该文件复原信息。在运用系统计算机停止时，使用待机系统计算机保持的文件复原信息，将文件复原为最后采集的校验点时的状态。



1. 一种应用于计算机系统的文件系统，所述计算机系统是在共有 RAID 的运用系统计算机和待机系统计算机中双重化的计算机系统，具有定期地采集用于使中断的进程再次开始执行的校验点、并在发生故障时通过将系统复原为在故障发生前最后采集的校验点的状态而再次开始执行所中断的进程的故障恢复功能，该文件系统的特征在于，具有：

第 1 文件复原信息保持单元，根据执行中的进程，指示了更新配置在上述 RAID 上的文件时，从该文件中读出更新前的数据，并作为用于将上述文件复原为对该读出的更新前的数据最后进行采集的校验点的状态的文件复原信息，在上述运用系统计算机上进行保持；

文件复原信息传送单元，将上述保持的文件复原信息向上述待机系统计算机传送；

第 2 文件复原信息保持单元，接收由上述单元传送来的文件复原信息，并保持在上述待机系统计算机上；和

校验点处理单元，在采集了新的校验点时，丢弃上述第 1 和第 2 文件复原信息保持单元所保持的文件复原信息。

2. 按照权利要求 1 所述的文件系统，其特征在於，还具有文件复原单元，在进程失败时，根据上述第 1 文件复原信息保持单元所保持的文件复原信息，将文件复原为最后采集的校验点的状态，并在复原后丢弃上述第 1 和第 2 文件复原信息保持单元所保持的文件复原信息。

3. 按照权利要求 1 所述的文件系统，其特征在於，还具有文件复原单元，由于硬件或操作系统的故障而使上述运用系统计算机停止时，根据上述第 2 文件复原信息保持单元所保持的文件复原信息，将文件复原为最后采集的校验点的状态，并在复原后丢弃上述第 2 文件复原信息保持单元所保持的文件复原信息。

4. 按照权利要求 1 所述的文件系统，其特征在於，所述文件复原信息传送单元具有中断单元，在由于硬件或操作系统的故障而使待机系统计算机停止时，中断上述文件复原信息向上述待机系统计算机的传送。

5. 按照权利要求 4 所述的文件系统, 其特征在于, 所述文件复原信息
传送单元进而还具有在上述文件复原信息向上述待机系统计算机的传送中
断时、将上述文件复原信息向共有上述 RAID 的上述运用系统计算机和待机
系统计算机以外的其他计算机传送的单元, 并具有接收上述传送来的文件
复原信息、在上述其他计算机上进行保持的第 3 文件复原信息保持单元。

6. 一种应用于在共有 RAID 的运用系统计算机和待机系统计算机中双
重化的计算机系统的文件系统, 所述计算机系统具有定期地采集用于使中
断的进程再次开始执行的校验点、并在发生故障时通过将系统复原为最后
采集的校验点的状态而再次开始执行所中断的进程的故障恢复功能, 该文
件系统的特征在于, 具有:

第 1 文件复原信息保持单元, 在采集校验点时, 取得配置在上述 RAID
上的、只进行添加更新的文件的大小, 并将该取得的文件大小作为用于将
文件复原为最后采集的校验点的状态的文件复原信息进行保持;

文件复原信息传送单元, 将上述保持的文件复原信息向上述待机系
统计算机传送; 和

第 2 文件复原信息保持单元, 接收上述传送的文件复原信息, 并在
上述待机系统计算机上保持。

7. 按照权利要求 6 所述的文件系统, 其特征在于, 还具有文件复原单
元, 在进程失败时, 根据上述第 1 文件复原信息保持单元所保持的文件复
原信息, 将文件大小重新设定为最后采集的校验点的文件大小。

8. 按照权利要求 6 所述的文件系统, 其特征在于, 还具有文件复原单
元, 在由于硬件或操作系统的故障而使上述运用系统计算机停止时, 根据
上述第 2 文件复原信息保持单元所保持的文件复原信息, 将文件大小重新
设定为最后采集的校验点的文件大小。

9. 按照权利要求 6 所述的文件系统, 其特征在于, 所述文件复原信息
传送单元具有中断单元, 在由于硬件或操作系统的故障而使上述待机系统
计算机停止时, 中断上述文件复原信息向上述待机系统计算机的传送。

10. 按照权利要求 9 所述的文件系统, 其特征在于, 所述文件复原信息
传送单元具有传送单元, 在中断上述文件复原信息向上述待机系统计算机

的传送时，将上述文件复原信息向共有上述 RAID 的上述运用系统计算机及待机系统计算机以外的其他计算机传送；此外，还具有第 3 文件复原信息保持单元，接收上述传送的文件复原信息，并在上述其他计算机上保持。

11. 一种应用于计算机系统的文件管理方法，所述计算机系统是在共有 RAID 的运用系统计算机和待机系统计算机中双重化的计算机系统，具有定期地采集用于使中断的进程再次开始执行的校验点、并在发生故障时通过将系统复原为最后采集的校验点的状态而再次开始执行所中断的进程的故障恢复功能，该文件管理方法的特征在于，包括如下步骤：

在由执行中的进程指示了更新配置在上述 RAID 上的文件时，从该文件中读出更新前的数据，并作为用于将上述文件复原为对该读出的更新前的数据进行最后采集的校验点的状态的文件复原信息在上述运用系统计算机上进行保持；

将上述保持的文件复原信息向上述待机系统计算机传送；

接收上述传送的文件复原信息，并在上述待机系统计算机上进行保持；

采集了下一个校验点时，丢弃在上述运用系统计算机和待机系统计算机上保持的文件复原信息；

在进程失败时，根据上述运用系统计算机上保持的文件复原信息，将文件复原为最后采集的校验点的状态，并在复原后丢弃在上述运用系统计算机和待机系统计算机上保持的文件复原信息。

12. 一种应用于计算机系统的文件管理方法，所述计算机系统是在共有 RAID 的运用系统计算机和待机系统计算机中双重化的计算机系统，具有定期地采集用于使中断的进程再次开始执行的校验点、并在发生故障时通过将系统复原为最后采集的校验点的状态而再次开始执行所中断的进程的故障恢复功能，该文件管理方法的特征在于包括如下步骤：

在由执行中的进程指示了更新配置在上述 RAID 上的文件时，从该文件中读出更新前的数据，并作为用于将上述文件复原为对该读出的更新前的数据进行最后采集的校验点的状态的文件复原信息在上述运用系统计算机上进行保持；

将上述保持的文件复原信息向上述待机系统计算机传送;

接收上述传送的文件复原信息并在上述待机系统计算机上进行保持;

5 采集了新的校验点时, 丢弃在上述运用系统计算机和待机系统计算机上保持的上述文件复原信息;

在由于硬件或操作系统的故障而使上述运用系统计算机停止时, 根据上述待机系统计算机上保持的文件复原信息, 将文件复原为最后采集的校验点的状态, 并在复原后丢弃在上述待机系统计算机上保持的文件复原信息。

10 13. 一种应用于计算机系统的文件管理方法, 所述计算机系统是在共有 RAID 的运用系统计算机和待机系统计算机中双重化的计算机系统, 具有定期地采集用于使中断的进程再次开始执行的校验点、并在发生故障时通过将系统复原为最后采集的校验点的状态而再次开始执行所中断的进程的故障恢复功能, 该文件管理方法的特征在于, 包括如下步骤:

15 在采集了校验点时, 取得配置在上述 RAID 上的、只进行添加更新的文件的大小并将该取得的文件大小作为用于将上述文件复原为最后采集的校验点的状态的文件复原信息进行保持;

将上述保持的文件复原信息向上述待机系统计算机传送;

接收上述传送的文件复原信息, 并在上述待机系统计算机上保持;

20 和

在进程失败时, 根据在上述运用系统计算机上保持的文件复原信息, 将上述文件大小重新设定为最后采集的校验点的文件大小。

25 14. 一种应用于计算机系统的文件管理方法, 所述计算机系统是在共有 RAID 的运用系统计算机和待机系统计算机中双重化的计算机系统, 具有定期地采集用于使中断的进程再次开始执行的校验点、并在发生故障时通过将系统复原为最后采集的校验点的状态而再次开始执行所中断的进程的故障恢复功能, 该文件管理方法的特征在于, 包括如下步骤:

在采集了校验点时, 取得配置在上述 RAID 上的、只进行添加更新的文件的大小并将该取得的文件大小作为用于将上述文件复原为最后采集的

-
- 校验点的状态的文件复原信息进行保持;
- 将上述保持的文件复原信息向上述待机系统计算机传送;
- 接收上述传送的文件复原信息,并在上述待机系统计算机上保持;
- 在由于硬件或操作系统的故障而使上述运用系统计算机停止时,根
- 5 据在上述待机系统计算机上保持的文件复原信息,将上述文件大小重新设定为最后采集的校验点的文件大小。

文件系统和文件管理方法

5 本发明涉及适用于计算机系统的文件系统和文件管理方法，特别是可以极大地提高共有的冗长式排列盘(RAID)的盘利用效率的文件系统和文件管理方法，其中计算机系统在利用共有冗长式排列盘(非昂贵盘的冗余阵列(Redundant Arrays of Inexpensive Disks)，以下简称为RAID)的网络连接的多个计算机中多重化，并且具有校验点/恢复方式的故障恢复功能。

10 现在，已存在定期地采集进程的地址空间及上下文和文件等的状态(称之为校验点)、在发生故障时、通过将系统恢复为最后采集的校验点的状态使进程重新开始、实现从故障中恢复的计算机系统的高可靠性技术。

15 另外，还适用于将该校验点·重新开始技术双重化的系统(例如热备用型系统)，存在着将由2个计算机中的一个主计算机(以后称为运用系统计算机)采集的校验点向备用计算机(以后，称为待机系统计算机)发送、在发生故障时使进程从校验点重新开始的恢复处理技术。当然，该恢复处理不论在运用系统计算机中还是待机系统计算机中都可以进行。

20 下面，参照图10简单地说明这些技术。

如图10(a)所示，在该系统中，由主计算机在执行中的时间 t_1 、 t_2 …定期地采集进程的校验点并保持该校验点，同时，也向备用计算机传送。

25 图10(b)是表示在上述系统中在主计算机上执行中的进程失败时，进程在主计算机上从校验点重新开始的情形。

图10(c)是表示在主计算机的硬件及操作系统中发生故障时该进程在备用计算机上从校验点重新开始的情形。这些技术在申请人的日本专利申请特愿平8-233021号等中详细地进行了说明。

30 如图11所示，以往，在这种系统中，主计算机51a和备用计算机51b分别具有独立的盘(53a、53b)。但是，如果这样采用双重化的盘，则盘的利用效率将降低为50%。

因此，作为不将双重化盘而从盘的故障中保护数据的机构，存在冗长式排列盘(RAID)。下面，对于该RAID，在《计算机体系数值逼近》(第二版)(COMPUTER ARCHTECTURE A QUANTTATIVE APPROACH SECOND EDITION)(JOHN L HENNESSY & DAVID PATTERNSON 公司出版)等中有详尽的说明。

如图12所示，在第四级冗长式排列盘RAID4和第五级RAID5中，例如利用5张盘提供即使其中的某一张盘坏了数据也可以复原的4张的盘容量。图中，「PA-D」表示用于复原「A」、「B」、「C」和「D」的奇偶数据。同样，「PE-H」表示用于复原「E」、「F」、「G」和「H」的奇偶数据。

例如，在单纯的双重化盘中，为了提供4张的盘容量，就需要8张盘，所以，盘利用效率 = $1 / 2 = 50\%$ 。但是，在RAID4和RAID5中，例如利用5张盘就可以提供可以复原的4张的盘容量，所以，盘利用效率 = $4 / 5 = 80\%$ ，与双重化盘相比，盘的利用效率高。以后，只要没有特别指明，就将RAID4或RAID5简称为RAID。

在上述特愿平8-233021号等所示的技术中，提出了运用系统计算机和待机系统计算机不是分别具有独立的盘、而在运用系统计算机和待机系统计算机中应用共有的RAID(以后，称为「共有型RAID」)的方法。但是，这种技术是以利用主文件进行的更新在经过校验点之前不反映在备用文件中为前提的。因此，如图13所示，即使运用系统计算机和待机系统计算机具有共有型RAID时，也必须采用在共有型RAID55上配置主文件54a和备用文件54b的结构。因此，双重保持文件的结果，就是盘利用效率 = $4 / 5 \times 1 / 2 = 40\%$ ，盘的利用效率非常差。

本发明就是鉴于这种情况而提案的，目的在于提供即使使用共有型冗长式排列盘RAID时也可以将盘利用效率提高到RAID原来的盘利用效率的文件系统和文件管理方法。

本发明的特征在于：更新配置在共有型冗长式排列盘(RAID)上的文件时，读出该更新前的数据，预先由运用系统计算机和待机系统计算机双方进行保持，在采集了校验点时，放弃该更新前的数据。

按照本发明，在进程失败时，使用在运用系统计算机上保持的更新前的数据，将文件复原为最后采集的检验点的状态，在文件复原后放弃更新前的数据。此外，在运用系统计算机停止时，就使用在待机系统计算机上保持的更新前的数据，将文件复原为最后采集的校验点的状态，在该文件复原后放弃更新前的数据。

即，本发明是，在具有例如利用校验点/恢复方式的故障恢复功能的热待机型计算机系统中，不采用双重化方法就可以将配置在运用系统计算机和待机系统计算机所共有的冗长式排列盘上的文件复原为校验点采集的状态。结果，可以极大地提高共有型 R A I D 的盘利用效率。

另外，本发明的特征在于：在每次进行校验点的采集时，获取只进行配置在共有型 R A I D 上的添加更新的文件大小，由运用系统计算机和待机系统计算机双方保持。

进而，本发明的特征在于：在进程失败时，就使用在运用系统计算机上保持的大小，将文件的大小重新设定为在最后采集的校验点的大小。

此外，在运用系统计算机停止时，就使用在待机系统计算机上保持的大小，将文件的大小重新设定为最后采集的校验点的文件大小。

即，本发明在应用利用校验点/恢复方式的故障恢复功能的备用型计算机系统中，不采用双重化方式就可以将配置在由运用系统计算机和待机系统计算机所共有的冗长式排列盘上的、只进行添加更新的文件复原为最后采集的校验点的状态，所以，可以极大地提高共有型 R A I D 的盘利用效率。

图 1 是表示本发明实施例的计算机系统的结构。

图 2 表示上述实施例的计算机系统的文件更新时的动作原理。

图 3 表示上述实施例的计算机系统发生故障时由运用系统（主）计算机使进程再次开始进行时的动作原理。

图 4 表示上述实施例的计算机系统发生故障时由待机系统（备用）计算机使进程再次开始进行时的动作原理。

图 5 表示在上述实施例的计算机系统的运用系统（主）计算机中更新文件时的处理的流程。

图 6 是表示与在上述实施例的计算机系统的运用系统（主）计算

机中的文件操作有关的校验点处理的流程图。

图7表示与由上述实施例的计算机系统的运用系统(主)计算机使进程再次开始进行时的文件操作有关的重算处理的流程。

图8表示与由上述实施例的计算机系统的待机系统(备用)计算机使进程再次开始进行时的文件操作有关的重算处理的流程。

图9表示在上述实施例的计算机系统的共有型冗长式排列盘上的文件访问仅是添加写入时的重算处理的动作。

图10是说明应用先有的校验点/恢复方式的故障恢复功能的双重化系统的图。

图11是说明在应用先有的校验点/恢复方式的故障恢复功能的双重化系统中的文件配置的图。

图12是用于说明冗长式排列盘(RAID)的概念图。

图13是用于说明在应用先有的校验点/恢复方式的故障恢复功能的双重化系统中具有共有型RAID时的文件配置的图。

下面,参照附图说明本发明的实施例。

图1是表示本发明的计算机系统结构的实施例。

本实施例的计算机系统是由运用系统计算机(主)计算机1和待机系统(备用)计算机2构成的双重化系统(例如,热备用型计算机系统)。2个计算机通过网络3而连接,进而,2个计算机具有可以访问的共有型冗长式排列盘(RAID)4。

在运用系统计算机1中,执行中的进程9进行文件操作时,通过由操作系统内部的文件系统10管理的文件缓冲器13(有时也称为高速缓冲存储器)访问配置在共有型RAID4上的文件,

进程9更新文件时,文件复原信息保存部11从文件中读出用于复原文件的信息,例如进行写入时就读出进行写入前的数据,并将该数据与读入的文件名及位置、长度等一起作为文件复原信息14进行保存。该文件复原信息保存部11将文件复原信息14向待机系统(备用)计算机2的文件复原信息接收部19传送。文件复原信息接收部19接收到文件复原信息时,就将其作为文件复原信息18另外保存。

校验点控制部5定期地要求进行进程9的校验点采集。校验点采集的指示同时从校验点控制部5向校验点信息保存部6和文件复原

信息保存部发送。

校验点信息保存部 6 从校验点控制部 5 接收到校验点采集的指示时，就将进程 9 的地址空间及处理上下文的内容作为校验点信息 8 进行保存。校验点信息保存部 6 将该校验点信息 8 向待机系统计算机 2 的校验点信息接收部 2 3 传送。校验点信息接收部 2 3 接收到该校验点信息 8 时，就将其作为校验点信息 2 2 另外保存。

另一方面，文件复原信息保存部 1 1 从校验点控制部 5 接收到校验点采集的指示时，通过将在文件缓冲器 1 3 中存在的还未写回到共有型冗长式排列盘 (RAID) 4 中的更新数据写回到共有型冗长式排列盘 (RAID) 4 中，在共有型 RAID 4 上反映数据更新。在该数据的反映结束后，就清除文件复原信息 1 4。

文件复原信息保存部 1 1 将该校验点采集的指示向待机系统计算机 2 的文件复原信息接收部 1 9 传送。

文件复原信息接收部 1 9 接收到校验点采集的指示时，就清除文件复原信息 1 8。

校验点信息保存部 6 和文件复原信息保存部 1 1 在待机系统计算机 2 下行时，就中断校验点信息 8 和文件复原信息 1 4 的传送。

在运用系统计算机 1 中，执行中的进程 9 失败时，运用系统计算机 1 使进程 9 从最后采集的校验点再次开始进行时，重新开始控制部 1 5 就向校验点复原部 7 和文件复原部 1 2 同时发送重算处理的指示。

校验点复原部 7 从重新开始控制部 1 5 接收到重算处理的指示时，就使用校验点信息 8 使进程 9 从校验点采集时再次开始进行。另外，文件复原部 1 2 从重新开始控制部 1 5 接收到重算处理的指示时，就使用文件复原信息 1 4 使文件复原为在校验点的状态。

然后，重新开始控制部 1 5 就通知待机系统计算机 2 的再次开始控制部 2 1 进行文件的重算处理。待机系统计算机的再次开始控制部 2 1 接收到进行文件的重算处理的通知时，就向文件复原部 2 0 发出指示，并清除文件复原信息 1 8。

另外，由于硬件或操作系统的故障等而使运用系统计算机 1 不能工作时，为了利用待机系统 (备用) 计算机 2 使进程 9 从校验点再次开始进行，重新开始控制部 2 1 同时向校验点复原部 2 4 和文件复原

部 2 0 发送重算处理的指示。

校验点复原部 2 4 从重新开始控制部 2 1 接收到重算处理的指示时，就使用校验点信息 2 2，使进程 9 作为进程 2 5，从校验点采集时再次开始进行。另外，文件复原部 2 0 从重新开始控制部 2 1 接收
5 收到重算处理的指示时，就使用文件复原信息 1 8，将文件复原为校验点的状态。

图 2 是表示本实施例的计算机系统的文件更新的动作原理的图。

在时刻 t_0 ，进程 9 从文件 x 的第 1 字节开始向 1 字节中写入
10 “B”。在进行更新之前，从文件 x 的第 1 字节开始读取 1 字节的数据，并将其作为文件复原信息 1 4 进行保存，并且，向备用计算机 2 传送，作为文件复原信息 1 8 进行保存。

其次，在时刻 t_1 ，采集校验点。这样，将还未在共有型 R A I D
15 4 中反映的文件缓冲器 1 3 中的更新数据写回到共有型 R A I D 4 中，然后，清除运用系统（主）计算机 1 的文件复原信息 1 4 和待机系统（备用）计算机 2 的文件复原信息 1 8。

图 3 是表示在本实施例的计算机系统中发生故障时在运用系统计算机使进程再次开始进行时的动作原理的图。

在时刻 t_1 ，采集校验点。然后，在时刻 t_2 ，进程 9 从文件 x
20 的第 1 字节开始向 1 字节中写入“B”。在进行更新之前，从文件 x 的第 1 字节开始读取 1 字节的数据，将其作为文件复原信息 1 4 进行保存，并且向备用计算机 2 传送，作为文件复原信息 1 8 进行保存。

然后，在时刻 t_3 ，将文件缓冲器 1 3 的更新数据写回到共有型
25 冗长式排列盘（R A I D）4 中。然后，在时刻 t_4 ，由于发生故障，所以，在运用系统计算机 1 中使进程 9 重新开始。因此，在运用系统（主）计算机 1 和待机系统（备用）计算机 2 中，必须使文件 x 在校验点的状态重新运行。

为了使文件在校验点的状态重新运行，（1）将作为文件复原信息 1 4 保存的数据写入文件缓冲器 1 0，（2）可以将文件复原信息
30 1 4 和文件复原信息 1 8 清除。

这时，共有型 R A I D 上的文件 X 的数据还未重新运行，但是，由于已在访问的文件缓冲器 1 3 上重新运行了数据，所以，没有问

题。

然后，共有型 R A I D 4 上的文件 X 的数据，还要从文件缓冲器 1 3 写回，但是，这时，也可以强制地写回。

图 4 表示在本实施例的计算机系统发生故障时在待机系统（备用）计算机中使进程重新开始进行时的动作原理。

和图 3 的情况一样，在时刻 t 1 采集校验点。然后，在时刻 t 2，进程 9 从文件 X 的第 1 字节开始将“B”写入 1 字节。在进行更新之前，从文件 X 的第 1 字节开始读取 1 字节的数据，将其作为文件复原信息 1 4 进行保存，并且也向备用计算机 2 传送，作为文件复原信息 1 8 进行保存。

然后，在时刻 t 3，将文件缓冲器 1 3 中的更新数据写回到共有型 R A I D 4 中。并且，在时刻 t 4 发生故障时，为了使进程 9 在待机系统（备用）计算机中再次开始进行，在待机系统（备用）计算机 2 中，使文件 X 恢复为校验点的状态。恢复的方式和图 3 的情况相同，（1）将作为文件复原信息 1 8 而保存的数据写入文件缓冲器 1 7，（2）清除文件复原信息 1 8。

图 5 是由本实施例的计算机系统的运用系统（主）计算机对文件进行更新处理的流程图。

文件复原信息保存部 1 1 挂上进程 9 的文件更新操作，读出文件更新前的数据，作为文件复原信息 1 4 进行保存（步骤 A 1）。同时，将该文件复原信息 1 4 向待机系统（备用）计算机 2 发送（步骤 A 2）。发送给待机系统（备用）计算机 2 的文件复原信息 1 4 由文件复原信息接收部 1 9 接收，作为文件复原信息 1 8 进行保存（步骤 A 3）。

图 6 是关于本实施例的计算机系统的运用系统（主）计算机中的文件操作的校验点处理的流程图。

校验点控制部 5 向文件复原信息保存部 1 1 指示执行校验点时的处理（步骤 B 1）。文件复原信息保存部 1 1 将文件缓冲器 1 3 写回到共有型 R A I D 4 中，并清除文件复原信息 1 4（步骤 B 2）。进而，文件复原信息保存部 1 1 向待机系统（备用）计算机 2 的文件复原信息接收部 1 9 指示执行校验点时的处理（步骤 B 3）。接收到该执行指示的文件复原信息接收部 1 9 清除文件复原信息 1 8（步骤

B 4) 。

图 7 是与在本实施例的计算机系统的运用系统(主)计算机中使进程重新开始进行时的文件操作有关的重新处理的流程图。

运用系统(主)计算机 1 的重新开始控制部 1 5 向文件复原部 1 2 指示执行重新处理(步骤 C 1)。文件复原部 1 2 根据文件复原信息 1 4 将文件复原(步骤 C 2)后,清除文件复原信息 1 4(步骤 C 3)。

重新开始控制部 1 5 通过待机系统计算机 2 的重新开始控制部 2 1 向文件复原部 2 0 通知文件已复原(步骤 C 4)。接收到该通知的文件复原部 2 0 清除文件复原信息 1 8(步骤 C 5)。

图 8 是与在本实施例的计算机系统的待机系统(备用)计算机中使进程重新开始进行时的文件操作有关的重新处理的流程图。

待机系统(备用)计算机 2 的重新开始控制部 2 1 向文件复原部 2 0 指示执行重新处理(步骤 D 1)。文件复原部 2 0 根据文件复原信息将文件复原(步骤 D 2)后,清除文件复原信息 1 8(步骤 D 3)。

至此,对文件的访问全部是以通过文件缓冲器 1 3 或文件缓冲器 1 7 为前提进行说明的,但是,也可以不通过文件缓冲器 1 3 或文件缓冲器 1 7 而直接访问共有型 R A I D 4。直接进行访问时,在进行校验点处理时,不需要进行将文件缓冲器 1 3 或文件缓冲器 1 7 写回到共有型 R A I D 4 中的处理。

另外,在只添加写入的情况下,在进行文件更新时,对共有型 R A I D 4 上的文件的访问不需要保存文件复原信息,代之以在进行校验点采集时,将文件大小作为文件复原信息 1 4 在运用系统计算机 1 上进行保存,同时也将其向待机系统(备用)计算机 2 传送,可以在待机系统(备用)计算机 2 上作为文件复原信息 1 8 进行保存。其理由在于,在进程失败时,通过根据在运用系统计算机 1 上保存的文件复原信息 1 4 重新设定进行校验点采集时保存的文件大小,在进行校验点采集时,可以复原文件。

同样,由于硬件或操作系统的故障等而使运用系统计算机 1 不能工作时,通过根据在待机系统计算机 2 上保存的文件复原信息 1 8 重新设定进行校验点采集时保存的文件大小,在进行校验点采集时,可

以复原文件。

图9是表示这种情况下的动作。假定文件X是只进行添加写入的文件。

在时刻t1,对文件x添加写入“aa”,进而,在时刻t2,
5 对文件X添加写入“bbb”。

在时刻t3,采集校验点,作为文件复原信息14保存「文件X是大小5(“aabbb”)」。

在时刻t4,对文件X添加写入“ccc”后,在时刻t5发生
故障。因此,重新恢复为在时刻t3采集的校验点的状态。这时,如
10 果文件X根据文件复原信息14将文件大小设定为5,就复原校验点的
状态。

作为这种只进行添加写入的文件的例子,有日志文件(记录文件)等。

在以上说明的实施例,作为文件的更新操作,只表示出了「写入」和「添加写入」,另外,作为文件复原信息,只表示出了「写入前的数据」和「添加写入前的数据」。但是,除此之外,作为文件更新信息,例如,例如也可以如「文件制作」或「文件删除」那样进行文件的索引操作。这时,文件的复原信息就是例如「文件删除」和「文件制作·删除前的全部数据」。

20 由于待机系统(备用)计算机2的故障而使校验点信息保存6和文件复原信息保存部11的校验点信息和文件复原信息的传送中断时,如果使待机系统(备用)计算机2以外的其他计算机代替待机系统(备用)计算机2的工作,校验点信息保存部6和文件复原信息保存部11向其他计算机传送校验点信息和文件复原信息,就可以进一步
25 提高系统的可靠性。

在上述实施例中记载的文件管理方法,作为可以由计算机执行的程序,可以存储到磁带、光盘和半导体存储器等记录媒体上。

如上所述,按照本发明,在应用校验点/恢复方式的故障恢复功能的热备用型系统等中,不使配置在运用系统(主)计算机和待机系统(备用)计算机所共有的冗长式排列盘(RAID4)上的文件双重化,就可以复原为所采集的校验点的状态,所以,可以极大地提高
30 共有型RAID的盘利用效率。

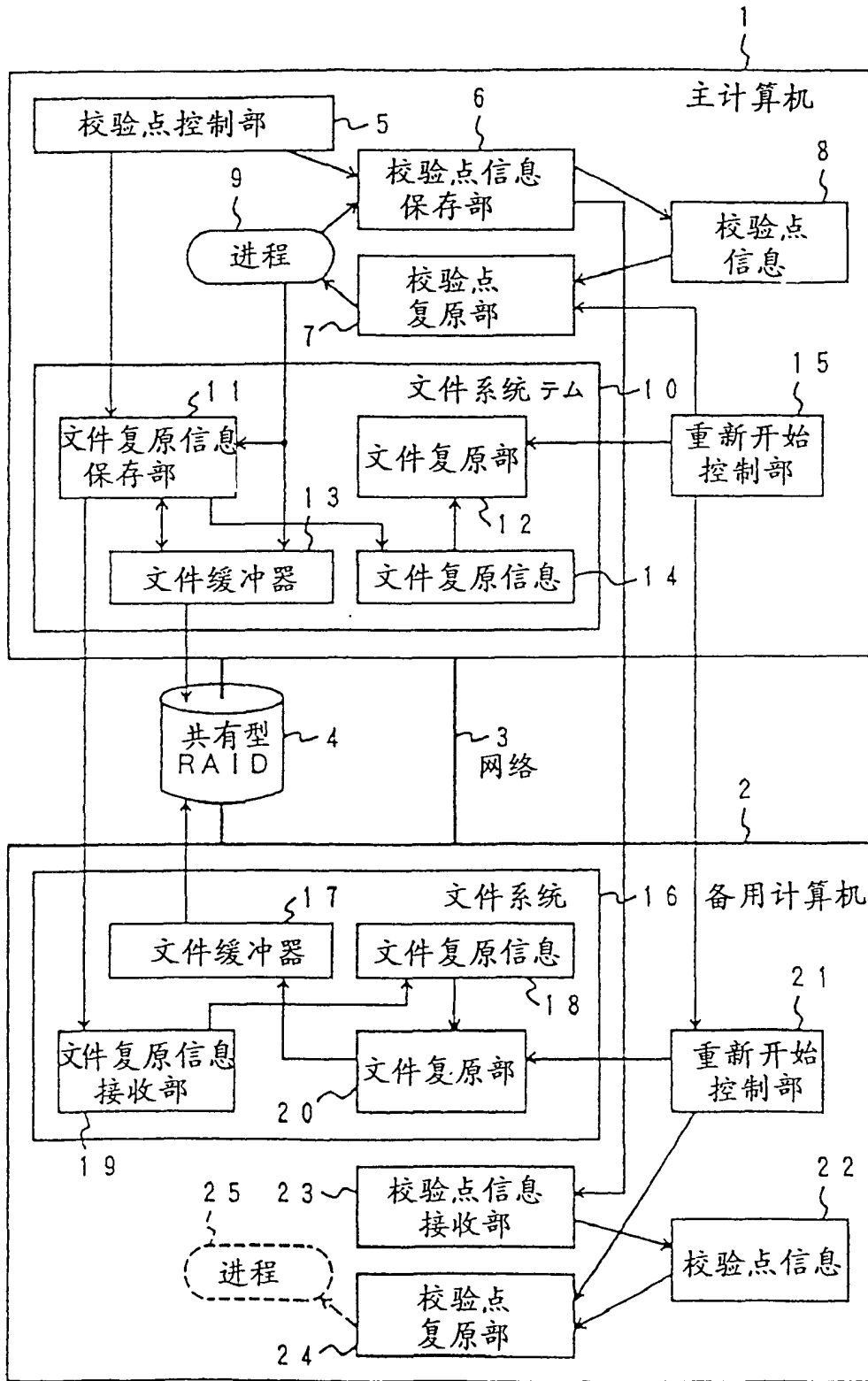


图 1

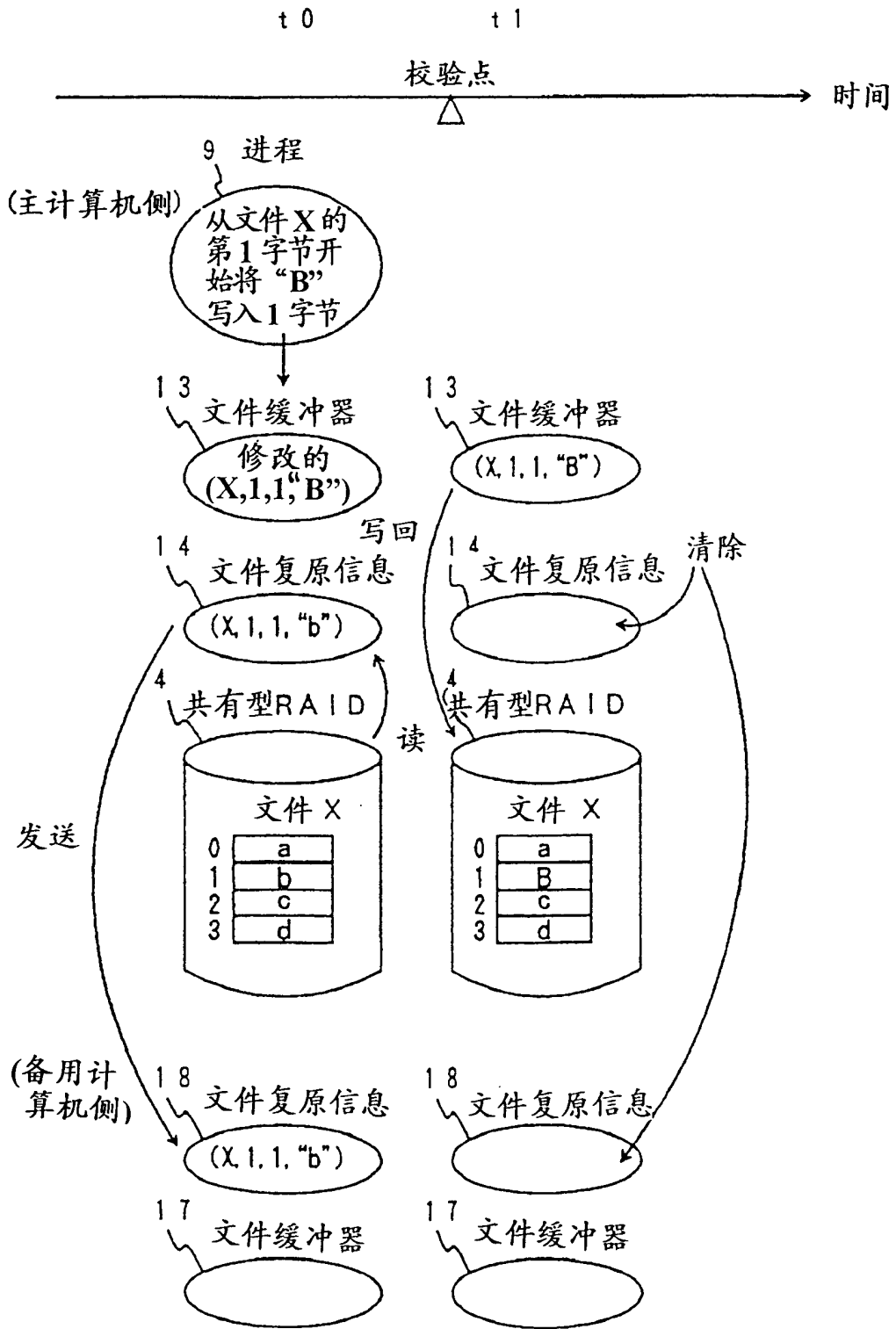


图 2

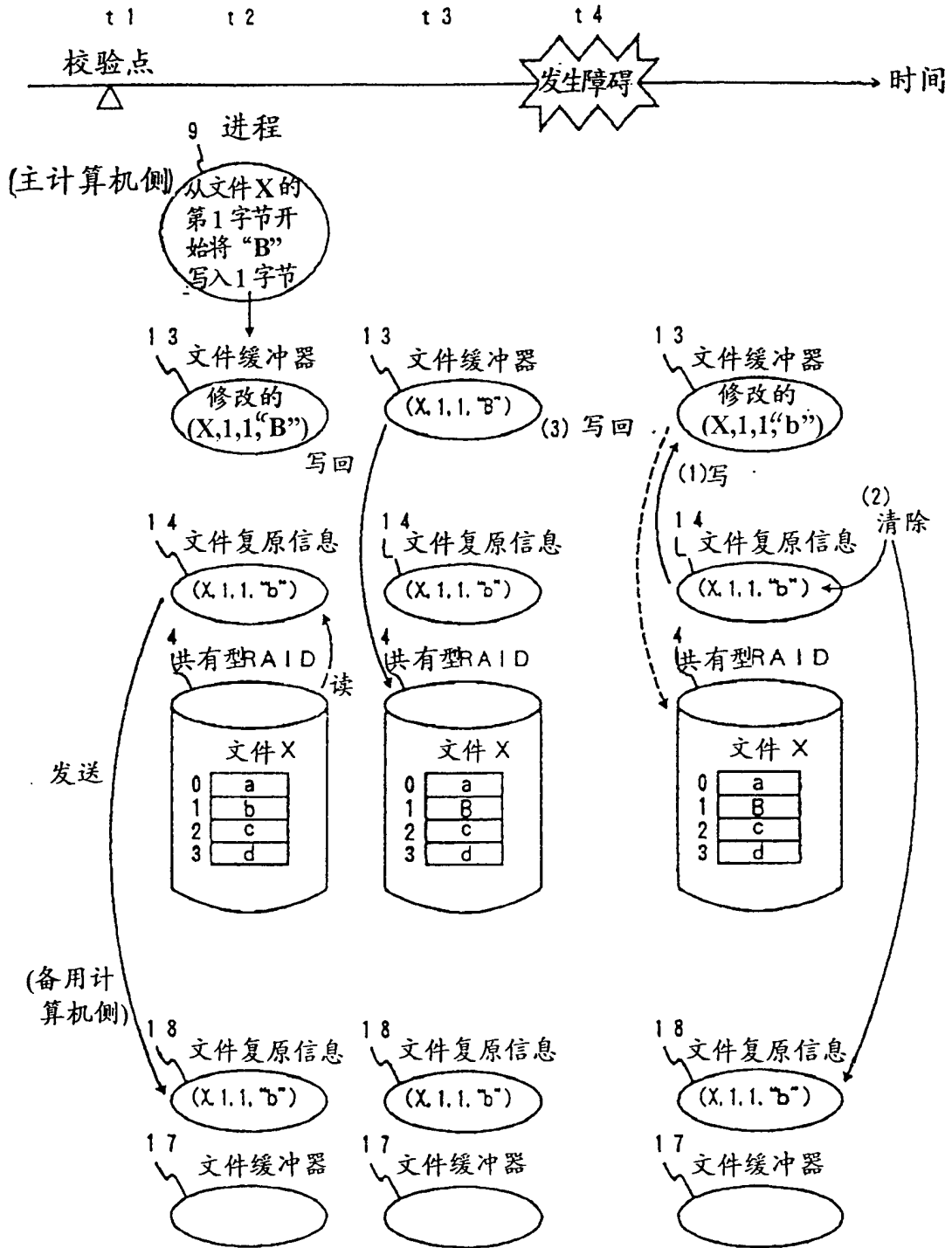


图 3

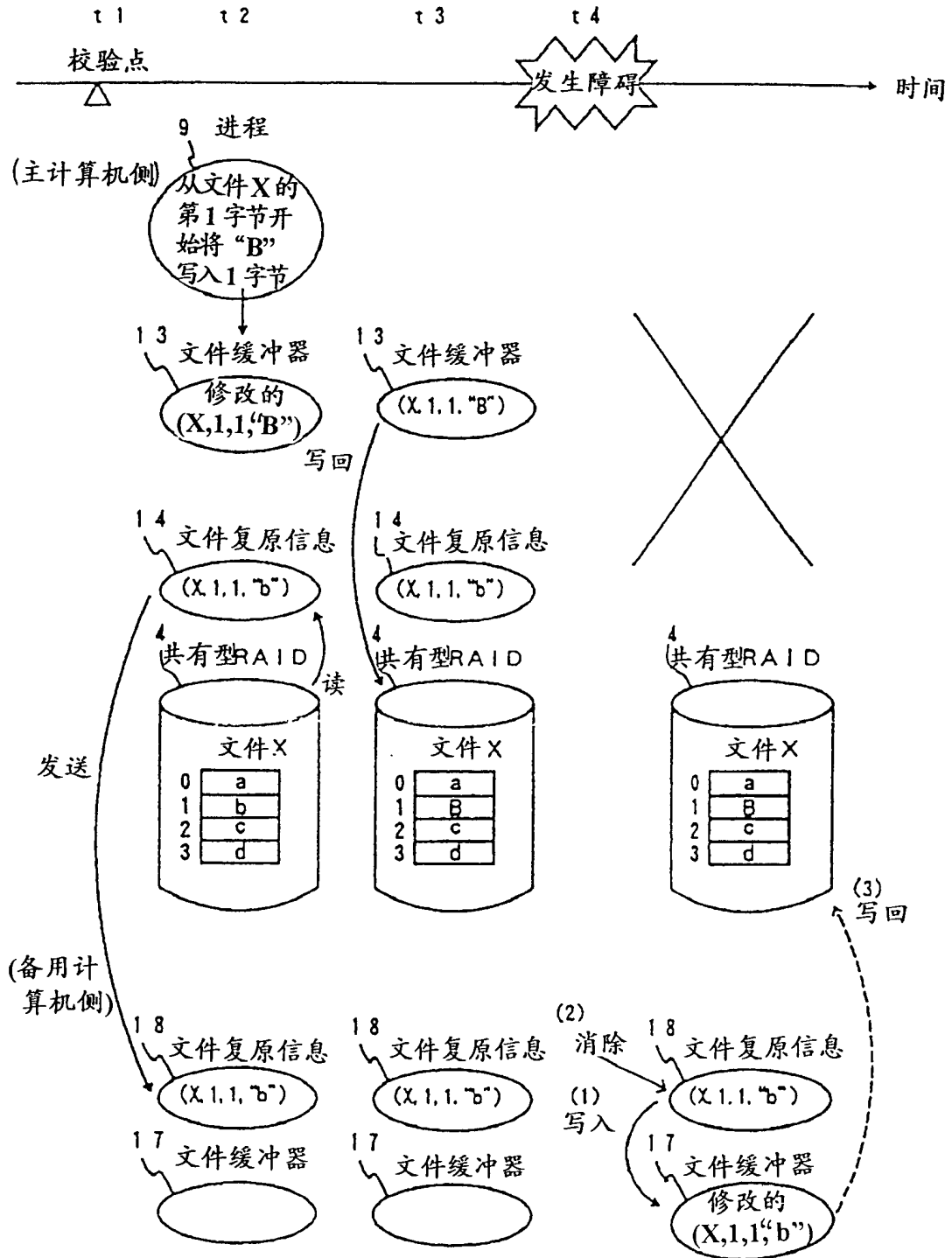


图 4

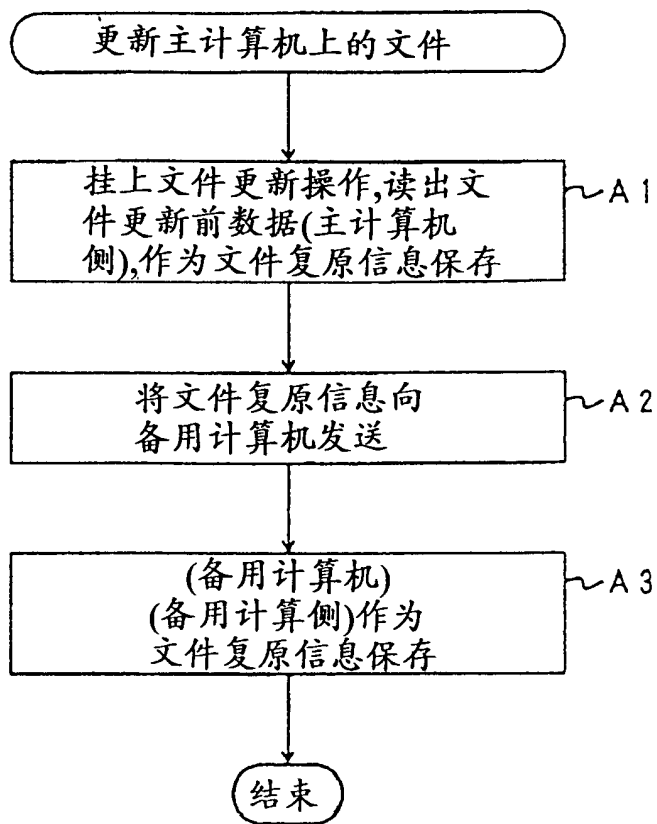


图 5

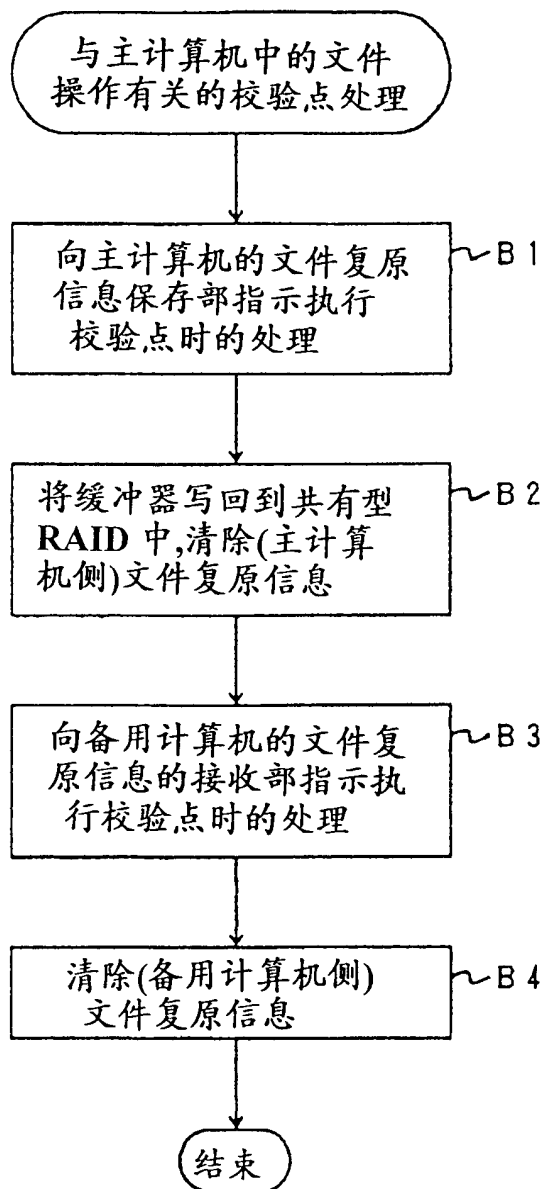


图 6

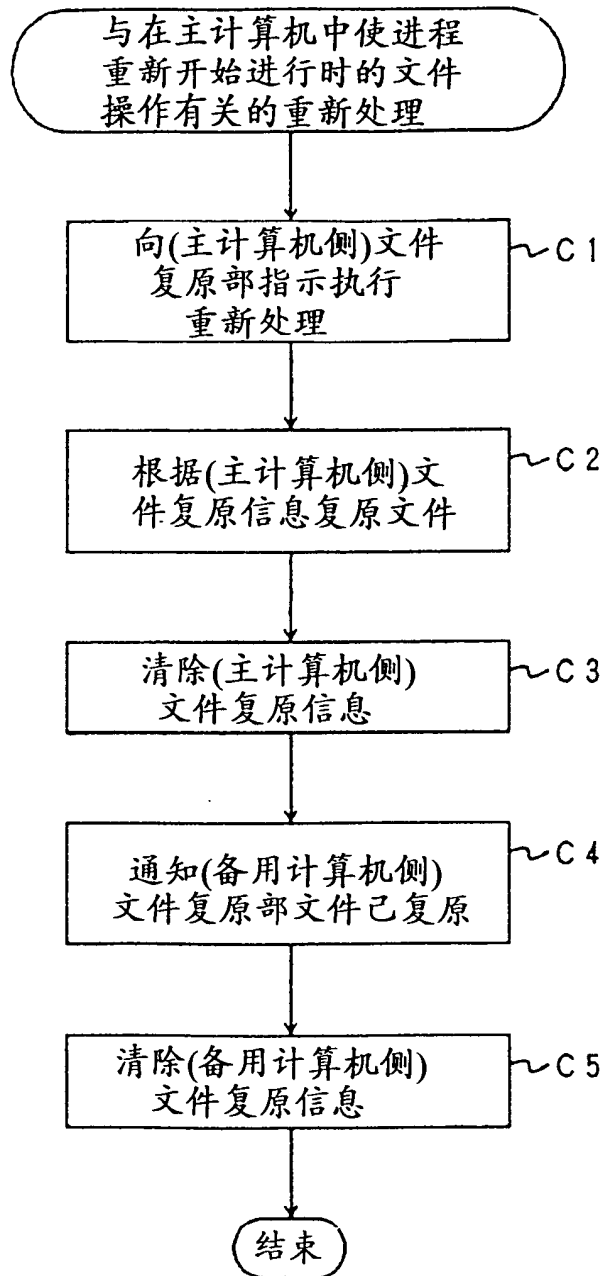


图 7

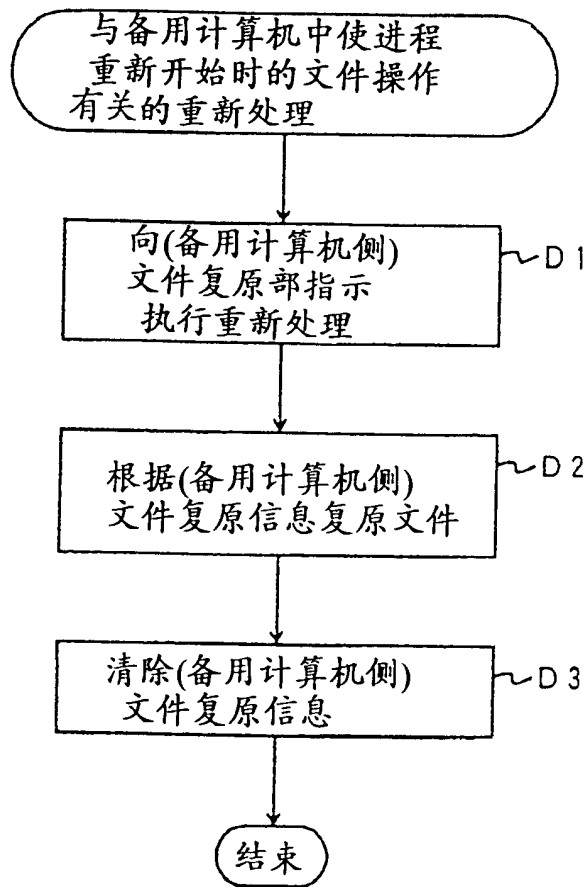


图 8

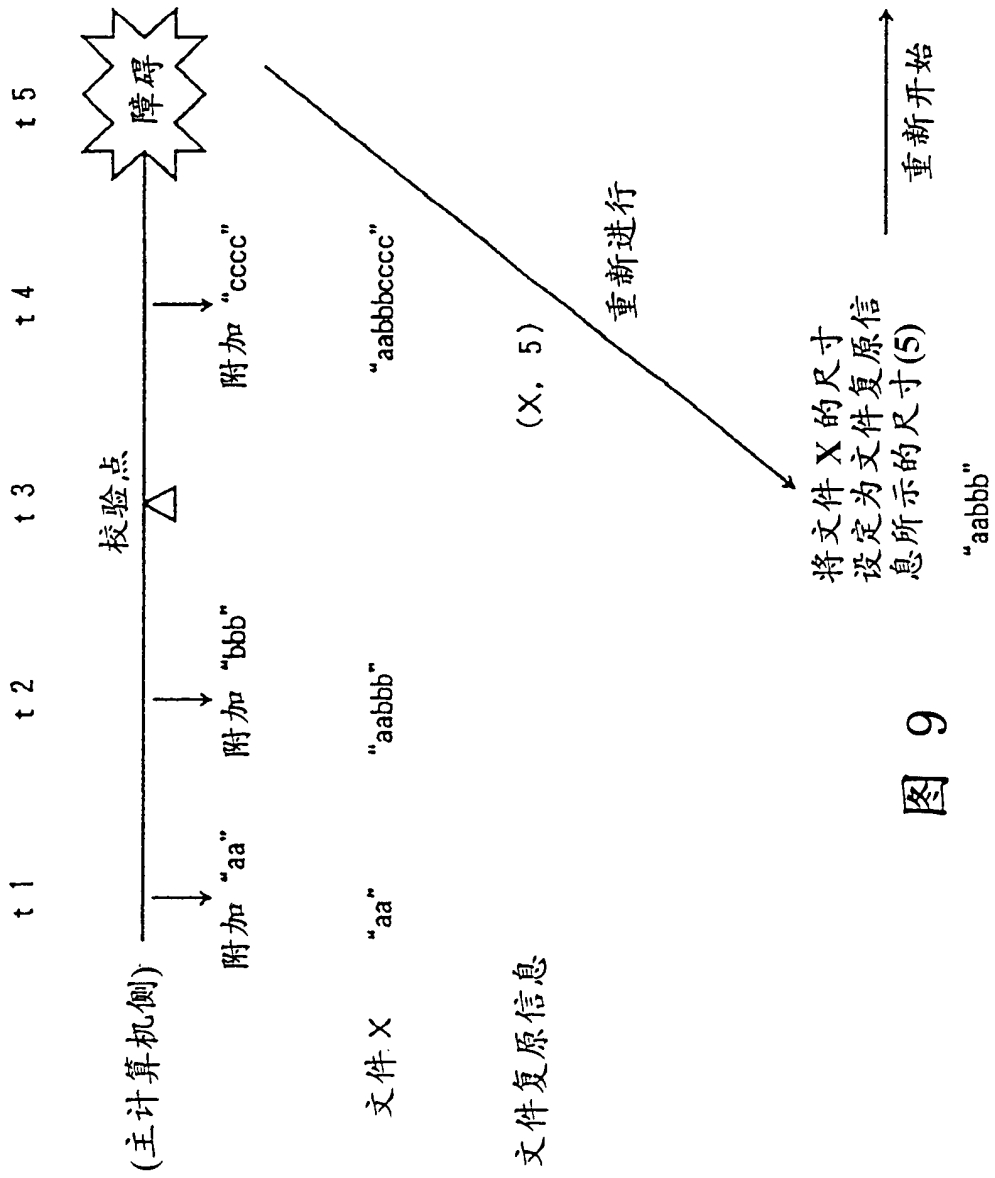


图 9

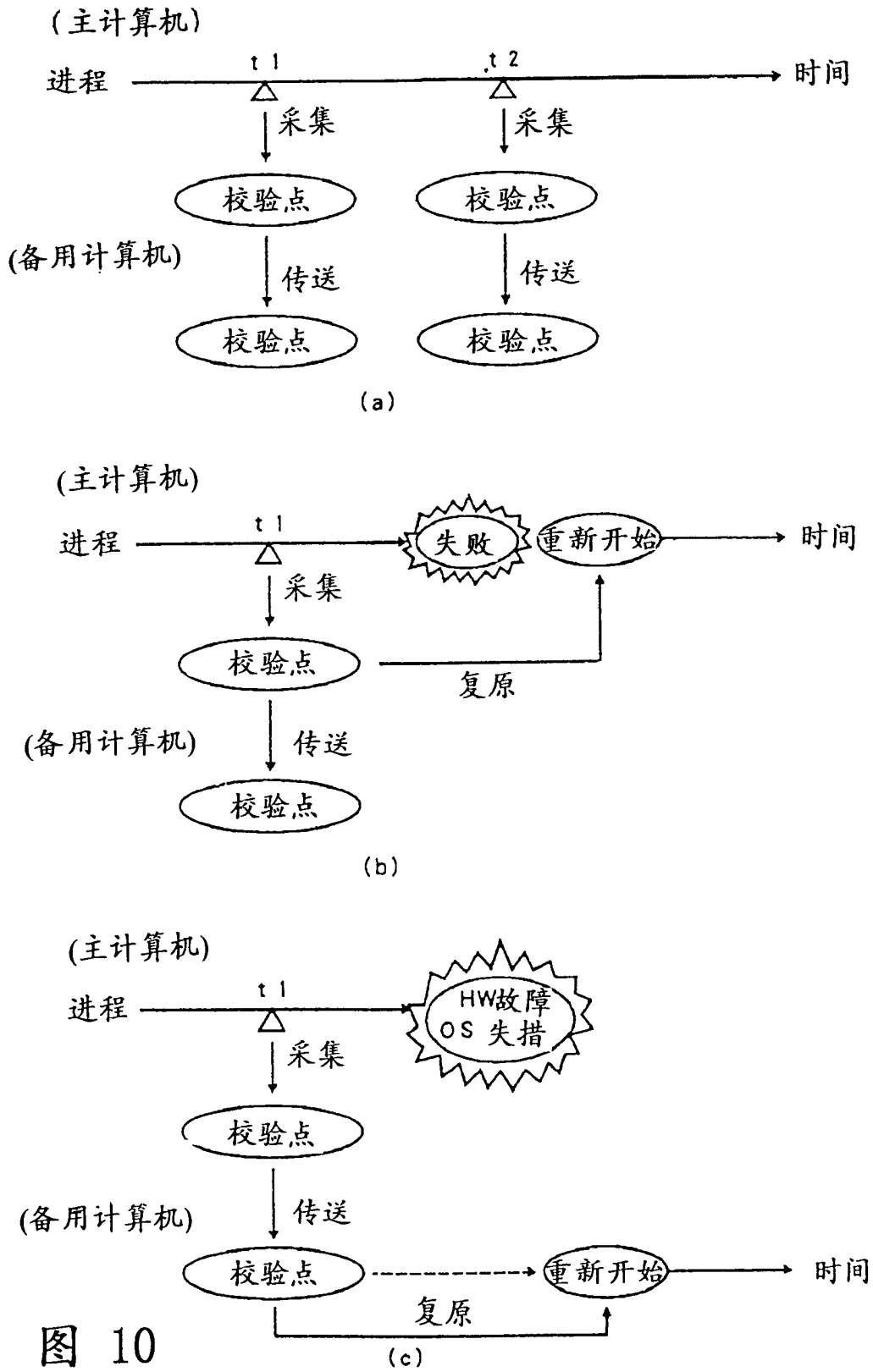


图 10

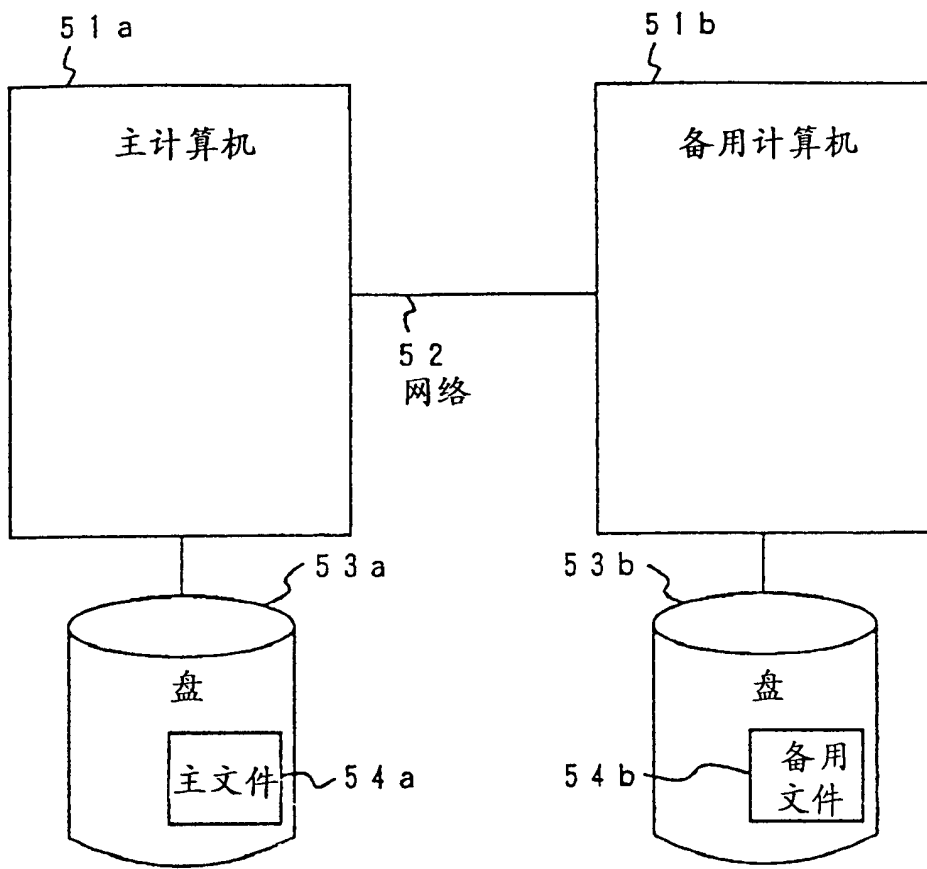


图 11

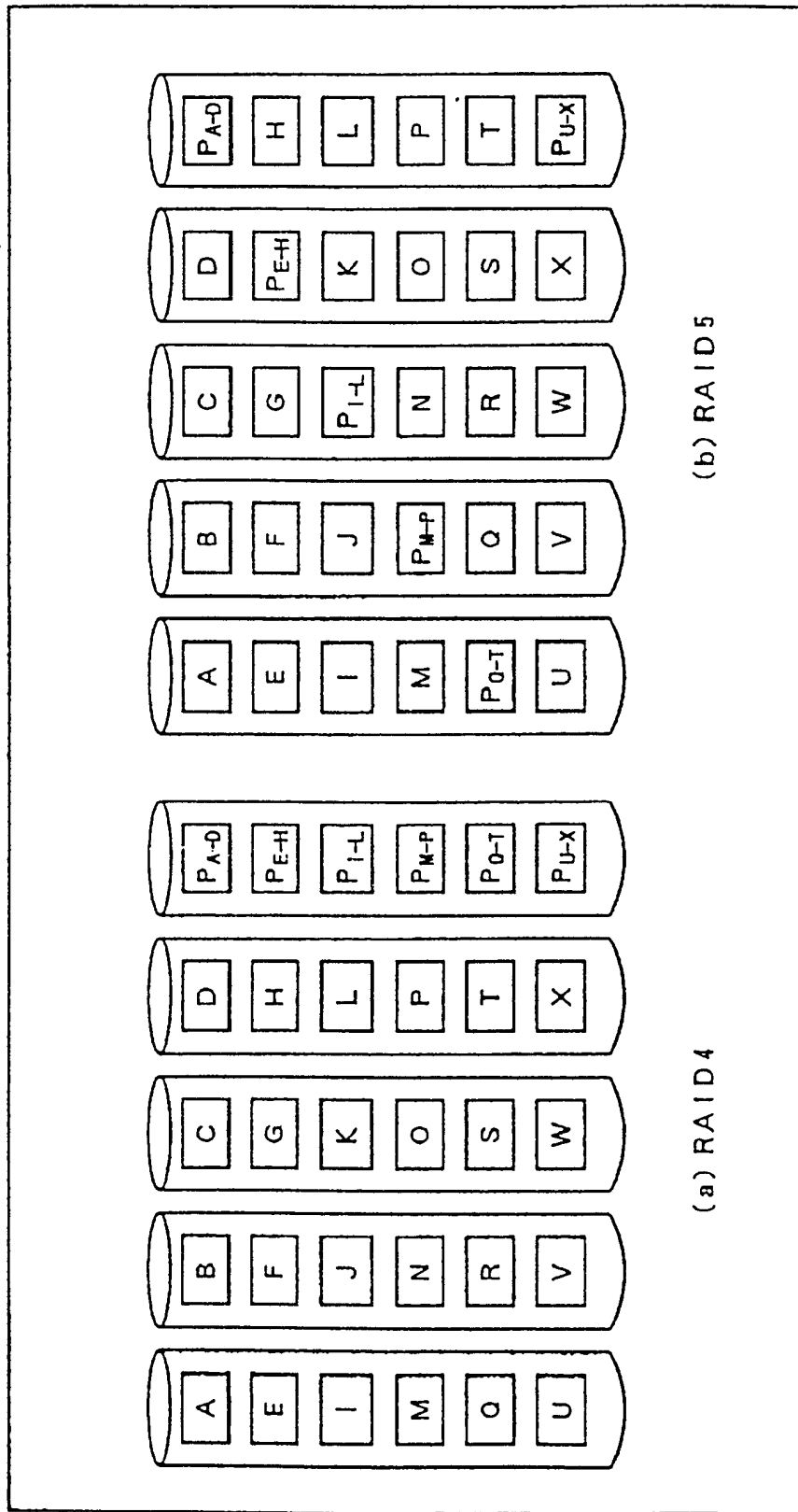


图 12

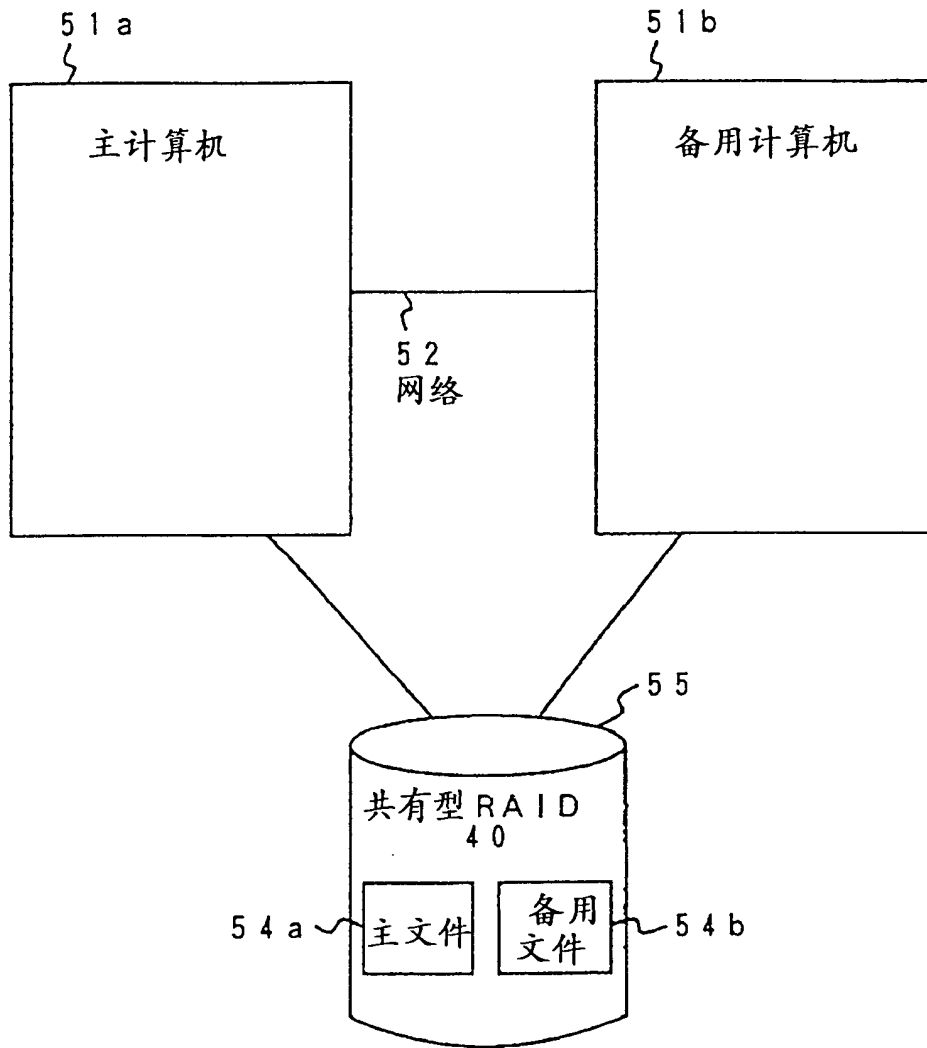


图 13