

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

G06F 11/07 (2006.01)

H04L 29/06 (2006.01)



[12] 发明专利说明书

专利号 ZL 200410085894.5

[45] 授权公告日 2009 年 11 月 4 日

[11] 授权公告号 CN 100557574C

[22] 申请日 2004.11.5

[21] 申请号 200410085894.5

[30] 优先权

[32] 2003.11.6 [33] US [31] 10/702,863

[73] 专利权人 微软公司

地址 美国华盛顿州

[72] 发明人 刘 海 L·N·安托诺夫

[56] 参考文献

CN1213800A 1999.4.14

JP10-11270A 1998.1.16

US4641274 1987.2.3

审查员 范 丹

[74] 专利代理机构 上海专利商标事务所有限公司

代理人 陈 斌

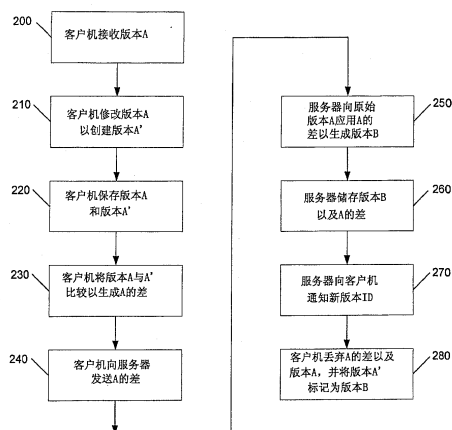
权利要求书 5 页 说明书 12 页 附图 6 页

[54] 发明名称

使用二进制比较优化文件复制

[57] 摘要

当对文件作出改变时，将文件的基于客户机和基于服务器的版本保持同步。将数据与对客户机和服务器都已知的前一版本相比较，并生成两者之间的差异的高度压缩的表示。差异，或“差”然后被发送，并可使用 HTTP(超文本传输协议)协议的扩充。



1. 一种维护更新的文件的方法，其特征在于，它包括：

在第一装置上储存一基础文件的第一副本和第二副本，其中所述第一副本和所述第二副本是相同的；

在所述第一装置上接收对所述第一副本的改变；

确定所述经改变的第一副本和所述第二副本之间的第一差异；

向服务器发送所述第一差异；以及

如果所述服务器上的基础文件的最近版本与储存在所述第一装置上的基础文件相同，则在所述服务器上接受所述第一差异，否则，在服务器上拒绝所述第一差异。

2. 如权利要求1所述的方法，其特征在于，它还包括在所述第一装置上从所述服务器接收所述文件的第一副本，并在所述第一装置上储存所述文件的第一副本和第二副本之前，在所述第一装置上制作所述文件的第二副本。

3. 如权利要求1所述的方法，其特征在于，确定所述第一差异包括使用所述经改变的第一副本和所述第二副本之间的二进制比较。

4. 如权利要求1所述的方法，其特征在于，所述第一差异是二进制差。

5. 如权利要求1所述的方法，其特征在于，如果在所述服务器上拒绝所述差异，则确定在所述服务器上的基础文件的最近版本与储存在所述第一装置上的基础文件之间的一第二差异，并从所述服务器向所述第一装置发送所述第二差异，以及向储存在所述第一装置上的基础文件的第二副本应用所述第二差异。

6. 如权利要求1所述的方法，其特征在于，它还包括确定所述服务器上的基础文件的最近版本是否与储存在所述第一装置上的基础文件相同。

7. 如权利要求1所述的方法，其特征在于，发送所述第一差异包括使用 HTTP 协议的扩充。

8. 如权利要求1所述的方法，其特征在于，它还包括：

在第二装置上储存所述基础文件的第一副本和第二副本；

在所述第二装置上接收对所述第一副本的改变；

确定在所述第二装置上的所述经改变的第一副本和所述第二副本之间的第二差异；

向所述服务器发送所述第二差异；以及

如果所述服务器上的基础文件的最近版本与储存在所述第二装置上的基础文件相同，则在服务器上接受所述第二差异，否则，在服务器上拒绝所述第二差异。

9. 如权利要求 8 所述的方法，其特征在于，它还包括，如果在所述服务器上拒绝所述第二差异，则向所述第二装置发送第三差异，并向所述第二装置上的基础文件的第二副本应用所述第三差异以生成经改变的第二副本，其中所述第三差异是在所述服务器上的基础文件的最近版本与在所述第二装置上的基础文件之间的差异，并且确定在所述第二装置上的所述经改变的第一副本和所述经改变的第二副本之间的一第四差异，并将所述第四差异发送到所述服务器。

10. 如权利要求 8 所述的方法，其特征在于，它还包括在所述第二装置上从所述服务器接收所述文件的第一副本，并在所述第二装置上储存所述文件的第一副本和第二副本之前，在所述第二装置上制作所述文件的第二副本。

11. 如权利要求 8 所述的方法，其特征在于，确定所述第二差异包括使用所述第二装置上所述经改变的第一副本和所述第二副本之间的二进制比较。

12. 如权利要求 8 所述的方法，其特征在于，所述第二差异是二进制差。

13. 如权利要求 8 所述的方法，其特征在于，它还包括确定所述服务器上的基础文件的最近版本是否与储存在所述第二装置上的基础文件相同。

14. 如权利要求 8 所述的方法，其特征在于，发送所述第二差异包括使用 HTTP 协议的扩充。

15. 一种用于执行维护更新的文件的系统，其特征在于，所述系统包括：

用于在第一装置上储存一基础文件的第一副本和第二副本的装置，其中所述第一副本和所述第二副本相同；

用于在所述第一装置上接收对所述第一副本的改变的装置；

用于确定所述经改变的第一副本和所述第二副本之间的第一差异的装置；

用于向服务器发送所述第一差异的装置；以及

用于如果所述服务器上的基础文件的最近版本与储存在所述第一装置上的基础文件相同，则在所述服务器上接受所述第一差异，否则，在所述服务器上拒绝所述第一差异的装置。

16. 如权利要求 15 所述的系统，其特征在于，它还包括用于在所述第一装置上从所述服务器接收所述文件的第一副本，并在所述第一装置上储存所述文件的第

一副本和第二副本之前，在所述第一装置上制作所述文件的第二副本的装置。

17. 如权利要求 15 所述的系统，其特征在于，用于确定所述第一差异的装置包括用于使用所述经改变的第一副本和所述第二副本之间的二进制比较的装置。

18. 如权利要求 15 所述的系统，其特征在于，所述第一差异是二进制差。

19. 如权利要求 15 所述的系统，其特征在于，它还包括用于如果在所述服务器上拒绝所述差异，则确定在所述服务器上的基础文件的最近版本与储存在所述第一装置上的基础文件之间的一第二差异，并从所述服务器向所述第一装置发送所述第二差异，以及向储存在所述第一装置上的基础文件的第二副本应用所述第二差异的装置。

20. 如权利要求 15 所述的系统，其特征在于，它还包括用于确定所述服务器上的基础文件的最近版本是否与储存在所述第一装置上的基础文件相同的装置。

21. 如权利要求 15 所述的系统，其特征在于，用于发送所述第一差异的装置包括用于使用 HTTP 协议的扩充的装置。

22. 如权利要求 15 所述的系统，其特征在于，它还包括：

用于在第二装置上储存所述基础文件的第一和第二副本的装置；

用于在所述第二装置上接收对所述第一副本的改变的装置；

用于确定在所述第二装置上的所述经改变的第一副本和所述第二副本之间的第二差异的装置；

用于向所述服务器发送所述第二差异的装置；以及

用于如果所述服务器上的基础文件的最近版本与储存在所述第二装置上的基础文件相同，则在所述服务器上接受所述第二差异，否则，在所述服务器上拒绝所述第二差异的装置。

23. 如权利要求 22 所述的系统，其特征在于，它还包括用于如果在所述服务器上拒绝所述第二差异，则向所述第二装置发送第三差异，并向储存在所述第二装置上的基础文件的第二副本应用所述第三差异以生成经改变的第二副本的装置，其中所述第三差异是在所述服务器上的基础文件与在所述第二装置上的基础文件之间的差异，并且所述系统还包括用于确定在所述第二装置上的所述经改变的第一副本和所述经改变的第二副本之间的一第四差异，并将所述第四差异发送到所述服务器的装置。

24. 如权利要求 22 所述的系统，其特征在于，它还包括用于在所述第二装置

上从所述服务器接收所述文件的第一副本,并在所述第二装置上储存所述文件的第一副本和第二副本之前,在所述第二装置上制作所述文件的第二副本的装置。

25. 如权利要求 22 所述的系统,其特征在于,用于确定所述第二差异的装置包括用于使用所述第二装置上的所述改变的第一副本和所述第二副本之间的二进制比较的装置。

26. 如权利要求 22 所述的系统,其特征在于,所述第二差异是二进制差。

27. 如权利要求 22 所述的系统,其特征在于,它还包括用于确定所述服务器上的基础文件的最近版本是否与储存在所述第二装置上的基础文件相同的装置。

28. 如权利要求 22 所述的系统,其特征在于,用于发送所述第二差异的装置包括用于使用 HTTP 协议的扩充的装置。

29. 一种在文件同步系统中使用的系统,其特征在于,它包括:

一存储装置,用于储存一基础文件的第一副本和第二副本,其中所述基础文件驻留在服务器上,其中所述第一副本和所述第二副本相同;

一输入装置,用于接收对所述基础文件的第一副本的改变;

一处理器,用于基于所接收的改变来改变所述第一副本,并确定所述经改变的第一副本和所述第二副本之间的差异;以及

一输出装置,用于向所述服务器发送所述差异;

并且其中如果所述服务器上的基础文件的最近版本与储存在所述存储装置上的基础文件相同,则在所述服务器上接受所述差异,否则,在所述服务器上拒绝所述差异。

30. 如权利要求 29 所述的系统,其特征在于,所述差异是二进制差异。

31. 如权利要求 29 所述的系统,其特征在于,所述差异使用 HTTP 协议的扩充来发送到所述服务器。

32. 如权利要求 29 所述的系统,其特征在于,它还包括一用于接收在所述服务器上的基础文件的最近版本与储存在所述存储装置上的基础文件之间的第二差异的装置,其中,所述处理器向所述基础文件的第二副本应用所述第二差异以生成一修改的第二副本,并向所述修改的第二副本应用所述所接收的改变以生成第三副本,并确定所述第三副本和所述经修改的第二副本之间的第三差异。

33. 如权利要求 32 所述的系统,其特征在于,它还包括一用于向服务器发送所述第三差异的输出装置。

34. 一种在文件同步系统中使用的系统，其特征在于，它包括：

一存储装置，用于储存一基础文件；

一输入装置，用于从客户机装置接收第一差异，其中所述第一差异是对于在第一设备上的基础文件的第一副本的改变与在所述第一设备上的基础文件的第二副本之间的差别，其中所述第一副本和所述第二副本相同；以及

一处理器，用于通过确定所述基础文件与所述客户机装置用于生成所述第一差异的客户机基础文件是否相同来确定所述第一差异是否能应用到所储存的基础文件，并且如果能，则向所述基础文件应用所述第一差异，否则，向所述客户机装置指示所述第一差异被拒绝。

35. 如权利要求 34 所述的系统，其特征在于，所述第一差异是二进制差异。

36. 如权利要求 34 所述的系统，其特征在于，所述第一差异以 HTTP 协议的扩充的形式接收。

37. 如权利要求 34 所述的系统，其特征在于，如果所述第一差异被所述处理器拒绝，则从所述存储装置检索在所储存的基础文件与所述客户机装置用于生成所述第一差异的客户机基础文件之间的第二差异，并通过一输出装置向所述客户机装置发送所述第二差异。

38. 如权利要求 37 所述的系统，其特征在于，所述第二差异从第二客户机装置接收。

使用二进制比较优化文件复制

技术领域

本发明一般涉及计算机文件复制领域，尤其涉及使用二进制比较的文件复制。

背景技术

复制通过令文件或文档的客户机和服务器版本保持同步而使本地和远程访问数据得以进行。尽管这一功能性对应用程序是重要且关键的，然而鉴于必须在客户机和服务器之间传输的数据量，它也是高成本的。整个文件和文档连同其各种版本一起储存在服务器上，并在服务器及其客户机之间传输。由此，许多复制系统试图通过在发送之前压缩数据来节省带宽。然而，这一传统的压缩形式为整个文件编码数据，即使该数据的大部分是作为前一版本的一部分曾传输过的。由此，文件或文档中的细小变化仍需要压缩并传输整个文件或文档，尽管大多数数据已以较早接收到的版本的形式驻留在目标地上。

鉴于上述原因，需要一种克服现有技术的局限和缺点的系统和方法。

发明内容

本发明提供一种用于当对文件作出改变时保持文件的基于客户机和基于服务器的副本同步的机制。将数据与对客户机和服务器都已知的前一版本相比较，并生成两者之间的差异的高度压缩表示。

依照一个实施例，在客户机处接收并储存一基础文件的第一副本和第二副本。这两个副本是相同的一客户机接收一个副本并储存该副本的两个实例。客户机然后对第一副本作出改变，然后确定改变的第一副本和第二副本之间的差异（如，二进制差异）。差异被发送到维护该基础文件的服务器。如果服务器上的基础文件与储存在第一装置上的基础文件相同，则服务器接受该差异；否则，服务器拒绝该差异。

依照本发明的各方面，如果服务器拒绝差异，则服务器向客户机发送第二差异。客户机然后向储存在第一装置处的基础文件的第二副本应用该第二差异。这令客户机的基础文件变得对于驻留在服务器上的基础文件为最新。客户机然后可对该

更新的基础文件作出改变，生成一新差异，然后向服务器发送该新差异。

结合附图阅读以下说明性实施例的详细描述，可以清楚本发明的另外的特征和优点。

附图说明

当结合附图阅读时，可以更好地理解以上概述以及以下较佳实施例的详细描述。为说明本发明的目的，附图中示出了本发明的示例性构造；然而，本发明不限于所揭示的具体方法和手段。附图中：

图 1 所示是可实现本发明的各方面的示例性计算环境的框图；

图 2 是依照本发明维护更新的文件的示例性方法的流程图；

图 3 是依照本发明维护更新的文件的另一示例性方法的流程图；

图 4 示出了有助于描述本发明的各方面的示例性系统的框图；以及

图 5 和 6 是依照本发明维护更新的文件的另一示例性方法的流程图。

具体实施方式

综述

本发明针对当对文件作出改变时保持文件的本地（也称为“客户机”）副本和基于服务器的副本同步。本发明描述的示例性系统和方法比现有技术更有效，并将带宽需求和执行同步的时间元素都最小化。

将数据与对客户机和服务器都已知的前一版本相比较，并生成两者之间的差异的高度压缩表示。这些差异，或“差”然后可被发送，并可使用对 HTTP（超文本传输协议）协议的扩充。

示例性计算环境

图 1 示出了可在其中实现本发明的合适的计算系统环境 100 的一个示例。计算系统环境 100 仅为合适的计算环境的一个示例，并非建议对本发明的使用或功能的范围的局限。也不应将计算环境 100 解释为对示例性操作环境 100 中示出的任一组件或其组合具有依赖或需求。

本发明可以使用众多其它通用或专用计算系统环境或配置来操作。适合使用本发明的众所周知的计算系统、环境和/或配置包括但不限于：个人计算机、服务

器计算机、手持式或膝上设备、多处理器系统、基于微处理器的系统、机顶盒、可编程消费者电子设备、网络 PC、小型机、大型机、包括任一上述系统或设备的分布式计算环境等等。

本发明可在诸如由网络环境中的计算机执行的程序模块等计算机可执行指令的一般上下文环境中描述。一般而言，程序模块包括例程、程序、对象、组件、数据结构等等，执行特定的任务或实现特定的抽象数据类型。本发明也可以在分布式计算环境中实践，其中，任务由通过通信网络连接的远程处理设备来执行。在分布式计算环境中，程序模块可以位于本地和远程计算机存储媒质中，包括存储器存储设备。

参考图 1，用于实现本发明的示例性系统包括以计算机 110 形式的通用计算装置。计算机 110 的组件包括但不限于，处理单元 120、系统存储器 130 以及将包括系统存储器的各类系统组件耦合至处理单元 120 的系统总线 121。系统总线 121 可以是若干种总线结构类型的任一种，包括存储器总线或存储器控制器、外围总线以及使用各类总线体系结构的局部总线。作为示例而非局限，这类体系结构包括工业标准体系结构（ISA）总线、微通道体系结构（MCA）总线、增强 ISA（EISA）总线、视频电子技术标准协会（VESA）局部总线以及外围部件互连（PCI）总线（也称为 Mezzanine 总线）。

计算机 110 通常包括各种计算机可读媒质。计算机可读媒质可以是可由计算机 110 访问的任一可用媒质，包括易失和非易失媒质、可移动和不可移动媒质。作为示例而非局限，计算机可读媒质包括计算机存储媒质和通信媒质。计算机存储媒质包括以用于储存诸如计算机可读指令、数据结构、程序模块或其它数据等信息的任一方法或技术实现的易失和非易失，可移动和不可移动媒质。计算机存储媒质包括但不限于，RAM、ROM、EEPROM、闪存或其它存储器技术、CD-ROM、数字多功能盘（DVD）或其它光盘存储、磁盒、磁带、磁盘存储或其它磁存储设备、或可以用来储存所期望的信息并可由计算机 110 访问的任一其它媒质。通信媒质通常在诸如载波或其它传输机制的已调制数据信号中包含计算机可读指令、数据结构、程序模块或其它数据，并包括任一信息传送媒质。术语“已调制数据信号”指以对信号中的信息进行编码的方式设置或改变其一个或多个特征的信号。作为示例而非局限，通信媒质包括有线媒质，如有线网络或直接连线连接，以及无线媒质，如声学、RF、红外和其它无线媒质。上述任一的组合也应当包括在计算机可读媒

质的范围之内。

系统存储器 130 包括以易失和/或非易失存储器形式的计算机存储媒质，如只读存储器（ROM）131 和随机存取存储器（RAM）132。基本输入/输出系统 133（BIOS）包括如在启动时帮助在计算机 110 内的元件之间传输信息的基本例程，通常储存在 ROM 131 中。RAM 132 通常包含处理单元 120 立即可访问或者当前正在操作的数据和/或程序模块。作为示例而非局限，图 1 示出了操作系统 134、应用程序 135、其它程序模块 136 和程序数据 137。

计算机 110 也可包括其它可移动/不可移动、易失/非易失计算机存储媒质。仅作示例，图 1 示出了对不可移动、非易失磁媒质进行读写的硬盘驱动器 141、对可移动、非易失磁盘 152 进行读写的磁盘驱动器 151 以及对可移动、非易失光盘 156，如 CD ROM 或其它光媒质进行读写的光盘驱动器 155。可以在示例性操作环境中使用的其它可移动/不可移动、易失/非易失计算机存储媒质包括但不限于，磁带盒、闪存卡、数字多功能盘、数字视频带、固态 RAM、固态 ROM 等等。硬盘驱动器 141 通常通过不可移动存储器接口，如接口 140 连接到系统总线 121，磁盘驱动器 151 和光盘驱动器 155 通常通过可移动存储器接口，如接口 150 连接到系统总线 121。

图 1 讨论并示出的驱动器及其关联的计算机存储媒质为计算机 110 提供了计算机可读指令、数据结构、程序模块和其它数据的存储。例如，在图 1 中，示出硬盘驱动器 141 储存操作系统 144、应用程序 145、其它程序模块 146 和程序数据 147。注意，这些组件可以与操作系统 134、应用程序 135、其它程序模块 136 和程序数据 137 相同，也可以与它们不同。这里对操作系统 144、应用程序 145、其它程序模块 146 和程序数据 147 给予不同的标号来说明至少它们是不同的副本。用户可以通过输入设备，如键盘 162 和定位设备 161（通常指鼠标、跟踪球或触模板）向计算机 110 输入命令和信息。其它输入设备（未示出）可包括麦克风、操纵杆、游戏垫、圆盘式卫星天线、扫描仪等等。这些和其它输入设备通常通过耦合至系统总线的用户输入接口 160 连接至处理单元 120，但是也可以通过其它接口和总线结构连接，如并行端口、游戏端口或通用串行总线（USB）。监视器 191 或其它类型的显示设备也通过接口，如视频接口 190 连接至系统总线 121。除监视器之外，计算机也包括其它外围输出设备，如扬声器 197 和打印机 196，通过输出外围接口 195 连接。

计算机 110 可以在使用到一个或多个远程计算机, 如远程计算机 180 的逻辑连接的网络化环境中操作。远程计算机 180 可以是个人计算机、服务器、路由器、网络 PC、对等设备或其它公用网络节点, 并通常包括许多或所有上述与计算机 110 相关的元件, 尽管在图 1 中仅示出了存储器存储设备 181。图 1 描述的逻辑连接包括局域网 (LAN) 171 和广域网 (WAN) 173, 但也可包括其它网络。这类网络环境常见于办公室、企业范围计算机网络、内联网以及因特网。

当在 LAN 网络环境中使用时, 计算机 110 通过网络接口或适配器 170 连接至 LAN 171。当在 WAN 网络环境中使用时, 计算机 110 通常包括调制解调器 172 或其它装置, 用于通过 WAN 173, 如因特网建立通信。调制解调器 172 可以是内置或外置的, 通过用户输入接口 160 或其它合适的机制连接至系统总线 121。在网络化环境中, 描述的与计算机 110 相关的程序模块或其部分可储存在远程存储器存储设备中。作为示例而非局限, 图 1 示出了远程应用程序 185 驻留在存储器设备 181 中。可以理解, 示出的网络连接是示例性的, 也可以使用在计算机之间建立通信链路的其它装置。

示例性分布式计算框架或体系结构

鉴于个人计算活动与因特网的交汇, 已经发展并且正在发展各种分布式计算框架。个人和商业用户同样地拥有用于应用程序和计算设备的无缝的互操作和 web 激活的接口, 使得计算活动越来越面向 web 浏览器和网络。

例如, MICROSOFT®的.NET 平台包括服务器、构件服务如基于 web 的数据存储、以及可下载设备软件。通常来说, .NET 平台提供 (1) 令整个范围的计算设备共同工作并在所有设备上自动更新并同步化用户信息的能力, (2) 提高的 web 站点交互能力, 通过大量使用 XML 而不是 HTML 来实现, (3) 从用于各种应用, 如电子邮件, 或软件, 如 Office.NET 的管理的中央起点到用户的具有产品和服务的定制访问和传送的特点的在线服务, (4) 中央化数据存储, 将增加对信息访问以及用户和设备间的信息同步的效率和简易性, (5) 整合各种通信媒质, 如电子邮件、传真和电话的能力, (6) 对开发员来说, 创建可重复使用模块的能力, 借此提高生产力并降低编程错误数, 以及 (7) 还有其它跨平台整合特征。

尽管这里的示例性实施例是结合驻留在计算设备上的软件描述的, 也可以通过操作系统、API 或协处理器和请求对象之间的“中间人”对象来实现本发明的一

个或多个部分，使服务可以由所有.NET 的语言和服务、以及在其它分布式计算框架中来完成、支持或访问。

示例性实施例

图 2 是依照本发明维护更新的文件的示例性方法的流程图。在本示例性实施例中，客户机修改文件并将变化上传到服务器。在步骤 200，客户机接收储存在服务器上的基础文件的最近版本（“版本 A”）的副本。在步骤 210，客户机对版本 A 作出其改变以创建版本 A'。在步骤 220，客户机保存原始版本 A 的副本和新版本 A'。由此，即使用户更新该文件，客户机也维护最后一个已知的服务器状态的副本。预期版本 A 的副本可以在步骤 210 之前或之后储存在客户机上。

然后，在步骤 230，通过比较版本 A 和 A'生成差异，或“差”。差是一种将文件的两个版本进行比较以生成可应用到较旧的文件来生成较新文件的压缩差的机制。可通过对本领域的技术人员已知的用于确定基础形式和修改形式之间的差异的任一方法、技术或系统来执行差异生成。生成的一种较佳的差异是二进制差异。文件被认为是一系列字节。常规的压缩算法被用于通过计算阴影或基础副本与修正的副本之间的差异来生成二进制差异。该差异然后被发送到服务器，服务器拒绝差异或接受差异。如果服务器上的基础已改变，则出现拒绝，在此情况下，该差异对服务器无用。预期依照本发明可使用任一差异引擎或技术。在本发明中，为示例性目的提供了二进制差异技术的使用。

具体地，在步骤 240，客户机向服务器发送该差。在检查以确保其基础文件的最近版本并未从客户机用以作出修改的版本 A 有所改变之后，在步骤 250，服务器向版本 A 应用该差，以生成该文件的新的、最近的版本—版本 B。服务器执行的版本检查在后文参考图 3-6 进一步描述。

在步骤 260，服务器储存新版本 B 以及客户机提供的差（可任选）。新版本 B 被认为是基础文件的最近更新，并保留该差以由可对原始版本 A 作出改变的其它客户机使用，这在后文参考图 3-6 进一步描述。服务器可任选地储存该差，以向其它客户机提供优化的更新。如果作出了多个修订，则需要获取从较旧版本到最新版本的多个差。注意，该差也可对不打算作出改变，但是希望读取最近的版本 B 并已具有版本 A 的客户机有用。

在步骤 270，服务器向客户机通知新版本标识符（如，“版本 B”）。客户机

然后丢弃它在步骤 230 所确定的差以及它所储存的版本 A, 并且在步骤 280 用新版本标识符标记其版本 A'。由此, 客户机将版本 A'更名为版本 B。

图 3 是维护更新的文件的另一示例性方法的流程图。在本示例中, 服务器向客户机提供差文件形式的最近变化。在步骤 300, 具有文件的版本 A 的客户机请求文件更新。例如, 客户机因为期望对该文件的最近版本作出改变而可能作出这一请求。客户机向服务器通知它具有版本 A, 并且作为响应, 在步骤 310, 服务器返回该差。服务器可能在维护的是来自前一客户机更新(如, 图 2 的步骤 260)的 A 的差异。在步骤 320, 客户机向其储存的版本 A 应用 A 的差以生成该文件的最近版本(如, “版本 B”)。

图 4 示出了有助于描述本发明的各方面的示例性系统的框图, 图 5 和 6 示出了当两个用户对同一基础文件作出改变时维护更新的文件的示例性方法的流程图。在本示例中, 假定服务器 400 维护基础文件(版本 A), 两个客户机 410、420(此处分别称为客户机 1 和客户机 2)都期望对同一基础文件作出改变。

在步骤 500, 客户机 1 和客户机 2 都从服务器 400 请求并接收最近版本(“版本 A”) (即, 客户机 1 和 2 下载基础文件)。预期客户机 1 和 2 可并发地或在时间上顺序地对基础文件作出改变。然而, 仅一个客户机将首先把其对原始基础文件的改变送到服务器。这些改变被应用到原始基础文件。由此, 向服务器发送差异的第一客户机令服务器接受该差异。基于基础文件的随后的客户机差异将被服务器拒绝。因此, 稍后发送其对原始基础文件的改变的客户机必须首先接收更新的基础文件, 然后对该更新的基础文件作出改变, 如后文更详细描述。

假定客户机 1 首先作出改变, 方法按似于图 2 中所描述的步骤 200 到 260 前进。即, 在步骤 505, 客户机 1 对版本 A 作出其改变以创建版本 A'。在步骤 510, 客户机 1 保存原始版本 A 的副本和新版本 A'。预期可在步骤 505 之前或之后在客户机 1 上储存版本 A 的副本。然后在步骤 515, 通过比较版本 A 和 A'生成差(较佳地为二进制差)。

在步骤 520, 客户机 1 向服务器 400 发送该差。当向服务器同步回改变时, 客户机核实服务器支持差机制, 然后将“差”连同指定原始文件的版本的版本信息一起上传。在检查以确保其基础文件的最近版本未被客户机用于作出修改的版本 A 上进行改变之后, 在步骤 525, 服务器 400 应用客户机 1 提供的版本 A 的差以生成该文件的新的、最近的版本一版本 B。

在步骤 530, 服务器储存该新版本 B 以及客户机 1 提供的差。新版本 B 被认为是基础文件的最近更新, 并保留差以由可对原始版本 A 作出改变的其它客户机 (如, 客户机 2) 使用。

类似于步骤 270 和 280, 尽管在图 5 中未示出, 服务器 400 向客户机 1 通知新版本标识符 (如, “版本 B”)。客户机 1 然后丢弃它所确定的差以及它所储存的版本 A, 并用新版本标识符标记其版本 A'。由此, 客户机将版本 A' 更名为版本 B。

同时, 在步骤 535, 客户机 2 修改它所接收到的原始基础文件版本 A 以创建新版本—版本 A"。在步骤 540, 客户机 2 保存原始版本 A 的副本和新版本 A"。预期可在步骤 540 之前或之后在客户机 2 上储存版本 A 的副本。然后在步骤 545, 通过比较版本 A 和 A" 生成差。

在 550, 客户机 2 向服务器 400 发送其版本 A 的差。服务器 400 检查它所储存的基础文件是否在客户机 2 用作客户机 2 的修改的基础的基础文件上做了改变。

如果基础文件的服务器状态被高速缓存到与请求方客户机关联的本地存储中, 则在本地存储中的文件和服务器上对应的文件状态之间作出比较。可进行该比较以确定储存在本地存储中的文件的副本是否为最近版本, 或者在服务器上是否存在更近的版本。换言之, 该比较着眼于自从请求方客户机获取该文件的副本以来, 另一客户机修改并更新了所请求的文件的可能性。应当注意, 期望该比较包括表示文件状态的标识符的传输, 而无需在客户机和服务器之间传输整个文件。以这一方式, 该比较减少了否则将需要的网络通信量, 并避免将文件的同一版本发送一次以上。

由此, 在步骤 555, 服务器检查以确保基础文件的最近版本未在客户机 2 用于作出修改的版本 A 上做过改变。如果基础文件未改变, 则在步骤 590, 服务器 400 向服务器储存的基础文件应用客户机 2 提供的差以生成该文件的新的、最近版本, 在步骤 595, 服务器储存该版本以及客户机 2 提供的差。服务器 400 向客户机 2 通知新版本标识符, 客户机 2 然后可丢弃它所确定的差以及它所储存的版本 A, 并用新版本标识符来标记其修改的版本 A"。

然而, 在本示例中, 客户机 1 已向服务器 400 提供了改变, 因此储存在服务器上的基础文件变为了版本 B。客户机 2 没有版本 B 的副本, 并对文件的版本 A 作出修改。因此, 由于基础文件已改变, 在步骤 560, 服务器 400 拒绝客户机 2 提供的差, 并向客户机 2 发送客户机 1 提供的、服务器 400 较早接收并储存 (步骤

530) 的版本 A 的差。

在步骤 565, 客户机 2 向它所储存的版本 A 应用客户机 1 提供的差, 以获取该文件的最近的服务器储存版本 (这里为版本 B)。客户机 2 然后分别在步骤 570 和 575 确定最近版本与其修改的版本 A"之间的差, 并向服务器 400 发送该新的差。在步骤 580, 服务器 400 向其储存的最近版本 (版本 B) 应用该新的差, 以生成新的最近版本 (这里为版本 C)。在步骤 585, 服务器 400 储存该新的、最近版本以及新接收的差。类似于步骤 270 和 280, 尽管未在图 5 中示出, 服务器 400 向客户机 2 通知新版本标识符 (如, “版本 C”)。客户机 2 然后丢弃它所确定的差以及它所储存的版本, 并用新版本标识符来标记其版本 A"。由此, 客户将版本 A"更名为版本 C。

与在服务器上自动保存修改的版本相反, 预期诸如管理员等用户可确定应当如何把改变集成。这可避免与前一用户改变的内容冲突。

注意, 可在服务器表明批准接受差之前或之后确定该差。由此, 为提高效率, 客户机可等到服务器已指示客户机已对服务器当前作为最近版本维护的基础文件的同一版本作出改变。仅在那时客户机才能确定差并向服务器提供该差。理想地, 服务器不计算差异, 而仅应用差异。

预期到一种情况, 一个客户机在另一客户机连接到服务器以提供其改变之前作出了多次上传。例如, 假定原始基础文件是版本 A。客户机 1 然后作出改变并且这些改变被接受为版本 B。如果客户机 1 作出进一步的改变, 并将它们提供给服务器, 则这一新的最近版本被保存为版本 C。理想地, 服务器保存版本 A 和 B 之间的差异, 以及版本 B 和 C 之间的差异。由此, 当另一客户机作出其改变时, 服务器较佳地将在同一消息中向该客户机发送版本 A 和 B 之间的差异, 以及版本 B 和 C 之间的差异。客户机然后重新创建版本 B 和版本 C, 并确定版本 C 和其改变之间的差异, 并向服务器提供该差异。

如果过时的客户机连接到服务器以获取最近版本, 它告诉服务器它具有哪一版本, 并且如果储存的差追溯到该版本, 则返回适当的一个或多个差以及当前版本 ID。较佳地, 服务器保留它在处理过程中接收的各种版本之间的所有差, 以容纳哪些仍希望对旧版本作出改变的客户机 (即, 对基础文件的旧版本“向后兼容”)。然而, 在某一点上, 服务器可删除或移除先前储存的它所保留的差。例如, 这些行动可由日期或存储容量提示。

理想地，使用超文本传输协议（HTTP）来传输差。具体地，可使用协议扩充来警告服务器正在传输“差”，或在消息内实现或结合进该差。

HTTP 已成为一种信息可借以在诸如因特网、内联网和外联网等兼容 TCP/IP（传输控制协议/因特网协议）的网络上传输的标准机制。HTTP 更特别地是一种用于分布式、协作、超媒体信息系统的应用程序级协议。它是一种一般性、无国界协议，可通过其请求方法、错误代码和头部的扩充，除其对超文本的使用之外还可用于许多任务，如名字服务器和分布式对象管理系统。由于信息依照其规范来传输，因此它被称为一种传输协议，而由于信息由作出对生成响应的服务器的请求的客户机来交换，因此它也被称为一种请求—响应协议。本发明所引用的 HTTP 一般涉及 HTTP 的任一标准，并可在网站 <http://www.w3.org> 上找到。

HTTP 的一种常见使用是依照标记语言格式化的信息的传输。例如，因特网的一种普遍应用是浏览其万维网页。在这些实例中，通常所检索到的信息是依照 HTTP 传输的超文本标记语言（HTML）格式。然而，也出现了其它标准标记语言。一个这样的标记语言是可扩展标记语言（XML）。XML 描述了一类被称为是 XML 文档的数据对象，并部分地描述了处理它们的计算机程序的行为。HTML 和 XML 之间的主要差异是，在前者内，信息内容与内容的布局相互缠结，例如，很难分离它们。相反，在 XML 内，与内容本身分离地维护内容的存储布局和逻辑结构的描述。然而，XML 和 HTML 都是被称为标准通用标记语言（SGML）的标记语言的派生。本发明所引用的 XML 一般涉及 XML 的任一标准，并可在网站 <http://www.w3.org> 上找到。

例如，为维护向后兼容性和互操作性，可使用 OPTIONS（选项）响应中的已扩充 HTTP 头部来允许客户机发现服务器支持二进制差。GET（获取）请求中的已扩充头部向服务器通知客户机接受该差。

客户机和服务器皆可选择不使用二进制差。在某些情况下，发送差的消息（如，HTTP 二进制差头部）可能大于文件本身。在这一情况下，更理想的是发送文档而非服务器和客户机之间的差。客户机可确定差的尺寸大于新文件。例如，如果新文件具有 0 字节，则会发生这一情况。服务器可决定丢弃差来节省存储空间。如果未使用差，则发送整个文件。为发信号通知何时发送差，客户机可用其 PUT（放置）请求发送已扩充头部，以指示正文中二进制差的存在以及从其生成差的基础文件的版本号。服务器用其 GET 响应发送已扩充头部，以指示二进制差链的存在、基础

文件的版本号以及链中差的数量。

当需要多个差来将客户机文件变为最新时，服务器可选择在单个回复中将差链接在一起，或如果差的总和大于新版本时，发送新版本本身。

计算并应用差的引擎较佳地在客户机和/或服务器内提供。较佳地也实现用于发现和差/版本管理的协议。已扩充 HTTP 头部允许客户机和服务器解释其生成差异的能力。例如，客户机用其 PUT 请求发送已扩充头部，以指示正文中二进制差的存在以及从其生成差的基础文件的版本号。服务器用其 GET 响应发送已扩充头部，以指示二进制差链的存在、基础文件的版本号以及链中差的数量。

服务器代码理想地能够管理差以应用（使用引擎）、储存并返回差，每一客户机较佳地具有维护服务器状态、生成差、向服务器发送差以及应用返回的差的能力。

复制在各种各样的应用程序广泛使用，但是这些系统的成本和性能永远都是一种挑战。本发明调节了复制的特性（已知的客户机/服务器状态）以在系统的效率中产生一个重要飞跃。本发明可应用到需复制例行更新的大文件的系统，诸如允许离线访问基于服务器的文档或文件的产品。

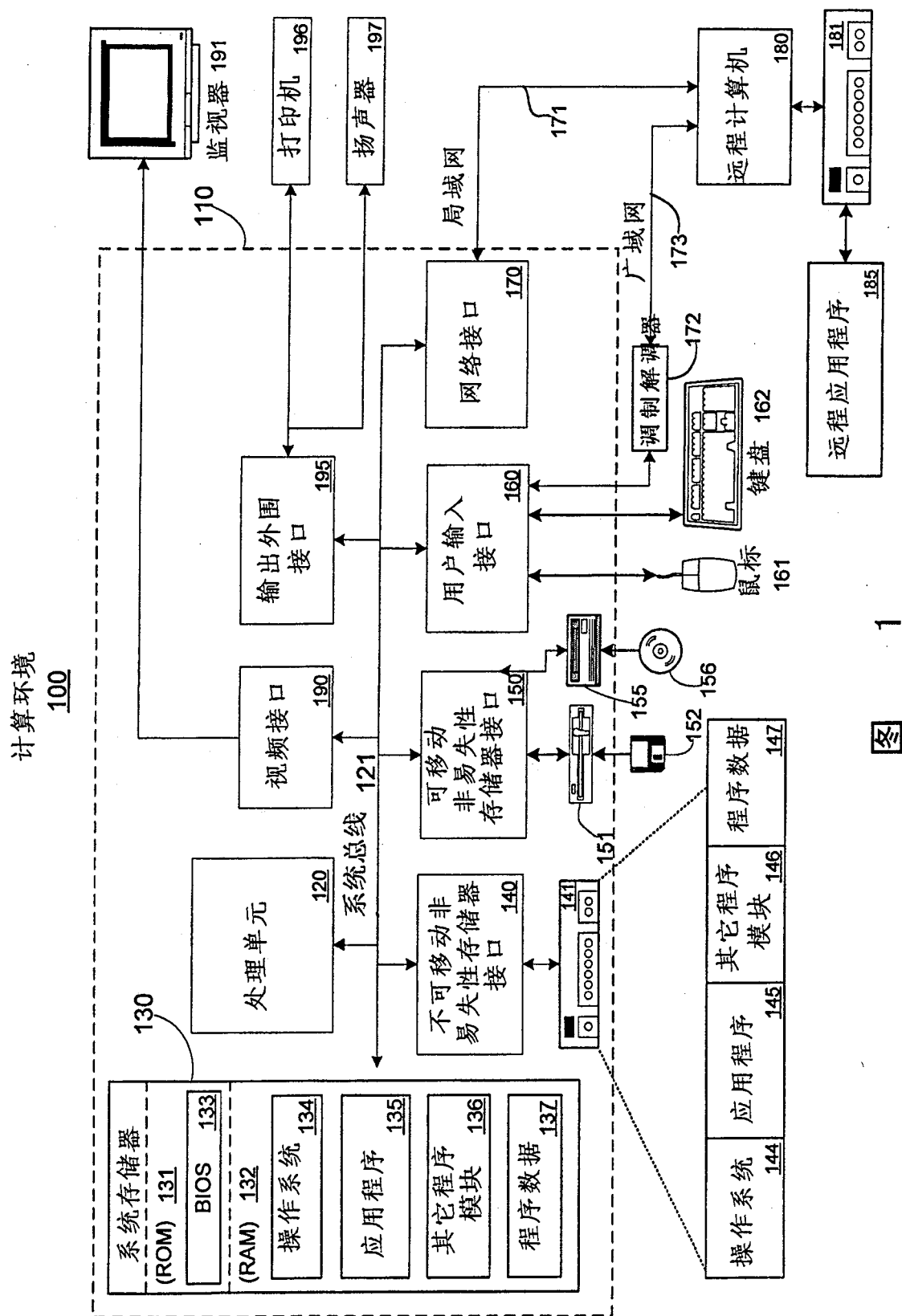
如上所述，尽管是结合各种计算装置描述了本发明的示例性实施例，然而其基础概念可应用到任何计算装置或系统。

本发明所描述的各种技术可结合硬件或软件实现，或者在适当时用两者的组合来实现。由此，本发明的方法和装置，或其某些方面或部分，可采用包含在诸如软盘、CD-ROM、硬盘或任一其它及其可读存储媒质等有形媒质中的程序代码（即，指令）的形式，其中，当程序代码被加载到诸如计算机等机器中并由其执行时，该机器变成用于实践本发明的装置。在可变成计算机上执行程序代码的情况下，计算装置一般包括处理器、可由处理器读的存储媒质（包括易失和非易失存储器和/或存储元件）、至少一个输入设备和至少一个输出设备。如需要，程序可以汇编语言或机器语言实现。在任一情况下，语言可以是编译或解释语言，并可与硬件实现相组合。

本发明的方法和装置也可通过以程序代码的形式实施的通信来实践，这些程序代码通过诸如电线或电缆、光纤或任一其它形式的传输等某些传输媒质来传输，其中，当由诸如 EPROM、门阵列、可变成逻辑器件（PLD）、客户机计算机等机器接收程序代码，并加载到其中由其执行时，该机器变成用于实践本发明的装置。

当在通用处理器上实现时,程序代码与处理器组合以提供一种操作来调用本发明的功能的唯一装置。另外,结合本发明使用的任何存储技术可不变地为硬件和软件的组合。

尽管结合各个附图的较佳实施例描述了本发明,然而可以理解,可在不脱离本发明的情况下使用其它类似的实施例,或对所描述的实施例作出修改和添加,以执行本发明的同一功能。因此,本发明不应当限于任一单个的实施例,而是相反,应当在依照所附权利要求书的宽度和范围内解释。



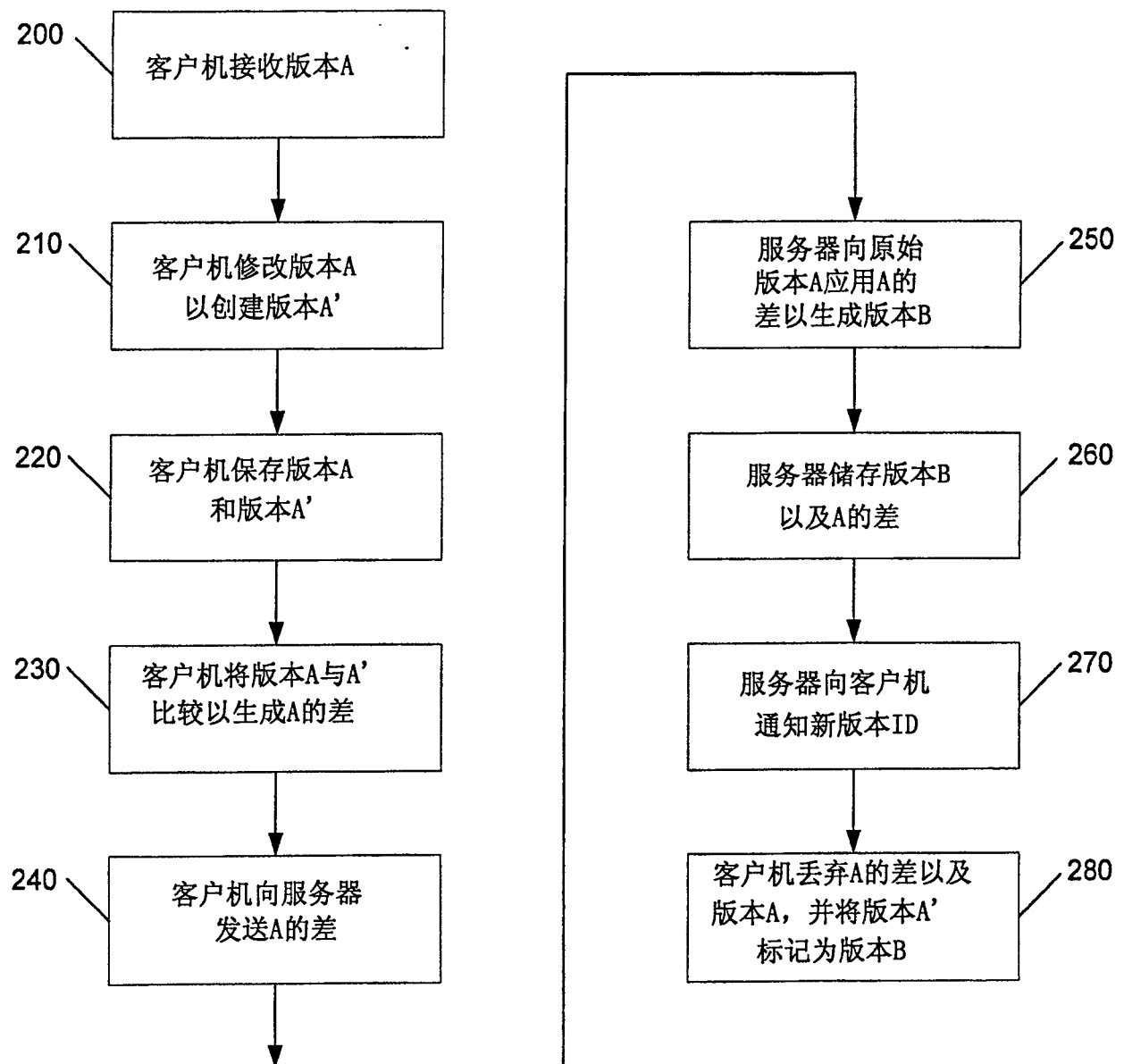


图 2

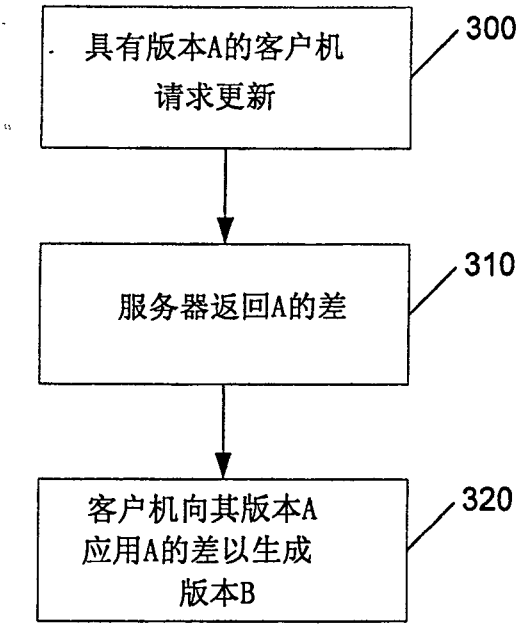


图 3

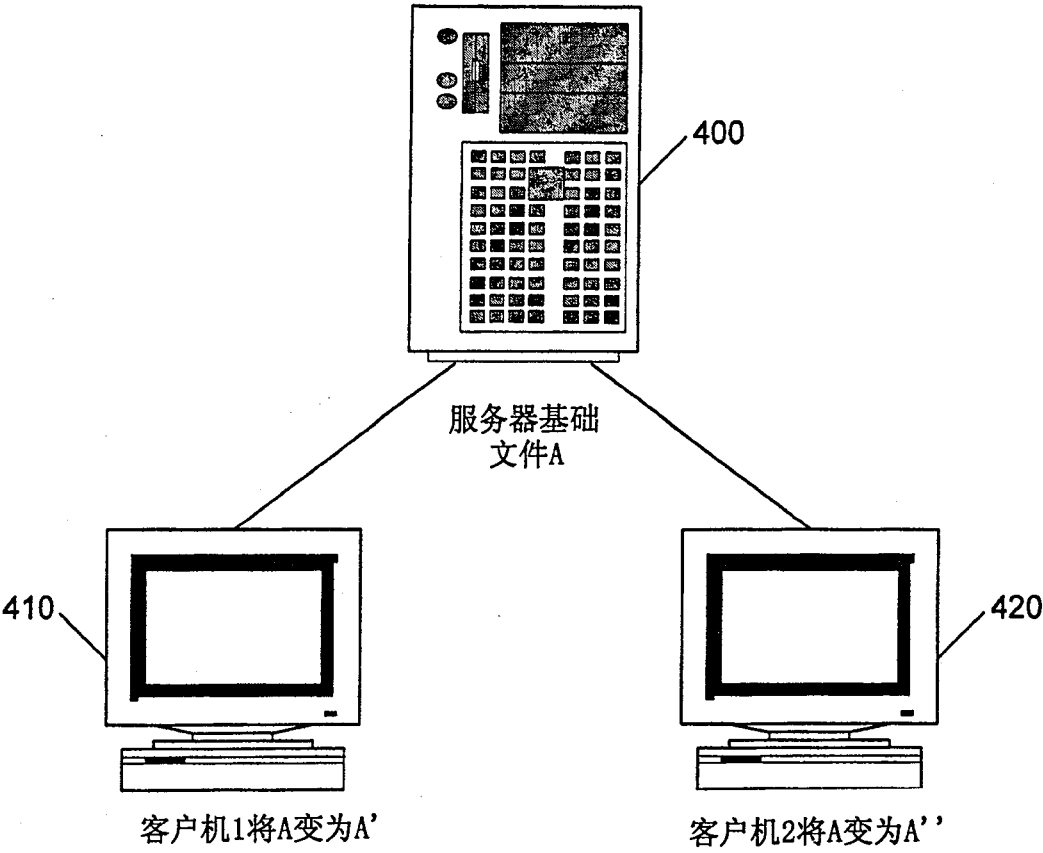
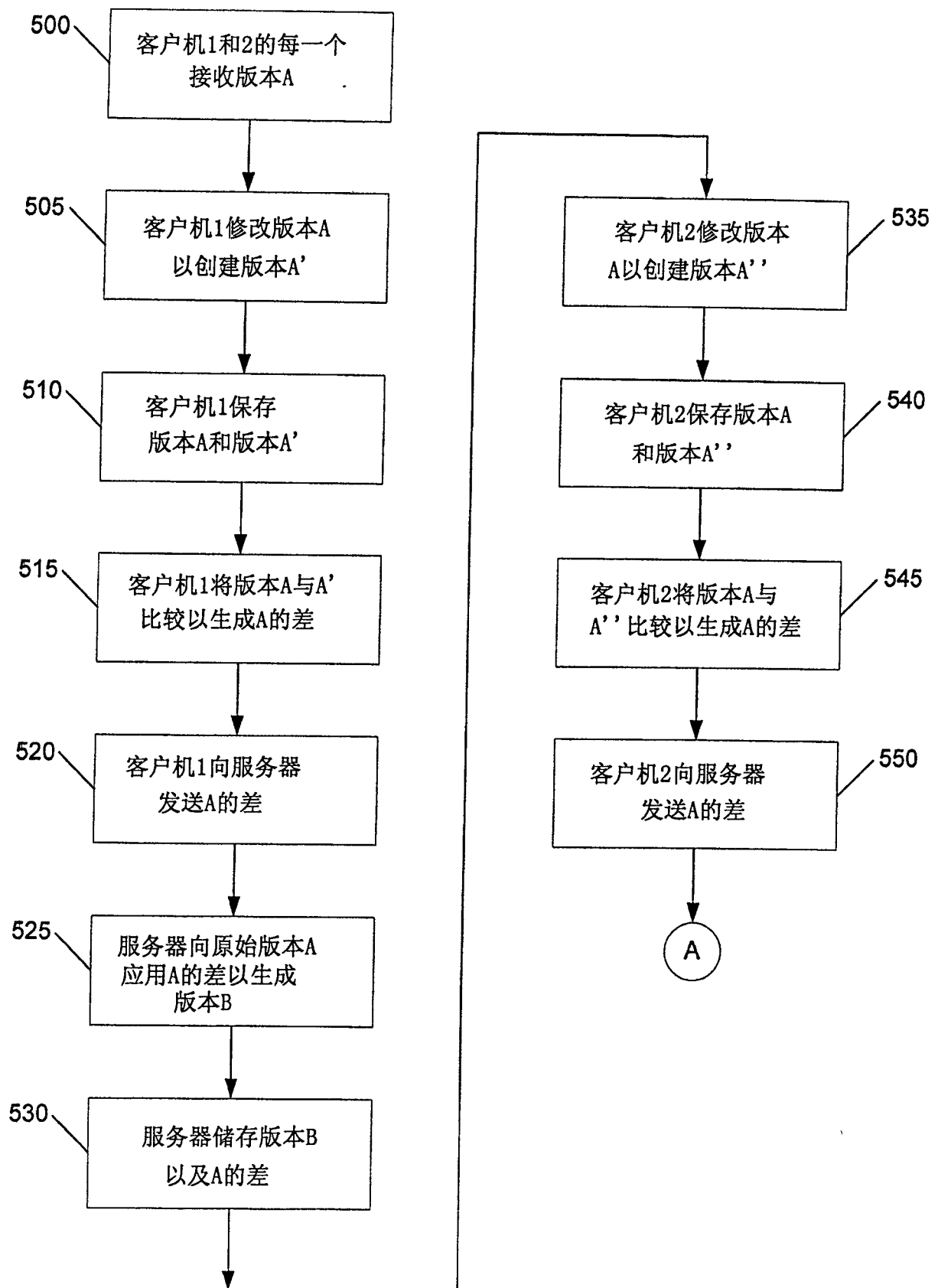


图 4



图

5

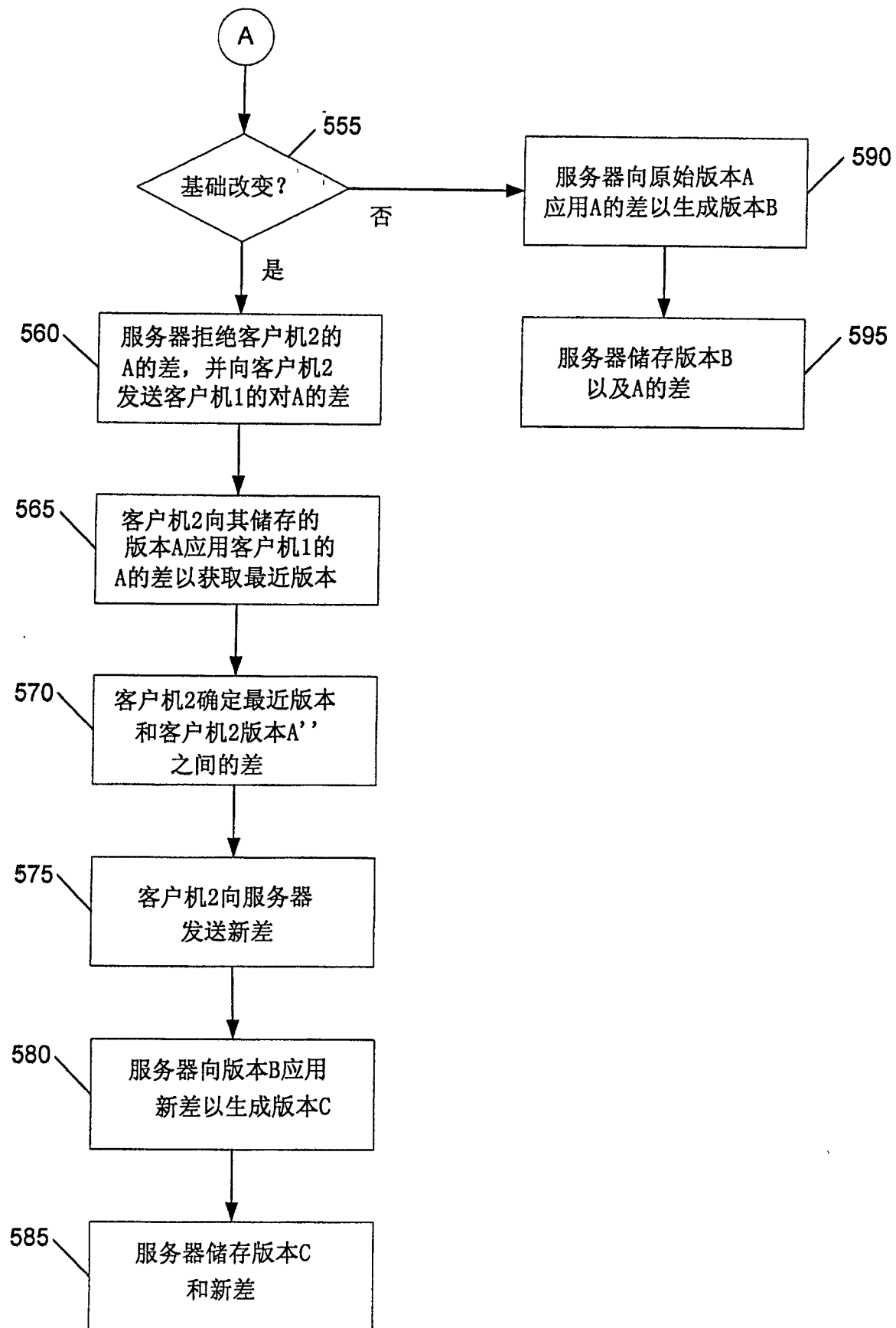


图 6