



(12)发明专利申请

(10)申请公布号 CN 106663064 A

(43)申请公布日 2017.05.10

(21)申请号 201480081655.4

(51)Int.Cl.

(22)申请日 2014.09.04

G06F 13/00(2006.01)

G06F 3/06(2006.01)

(85)PCT国际申请进入国家阶段日
2017.03.01

(86)PCT国际申请的申请数据
PCT/JP2014/073379 2014.09.04

(87)PCT国际申请的公布数据
W02016/035194 JA 2016.03.10

(71)申请人 富士通株式会社
地址 日本神奈川县

(72)发明人 佐沢真一 龟山裕亮

(74)专利代理机构 北京集佳知识产权代理有限公司 11227

代理人 舒艳君 李洋

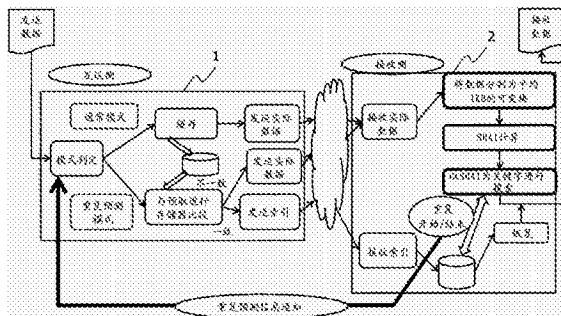
权利要求书3页 说明书14页 附图14页

(54)发明名称

信息处理系统、信息处理装置、信息处理方法、以及信息处理程序

(57)摘要

本发明提供在重复数据的去除中,降低发送侧装置的处理负荷的信息处理系统。接收侧装置具备:第一存储部,其对有接收实绩的第一数据进行存储;重复检测部,其判定是否在第一存储部存储有与来自发送侧装置的接收数据的至少一部分重复的数据;以及通知部,其在第一存储部存储有重复的数据的情况下,向发送侧装置通知第一数据的重复接收的预测信息,发送侧装置具备:第二存储部,其对有发送实绩的第一数据进行存储;判定部,其在从接收侧装置接收到预测信息的情况下,基于重复接收的预测信息来对预发送数据与第一数据的一部分进行比较,并判定与预发送数据一致的第二存储部内的第一数据的一部分的有无;以及发送部,其在第二存储部内有与预发送数据一致的第一数据的一部分的情况下,发送该预发送数据的概要信息。



1. 一种信息处理系统,是包含发送侧装置和接收侧装置的信息处理系统,其中,
上述接收侧装置具备:
第一存储部,其对有接收实绩的第一数据进行存储;
接收部,其从上述发送侧装置接收数据;
重复检测部,其判定是否在上述第一存储部存储有与上述接收数据的至少一部分重复的数据;以及
通知部,其在上述第一存储部存储有与上述接收数据的至少一部分重复的数据的情况下,向上述发送侧装置通知上述第一数据的重复接收的预测信息,
上述发送侧装置具备:
第二存储部,其对有发送实绩的上述第一数据进行存储;
判定部,其在从上述接收侧装置接收到上述预测信息的情况下,基于上述重复接收的预测信息来对上述预发送数据与上述第一数据的一部分进行比较,并判定与上述预发送数据一致的上述第二存储部内的上述第一数据的一部分的有无;以及
发送部,其在上述第二存储部内有与上述预发送数据一致的上述第一数据的一部分的情况下,代替该预发送数据而发送上述预发送数据的概要信息。
2. 根据权利要求1所述的信息处理系统,其中,
上述第一存储部存储上述第一数据的分割为规定长度的多个第一数据块、和该多个第一数据块的识别信息,
上述重复检测部具备:
分割部,其将上述接收数据分割为上述规定长度的多个第二数据块;
计算部,其计算出上述第二数据块的上述识别信息;以及
搜索部,其进行与上述第二数据块的上述识别信息一致的上述第一数据块的搜索。
3. 根据权利要求1或者2所述的信息处理系统,其中,
上述接收侧装置的上述通知部在上述预测信息包含从接收开始起的上述接收数据的位置、和上述接收数据的上述第一数据中的位置而进行通知。
4. 根据权利要求2或者3所述的信息处理系统,其中,
上述接收侧装置还具备:
读出部,其在从上述发送侧装置接收到上述预发送数据的概要信息的情况下,从上述第一存储部读出与上述预发送数据相应的量的第一数据块,并根据读出的第一数据块恢复上述预发送数据;以及
转送部,其将该恢复的上述预发送数据转送给目的地。
5. 根据权利要求1~4中任意一项所述的信息处理系统,其中,
上述发送侧装置的上述发送部在上述第二存储部不存在与上述预发送数据一致的上述第一数据的一部分的情况下,发送上述预发送数据。
6. 根据权利要求5所述的信息处理系统,其中,
上述接收侧装置的上述通知部在上述第一存储部未存储有与上述接收数据的至少一部分一致的上述第一数据的一部分的情况下,将上述第一数据的重复接收的结束的预测信息通知给上述发送侧装置,
上述发送侧装置的上述发送部在接收到上述第一数据的重复接收的结束的预测信息

的情况下,发送上述预发送数据。

7. 一种信息处理装置,具备:

存储部,其对有接收实绩的第一数据进行存储;

接收部,其从发送侧装置接收数据;

重复检测部,其判定是否在上述存储部存储有与上述接收数据的至少一部分重复的数据;以及

通知部,其在上述存储部存储有与上述接收数据的至少一部分重复的数据的情况下,向上述发送侧装置通知上述第一数据的重复接收的预测信息。

8. 一种信息处理装置,具备:

存储部,其对有发送实绩的第一数据进行存储;

判定部,其在从接收侧装置接收到上述第一数据的重复接收的预测信息的情况下,对基于上述重复接收的预测信息的上述预发送数据与上述第一数据的一部分进行比较,判定与上述预发送数据一致的上述存储部内的上述第一数据的一部分的有无;以及

发送部,其在上述存储部内有与上述预发送数据一致的上述第一数据的一部分的情况下,代替该预发送数据而发送上述预发送数据的概要信息。

9. 一种信息处理方法,其中,

在包含分别具备处理器和存储器的发送侧装置和接收侧装置的信息处理系统中,上述接收侧装置的处理器进行以下处理,即

将有接收实绩的第一数据存储于上述接收侧装置的存储器;

从上述发送侧装置接收数据;

判定是否在上述存储器存储有与上述接收数据的至少一部分重复的数据;以及

在上述存储器存储有与上述接收数据的至少一部分重复的数据的情况下,向上述发送侧装置通知上述第一数据的重复接收的预测信息,上述发送侧装置的处理器进行以下处理,即

将有发送实绩的上述第一数据存储于上述发送侧装置的存储器;

在从上述接收侧装置接收到上述预测信息的情况下,基于上述重复接收的预测信息来对上述预发送数据与上述第一数据的一部分进行比较,判定与上述预发送数据一致的上述存储器内的上述第一数据的一部分的有无;以及

在上述存储器内有与上述预发送数据一致的上述第一数据的一部分的情况下,代替该预发送数据而发送上述预发送数据的概要信息。

10. 一种信息处理方法,其中,

在包含处理器和存储器的信息处理装置中,上述处理器进行以下处理,即

将有接收实绩的第一数据存储于上述存储器;

从发送侧装置接收数据;

判定是否在上述存储器存储有与上述接收数据的至少一部分重复的数据;以及

在上述存储器存储有与上述接收数据的至少一部分重复的数据的情况下,向上述发送侧装置通知上述第一数据的重复接收的预测信息。

11. 一种信息处理方法,其中,

在包含处理器和存储器的信息处理装置中,上述处理器进行以下处理,即

将有发送实绩的第一数据存储于上述存储器；

在从接收侧装置接收到上述第一数据的重复接收的预测信息的情况下，对基于上述重复接收的预测信息的上述预发送数据与上述第一数据的一部分进行比较，判定与上述预发送数据一致的上述存储器内的上述第一数据的一部分的有无；以及

在上述存储器内有与上述预发送数据一致的上述第一数据的一部分的情况下，代替该预发送数据而发送上述预发送数据的概要信息。

12. 一种信息处理程序，其中，

用于在包含发送侧装置和接收侧装置的信息处理系统中，

使上述接收侧装置进行以下处理，即

将有接收实绩的第一数据存储于第一存储部；

从上述发送侧装置接收数据；

判定是否在上述第一存储部存储有与上述接收数据的至少一部分重复的数据；以及

在上述第一存储部存储有与上述接收数据的至少一部分重复的数据的情况下，向上述发送侧装置通知上述第一数据的重复接收的预测信息，

并使上述发送侧装置进行以下处理，即

将有发送实绩的上述第一数据存储于第二存储部；

从上述接收侧装置接收到上述预测信息的情况下，基于上述重复接收的预测信息来对上述预发送数据与上述第一数据的一部分进行比较，判定与上述预发送数据一致的上述第二存储部内的上述第一数据的一部分的有无；以及

在上述第二存储部内有与上述预发送数据一致的上述第一数据的一部分的情况下，代替该预发送数据而发送上述预发送数据的概要信息。

13. 一种信息处理程序，其中，

用于使信息处理装置进行以下处理，即

将有接收实绩的第一数据存储于存储部；

从发送侧装置接收数据；

判定是否在上述存储部存储有与上述接收数据的至少一部分重复的数据；以及

在上述存储部存储有与上述接收数据的至少一部分重复的数据的情况下，向上述发送侧装置通知上述第一数据的重复接收的预测信息。

14. 一种信息处理程序，其中，

使信息处理装置进行以下处理，即

将有发送实绩的第一数据存储于存储部；

从上述接收侧装置接收到上述第一数据的重复接收的预测信息的情况下，对基于上述重复接收的预测信息的上述预发送数据与上述第一数据的一部分进行比较，判定与上述预发送数据一致的上述存储部内的上述第一数据的一部分的有无；以及

在上述存储部有与上述预发送数据一致的上述第一数据的一部分的情况下，代替该预发送数据而发送上述预发送数据的概要信息。

信息处理系统、信息处理装置、信息处理方法、以及信息处理程序

技术领域

[0001] 本发明涉及信息处理系统、信息处理装置、信息处理方法、以及信息处理程序。

背景技术

[0002] 实现通信的高速化的方法之一有削减转送数据量的方法。削减转送数据量的方法由于使用频带减少,所以对低频带的线路有效。作为削减转送数据量的方法,例如,有去除重复发送的数据的方法、对发送数据进行压缩的方法。

[0003] 在去除重复发送的数据的方法中,例如,采用分别在发送侧网络以及接收侧网络配置发送侧装置和接收侧装置的系统构成较多。发送侧装置以及接收侧装置预先缓存有处理实绩的数据,发送侧装置使用缓存的数据进行发送数据的重复的确认,在发送数据重复的情况下,代替发送数据而发送发送数据的概要信息。若接收侧装置接收数据的概要信息,则从缓存读出相应的数据,并转送给目的地。由于与发送数据的尺寸相比发送数据的概要信息的尺寸较小,所以能够削减转送数据量,能够抑制发送侧装置与接收侧装置之间的使用频带。

[0004] 专利文献1:日本特开2009—21906号公报

[0005] 专利文献2:日本特开2001—331661号公报

[0006] 专利文献3:日本特开2000—357121号公报

[0007] 然而,在发送侧装置进行的发送数据的重复判定处理负荷较大,而成为开销。在有发送数据的重复的情况下,对于发送数据的重复判定处理所引起的负荷、延迟来说,发送数据的重复判定处理的开销被数据削减所带来的使用频带的减少效果抵消。然而,在没有发送数据的重复的情况下,发送数据的重复判定处理成为不能够忽略的开销。

发明内容

[0008] 本发明的一方式的目的在于提供在重复数据的去除中,能够降低发送侧装置的处理负荷的信息处理系统、信息处理装置、信息处理方法、以及信息处理程序。

[0009] 本发明的方式之一是包含发送侧装置和接收侧装置的信息处理系统。接收侧装置具备:第一存储部,其对有接收实绩的第一数据进行存储;接收部,其从发送侧装置接收数据;重复检测部,其判定是否在第一存储部存储有与接收数据的至少一部分重复的数据;以及通知部,其第一存储部存储有与接收数据的至少一部分重复的数据的情况下,向发送侧装置通知第一数据的重复接收的预测信息。发送侧装置具备:第二存储部,其对有发送实绩的第一数据进行存储;判定部,其在从接收侧装置接收到第一数据的重复接收的预测信息的情况下,基于该重复接收的预测信息来对预发送数据与第一数据的一部分进行比较,判定与预发送数据一致的第二存储部内的第一数据的一部分的有无;以及发送部,其在第二存储部内有与预发送数据一致的第一数据的一部分的情况下,代替该预发送数据而对预发送数据的概要信息进行发送。

[0010] 根据公开的信息处理系统、信息处理装置、信息处理方法、以及信息处理程序,在重复数据的去除中,能够降低发送侧装置的处理负荷。

附图说明

- [0011] 图1是表示数据去重系统的一个例子的图。
- [0012] 图2是表示第一实施方式的数据去重系统的构成的一个例子的图。
- [0013] 图3是表示第一实施方式的处理的一个例子的图。
- [0014] 图4是表示发送侧装置的硬件构成的一个例子的图。
- [0015] 图5是表示发送侧装置的功能构成的一个例子的图。
- [0016] 图6是表示发送侧缓存内的数据结构的一个例子的图。
- [0017] 图7是表示接收侧装置的功能构成的一个例子的图。
- [0018] 图8是表示接收侧缓存的数据结构的一个例子的图。
- [0019] 图9是表示重复预测信息的一个例子的图。
- [0020] 图10是发送侧装置的处理的流程图的一个例子。
- [0021] 图11是接收侧装置的处理的流程图的一个例子。
- [0022] 图12是表示具体例中的处理对象数据的设定的图。
- [0023] 图13A是具体例中的时序图的一个例子。
- [0024] 图13B是具体例中的时序图的一个例子。
- [0025] 图14是表示第一实施方式所涉及的数据去重系统的效果的图。

具体实施方式

[0026] 以下,基于附图,对本发明的实施方式进行说明。以下的实施方式的构成是例示的构成,本发明并不限定于实施方式的构成。

[0027] <数据去重的一个例子>

[0028] 图1是表示数据去重系统的一个例子的图。数据去重系统P100包含发送侧装置P1和接收侧装置P2。发送侧装置P1以及接收侧装置P2分别例如是位于网络的边界的装置。

[0029] 发送侧装置P1例如,若从应用程序接收数据,则将该数据分割为被称为组块(Chunk)的1~4KB的可变长块。发送侧装置P1对各组块进行SHA1(Secure Hash Algorithm:安全哈希算法)的计算获取哈希值,并将该哈希值作为关键字(key),进行缓存内的哈希搜索。在发送侧装置P1的缓存保持了有发送实绩的数据的组块和SHA1的计算结果。

[0030] 在作为哈希搜索的结果,未检测到SHA1的计算结果一致的组块的情况下,即、在组块未重复的情况下,发送侧装置P1将该组块保存于缓存,并发送实际数据。在哈希搜索的结果是检测到SHA1的计算结果一致的组块的情况下,即、在组块重复的情况下,发送侧装置P1发送该组块的ID数据。在组块的ID数据例如包含有应用程序ID、该组块的位置信息、以及该组块的尺寸。

[0031] 接收侧装置P2在接收了实际数据的情况下,将该实际数据转送至目的地,并且将该实际数据保存于缓存。接收侧装置P2在接收了组块的ID的情况下,从缓存读出相应的实际数据对数据进行恢复,并发送至目的地。

[0032] 在图1所示的数据去重系统P100中,发送侧装置进行组块分割、SHA1计算、哈希搜

索的处理,在组块重复的情况下,通过发送该组块的ID数据,来削减数据转送量。然而,组块分割、SHA1计算、哈希搜索的处理负荷较大,所以例如,在发送侧装置P1是终端装置等性能比较低的装置的情况下,有数据去重的处理不会有助于通信的高速化的可能性。

[0033] <第一实施方式>

[0034] 图2是表示第一实施方式的数据去重系统100的构成的一个例子的图。例如,数据去重系统100包含去重客户端1、和去重服务器2。去重客户端1例如是智能手机1A、PC(Personal Computer:个人计算机)1B等终端装置。另外,去重客户端1也可以是服务器等专用计算机。去重客户端1也称为发送侧装置1。去重服务器2也称为接收侧装置2。

[0035] 在第一实施方式中,接收侧装置2对来自发送侧装置1的有接收实绩的数据进行缓存,并使用该缓存检测接收数据的重复的开始。若接收侧装置2检测到接收数据的重复的开始,则向发送侧装置1通知重复预测信息。若发送侧装置1从接收侧装置2接收到重复预测信息,则切换动作模式,基于该重复预测信息,对预发送数据进行通过负荷比较轻的存储器比较的重复判定处理,并发送数据。

[0036] 图3是表示第一实施方式的处理的一个例子的图。发送侧装置1具有通常模式和重复预测模式两个动作模式。在通常模式中,发送侧装置1将预发送数据储存于缓存,并发送实际数据。

[0037] 若发送侧装置1从接收侧装置2接收到重复预测信息的通知,则动作模式切换为重复预测模式。在重复预测模式中,发送侧装置1基于重复预测信息从缓存数据预取部分数据,并进行预发送数据与存储器的比较,判定重复的有无。在部分数据与预发送数据不一致的情况下,预发送数据不是重复的数据,所以发送侧装置1发送实际数据。在部分数据与预发送数据一致的情况下,预发送数据重复,所以发送侧装置1发送预发送数据的索引。

[0038] 接收侧装置2在接收到实际数据的情况下,将实际数据分割为组块,对各组块计算SHA1,并将SHA1的计算结果作为关键字搜索缓存。在缓存内检测到重复的组块的情况下,接收侧装置2检测到重复接收的开始,并生成重复预测信息通知给发送侧装置1。

[0039] 接收侧装置2在接收到索引的情况下,从缓存读出索引所示的组块,恢复数据,并发送给目的地的装置。

[0040] 在第一实施方式中,接收侧装置2进行组块分割、SHA1计算、哈希搜索等负荷较高的处理,发送侧装置1进行预取的数据与预发送数据的比较这样的负荷比较轻的处理。由此,能够降低发送侧装置1所涉及的处理负荷。

[0041] <装置构成>

[0042] <发送侧装置>

[0043] 图4是表示发送侧装置1的硬件构成的一个例子的图。发送侧装置1例如是移动电话终端、智能手机、平板终端、通用或者专用的计算机。

[0044] 发送侧装置1具备CPU(Central Processing Unit:中央处理器)101、主存储装置102、输入装置103、输出装置104、辅助存储装置105、以及网络接口107。另外,它们通过总线109相互连接。

[0045] 输入装置103例如是操作按钮、触摸面板、键盘、小键盘等。从输入装置103输入的数据输出至CPU101。输入装置例如也可以包含麦克那样的声音输入装置。

[0046] 辅助存储装置105储存各种程序、在各程序的执行时CPU101使用的数据。辅助存储

装置105例如是EPROM (Erasable Programmable ROM:可擦除可编程只读存储器)、闪存、或者硬盘驱动器(Hard Disk Drive)等非易失性的存储器。辅助存储装置105例如保持操作系统(OS)、去重客户端程序、其它各种应用程序。去重客户端程序是用于使发送侧装置1作为去重客户端发挥作用的程序。

[0047] 主存储装置102对CPU101提供下载储存于辅助存储装置105的程序的存储区域以及工作区域,或者作为缓冲区使用。主存储装置102例如,包含ROM(Read only Memory:只读存储器)、RAM(Random Access Memory:随机存储器)那样的半导体存储器。

[0048] CPU101通过将保持于辅助存储装置105的OS、各种应用程序下载至主存储装置102并执行,来执行各种处理。CPU101并不限定于一个,也可以具备多个。

[0049] 网络接口107是进行与网络的信息的输入输出的接口。网络接口107包含与有线的网络连接的接口、与无线的网络连接的接口。网络接口107例如是NIC(Network Interface Card:网卡)、无线LAN(Local Area Network:局域网)卡等。在网络接口107接收到的数据等输出给CPU101。

[0050] 输出装置104输出CPU101的处理的结果。输出装置104包含扬声器等声音输出装置、显示器、打印机。

[0051] 此外,图4所示的发送侧装置1的硬件构成是一个例子,并不限定于上述构成,而能够根据实施方式适当地进行构成要素的省略、置换、追加。例如,也可以是发送侧装置1具备便携式记录介质驱动装置,并执行记录于便携式记录介质的程序。便携式记录介质例如是SD卡、miniSD卡、microSD卡、USB(Universal Serial Bus:通用串行总线)闪存、CD(Compact Disc:光盘)、DVD(Digital Versatile Disc:数字通用光盘)、Blu-ray(注册商标)Disc、或者闪存卡那样的记录介质。

[0052] 另外,在发送侧装置1为服务器的情况下,例如,发送侧装置1也可以不具备输入装置103、输出装置104。

[0053] 图5是表示发送侧装置1的功能构成的一个例子的图。发送侧装置1作为功能构成,具备数据接收部11、重复判定部12、缓存保存部13、数据发送部14、以及重复预测接收部15。这些功能构成是通过CPU101执行储存于辅助存储装置105的去重客户端程序来实现的功能构成。另外,通过执行去重客户端程序,在辅助存储装置105的存储区域确保发送侧缓存16用的区域。另外,在主存储装置102的存储区域确保储存重复预测表格17的区域。去重客户端程序例如是应用层程序。

[0054] 数据接收部11从成为数据的发送源的应用程序接收数据,并终止该应用程序的会话。例如,在作为数据的发送源的应用程序是利用TCP(Transmission Control Protocol:传输控制协议)进行通信的应用程序的情况下,若数据接收部11从数据的发送源接收到连接要求,则与该数据的发送源确立TCP会话。数据接收部11使用作为IP地址与端口编号的组合的套接字(socket)与数据的发送源进行通信。因此,数据接收部11能够根据套接字识别应用程序。

[0055] 此外,执行成为数据的发送源的应用程序的实体既可以是发送侧装置1自身,也可以是其它的装置。例如,在发送侧装置1是服务器的情况下,执行成为数据的发送源的应用程序的实体是其它的装置,数据接收部11通过网络接口107,从该其它的装置接收数据。例如,在发送侧装置1为终端装置的情况下,执行成为数据的发送源的应用程序的实体为发送

侧装置1自身。在第一实施方式中,后面不特定执行成为数据的发送源的应用程序的实体,而仅称为“发送源”。

[0056] 数据接收部11在动作模式为通常模式的情况下,将预发送数据输出给数据发送部14。预发送数据是从发送源的应用程序接收的数据。数据接收部11在动作模式为重复预测模式的情况下,将预发送数据输出给重复判定部12。

[0057] 数据接收部11例如保持规定尺寸的处理缓冲区。从发送源的应用程序接收的数据在发送侧装置1中以该处理缓冲区的尺寸的单位进行处理。处理缓冲区的尺寸例如为50~100KB。但是,处理缓冲区的尺寸并不限于此。即,预发送数据是与处理缓冲区相同或者小于处理缓冲区的尺寸的数据块。

[0058] 另外,数据接收部11监视从与发送源的应用程序的通信开始的预发送数据的偏移。从与发送源的应用程序的通信开始的预发送数据的偏移是表示预发送数据的前端的位置的值。例如,偏移是从0开始,且每隔数据的尺寸一个字节增加1的值。例如,在通信开始后,在第一个预发送数据的尺寸为80KB的情况下,该第一个预发送数据的偏移为0,下一个预发送数据的偏移为80000。

[0059] 在动作模式为重复预测模式的情况下,在重复判定部12从数据接收部11输入预发送数据。此外,动作模式的初始值为通常模式。重复判定部12参照预发送数据的应用程序的重复预测表格17,从重复预测表格17所示的应用程序的发送侧缓存16预取相应的部分数据并进行与预发送数据的比较。此外,预发送数据的应用程序与重复预测表格17所示的应用程序既有一致的情况,也有不一致的情况。后述部分数据的获取方法的详细内容。

[0060] 在该部分数据与预发送数据不一致的情况下,重复判定部12判定为预发送数据是没有重复的新数据。该情况下,重复判定部12将预发送数据本身输出给缓存保存部13、和数据发送部14。

[0061] 在部分数据与预发送数据一致的情况下,重复判定部12判定为预发送数据是重复发送的数据。将与预发送数据一致的部分数据称为重复数据。在存在重复数据的情况下,重复判定部12将预发送数据的索引输出给数据发送部14。重复判定部12是“判定部”的一个例子。

[0062] 预发送数据的索引例如包含表示是索引的标志、预发送数据的应用程序的识别信息、从通信开始的预发送数据的偏移、重复数据的应用程序的识别信息、重复数据的应用程序的发送侧缓存16内的重复数据的偏移、以及预发送数据的尺寸。另外,重复数据的偏移从相应应用程序的发送侧缓存16获取。预发送数据的索引是“预发送数据的概要信息”的一个例子。

[0063] 缓存保存部13在预发送数据为新数据的情况下从重复判定部12接受该预发送数据的输入。缓存保存部13将该预发送数据追加到储存在该预发送数据的应用程序的发送侧缓存16内的数据的最后尾。

[0064] 数据发送部14例如将来自数据的发送源的连接要求转送给接收侧装置2,并且在与接收侧装置2之间确立TCP的会话。数据发送部14使用该会话将输入的数据发送给接收侧装置2。

[0065] 例如,在从重复判定部12接受了预发送数据本身的输入的情况下,数据发送部14将预发送数据发送给接收侧装置2。例如,在从重复判定部12接受了预发送数据的索引的输

入的情况下,数据发送部14将预发送数据的索引发送给接收侧装置2。此外,从数据发送部14发送的数据添附有应用程序的识别信息、和数据的本来的目的地的信息,即、套接字的信息。数据发送部14是“发送部”的一个例子。

[0066] 重复预测接收部15从接收侧装置2接收重复预测信息。重复预测接收部15利用与数据发送部14相同的会话。重复预测接收部15使接收的重复预测信息反映于相应应用程序的重复预测表格17。

[0067] 重复预测表格17例如储存于主存储装置102的存储区域。重复预测表格17所包含的信息与重复预测信息所包含的信息相同。重复预测表格17被由重复预测接收部15新接收的重复预测信息覆盖。后述重复预测表格17以及重复预测信息的详细内容。

[0068] 图6是表示发送侧缓存16内的数据结构的一个例子的图。发送侧缓存16按照每个应用程序生成在辅助存储装置105内的存储区域。例如,若缓存保存部13接受新应用程序的预发送数据的输入,则在辅助存储装置105的确保为发送侧缓存16用的存储区域生成该新应用程序的发送侧缓存16。

[0069] 发送侧缓存16针对有发送实绩的数据,包含发送偏移和发送数据。发送数据是数据接收部11的处理缓冲区的尺寸或者小于处理缓冲区的尺寸的数据块。发送偏移是数据的相应应用程序的发送侧缓存16内的数据块的前端位置。例如,偏移从0开始,在相应应用程序的发送侧缓存16内已经储存有80KB的数据的情况下,新储存的数据的偏移为80000。预发送数据的索引所包含的重复数据的偏移是发送侧缓存16内的相应数据的发送偏移。发送侧缓存16是“第二存储部”的一个例子。

[0070] <接收侧装置>

[0071] 接收侧装置2例如是专用的服务器。接收侧装置2的硬件构成与发送侧装置1的硬件构成类似,省略图示。即,接收侧装置2具备CPU、主存储装置、辅助存储装置、以及网络接口。另外,接收侧装置2也可以具备输入装置、输出装置、便携式记录介质驱动装置等。

[0072] 在接收侧装置2的辅助存储装置储存有OS、去重服务器程序、以及其它应用程序等。去重服务器程序是用于使接收侧装置2作为去重服务器动作的程序。

[0073] 图7是表示接收侧装置2的功能构成的一个例子的图。接收侧装置2作为功能构成,具备数据接收部21、组块分割部22A、SHA1计算部22B、哈希搜索部22C、缓存读出部23、缓存保存部24、表格管理部25、数据发送部26、以及重复预测发送部27。这些功能构成例如是通过由CPU执行储存于辅助存储装置的去重服务器程序来实现的功能构成。另外,通过执行去重服务器程序,在辅助存储装置的存储区域确保接收侧缓存28用的区域。另外,在主存储装置的存储区域确保用于储存重复预测表格的区域。此外,去重服务器程序例如是应用层程序。

[0074] 数据接收部21例如若从发送侧装置1接收连接要求,则确立TCP会话,并通过该TCP会话接收数据。在接收数据为实际数据的情况下,数据接收部21将接收数据输出给组块分割部22A。在接收数据为索引的情况下,数据接收部21将接收数据输出给缓存读出部23。例如,在索引内包含有表示是索引的标志,数据接收部21根据该标志判定接收数据是索引。

[0075] 此外,数据接收部21保持规定尺寸的处理缓冲区,接收数据在接收侧装置2中以该处理缓冲区的尺寸的单位进行处理。处理缓冲区的尺寸例如是50~100KB。但是,处理缓冲区的尺寸并不限定于此。另外,数据接收部21按照接收数据的应用程序,监视从接收开始的

数据的偏移。数据接收部21是“接收部”的一个例子。

[0076] 组块分割部22A将接收数据分割为组块。组块的尺寸由规定的算法决定。组块分割部22A将根据接收数据生成的多个组块输出给SHA1计算部22B。组块分割部22A是“分割部”的一个例子。

[0077] SHA1计算部22B针对根据接收数据生成的多个组块的各个进行SHA1的计算。SHA1计算部22B将各组块的SHA1的计算结果输出给哈希搜索部22C。SHA1计算部22B是“计算部”的一个例子。

[0078] 哈希搜索部22C将接收的各组块的SHA1的计算结果(哈希值)作为关键字,搜索所有应用程序的接收侧缓存28。哈希搜索部22C将搜索结果输出给缓存保存部24、表格管理部25。哈希搜索部22C是“搜索部”的一个例子。组块分割部22A、SHA1计算部22B、哈希搜索部22C合起来为“重复检测部”的一个例子。

[0079] 表格管理部25在对于接收数据的组块中至少一个组块,在接收侧缓存28检测到SHA1的计算结果一致的组块的情况下,覆盖更新接收数据的应用程序的重复预测表格29。此时的更新内容的一个例子是动作模式向重复预测模式的变更。

[0080] 表格管理部25在对于接收数据的组块,在接收侧缓存28都没有SHA1的计算结果一致的组块的情况下,不进行接收数据的应用程序的重复预测表格29的更新。

[0081] 缓存保存部24将接收数据的组块中,新的组块追加到相应应用程序的接收侧缓存28的最后尾。此外,既有接收数据的组块的全部为新数据的情况,也有一部分组块为新数据的情况。

[0082] 缓存读出部23从数据接收部21接受索引的输入。缓存读出部23基于该索引的信息,从相应的应用程序的接收侧缓存28读出相应的组块,恢复数据并输出给数据发送部26。缓存读出部23是“转送部”的一个例子。

[0083] 数据发送部26将接收数据或者恢复的数据发送给目的地的应用程序。数据发送部26将从发送侧装置1接收的连接要求转送给目的地,与目的地的应用程序例如确立TCP会话,并通过该TCP会话发送数据。目的地的应用程序既可以存在于接收侧装置2内,也可以存在于其它的装置内。在第一实施方式中,不特定目的地的应用程序的执行实体,而仅称为“目的地”。数据发送部26是“发送部”的一个例子。

[0084] 重复预测发送部27在生成或者更新重复预测表格的情况下,将该重复预测表格作为重复预测信息发送给发送侧装置1。重复预测发送部27是“发送部”的一个例子。

[0085] 图8是表示接收侧缓存28的数据结构的一个例子的图。接收侧缓存28例如按照每个应用程序准备。在接收侧缓存28中,针对各组块,储存哈希值、应用程序ID、接收偏移、以及接收数据。哈希值是SHA1的计算结果的值。接收偏移是相应组块的前端的应用程序中的位置。接收侧缓存是“第一存储部”的一个例子。

[0086] 图9是表示重复预测信息的一个例子的图。如上述,在发送侧装置1保持的重复预测表格17与在接收侧装置2保持的重复预测表格29的格式相同,与重复预测信息的格式相同。接收侧装置2的重复预测表格29由接收侧装置2的表格管理部25生成以及更新。

[0087] 重复预测信息例如包含当前应用程序种类、重复预测模式、重复应用程序种类、当前应用程序重复开始位置、重复应用程序开始位置、以及重复预测长度的项目。

[0088] 在当前应用程序种类的项目储存有接收侧装置2的接收数据的应用程序的识别信

息。在重复预测模式的项目设定有“开启(on)”、“关闭(off)”的任意一个。在重复预测模式的项目设定了“开启”的情况下,表示发送侧装置1的动作模式指定为重复预测模式。在重复预测模式的项目设定了“关闭”的情况下,表示发送侧装置1的动作模式指定为通常模式。

[0089] 在重复应用程序种类的项目储存有与接收侧装置2的接收数据的组块重复的接收侧缓存28内的组块的应用程序的识别信息。也有相同的数据被不同的应用程序使用的情况,所以接收数据的应用程序与储存与接收数据重复的数据的缓存的应用程序并不一定一致。以后,将储存与接收数据或者预发送数据重复的数据的缓存的应用程序称为重复应用程序。

[0090] 在当前应用程序重复开始位置的项目储存有从与接收侧装置2中的接收数据的应用程序的通信开始的、检测到重复的接收数据的组块的偏移。储存于当前应用程序重复开始位置的项目的值例如从数据接收部21获取。

[0091] 在重复应用程序重复开始位置的项目储存有重复应用程序的接收侧缓存28内的、与接收数据的组块一致的组块的接收偏移。储存于重复应用程序重复开始位置的项目的值从重复应用程序的接收侧缓存28获取。

[0092] 在重复预测长度的项目储存有重复接收的数据的预测的数据的长度。储存于重复预测长度的项目的值例如是从储存于重复应用程序的接收侧缓存28的数据的量减去与接收数据的组块一致的组块的接收偏移后的值。

[0093] 重复应用程序种类、当前应用程序重复开始位置、重复应用程序重复开始位置、重复预测长度的项目在重复预测模式的项目的值为“开启”以外的情况下无效。此外,图9所示的重复预测信息的格式是一个例子,并不限定于此。例如,重复预测信息也可以不包含重复预测长度的项目。

[0094] 例如,检测到从应用程序A的通信开始的偏移80000的接收数据的组块与应用程序B的接收侧缓存内的接收偏移100000的组块重复的情况下生成的重复预测信息如以下。此外,在以下的重复预测信息省略重复预测长度的项目。

[0095] (生成的重复预测信息)

[0096] 应用程序种类:应用程序A

[0097] 重复预测模式:开启

[0098] 重复应用程序种类:应用程序B

[0099] 当前应用程序重复开始位置:80000

[0100] 重复应用程序重复开始位置:100000

[0101] 接下来,在发送侧装置1接收了上述重复预测信息之后,发送从应用程序A的通信开始的偏移90000的预发送数据的情况下,例如,进行以下的处理。重复判定部12参照应用程序种类的项目为应用程序A的重复预测表格17,并参照重复应用程序种类的项目示出的应用程序B的发送侧缓存16。

[0102] 重复判定部12基于重复预测表格17,从应用程序B的发送侧缓存16读出部分数据,并进行该部分数据与预发送数据的比较。部分数据是接收偏移具有对重复应用程序重复开始位置的项目的值加上从应用程序A的通信开始的预发送数据的偏移减去当前应用程序重复开始位置的项目的值的值后的值的数据。此时,部分数据的接收偏移=重复应用程序重复开始位置100000+(90000-80000)=110000。

[0103] 在应用程序A的预发送数据与应用程序B的部分数据一致的情况下,发送侧装置1的数据发送部14发送预发送数据的索引。在预发送数据的索引包含有以下信息。

[0104] (预发送数据的索引的信息)

[0105] 索引标志:开启(on)

[0106] 当前应用程序种类:应用程序A

[0107] 当前应用程序发送偏移:90000

[0108] 重复应用程序种类:应用程序B

[0109] 重复应用程序偏移:110000

[0110] 数据长:80KB

[0111] 若接收侧装置2接收上述预发送数据的索引,则缓存读出部23基于索引所包含的信息进行处理。具体而言,缓存读出部23从应用程序B的接收侧缓存28读出偏移11000~19000所包含的组块,并恢复数据。

[0112] <处理的流程>

[0113] 图10是发送侧装置1的处理的流程图的一个例子。图10所示的处理例如在确立与数据发送源的应用程序的TCP会话时开始。

[0114] 在OP1中,缓存保存部13将确立了会话的应用程序的发送侧缓存16的发送偏移设定为0,进行初始化。另外,重复预测接收部15对重复预测表格17进行初始化。接下来处理进入OP2。

[0115] 在OP2中,发送侧装置1接收数据。在接收数据为重复预测信息的情况下(OP3:是),处理进入OP4。在接收数据不为重复预测信息的情况下(OP3:否),即、在接收数据是来自应用程序的预发送数据的情况下,处理进入OP5。

[0116] 在OP4中,由于重复预测接收部15接收了重复预测信息,所以将该重复预测信息复印至重复预测表格17。其后,图10所示的处理结束。此外,接下来,在发送侧装置1接收了数据的情况下,处理从OP2开始。

[0117] 在OP5中,数据接收部11参照重复预测表格17,判定动作模式是否为重复预测模式。数据接收部11与发送源的应用程序进行套接字通信,所以能够获取接收数据的应用程序的识别信息,参照与接收数据的应用程序相应的重复预测表格17。

[0118] 在动作模式为重复预测模式的情况下(OP5:是),处理进入OP6。在动作模式不为重复预测模式的情况下(OP5:否),处理进入OP9。

[0119] OP6~OP9是动作模式为重复预测模式的情况下的处理。在OP6中,重复判定部12基于重复预测表格17,从重复应用程序的发送侧缓存16读出与预发送数据进行比较的部分数据。部分数据的读出方法如上述。接下来处理进入OP7。

[0120] 在OP7中,重复判定部12对预发送数据与部分数据的存储器进行比较,判定重复的有无。在预发送数据与部分数据一致的情况下(OP7:是),处理进入OP8。在预发送数据与部分数据不一致的情况下(OP7:否),处理进入OP9。

[0121] 在OP8中,预发送数据与部分数据一致,所以数据发送部14将预发送数据的索引发送给接收侧装置2。其后,图10所示的处理结束,接下来若从相同的应用程序接收数据则从OP2开始处理。

[0122] OP9~OP11是动作模式为通常模式的情况下、以及动作模式为重复预测模式,并

且,预发送数据与部分数据不一致的情况下的处理。在OP9中,缓存保存部13将预发送数据储存于相应应用程序的发送侧缓存16。接下来处理进入OP10。

[0123] 在OP10中,缓存保存部13将相应的发送侧缓存16内的下一个储存的数据的发送偏移更新为对预发送数据的发送偏移加上预发送数据的尺寸后的值。接下来处理进入OP11。

[0124] 在OP11中,数据发送部14将预发送数据的实际数据发送给接收侧装置2。其后,图10所示的处理结束,接下来若从相同的应用程序接收数据则从OP2开始处理。

[0125] 图11是接收侧装置2的处理的流程图的一个例子。图11所示的处理在发送侧装置1与接收侧装置2之间例如确立了TCP的会话时开始。

[0126] 在OP21中,在接收侧装置2中,进行初始化处理。具体而言,缓存保存部24将确立了会话的应用程序的接收侧缓存28的接收偏移设定为0。表格管理部25对重复预测表格29进行初始化。此外,动作模式的初始值为通常模式,所以重复预测表格29内的重复预测模式的项目关闭。接下来处理进入OP22。

[0127] 在OP22中,数据接收部21接收数据。在接收数据为索引的情况下(OP23:是),处理进入OP24。在接收数据为实际数据的情况下(OP23:否),处理进入OP26。

[0128] 在OP24中,缓存读出部23基于接收的索引所包含的信息,从重复应用程序的接收侧缓存28读入相应的组块,并恢复数据。接下来处理进入OP25。

[0129] 在OP25中,数据发送部26向目的地发送数据。其后,图11所示的处理结束,在接收下一个数据的情况下,从OP22开始处理。

[0130] OP26以后的处理是接收数据为实际数据的情况下的处理。在OP26中,组块分割部22A将接收数据分割为组块。接下来处理进入OP27。

[0131] 在OP27中,SHA1计算部22B针对各组块进行SHA1的计算。接下来处理进入OP28。

[0132] 在OP28中,哈希搜索部22C针对各组块,搜索各应用程序的接收侧缓存28。在储存于接收侧缓存28的组块中至少有一个SHA1的计算结果一致的组块的情况下(OP29:是),处理进入OP33。在储存于接收侧缓存28的组块中没有SHA1的计算结果一致的组块的情况下(OP29:否),处理进入OP30。

[0133] OP30、OP31的处理是在储存于接收侧缓存28的组块中没有SHA1的计算结果一致的组块,即、接收数据为新的数据的情况下的处理。在OP30中,表格管理部25参照接收数据的应用程序的重复预测表格29,判定重复预测模式的项目是否为“开启”。在重复预测模式的项目为“开启”的情况下(OP30:是),处理进入OP31。在重复预测模式的项目为“关闭”的情况下(OP30:否),处理进入OP32。

[0134] 在OP31中,由于接收数据为新的数据,并且,重复预测模式的项目的值为“开启”,所以表格管理部25检测数据的重复接收的结束,并将接收数据的应用程序的重复预测表格29的重复预测模式的项目设定为“关闭”。通过该重复预测表格29的更新,重复预测发送部27向发送侧装置1发送重复预测模式的项目为“关闭”的重复预测信息。接下来处理进入OP32。

[0135] 在OP32中,缓存保存部24将接收数据的组块储存于相应的接收侧缓存28,并设定下一个接收偏移。下一个接收偏移是对储存的组块的偏移加上该组块的尺寸后的值。接下来处理进入OP25,向目的地发送接收数据。其后,图11所示的处理结束,在接收下一个数据的情况下,从OP22开始处理。

[0136] OP33~OP35是在储存于接收侧缓存28的组块中至少有一个SHA1的计算结果一致的组块的情况下,即、接收数据的一部分或者全部重复的情况下的处理。在OP33中,表格管理部25参照接收数据的应用程序的重复预测表格29,判定重复预测模式的项目是否为“关闭”。

[0137] 在重复预测模式的项目为“开启”的情况下(OP33:否),处理进入OP32。在OP32中,在接收数据的组块中存在新的组块的情况下,缓存保存部24将该新的组块储存于接收侧缓存28,并设定下一个接收偏移。其后,处理进入OP25,向目的地发送接收数据。

[0138] 在重复预测模式的项目为“关闭”的情况下(OP33:是),处理进入OP34。在OP34中,表格管理部25更新接收数据的应用程序的重复预测表格29。具体而言,表格管理部25将预测重复模式的项目设定为“开启”。另外,也适当地更新重复应用程序的项目、重复应用程序重复开始位置的项目、当前应用程序重复开始位置的项目。接下来处理进入OP35。

[0139] 在OP35中,重复预测发送部27将更新的重复预测表格的内容作为重复预测表格发送给发送侧装置1。接下来处理进入OP32。在OP32中,在接收数据的组块中存在新的组块的情况下,缓存保存部24将该新的组块储存于接收侧缓存28,并设定下一个接收偏移。其后,处理进入OP25,向目的地发送接收数据。

[0140] 此外,图11所示的流程图是一个例子,各处理的执行顺序等并不限于于此,也可以适当地变更执行顺序。例如,OP25的向目的地发送接收数据的处理也可以在OP31、OP35的重复预测信息的发送处理之前进行。

[0141] <具体例>

[0142] 图12是表示具体例中的处理对象数据的设定的图。在具体例中,发送侧装置1以及接收侧装置2均在应用程序X的缓存保持数据A、数据B、数据C。数据A、数据B、数据C分别为100KB。在发送侧装置1以及接收侧装置2中,数据A、数据B、数据C的偏移分别为0、100000(100K)、200000(200K)。此外,接收侧装置2以组块的形式保持各数据(参照图8),然而在图12中,为了方便,与发送侧装置1相同地示出。

[0143] 具体例假定发送侧装置1在应用程序Y的数据X的发送后,发送数据A、数据B的情况。数据X的尺寸为100KB。在该通信中数据X、数据A、数据B的偏移分别为0、100000(100K)、200000(200K)。

[0144] 图13A以及图13B是具体例中的时序图的一个例子。在图13A中,示出了发送侧装置1在应用程序Y的数据X之后发送数据A的情况。此外,若从客户端装置向服务器装置发送连接要求,且发送侧装置1接收该连接要求,则通过从发送侧装置1向接收侧装置2发送连接要求来进行发送侧装置1与接收侧装置2之间的会话的确立处理。另外,与此同时,也在客户端装置与发送侧装置1、接收侧装置2与服务器装置之间确立会话。

[0145] 在S1中,客户端装置发送应用程序Y的数据A。在S2中,发送侧装置1接收应用程序Y的数据A(图10、OP2)。

[0146] 在S3中,由于动作模式是初始值的通常模式(图10、OP5:否),所以发送侧装置1将数据A发送给接收侧装置2(图10、OP11)。

[0147] 在S4中,接收侧装置2接收数据A(图11、OP22)。在S5中,由于接收数据为实际数据A(图11、OP23:否),所以接收侧装置2将数据A分割为组块,计算各组块的SHA1,并搜索缓存,检测数据A的重复接收(图11、OP26~OP29)。

[0148] 此时,在应用程序X的接收侧缓存28检测到数据A(参照图12)。即,重复应用程序为应用程序X,重复应用程序重复开始位置为应用程序X的接收侧缓存28内的接收偏移0。

[0149] 在S6中,接收侧装置2更新重复预测表格(图11、0P33~0P34)。重复预测模式从关闭变更为开启。在S7中,接收侧装置2将重复预测信息发送给接收侧装置2(图11、0P35)。重复预测信息所包含的信息如以下。

[0150] (重复预测信息)

[0151] 当前应用程序种类:应用程序X

[0152] 重复预测模式:开启

[0153] 当前应用程序重复开始位置:100000

[0154] 重复应用程序种类:应用程序Y

[0155] 重复应用程序重复开始位置:0

[0156] 在S8中,发送侧装置1接收重复预测信息(图10、0P2、0P3:是)。在S9中,发送侧装置1将重复预测信息复印到重复预测表格17(图10、0P4)。在S10中,接收侧装置2向目的地发送数据A(图11、0P25)。

[0157] 图13B示出接着图13A的处理,发送侧装置1发送应用程序Y的数据B的情况。在S11中,客户端装置发送应用程序Y的数据B。在S12中,发送侧装置1接收应用程序Y的数据B(图10、0P2)。

[0158] 在S13中,由于动作模式为重复预测模式(图10、0P5:是),所以发送侧装置1对数据B进行重复判定。此时,从重复预测表格内的重复应用程序种类亦即应用程序X的发送侧缓存16获取部分数据。

[0159] 在重复预测信息中,重复应用程序种类:应用程序X,重复应用程序重复开始位置:0,当前应用程序重复开始位置:100000。另外,在应用程序Y中数据B的偏移为200000。因此,应用程序X的发送侧缓存16中的部分数据的偏移为 $0 + (200000 - 100000) = 100000$ 。

[0160] 即,作为部分数据,获取应用程序X的发送侧缓存16的发送偏移100000的数据。部分数据为数据B(参照图12),所以发送侧装置1检测到重复发送(图10、0P7:是)。

[0161] 在S14中,发送侧装置1发送数据B的索引。在数据B的索引包含有以下信息。

[0162] (数据B的索引)

[0163] 当前应用程序种类:Y

[0164] 重复应用程序种类:X

[0165] 重复应用程序重复开始位置:100000

[0166] 数据的长度:100KB

[0167] 在S15中,接收侧装置2接收数据B的索引(图11、0P22、0P23:是)。在S16中,接收侧装置2基于索引所包含的信息,读入应用程序X的接收侧缓存28的接收偏移100000~200000所包含的组块(图11、0P24)。在S17中,从组块恢复数据B,并向目的地发送(图11、0P25)。

[0168] <第一实施方式的作用效果>

[0169] 在第一实施方式中,接收侧装置2检测数据的重复接收,并向发送侧装置1通知重复预测信息,发送侧装置1基于重复预测信息,进行预发送数据与部分数据的比较这样的负荷比较轻的重复判定处理。因此,在第一实施方式中,在重复数据的去除处理中,能够降低发送侧装置1的负荷。

[0170] 由接收侧装置2进行组块分割、SHA1计算、哈希搜索等而不是由发送侧装置1进行，从而例如，能够使发送侧装置1的CPU的利用率大致为四分之一。

[0171] 在重复预测信息包含有重复应用程序、和重复应用程序中的重复开始位置等信息，所以发送侧装置1通过参照这些信息，能够容易地获取成为预发送数据的比较对象的部分数据。由此，能够进一步降低发送侧装置1的处理负荷。

[0172] 另外，发送侧装置1在以重复预测模式进行动作的情况下，也在预发送数据与重复应用程序的部分数据进行重复检测处理。在预发送数据内包含新的部分的情况下，发送侧装置1即使以重复预测模式进行动作，也向接收侧装置2发送预发送数据。在接收侧装置2接收了该预发送数据（对于接收侧装置2来说是接收数据）的情况下，该接收数据直接被转送到目的地（图11、OP25）。因此，即使在发送侧装置1以重复预测模式进行动作时在预发送数据包含新的数据的情况下，也能够正确地向目的地发送该新数据。

[0173] 另外，在接收侧装置2接收包含新数据的接收数据，且对于该接收数据未找到重复的组块的情况下，接收侧装置2将重复预测模式设定为“关闭”并将重复预测信息发送给发送侧装置1。若发送侧装置1接收该重复预测信息，则发送侧装置1以通常模式开始动作。由此，接收侧装置2能够控制发送侧装置1的动作模式。

[0174] 以上，通过降低发送侧装置1的负荷，作为发送侧装置1，例如，能够采用智能手机、平板终端、PC等性能比较低的终端装置。换句话说，能够将第一实施方式所涉及的数据去重客户端程序安装于终端装置。

[0175] 图14是表示第一实施方式所涉及的数据去重系统100的效果的图。在图14中，示出使用转送速度10Mbps的物理线路bw来转送100Mbit的流量数据X的情况下的、图1所示的数据去重系统P100和第一实施方式所涉及的数据去重系统100的转送时间。

[0176] 将发送侧装置P1中的数据去重处理所涉及的速度（去重速度）dv设为10Mbps。将发送侧装置1中的去重速度dw设为40Mbps。这是因为发送侧装置1的CPU的利用率与发送侧装置进行组块分割、SHA1计算、哈希搜索等的情况相比，大致为四分之一。

[0177] 另外，假设流量数据100Mbit中90%重复。该情况下，通过数据去重处理，以实际数据发送的数据的比例削减到10%。即，数据去重处理所带来的实际数据的削减率（去重削减率）r为0.1。

[0178] 在上述前提中，(1) 在未进行数据去重处理的情况下，转送时间为 $X/bw = 100M/10Mbps = 10s$ 。(2) 在数据去重系统P100中， $X/dv + X * r/bw = 100M/10Mbps + 100M * 0.1/10Mbps = 11s$ 。(3) 在数据去重系统100中， $X/dw + X * r/bw = 100M/40Mbps + 100M * 0.1/10Mbps = 3.5s$ 。

[0179] 即，第一实施方式所涉及的数据去重系统100的数据转送时间最短。因此，根据第一实施方式，在利用10Mbps等低速的线路的通信中，通过降低发送侧装置1的处理负荷，能够缩短数据转送所涉及的时间。

[0180] 另外，发送侧装置1以及接收侧装置2在进行任意的处理的情况下，均不管发送或者接收的数据的内容。即，发送侧装置1以及接收侧装置2无论应用程序处理何种数据，均不管数据的内容，而能够进行数据去重处理。因此，第一实施方式的数据去重系统100能够应用于进行任意的应用程序的通信的系统。

[0181] <变形例>

[0182] 例如,也可以在接收侧装置2的CPU所涉及的处理负荷比规定的阈值高的情况下,发送侧装置1进行组块分割、SHA1计算、哈希搜索的处理。例如,能够通过向发送侧装置1具备组块分割部、SHA1计算部、哈希搜索部,且对重复预测信息的重复预测模式追加“卸载(off load)”来实现。

[0183] 接收侧装置2在CPU负荷超过规定阈值的情况下,向发送侧装置1发送重复预测模式设定了“卸载”的重复预测信息。若接收该重复预测信息,则发送侧装置1将预发送数据输出给组块分割部、SHA1计算部、哈希搜索部,且例如与图1的发送侧装置P1同样地进行处理。

[0184] 由此,在接收侧装置2的处理负荷变高的情况下,能够使接收侧装置2的处理负荷降低。

[0185] 另外,在第一实施方式中,未提及识别应用程序的用户,但在存在多个应用程序的用户的情况下,分别按照每个用户并且按照每个应用程序生成发送侧缓存16以及接收侧缓存28。即,在存在多个应用程序的用户的情况下,以用户识别信息和应用程序识别信息的组合识别发送侧缓存16以及接收侧缓存28。

[0186] <记录介质>

[0187] 能够将使计算机、其它的设备、装置(以下,称为计算机等)实现上述任意一个功能的程序记录于计算机等能够读取的记录介质。通过使计算机等读入该记录介质的程序并执行,能够提供该功能。

[0188] 这里,计算机等能够读取的记录介质是指通过电、磁、光学、机械、或者化学的作用积蓄数据、程序等信息,且能够由计算机等读取的非临时的记录介质。作为这样的记录介质中能够从计算机等取下的记录介质,例如有软盘、光磁盘、CD-ROM、CD-R/W、DVD、蓝光盘、DAT、8mm磁带、闪存等存储卡等。另外,作为固定于计算机等的记录介质有硬盘、ROM(只读存储器)等。并且,SSD(Solid State Drive:固盘)既能够作为能够从计算机等取下的记录介质利用,也能够作为固定于计算机等的记录介质利用。

[0189] 符号说明

[0190] 1…发送侧装置,2…接收侧装置,11、21…数据接收部,12…重复判定部,13…缓存保存部,14、26…数据发送部,15…重复预测接收部,16…发送侧缓存,17、29…重复预测表格,22A…组块分割部,22B…SHA1计算部,22C…哈希搜索部,23…缓存读出部,24…缓存保存部,25…表格管理部,27…重复预测发送部,28…发送侧缓存,101…CPU,102…主存储装置,105…辅助存储装置,107…网络接口。

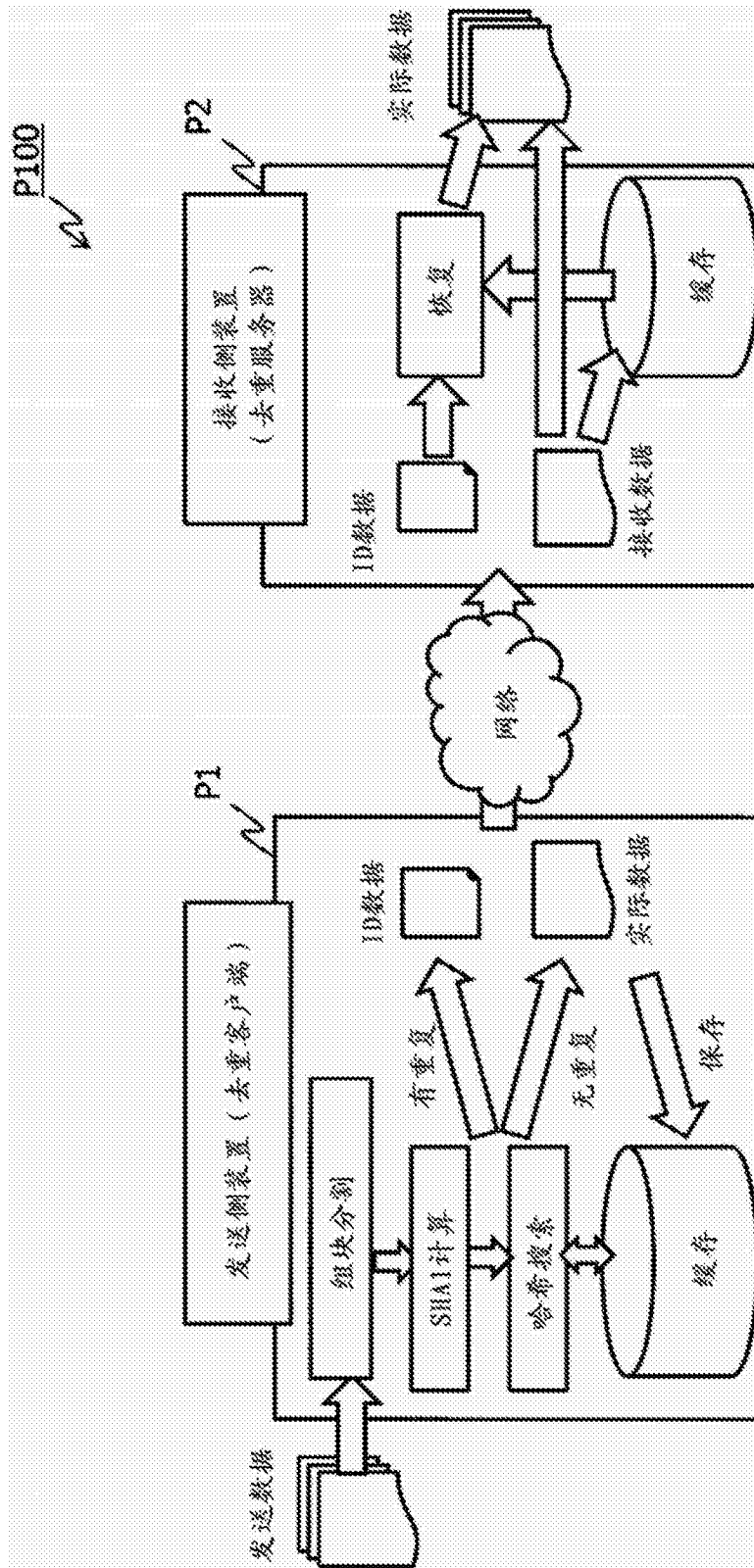


图1

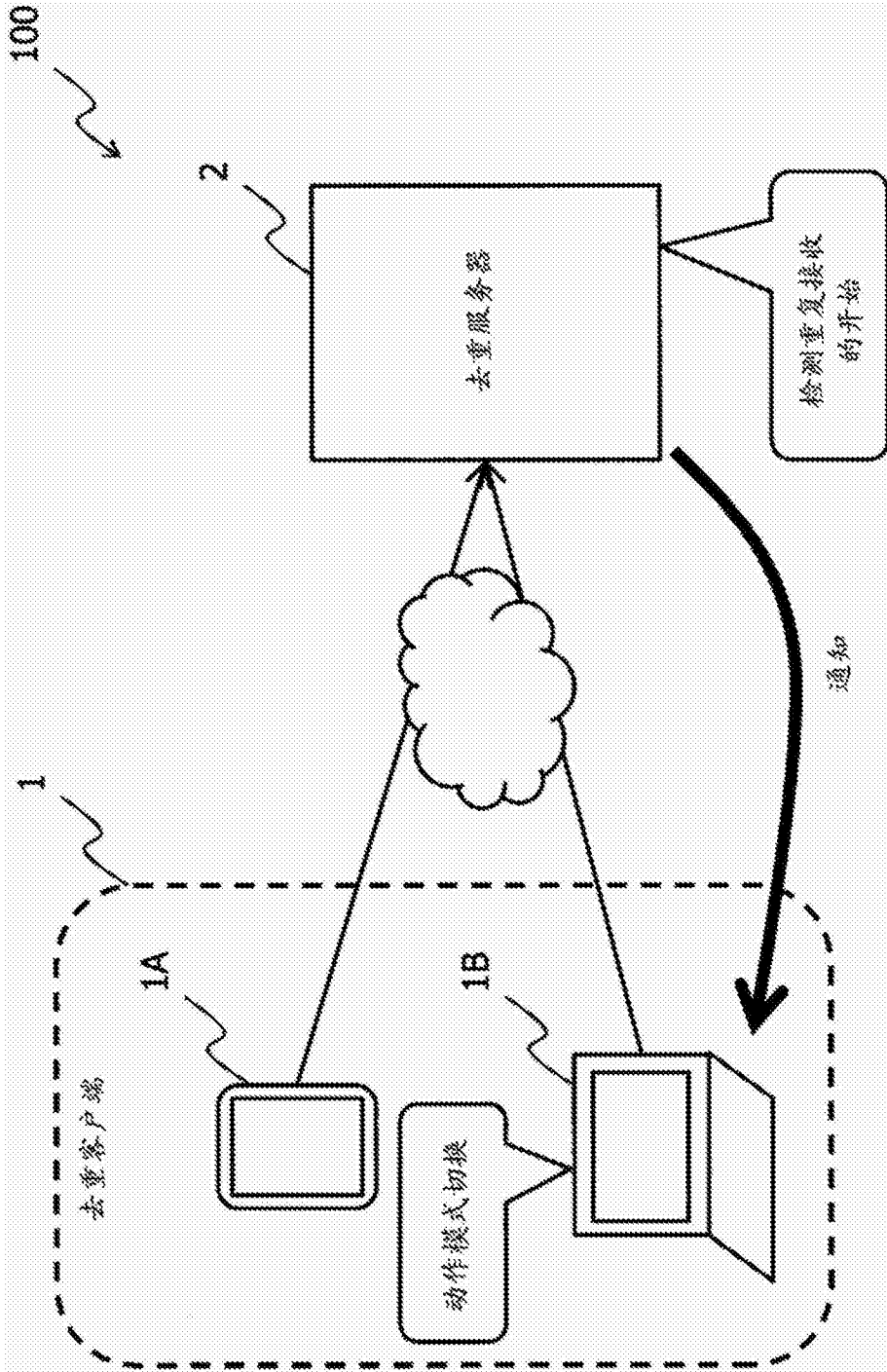


图2

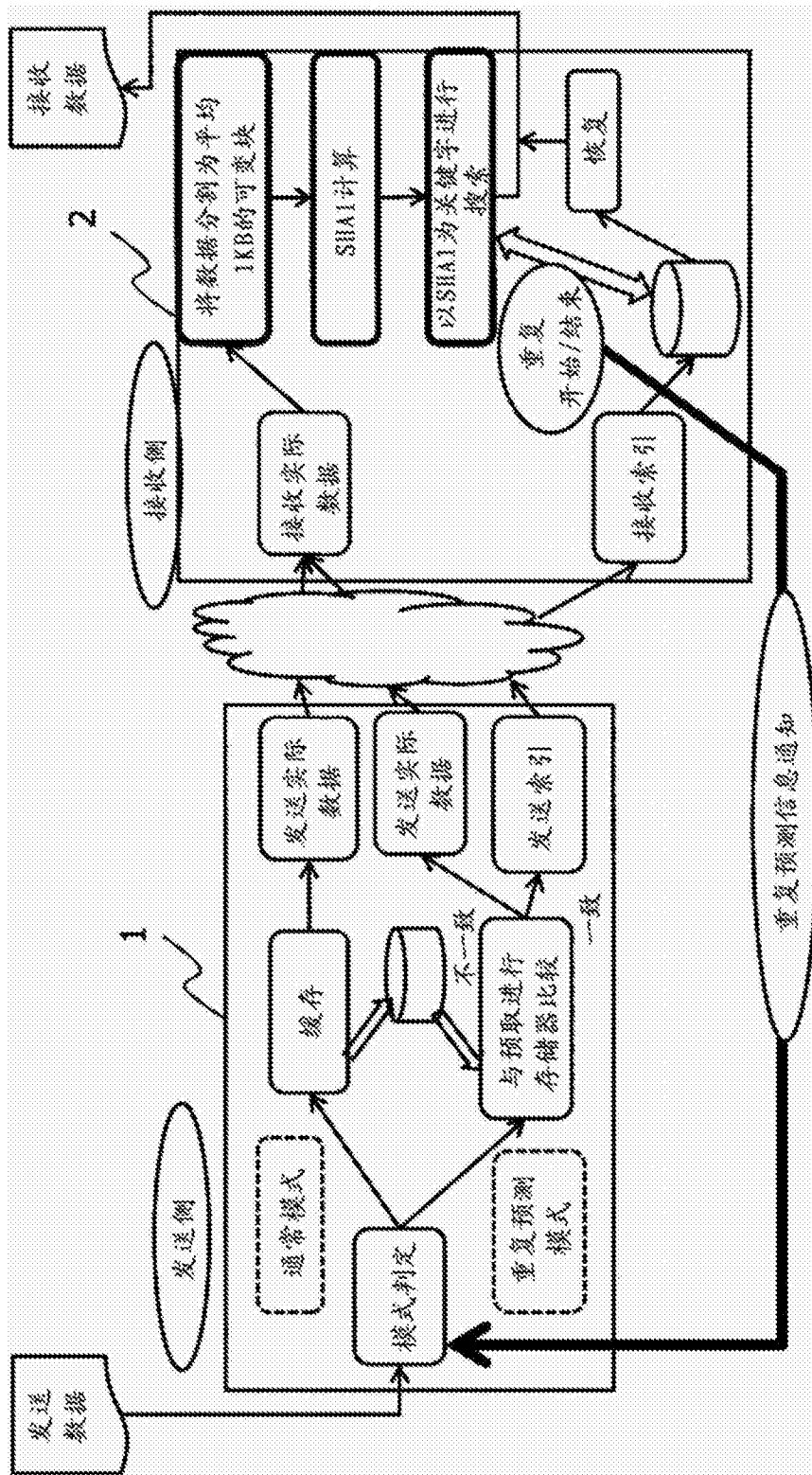


图3

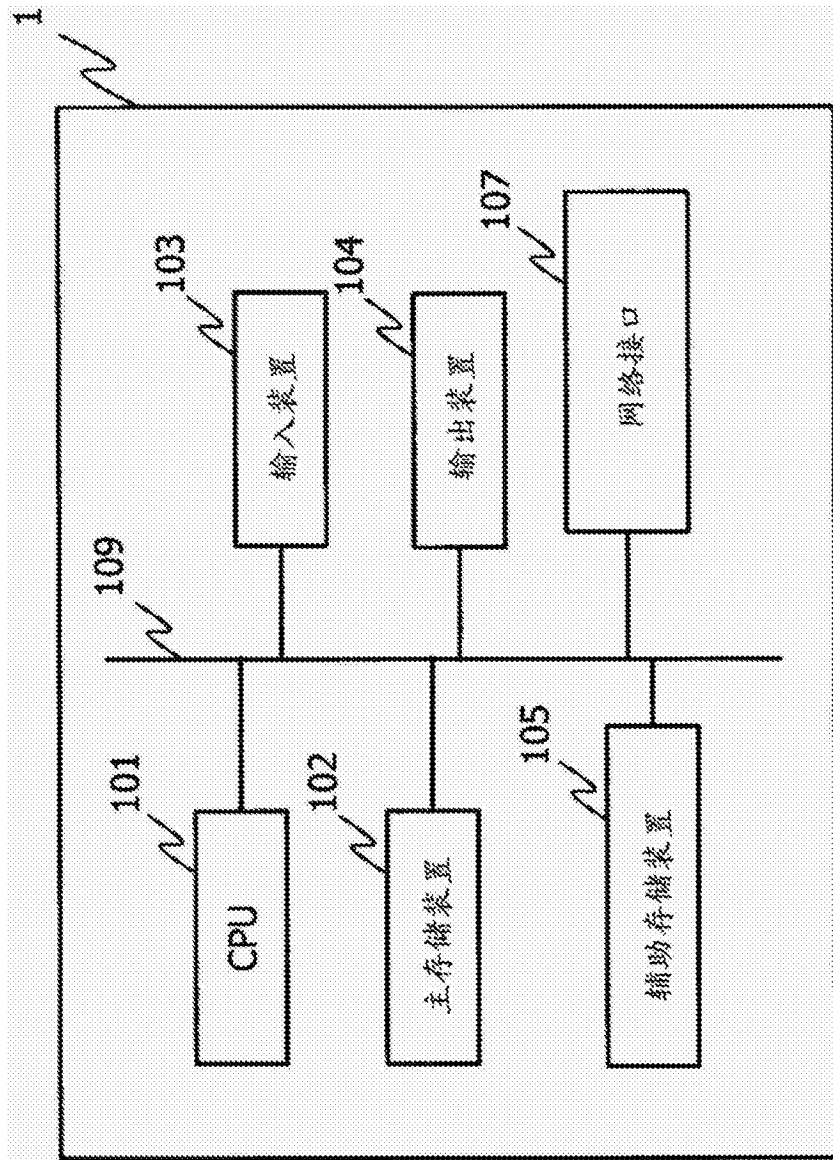


图4

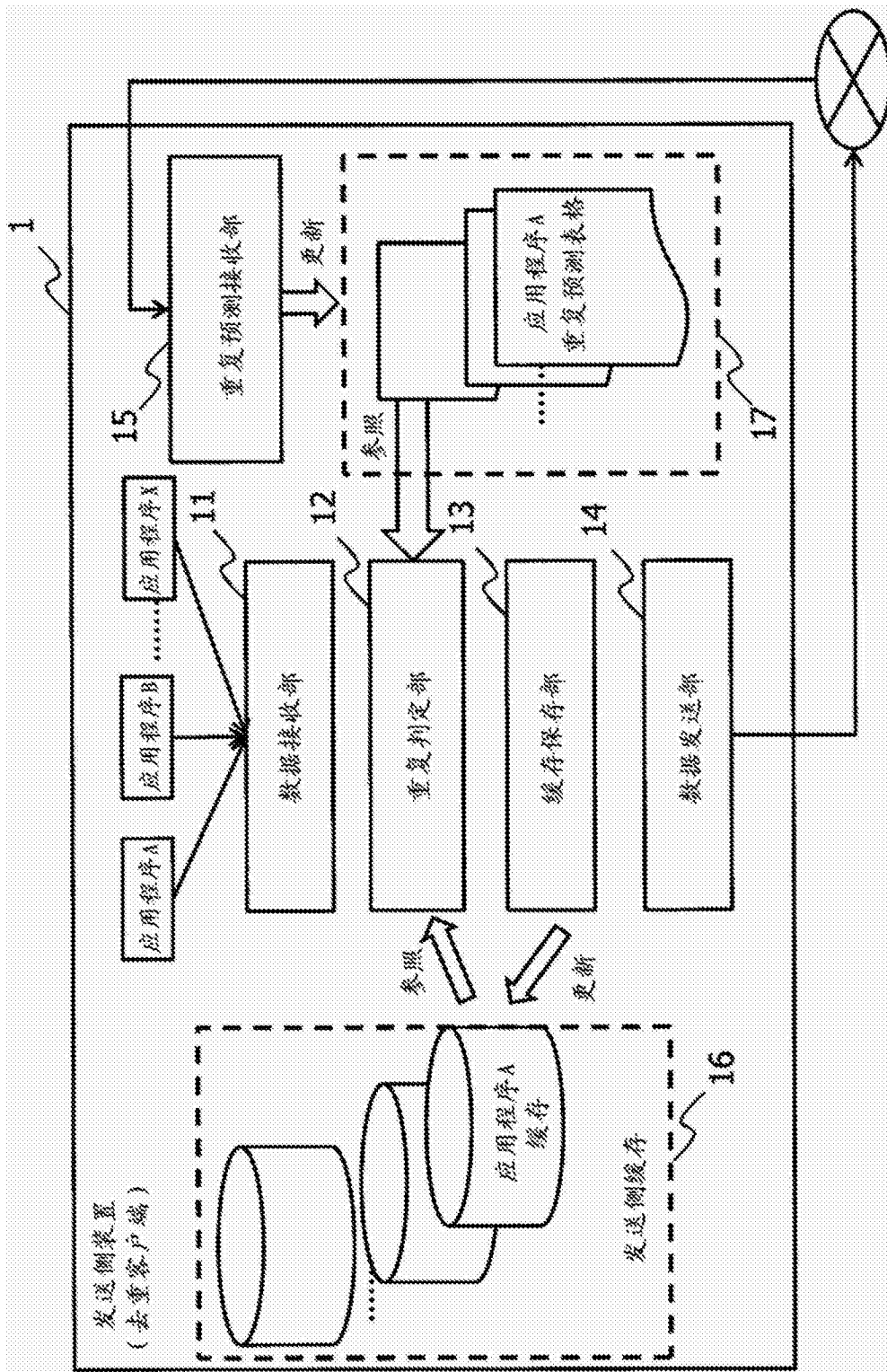


图5

发送侧缓存 (按照每个应用程序)

发送偏移	发送数据
0	xyzxyz....
80000	abcabc....

图6

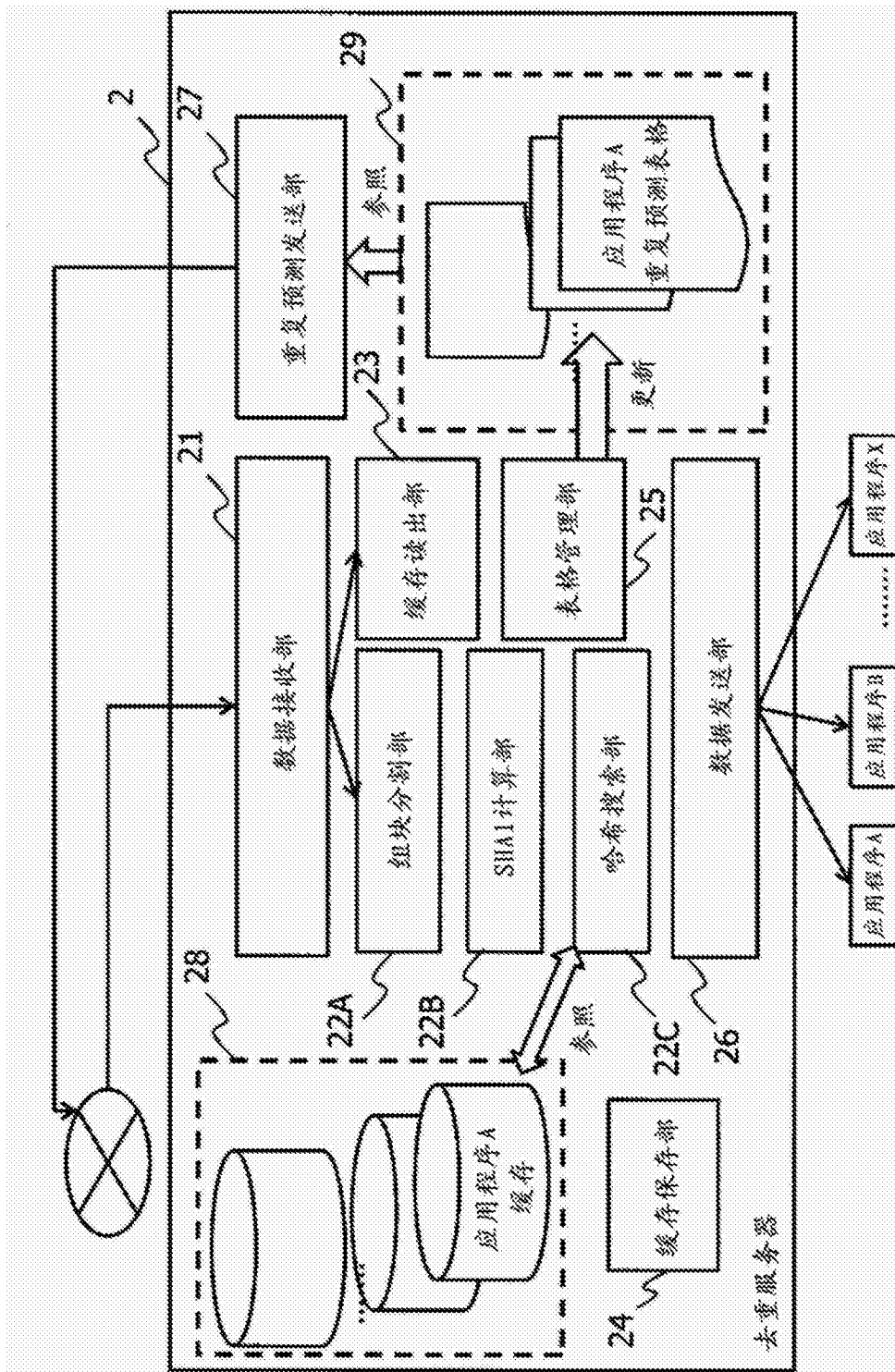


图7

接收侧缓存

哈希值	应用程序ID	接收偏移	接收数据
kkkkk	x	0	xyxyz....
hhhhh	x	1000	uvwuvw.....
.....			
ijjjj	x	80000	abcabc....

图8

重复预测信息/重复预测表格

项目编号	名称	值
1	当前应用程序种类	识别信息
2	重复预测模式	开启/关闭
3	重复应用程序种类	识别信息
4	当前应用程序重复开始位置	偏移
5	重复应用程序重复开始位置	偏移
6	重复预测长度	数据的长度

图9

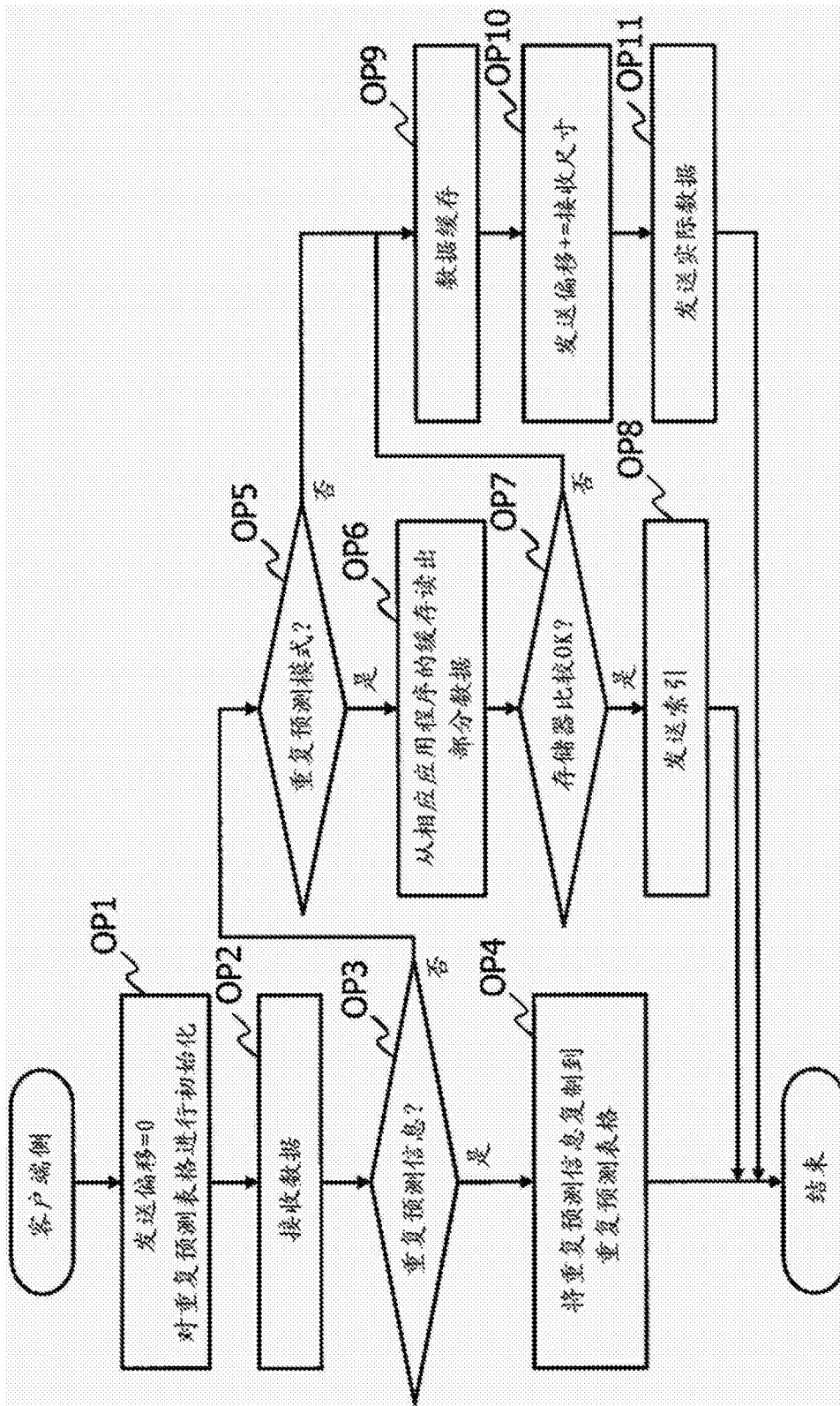


图10

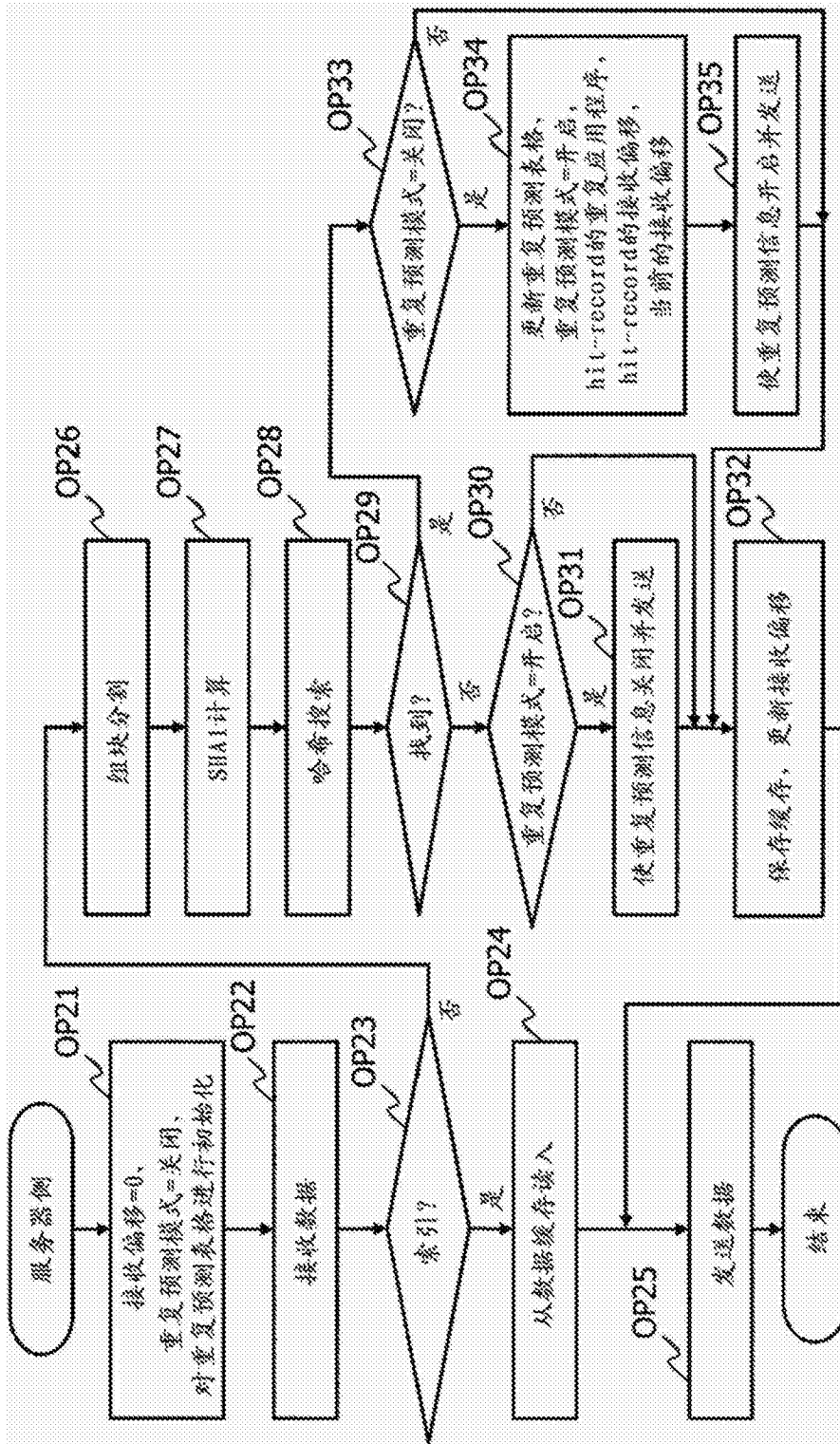


图11

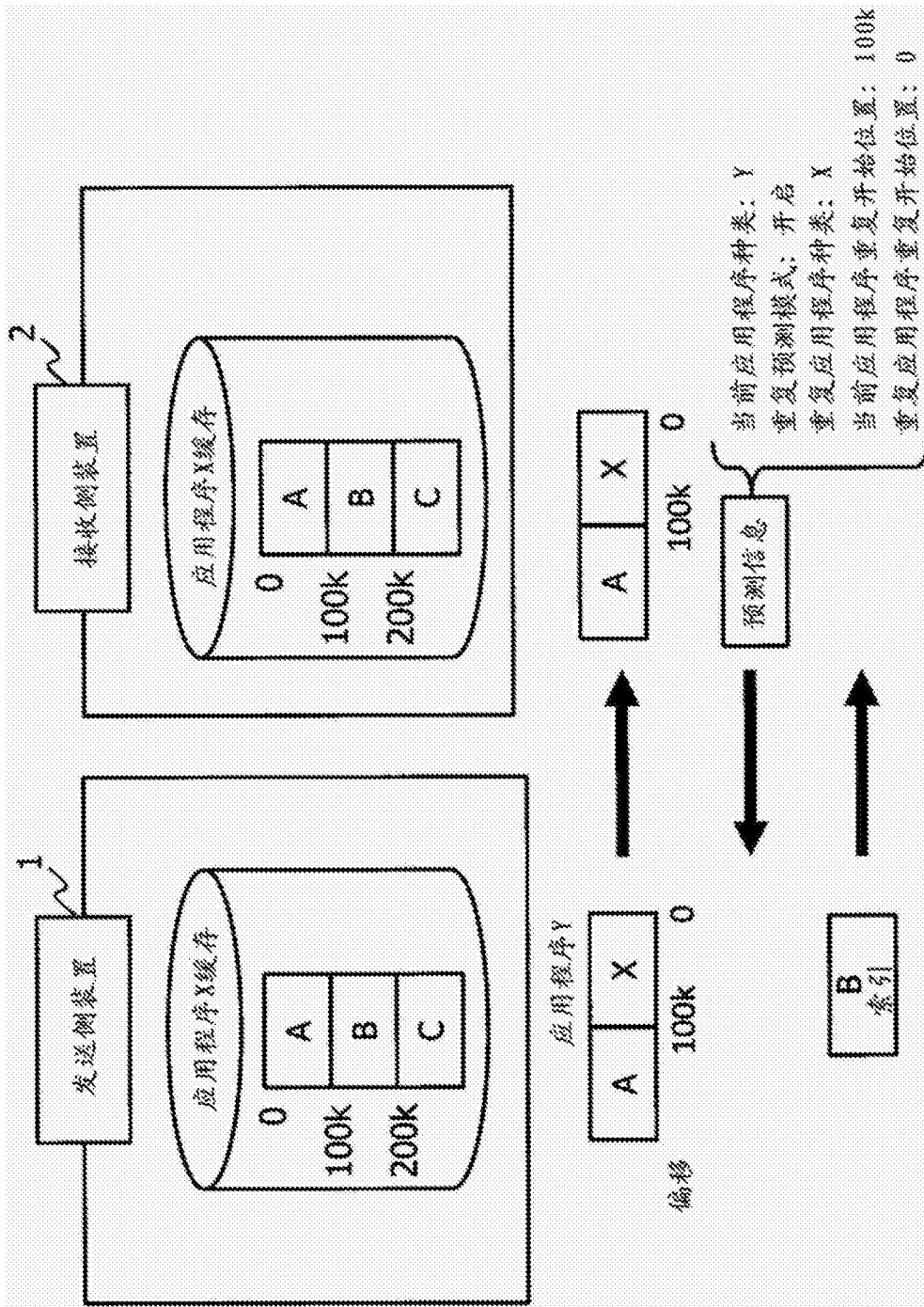


图12

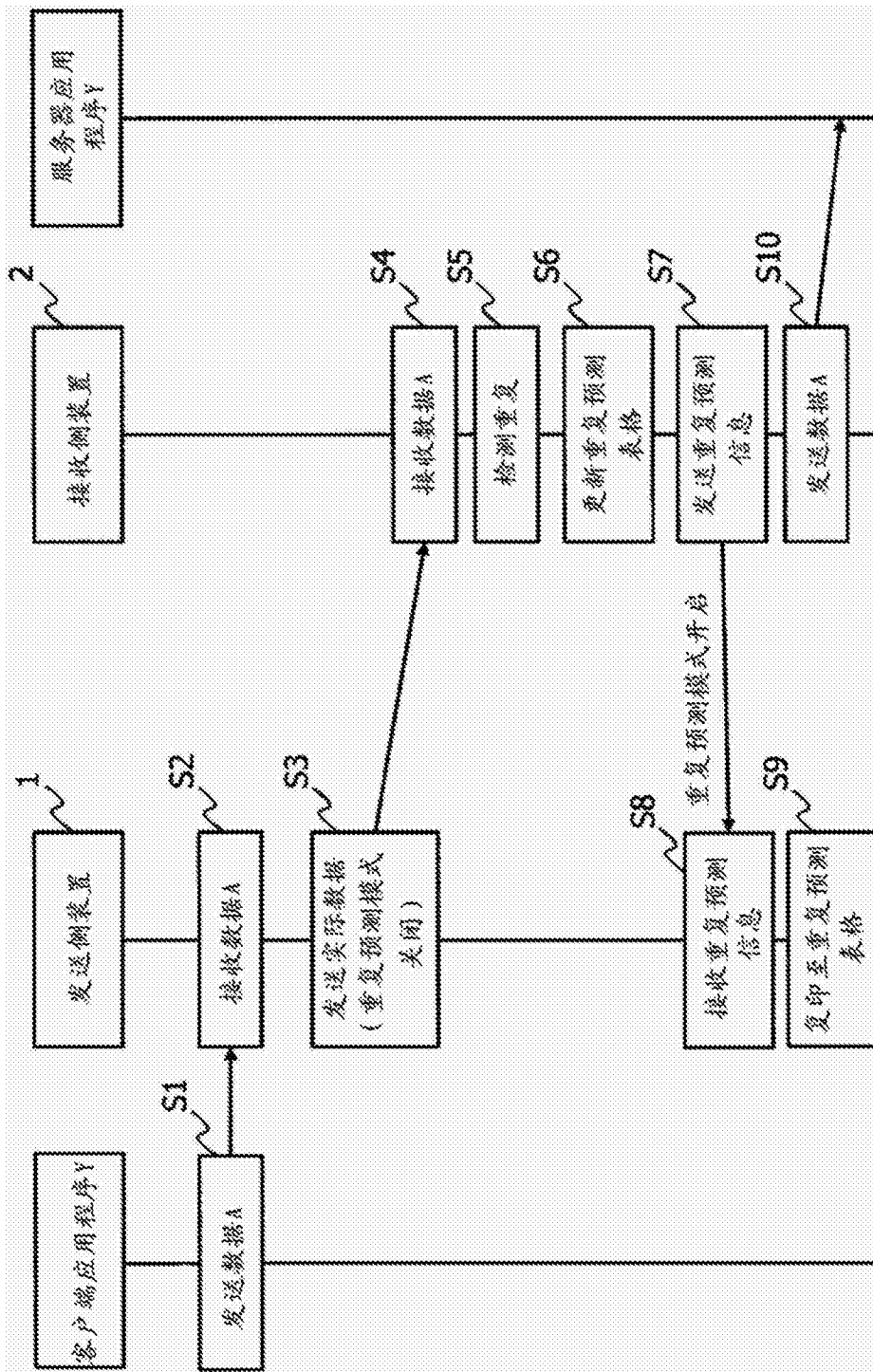


图13A

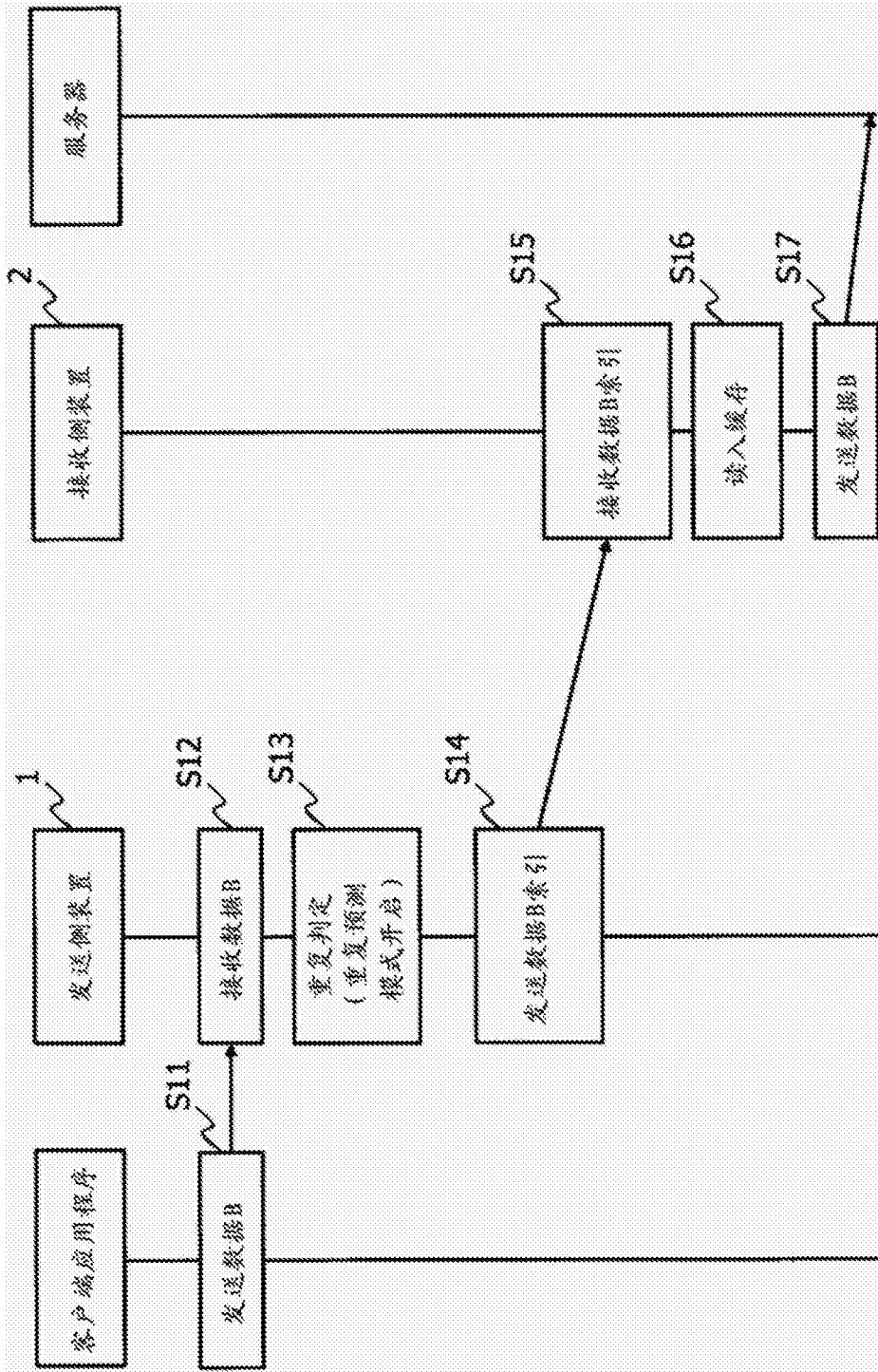


图13B

- 物理线路 : bw (e.g. 10Mbps)
- 发送侧装置P1的以往的去重速度 : dv (e.g. 10Mbps)
发送侧装置1的去重速度 : dw (e.g. 40Mbps)
- 流量数据 : X (e.g. 100Mbit)
- 流量的去重削减率: r : (e.g. 若重复90%则为0.1)

转送时间的比较

- (1) 无去重 = $X/bw = 100/10 = 10$
- (2) 去重系统P100 $\Rightarrow X/dv + X*r/bw = 100/10 + 100*0.1/10 = 11$
- (3) 去重系统100 $\Rightarrow X/dw + X*r/bw = 100/40 + 100*0.1/10 = 3.5$

图14