



(12)发明专利

(10)授权公告号 CN 104396197 B

(45)授权公告日 2017.10.27

(21)申请号 201380032211.7

(22)申请日 2013.04.10

(65)同一申请的已公布的文献号
申请公布号 CN 104396197 A

(43)申请公布日 2015.03.04

(30)优先权数据
13/452780 2012.04.20 US

(85)PCT国际申请进入国家阶段日
2014.12.19

(86)PCT国际申请的申请数据
PCT/IB2013/052859 2013.04.10

(87)PCT国际申请的公布数据
W02013/156903 EN 2013.10.24

(73)专利权人 瑞典爱立信有限公司
地址 瑞典斯德哥尔摩

(72)发明人 D.I.艾伦

(74)专利代理机构 中国专利代理(香港)有限公司 72001

代理人 徐予红 徐红燕

(51)Int.Cl.
H04L 12/703(2006.01)
H04L 12/761(2006.01)
H04L 12/707(2006.01)
H04L 12/721(2006.01)

(56)对比文件
US 2012057603 A1,2012.03.08,
CN 101965715 A,2011.02.02,
CN 101322355 A,2008.12.10,
IEEE.Bridges and Virtual Bridged
Local Area Networks Amendment 8:Shortest
Path Bridging.《INTERWORKING TASK GROUP OF
IEEE 802.1》.2011,(第802.1aq/D4.5期),第
28.5部分-28.8部分.

审查员 何思佳

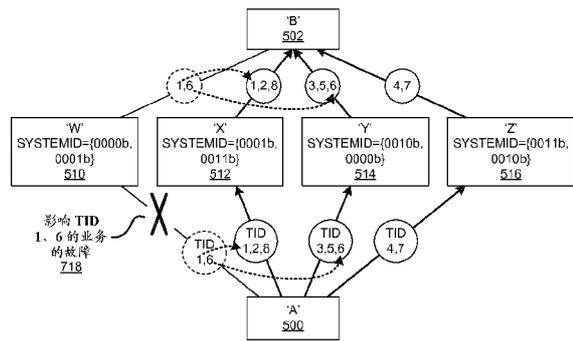
权利要求书4页 说明书15页 附图10页

(54)发明名称

在802.1aq网络中使用分离的平局打破器在等成本最短路径之间选择

(57)摘要

通信网络中的节点利用分离的平局打破器(split tiebreaker)在等成本最短路径(ECSP)之间选择。该节点公告该节点的多个系统标识符(ID),并且每个系统ID与区分网络中的不同虚拟网络的不同的业务ID集合相关联。该节点接收网络中的每个其它节点的系统ID集合。在两个节点之间构造多个不同的ECSP之后,该节点为在这两个节点之间所使用的每个业务ID选择ECSP之一。对于业务ID,该节点利用路径中的每个节点的与该业务ID相关联的一个系统ID构造每个ECSP的路径ID。由于网络中的系统ID和业务ID的配置,将跨多个残存ECSP分配受故障影响的业务。



1. 一种通过通信网络中的多个节点中的节点执行以便在所述网络中利用分离的平局打破器在等成本最短路径之间选择的方法,包括以下步骤:

公告所述节点的第一和第二系统标识符(ID),其中所述第一和第二系统标识符分别与第一和第二业务标识符集合相关联,其中业务标识符用于区分所述网络内的不同虚拟网络,并且其中每个业务标识符与平局打破变换相关联;

接收所述网络中的其它多个节点中的每个节点的第三和第四系统标识符,其中所述第三和第四系统标识符分别与所述第一和第二业务标识符集合相关联;

在所述网络的第一节点和第二节点之间构造多个不同的等成本最短路径;

通过以下步骤为每个业务标识符选择所述构造的等成本最短路径之一:

利用与该业务标识符相关联的所述平局打破变换以及包括所述系统标识符中与该业务标识符相关联的那些系统标识符的系统标识符集合为每个所述等成本最短路径构造路径标识符,并且其中所述系统标识符集合中的每个系统标识符与该等成本最短路径的不同节点相关联,以及

对所述路径标识符应用选择算法以选择所述等成本最短路径之一;以及

响应于影响一个或多个所选择的等成本最短路径上的多个业务标识符的业务故障,由于所述网络中的所述系统标识符和所述业务标识符的配置,跨多个其它等成本最短路径分配与那些业务标识符相关联的业务。

2. 如权利要求1所述的方法,其中构造所述多个不同的等成本最短路径包括计算所述网络中的所述多个节点的所有对最短路径。

3. 如权利要求1所述的方法,其中为每个所述等成本最短路径构造所述路径标识符包括:

对所述系统标识符集合中的每个系统标识符应用与该业务标识符相关联的所述平局打破变换;

将每个变换的系统标识符按字典顺序排序;以及

连结所述排序的变换的系统标识符。

4. 如权利要求1所述的方法,其中所述选择算法包括:

将所述路径标识符分等级;以及

选择所述分等级的路径标识符的最低值。

5. 如权利要求1所述的方法,其中所述通信网络是802.1aq网络,并且其中所述业务标识符是骨干虚拟局域网标识符(B-VID)。

6. 如权利要求1所述的方法,其中:

所述故障影响一个选择的等成本最短路径上的N个业务标识符的业务,其中N是大于1的整数;并且

跨N个不同的等成本最短路径分配所述N个业务标识符的业务。

7. 如权利要求6所述的方法,其中在引起所述故障的故障组件恢复时,通过所述一个选择的等成本最短路径转发所述N个不同业务标识符的所述业务。

8. 一种网络元件,其将耦合到通信网络中的其它网络元件的集合,配置成在所述网络中利用分离的平局打破器在等成本最短路径之间选择,所述网络元件包括:

链路状态协议模块,配置成:

管理链路状态数据库,所述链路状态数据库表示所述网络的拓扑并包括多个网络元件以及所述多个网络元件中的每个网络元件的多个系统标识符(ID),

利用所述网络传送链路状态消息以公告所述网络元件的第一和第二系统标识符,其中所述第一和第二系统标识符分别与第一和第二业务标识符集合相关联,其中每个业务标识符用于区分所述网络内的不同虚拟网络,并且其中每个业务标识符与平局打破变换相关联,

接收指示指派给所述网络中的其它多个网络元件中的每个网络元件的第三和第四系统标识符的链路状态消息,其中所述第三和第四系统标识符分别与所述第一和第二业务标识符集合相关联,

利用所述链路状态数据库中的信息在所述网络的第一网络元件和第二网络元件之间构造多个不同的等成本最短路径,

对于每个业务标识符,利用与该业务标识符相关联的所述平局打破变换以及包括所述系统标识符中与该业务标识符相关联的那些系统标识符的系统标识符集合构造所述多个不同的等成本最短路径中的每个等成本最短路径的路径标识符,并且其中所述系统标识符集合中的每个系统标识符与该等成本最短路径的不同网络元件相关联,

通过对每个构造的路径标识符应用选择算法来为所述业务标识符中的每个业务标识符选择对应的构造的等成本最短路径之一,以及

使得配置一个或多个转发条目以反映所述选择的等成本最短路径;以及

转发模块,配置成:

从端口接收多个分组,

在所述网络中发生影响一个或多个等成本最短路径上的多个业务标识符的业务的故障之前,根据所述转发条目通过所述一个或多个等成本最短路径传送所述多个业务标识符的业务,以及

在所述故障之后,由于所述网络中的所述系统标识符和所述业务标识符的配置,通过多个不同的等成本最短路径传送所述多个业务标识符的业务。

9. 如权利要求8所述的网络元件,其中所述链路状态协议模块配置成通过计算所述网络中的所述多个网络元件的所有对最短路径来构造所述多个不同的等成本最短路径。

10. 如权利要求8所述的网络元件,其中所述链路状态协议模块配置成通过以下步骤来构造所述多个等成本最短路径中的每个等成本最短路径的所述路径标识符:

对所述系统标识符集合中的每个系统标识符应用与该业务标识符相关联的所述平局打破变换;

将每个变换的系统标识符按字典顺序排序;以及

连结所述排序的变换的系统标识符。

11. 如权利要求8所述的网络元件,其中所述选择算法包括:

将所述构造的路径标识符分等级;以及

选择所述分等级的路径标识符的最低值。

12. 如权利要求8所述的网络元件,其中所述通信网络是802.1aq网络,并且其中所述业务标识符是骨干虚拟局域网标识符(B-VID)。

13. 如权利要求8所述的网络元件,其中:

所述故障影响一个选择的等成本最短路径上的N个业务标识符的业务,其中N是大于1的整数;并且

在所述故障之后,跨N个不同的等成本最短路径传送所述N个业务标识符的业务。

14. 如权利要求13所述的网络元件,其中在引起所述故障的故障组件恢复时,通过所述一个选择的等成本最短路径传送所述N个业务标识符的所述业务。

15. 一种通信网络,包括:

多个网络元件,其中每个网络元件在通信上耦合到至少一个其它网络元件,并且其中每个网络元件包括处理器和耦合到所述处理器并具有指令的非暂时性机器可读介质,所述指令在由所述处理器执行时使所述处理器利用分离的平局打破器通过执行以下操作来在所述网络的等成本最短路径之间选择:

公告所述网络元件的第一和第二系统标识符(ID),其中所述第一和第二系统标识符分别与第一和第二业务标识符集合相关联,其中业务标识符用于区分所述网络内的不同虚拟网络,并且其中每个业务标识符与平局打破变换相关联;

接收所述网络中的其它多个网络元件中的每个网络元件的第三和第四系统标识符,其中所述第三和第四系统标识符分别与所述第一和第二业务标识符集合相关联;

在所述网络的第一和第二网络元件之间构造多个不同的等成本最短路径;

通过以下步骤为每个业务标识符选择所述构造的等成本最短路径之一:

利用与该业务标识符相关联的所述平局打破变换以及包括所述系统标识符中与该业务标识符相关联的那些系统标识符的系统标识符集合为每个所述等成本最短路径构造路径标识符,并且其中所述系统标识符集合中的每个系统标识符与该等成本最短路径的不同网络元件相关联,以及

对所述路径标识符应用选择算法以选择所述等成本最短路径之一;以及

响应于影响一个或多个所选择的等成本最短路径上的多个业务标识符的业务的故障,由于所述网络中的所述系统标识符和所述业务标识符的配置,跨多个其它等成本最短路径分配与那些业务标识符相关联的业务。

16. 如权利要求15所述的通信网络,其中构造所述多个不同的等成本最短路径包括计算所述网络中的所述多个网络元件的所有对最短路径。

17. 如权利要求15所述的通信网络,其中为每个所述等成本最短路径构造所述路径标识符包括:

对所述系统标识符集合中的每个系统标识符应用与该业务标识符相关联的所述平局打破变换;

将每个变换的系统标识符按字典顺序排序;以及

连结所述排序的变换的系统标识符。

18. 如权利要求15所述的通信网络,其中所述选择算法包括:

将所述路径标识符分等级;以及

选择所述分等级的路径标识符的最低值。

19. 如权利要求15所述的通信网络,其中所述通信网络是802.1aq网络,并且其中所述业务标识符是骨干虚拟局域网标识符(B-VID)。

20. 如权利要求15所述的通信网络,其中在引起所述故障的故障组件恢复时,通过所述

一个或多个选择的等成本最短路径转发与所述受影响的多个业务标识符相关联的业务。

在802.1aq网络中使用分离的平局打破器在等成本最短路径之间选择

技术领域

[0001] 本发明的实施例涉及计算机联网领域;更具体来说,涉及在802.1aq网络中使用分离的平局打破器(split tiebreaker)。

背景技术

[0002] 以太网是由电气和电子工程师协会(IEEE)定义的联网协议,它包含规定物理层的802.3标准和规定网络层的802.1标准。发明以太网以允许局域网(LAN)并允许“即插即用”联网,而只需很少的配置。随着时间,以太网扩展为提供额外的功能性并支持备选网络拓扑。如本文所使用,术语“以太网网络”是指实现包括例如802.1、802.1Q、802.1ad、802.1ah、802.1Qay、802.1Qbp或802.1aq的802.1系列协议中的一个或多个协议的网络。

[0003] IEEE 802.1aq已经发展成支持高级虚拟化、使多路径路由成为可能、并简化载波、企业和云网络的创建和管理。在IEEE 802.1aq网络中,桥接器利用链路状态协议通过公告网络拓扑和逻辑网络成员来控制网络转发以太网帧。链路状态路由协议的两个示例包括开放最短路径优先(OSPF)和中间系统到中间系统(IS-IS),后者在802.1aq网络中使用。

[0004] 在链路状态路由网络中,形成网络的桥接器交换链路状态公告(LSA)以使得每个节点能够具有网络拓扑的同步视图。由于网络中的每个桥接器具有网络拓扑的同步视图并完全知道所需的单播和多播连接性,所以每个桥接器可以计算网络中的任何一对桥接器之间的一个或多个最短路径。利用该信息,每个桥接器相应地各自填充转发信息库(FIB)。

[0005] IEEE 802.1aq在等成本(equal cost)树(ECT)上基于边缘扩展的基础上分配负载,其中网络的单个全网格视为是一个ECT集合。通过与骨干VLAN标识符(B-VID)1:1相关联来在数据平面中标识每个ECT集合。

[0006] 众多联网应用对多路径网络设计具有增加的依赖性,在多路径网络设计中,在网络中的任何两个点之间将存在多个等成本路径。在大多数计算机网络中,适应网络中的一个或多个故障的需要对于确保网络的连续可用性是必不可少的。在“传统”802.1aq网络中,将通过一个或多个周围桥接器观察链路或桥接器的故障,并在整个网络中公告。网络中的每个桥接器将为受故障影响的业务重新计算新路径,并且转发将利用新路径自动继续。但是,当存在故障时,业务会作为一个块切换到故障转移路径。在具有稀少连接性的网络中,这些后果容易理解,因为备选路径的数量受到高度限制。但是,在具有丰富连接性的多路径网络中,受故障影响的两个给定点之间的负载只会切换到若干个可能路径之一而不是跨残存(surviving)的路径集合散布。这可能会降低网络的稳定性,因为故障转移路径现在拥有显著的业务增加,这可使该路径上已有的业务、切换到该路径的业务以及利用受影响的路径中的链路或桥接器的任何其它业务的性能降级。此外,到故障转移路径的业务的这种块切换会因此压倒故障转移路径中的链路和桥接器,从而有效地使那些元件失效。以这种方式,可能由于业务块的重复切换以及所导致的故障转移故障而导致级联网络故障。因此,希望具有这样一种机制,它在故障场景中散布网络负载而无需重新路由已有的未受影响的路

径,并且一旦校正了故障,便将移走的负载恢复回到原始路由。

发明内容

[0007] 根据本发明的一个实施例,一种方法由通信网络中的多个节点中的节点执行,以便在网络中利用分离的平局打破器在等成本最短路径之间选择。在网络中公告该节点的第一和第二系统标识符(ID)。第一和第二系统ID分别与第一和第二业务ID集合相关联。它们对应于802.1aq用语中的ECT集合。业务ID用于区分网络内的不同虚拟网络,并且与平局打破(tiebreaking)变换相关联。该方法还包括接收网络中的其它多个节点中的每个节点的第三和第四系统ID的步骤。第三和第四系统ID分别与第一和第二业务ID集合相关联。该方法还包括以下步骤:在网络的第一节点和第二节点之间构造多个不同的等成本最短路径;以及对于每个业务ID,选择所构造的等成本最短路径之一。选择部分地通过以下步骤来执行:利用与该业务ID相关联的平局打破变换以及包括与该业务ID相关联的那些系统ID的系统ID集合来为每个等成本最短路径构造路径ID。系统ID集合中的每个系统ID与该等成本最短路径的不同节点相关联。选择还部分地通过以下步骤来执行:对路径ID应用选择算法以选择等成本最短路径之一。该方法还包括以下步骤:响应于影响一个或多个所选择的等成本最短路径上的多个业务ID的业务故障,由于网络中系统ID和业务ID的配置,跨多个其它等成本最短路径分配与那些业务ID相关联的业务。

[0008] 根据本发明的另一个实施例,一种耦合到通信网络中的其它网络元件的集合的网络元件配置成在网络中利用分离的平局打破器在等成本最短路径之间选择。该网络元件包括配置成管理链路状态数据库的链路状态协议模块,链路状态数据库表示网络的拓扑并包括多个网络元件以及所述多个网络元件中的每个网络元件的多个系统标识符(ID)。链路状态协议模块还配置成利用网络来传送链路状态消息以公告该网络元件的第一和第二系统ID,其中第一和第二系统ID分别与第一和第二业务ID集合相关联,其中每个业务ID用于区分网络中的不同虚拟网络,并且其中每个业务ID与平局打破变换相关联。链路状态协议模块还配置成接收指示指派给网络中的所述多个其它网络元件中的每个网络元件的第三和第四系统ID的链路状态消息,其中第三和第四系统ID分别与第一和第二业务ID集合相关联。链路状态协议模块还配置成利用链路状态数据库中的信息在网络的第一网络元件和第二网络元件之间构造多个不同的等成本最短路径。链路状态协议模块还配置成对于每个业务ID利用与该业务ID相关联的平局打破变换以及包括与该业务ID相关联的那些系统ID的系统ID集合为所述多个不同的等成本最短路径中的每个路径构造路径ID,其中系统ID集合中的每个系统ID与该等成本最短路径的不同网络元件相关联。链路状态协议模块还配置成通过对所构造的每个路径ID应用选择算法来为每个业务ID选择对应的构造的等成本最短路径之一。链路状态协议模块还配置成使得配置一个或多个转发条目以反映所选择的等成本最短路径。该网络元件还包括配置成从端口接收多个分组的转发模块。该转发模块还配置成:在网络中发生影响一个或多个等成本最短路径上的多个业务ID的业务故障之前,根据转发条目通过这一个或多个等成本最短路径传送所述多个业务ID的业务。该转发模块还配置成:在故障之后,由于网络中系统ID和业务ID的配置,通过多个不同的等成本最短路径传送所述多个业务ID的业务。

附图说明

[0009] 通过参考以下描述以及用于说明本发明的实施例的附图,可以最好地理解本发明。图中:

[0010] 图1示出根据本发明一个实施例的示例性网络以及用于在该示例网络中使用分离的平局打破器的一些过程;

[0011] 图2示出根据本发明一个实施例用于实现802.1aq的分离的平局打破器的网络元件的示例性功能框图;

[0012] 图3示出根据本发明一个实施例用于封装网络业务的分组报头;

[0013] 图4示出根据本发明一个实施例在802.1aq网络中使用分离的平局打破器的流程图;

[0014] 图5示出根据本发明一个实施例用于说明最短路径和业务标识符的示例性网络拓扑;

[0015] 图6示出根据本发明一个实施例利用分离的平局打破器的最短路径的逻辑计算;

[0016] 图7示出根据本发明一个实施例在故障之后如图5所示的示例性网络拓扑;

[0017] 图8示出根据本发明一个实施例利用分离的平局打破器的最短路径的逻辑计算;

[0018] 图9示出根据本发明一个实施例在示例性网络上的多播业务的最短路径;以及

[0019] 图10示出根据本发明一个实施例在故障之后在图9的示例性网络上的多播业务的不同最短路径。

具体实施方式

[0020] 在以下描述中,阐述了众多具体细节。但是,将理解,没有这些具体细节也可以实践本发明的实施例。在其它情况下,没有详细示出公知的电路、结构和技术,以免使本描述晦涩难懂。通过所包含的描述,本领域技术人员将能够在无需过多试验的情况下实现合适的功能性。

[0021] 说明书中提到“一个实施例”、“实施例”、“示例实施例”等时表示,所描述的实施例可以包括特定特征、结构或特性,但不是每个实施例都一定要包括该特定特征、结构或特性。而且,这些短语不一定指相同的实施例。此外,当结合实施例描述特定特征、结构或特性时,认为本领域技术人员知道如何结合其它实施例来实现该特征、结构或特性,而不管是否有明确描述。

[0022] 在以下描述和权利要求书中,可以使用术语“耦合”和“连接”及其衍生词。应了解,这些术语不是要彼此同义。“耦合”用于指示彼此可以或者可以不直接物理或电接触的两个或两个以上元件彼此共同协作或交互。“连接”用于指示在彼此耦合的两个或两个以上元件之间建立通信。

[0023] 本发明的不同实施例可以利用软件、固件和/或硬件的不同组合来实现。因此,图中所示的技术可以利用在一个或多个电子装置(例如,端站、网络元件)上存储并执行的代码和数据来实现。这些电子装置利用计算机可读介质来存储和(在内部和/或通过网络与其它电子装置)通信代码和数据,计算机可读介质可以是例如非暂时性计算机可读存储介质(例如,磁盘、光盘、随机存取存储器、只读存储器、闪存存储器装置、相变存储器)和暂时性

计算机可读传输介质(例如,电、光、声或其它形式的传播信号,例如载波、红外信号、数字信号)。另外,这些电子装置通常包括耦合到诸如一个或多个存储装置(非暂时性机器可读存储介质)、用户输入/输出装置(例如,键盘、触摸屏和/或显示器)和网络连接的一个或多个其它组件的一个或多个处理器的集合。处理器集合与其它组件的耦合通常是通过一个或多个总线和桥接器(又称为总线控制器)。因此,给定电子装置的存储装置通常存储用于在该电子装置的一个或多个处理器的集合上执行的代码和/或数据。

[0024] 如本文所使用,网络元件(例如,路由器、交换机、桥接器或广义地指节点)是一台包括硬件和软件的联网设备,它在通信上互连网络上的其它设备(例如,其它网络元件、端站)。一些网络元件是对多个联网功能(例如,路由、桥接、交换、第2层聚合、会话边界控制、服务质量和/或订户管理)提供支持和/或对多个应用服务(例如,数据、语音和视频)提供支持的“多服务网络元件”。订户端站(例如,服务器、工作站、膝上型计算机、上网本、掌上型计算机、移动电话、智能电话、多媒体电话、互联网协议语音(VOIP)电话、用户设备、终端、便携式媒体播放器、GPS单元、游戏系统、机顶盒)存取通过互联网提供的内容/服务和/或在敷设在互联网上(例如,隧穿互联网)的虚拟专用网络(VPN)上提供的内容/服务。这些内容和/或服务通常由属于服务或内容供应商的一个或多个端站(例如,服务器端站)或参与对等服务的端站提供,并且可以包括例如公共网页(例如,免费内容、店面、搜索服务)、私人网页(例如,提供email服务的用户名/密码存取的网页)和/或通过VPN的公司网络。通常,订户端站(例如,通过(有线或无线)耦合到接入网络的客户驻地设备)耦合到边缘网络元件,边缘网络元件(例如,通过一个或多个核心网络元件)耦合到其它边缘网络元件,这些其它边缘网络元件耦合到其它端站(例如,服务器端站)。

[0025] 一些网络元件对于实现VPN(虚拟专用网络)(例如,第2层VPN和/或第3层VPN)提供支持。例如,位于耦合供应商网络和客户网络所在的点的网络元件分别称为PE(供应商边缘)和CE(客户边缘)。在第2层VPN中,通常在VPN的任一端的CE上执行转发,并跨网络(例如,借助于通过其它网络元件耦合的一个或多个PE)发送业务。第2层电路配置在CE和PE之间(例如,以太网端口、802.1Q永久虚拟电路(PVC)、按需802.1Q PVC、帧中继PVC、异步传送模式(ATM)PVC)。在第3层VPN中,通常通过PE来执行路由。举例来说,支持多个上下文的边缘网络元件可以作为PE部署;可以利用VPN协议来配置上下文,并且因此该上下文称为VPN上下文。

[0026] 网络元件通常分割为控制平面和数据平面(有时称为转发平面或媒体平面)。在网络元件是路由器(或实现路由功能性)的情况下,控制平面通常确定将如何路由数据(例如,分组)(例如,数据的下一跳以及该数据的输出端口),而数据平面负责转发该数据。例如,控制平面通常包括与其它网络元件通信以便交换路由并基于一个或多个路由度量选择那些路由的一个或多个路由协议(例如,边界网关协议(BGP)、内部网关协议(IGP)(例如,开放最短路径优先(OSPF)、路由信息协议(RIP)、中间系统到中间系统(IS-IS))、标签分配协议(LDP)、资源预留协议(RSVP))。

[0027] 路由和邻接性存储在控制平面上的一个或多个路由结构(例如,路由信息库(RIB)、标签信息库(LIB)、一个或多个邻接性结构)中。控制平面用基于这个(或这些)路由结构的信息(例如,邻接性和路由信息)对数据平面编程。例如,控制平面将邻接性和路由信息编程到数据平面上的一个或多个转发结构(例如,转发信息库(FIB)、标签转发信息库

(LFIB) 和一个或多个邻接性结构) 中。数据平面在转发业务时利用这些转发和邻接性结构。

[0028] 对于第2层转发, 网络元件可以存储用于基于存在于数据中的第2层信息来转发数据的一个或多个桥接表。

[0029] 通常, 网络元件包括一个或多个线卡的集合、一个或多个控制卡的集合以及可选的一个或多个服务卡(有时称为资源卡)的集合。这些卡通过一个或多个机制耦合在一起(例如, 第一全网格耦合线卡, 而第二全网格耦合所有卡)。线卡的集合构成数据平面, 而控制卡的集合提供控制平面并通过线卡与外部网络元件交换分组。服务卡的集合可以提供专门处理(例如, 第4层到第7层服务(例如, 防火墙、IPsec、IDS、P2P)、VoIP会话边界控制器、移动无线网关(GGSN、演进型分组系统(EPS)网关))。举例来说, 服务卡可用于终止IPsec隧道并执行伴随的认证和加密算法。

[0030] IEEE 802.1aq具有两种可能的操作模式。基于VLAN的网络的第一模式称为最短路径桥接VID (SPBV)。基于MAC的网络的第二模式称为最短路径桥接MAC (SPBM)。每个ECT集合通常与SPBV的最短路径VLAN标识符 (SPVID) 集合以及与SPBM的骨干VLAN标识符 (B-VID) 相关联。以太网网络可以同时支持数据平面中的多于一个ECT集合。

[0031] IEEE 802.1aq要求, 对于给定的ECT集合, 网络中的两个点之间的所有节点独立选择网络中的任何两个点之间的相同对称路径。因此, 规定了一种算法, 以使得IEEE 802.1aq网络中的路径选择利用平局打破过程, 使得任何两个节点之间的路径将解析到单个对称路径, 而不管计算的方向、计算的顺序或路径的任何子集的检查。这种特性备选地描述为“最短路径的任何部分也是最短路径”。在沿最短路径的任何部分出现平局以使得存在多个“相等”选择的情况下, 那些节点将为具有相同选择的路径的子集解决平局, 结果是最低成本最短路径树。这在本文中称为“共同算法平局打破(common algorithm tiebreaking)”过程。有可能操纵共同算法平局打破过程的输入以使得将生成多个ECT集合, 由此允许更加有效地使用多路径网络。

[0032] 在路径选择过程中, 链路状态(拓扑)数据库首遍利用共同算法平局打破过程导致生成第一树集合。该初始步骤确定网络中的每个节点对之间的最短路径。在发现在任何两个节点之间存在多于一个最短路径的情况下, 利用共同算法平局打破过程来进行平局打破以便在网络中的每个节点对之间生成唯一路径选择并生成一个或多个等成本转发树集合(在IEEE 802.1aq中称为“ECT集合”)。

[0033] 在传统的802.1aq SPBM网络中, 采用以下方式每个B-VID选择路径。首先, 每个节点利用在节点之间发送的IS-IS链路状态路由协议消息了解网络的拓扑。一旦所有节点都了解了网络的拓扑, 每个节点利用最短路径计算算法计算用于单播和多播业务的最短路径。这些算法在本领域中众所周知, 并且包括所有对(all-pairs)最短路径算法(例如, Floyd-Warshall算法、Johnson's算法等)和单源最短路径算法(例如, Dijkstra's算法、Bellman-Ford算法等)。如果一个节点确定它在参与共同服务(即, 传送利用共同B-VID的共同服务的业务)的一对节点之间的最短路径上, 那么该节点安装转发状态以使得它可以相应地转发该业务。

[0034] 但是, 当在参与共同服务的一对节点之间存在多个最短路径时, 网络中的每个节点执行平局打破过程以便为该ECT集合的业务选择最短路径中的共同最短路径。该平局打破过程确保关于该业务的这一个正确路径做出符合网络范围的决定以便保持路径一致性,

从而规定任何两个桥接器之间沿两个方向的路径都将共享共同路由。这个过程是对称平局打破算法,它通过以下步骤工作:利用来自路径中的每个节点的系统标识符(系统ID)为每个最短路径构造路径标识符(路径ID);然后根据特定平局打破算法选择这些路径ID之一。根据802.1aq规范,利用与该ECT集合的B-VID相关联的掩码来变换路径中的每个节点的系统ID。接着,连结(concatenate)路径的每个变换的系统ID,然后按字典顺序将其排序,以形成路径ID。将表示潜在路径的集合的这些路径ID分等级,并选择最低路径ID作为该ECT集合中的业务的路径。以此方式,实现相同逻辑的所有节点将从相同选项中选择相同路径,并且给定ECT集合的业务的准确路径因此是可预测的,即使在网络拓扑中存在不会直接影响路径的给定变化。该过程还允许先验业务工程,因为可以选择性地将业务的服务指派给B-VID,从而允许跨网络中的多个等成本路径分配系统负载。

[0035] 如果发生了链路或节点故障,那么网络中的每个节点经由IS-IS链路状态消息知道该故障,并为受该故障影响的业务(即,经过该故障链路或经过该故障节点的业务)重新计算最短路径。例如,如果具有特定B-VID的业务经过受故障影响的特定路由,那么网络中的每个节点将确定应当利用不同的路径来转发利用该B-VID的所有业务。这导致业务切换到不同的链路和节点上。这种方法的益处是,网络中中断的唯一业务局限于经过故障链路或节点的业务。

[0036] 但是,这种方法也具有显著的局限性。一旦网络中发生故障,给定最短路径上的受影响的B-VID的业务会作为一个块切换到新路径,而不是跨残存资源平均地分配。这可对受影响的业务的“新”路径中的链路和节点造成显著负担。例如,假设图1的网络在节点A 104和节点B 102之间具有四个物理等成本最短路径,那么这些路径之一(例如,A-Y-B)上的故障会导致利用该路径的业务共同地切换到不同路径(例如,A-Z-B)。因此,假设在故障之前所有业务平均分配在这四个等成本最短路径中,在故障之后,路径A-Z-B将经历业务加倍。这个结果会很棘手,因为它可造成利用该路径上的链路的业务的服务实质降级,并且可导致进一步的网络故障和进一步的服务降级。

[0037] 图1示出根据本发明一个实施例的示例性网络以及在示例网络中利用分离的平局打破器的过程的逻辑表示。该示例以太网网络包括通过链路互连以便说明分离的平局打破器的一个使用的六个节点(例如,桥接器、网络元件等)。但是,在其它实施例中,网络可以包括更多节点、更少节点、或诸如fat-tree或Clos网络体系结构的不同网络拓扑。

[0038] 在该网络中,节点A 104是网络的入口节点,并且因此从另一个网络(未示出)接收业务或生成即将发送到网络的其它节点的业务。这里,节点A 104配置成将业务发送到出口节点B 102,出口节点B 102可以将业务转发到另一个网络(未示出)上。作为网络的入口节点的节点A 104和B 102可以视为是骨干边缘桥接器(BEB)或供应商边缘网络元件。

[0039] 在节点A 104和节点B 102之间有四个节点。这四个节点-节点W 110、节点X 112、节点Y 114和节点Z 116-均利用链路的集合连接至节点A 104和节点B 102。这四个中间节点110-116中的每个节点均可视为是骨干核心桥接器(BCB)或供应商网络元件。

[0040] 在该配置中,存在从节点A 104到节点B 102的四个不同的物理路径:A-W-B、A-X-B、A-Y-B和A-Z-B。出于说明的目的,将这四个路径中的每个路径定义为等成本。但是,实际上,尽管这些路径中的每个路径从节点A 104只通过一个节点(即,节点W-Z 110-116之一)行进到节点B 102,但是这些路径中的每个路径都可通过各种配置或确定方法而视为是具

有不同成本。例如,可以将网络配置成确定最短路径是经过最少数量的节点的路径,或者基于路径中的链路的带宽或通过指派给每个链路或节点的配置的“权重”确定最短路径。

[0041] 因此,该网络在节点A 104和节点B 102之间包括四个物理等成本最短路径。根据802.1aq SPBM,对于从节点A 104前往节点B 102的分组,将根据与该分组相关联的标识相关联的ECT集合的B-VID选择这些最短路径之一。例如,一旦进入SPBM网络,便在包括B-VID报头字段的802.1ah报头340(参见图3,本文将稍后论述)中封装分组。该B-VID一般在本文中又称为业务标识符(业务ID或TID),它供网络的每个节点用于在存在多于一个等成本最短路径时为该B-VID的业务确定从源到目的地的一个最短路径。

[0042] 根据本发明的一个实施例,如果发生故障,那么通过使用分离的平局打破器,可以将网络业务快速且简单地分配在多个残存的等成本最短路径上。在诸多益处中,分离的平局打破器尤其提供这样的实质益处,因为在故障之后,受影响的业务的部分会切换到多个残存路径,从而减少对每个残存链路和节点的影响。分离的平局打破器的使用还保留了传统802.1aq网络的路径可确定性的益处,因为故障前路径和故障后路径均可通过将服务和B-VID选择性地指派给业务来确定和配置。这些益处的基本机制是配置每个集合中的系统ID的分配,以使得在平局打破期间的次级路径或“不同于优选路径”的选择将不同于其它系统ID集合。当发生故障时,多个ECT集合将受影响,并且由于以使得系统ID不具有共同“下一个最佳”选择的方式选择了系统ID,所以受故障影响的业务分配在多于一个残存路径上。该技术的先决条件是生成系统ID集合,以使得在任何单个故障上,与将造成选择特定路径的平局打破器集合相关联的受影响的ECT集合将均匀地分配在残存资源集合上。对于诸如3级CLOS体系结构的简单网络,这证实是轻易就能做到的;对于任何拓扑,可能需要更复杂的算法来确定每个集合中的值的分布。

[0043] 图1的示例简单网络中描绘了根据本发明一个实施例的分离的平局打破器的一个使用。如上文所描述,该网络包括六个节点以及位于节点A 104和节点B 102之间的四个物理等成本最短路径。在该示例中,利用解析到这四个物理等成本最短路径的八个虚拟等成本最短路径来将由八个不同业务ID(即,B-VID)所表示的业务从节点A 104转发到节点B 102。该示例示出在影响最短路径之一的故障118之前和之后具有分离的平局打破器的网络的操作。当然,分离的平局打破器也可以配置成在具有更多或更少节点、更多或更少业务ID的网络中以及对于在一对节点之间具有更多或更少最短路径的场景起作用。

[0044] 尽管传统802.1aq网络在构造路径标识符时使用指派给每个节点的系统ID,但是分离的平局打破器的实现将多于一个系统ID指派给网络中的节点。当为一个或多个业务ID的特定集合构造路径标识符时,使用节点的每个系统ID。因此,当为第一业务ID创建路径标识符时,使用每个节点的第一系统ID。但是,当为第二业务ID创建路径标识符时,将使用来自该路径中的每个节点的第二系统ID。根据本发明的一个实施例,每个节点的所述多个系统ID可以通过网络管理员手动地指派或者通过网络上的软件工具或硬件元件指派。类似地,即将在确定最短路径时使用的业务ID集合与节点的系统ID集合中的某个系统ID之间的映射也可以通过网络管理员手动地指派或者通过网络上的软件工具或硬件元件指派。

[0045] 在图1中,网络中的每个节点在节点A 104和节点B 102之间构造八个虚拟等成本最短路径。图1中所构造的这八个虚拟等成本最短路径包括对应于业务ID 1-8指派给这四个物理路径120 A-W-B、A-X-B、A-Y-B和A-Z-B中的每个路径的两个示例。为了为业务ID为1

的业务选择最短路径之一,利用与每个路径中的每个节点相关联的第一系统ID来构造122每个路径的路径标识符。在该网络中,每个节点的第一系统ID与业务ID 1-4相关联,而每个节点的第二系统ID与业务ID 5-8相关联。在备选实施例中,存在指派给每个节点的多于两个系统ID,并且与每个系统ID相关联的业务ID可以不同,或者在数量上可以更多或更少。对于不同的网络拓扑和场景,这些选项之间的选择及其配置可以不同,并且随网络可以产生不同的益处。

[0046] 对于第一路径A-W-B,选择122指派给每个节点的第一系统ID。对于节点A 104,这是1001b,对于节点W 110,这是0000b,而对于节点B 102,这是1000b。然后,变换这些系统ID。例如,图1示出利用与业务ID 1相关联的变换来变换123每个系统ID。在该示例中,变换是经由“异或”(XOR)操作对系统ID应用掩码,其中与业务ID 1相关联的掩码是掩码0000b。然后,按字典顺序将每个变换的系统ID 123排序并进行连结130以构造路径ID 0000/1000/1001b。对于其它三个等成本最短路径中的每个路径重复该过程。例如,对于路径A-Z-B,标识节点A 104的第一系统ID 1001b,标识节点Z 116的第一系统ID 0011b,并标识节点B 102的第一系统ID 1000b。将所标识的这些系统ID进行变换,按字典顺序排序,并进行连结,以形成路径ID 0011/1000/1001b。

[0047] 在根据该过程构造这四个路径ID中的每个路径ID之后,将它们分等级,选择具有最低值的路径ID,并确定与该路径ID相关联的等成本最短路径就是该业务ID的业务的路径。在该示例中,通过将每个位模式作为二进制无符号整数对待来比较路径ID,但是在本发明的其它实施例中,可以有其它方法来比较路径ID,并且它们在本领域中众所周知。这里,路径A-W-B的路径ID小于其它三个路径ID,因此将利用路径A-W-B来在该网络中路由业务ID 1的业务。对于业务ID 2-4中的每个业务ID重复该过程。

[0048] 另一方面,业务ID 5-8的平局打破算法依赖于不同的系统ID集合来表示每个等成本最短路径中的桥接器。为了为业务ID 5-8选择这四个物理等成本最短路径之一,选择124指派给每个节点的第二系统ID。例如,现在利用节点A 104的第二系统ID 1000b、节点W 110的第二系统ID 0001b和节点B 102的第二系统ID 1001b来构造路径A-W-B的路径ID。

[0049] 在网络中的故障118之前,对于业务ID 5-8中的每个业务ID,利用平局打破算法以及与该特定业务ID相关联的变换来选择所构造的这四个路径之一。在图1中,对于业务ID 6示出该过程,业务ID 6与掩码0001b相关联。因此,通过对照掩码0001b对系统ID执行XOR操作来变换125该路径中的每个第二系统ID。这导致将路径A-W-B的原始系统ID 1000/0001/1001b变换为1001/0000/1000b:由于 $1000b \text{ XOR } 0001b = 1001b$,所以出现该结果的前四个位;由于 $0001b \text{ XOR } 0001b = 0000b$,所以出现该结果的中间四个位;并且由于 $1001b \text{ XOR } 0001b = 1000b$,所以出现最后四个位。然后,将变换后的每个系统ID集合按字典顺序排序并进行连结132以形成即将在指定业务ID的平局打破中使用的路径ID。

[0050] 对于业务ID 6,对其它三个等成本最短路径中的每个路径进行该过程。在为这四个等成本最短路径中的每个路径构造路径ID之后,通过选择具有最低值的路径ID进行126平局打破,并确定与该路径ID相关联的等成本最短路径就是业务ID 6的路径。在该示例中,对于经过节点A-W-B的路径,最低路径ID为0000/1000/1001b。对于业务ID 6-8中的每个业务ID重复该过程,每次使用每个路径中的每个节点的第二系统ID以及与该特定业务ID相关联的变换。

[0051] 在故障前126为业务ID 1-4和5-8选择最短路径结束时,业务ID 1和6的业务经过路径A-W-B,业务ID 2和8的业务经过路径A-X-B,业务ID 3和5的业务经过路径A-Y-B,并且业务ID 4和7的业务经过路径A-Z-B。网络中的每个节点独立计算这些路径,并安装用于实现这些转发过程的正确的转发状态。

[0052] 在某个点,在网络中发生118影响节点A 104和W 110之间的链路的故障。节点A 104和节点W 110中至少有一个节点在整个网络中溢出指示丧失该链路的公告。利用该信息,系统中的每个节点从它的链路状态数据库移除该链路(或链路的有效性)。结果,业务ID 1和6的业务不再具有通过该网络的有效路径,并且通过每个节点确定每个业务ID的新路径。

[0053] 为了确定业务ID 1和6中的每个业务ID的新路径,每个节点再次构造等成本最短路径,利用该业务ID的变换为所构造的每个等成本最短路径构造路径ID,并应用共同平局打破算法以选择路径。例如,假设除了故障118以外没有任何其它网络变化,那么将如同在120中那样构造等成本最短路径,但是将不再存在第一路径A-W-B。对于业务ID 1,将利用这些路径中的每个节点的第一系统ID来选择系统ID(122)(同样地,将不构造路径A-W-B的路径ID)。将利用与业务ID 1相关联的变换(它是掩码0000b)来变换123每个系统ID。在该故障后状态中,将变换后的系统ID按字典顺序排序并进行连结134以创建路径ID,并选择128最低路径ID。业务ID 1的业务现在将经过路径A-X-B。

[0054] 类似地,对于业务ID 6,将如同在124中那样利用路径中的每个节点的第二系统ID来选择系统ID,但是A-W-B的路径将不存在。利用与业务ID 6相关联的掩码0001b来变换每个第二系统ID(125)。在该故障后状态中,将变换后的系统ID按字典顺序排序并进行连结以形成路径ID(136),然后选择128最低所得路径ID。业务ID 6的业务现在将经过路径A-Y-B。

[0055] 该结果说明由于使用分离的平局打破器而产生的一个益处。根据传统802.1aq的操作,一旦发生故障118,经过路径A-W-B的所有业务都可能会切换到一个其它残存的路径A-X-B。因此,所有受故障影响的业务切换到一个其它路径上的这个过程将会显著增加该路径上的负载。在最好的情况下,这会使该路径中的所有链路和节点的性能降级;在最差的情况下,它可使这些节点和链路变得如此拥挤以至于可能会丢弃业务或者可能会出现节点或链路故障。

[0056] 相反,通过为系统的节点有目的地选择系统ID并为网络的业务细心地选择业务ID和变换,实现分离的平局打破器的网络能够通过将受影响的业务分配在多个残存等成本最短路径上来从容且自动地适应故障。例如,在图1中,在发生故障118时,与业务ID 1和6相关联的业务不会切换到相同路径;而是,将该业务分配到两个路径(A-X-B和A-Y-B)中。另外,可能的是,分析网络的拓扑和业务指派以设计多路径和分离的平局打破器配置的组合,从而确保在发生链路故障时,将在网络中的残存资源中出现业务的显著散布。

[0057] 此外,当故障118被校正时,与业务ID 1和业务ID 6相关联的业务将回复到原始路径(A-W-B),而其它路径的业务将保持不受搅乱。这将拓扑变化诱导网络中的分组流重新排序的可能性降至最低。

[0058] 结果,逐跳等成本多路径路由(ECMP)的大多数益处得以重复,但是具有许多故障场景中的流排序、网络可预测性、以及操作、管理和维护(OAM)实用性和可测试性的改进的保存。另外,尽管采用分离的平局打破器的系统得益于使用比传统802.1aq网络中常用的业

务ID更多的业务ID,但是每个节点的转发数据库中的多播状态的量不会因为业务ID的此增加而受影响。此外,由于任何顺序的所有故障的确切效果可以有把握地预测,所以网络计划得以简化。

[0059] 图2示出根据本发明一个实施例的用于为802.1aq实现分离的平局打破器的示例性网络元件210的功能框图。如图2中所描绘,网络元件210包括控制平面212和数据平面214。数据平面214的作用是接收进入分组252并根据转发信息库236内的信息将这些分组作为外出分组254进行转发。在网络元件210用作BEB并且因此位于骨干网络的边缘上并与独立网络(例如,客户网络)对接的本发明的一个实施例中,进入分组252到达独立网络。在进入分组252是从独立网络接收的802.1ad分组的本发明的一个实施例中,转发模块234可以用802.1ah报头封装进入分组252,然后将它们作为外出分组254转发到骨干网络。(关于802.1ad分组和802.1ah分组的进一步细节将在本文中稍后关于图3呈现。)

[0060] 备选地,在本发明的一个实施例中,网络元件210可以用作骨干网络和独立网络的边缘上的BEB,并且可以从骨干网络接收已经用802.1ad报头封装的进入分组252。在该场景中,转发模块234将从分组解除(即,去除)802.1ad报头的封装,然后将分组作为外出分组254发送到独立网络。在网络元件210用作BCB并且因此在骨干网络的核心中的本发明的一个实施例中,进入分组252和外出分组254都将是802.1ad分组,并且不会通过转发模块234进一步封装或解封装。

[0061] 控制平面212配置成控制网络元件210的操作。控制平面212包括链路状态协议模块216,在本文稍后将论述的诸多功能性中,它通过在骨干网络的网络元件中交换(即,传送和接收)链路状态消息250来与网络的其它网络元件通信。在采用诸如IS-IS或OSPF的链路状态路由协议的网络中利用链路状态消息250来在网络的节点中通信拓扑信息,并且这些链路状态路由协议为本领域技术人员所熟知。链路状态消息250的交换使得网络中的每个网络元件能够发展网络拓扑的同步理解。在本发明的一个实施例中,根据IS-IS协议的版本传送链路状态消息,该IS-IS协议的版本可以是例如在2001年3月8日发布的题为“IS-IS Extensions Supporting IEEE 802.1aq Shortest Path Bridging”第5版(draft-ietf-isis-ieee-aq-05.txt)的互联网工程工作小组(IETF)互联网草案中所描述的版本。根据本发明的一个实施例,对基本IS-IS协议进一步修改以允许每个网络元件传送和接收指派给特定网络元件的多个系统ID,而不是只简单地传送一个系统ID。这种修改是为了分离的平局打破的目的而允许每个网络元件分配多个系统ID的一种方法;但是,也可以利用其它方法,例如在每个网络元件处手动地配置该信息。

[0062] 在从网络的其它网络元件接收链路状态消息250的集合之后,链路状态协议模块216利用来自这些链路状态消息250的信息确定网络的拓扑。链路状态协议模块216通过存取260链路状态数据库模块224来存储和检索关于网络的拓扑信息。链路状态数据库模块224包括链路状态数据库226,链路状态数据库226包含表示网络的网络元件以及这些网络元件之间的连接的条目。在本发明的一个实施例中,这些网络元件是以太网桥接器,而以太网桥接器之间的连接是以太网链路。此外,链路状态数据库226还存储用于在等成本最短路径之间进行分离的平局打破的每个网络元件228的多个系统ID。

[0063] 通过存取260存储在链路状态数据库模块224中的网络拓扑信息,链路状态协议模块216能够利用路径构造模块220来构造网络元件之间的最短路径。在本发明的一个实施例

中,路径构造模块220利用Dijkstra's算法的实现来在网络中的BEB对之间构造这一个或多个等成本最短路径,并且随后作为FIB构造的一部分求解指派给该特定B-VID的给定一对BEB之间的I-SID的交集。

[0064] 当通过路径构造模块220确定在这对BEB之间存在多于一个等成本最短路径时,链路状态协议模块216利用平局打破模块222来选择所确定的等成本最短路径之一。在本发明的一个实施例中,平局打破模块222对于特定业务ID为特定节点对之间的路径利用与该B-VID/业务ID相关联的变换以及每个网络元件228的多个系统ID中与该B-VID/业务ID相关联的特定系统ID来执行平局打破。

[0065] 在确定网络的网络元件之间的最短路径之后,链路状态协议模块216更新266转发模块234以便将数据平面214配置成根据最短路径转发分组。在本发明的一个实施例中,将最短路径信息存储在转发信息库(FIB) 236中,FIB 236包括用于将一个或多个输入接口、目的地地址和B-VID映射到一个或多个输出接口的条目。一旦接收到指示网络拓扑的变化的额外链路状态消息250,该过程将重复,因为链路状态协议模块216将更新链路状态数据库模块224,构造最短路径,并再次更新266转发模块234。

[0066] 图3示出根据本发明一个实施例用于封装网络业务的分组报头。如之前所论述,用作BEB装置的网络元件从外部网络(它可以是客户端网络)接收业务。在本发明的一个实施例中,客户端网络上的以太网业务将具有有效负载310和客户端报头330,客户端报头330包括客户端目的地MAC地址(C-DA) 318、客户端源MAC地址(C-SA) 316和至少一个虚拟局域网ID(VID)。根据本发明的一个实施例,如图3所示,客户端报头包括如IEEE 802.1ad所详述的S-VID 314和C-VID 312。在本发明的其它实施例中,如同在802.1Q中一样,客户端报头可以只包括C-VID。当然,客户端网络也可以使用其它以太网帧格式。一旦接收到客户端业务,BEB网络元件便将用IEEE 802.1ah骨干报头340封装业务350,这允许骨干网络的网络元件基于骨干基MAC地址空间转发业务,而不是基于客户端MAC地址空间进行转发。802.1ah报头330包括骨干(B)目的地MAC地址(B-DA) 326、骨干源MAC地址(B-SA) 324、供应商VLAN ID(B-VID) 322和服务标识符(I-SID) 320。

[0067] 将参考图1、2、5、7、9和10的示例性实施例描述图4中的流程图的操作。但是,应了解,图4中的流程图的操作可以通过不同于参考图1、2、5、7、9和10论述的实施例的本发明的其它实施例来执行,并且参考图1、2、5、7、9和10论述的实施例可以执行与参考图4中的流程图论述的操作不同的操作。

[0068] 图4示出根据本发明一个实施例在802.1aq网络中利用分离的平局打破器的流程图。在框402,802.1aq网络中的节点公告该节点的第一和第二系统标识符。第一和第二系统ID分别与第一和第二业务ID集合相关联。业务ID用于区分网络内的不同虚拟网络,并且其中每个业务ID与平局打破变换相关联。在本发明的一个实施例中,为该节点公告多于两个系统ID,其中每个系统ID与不同的业务ID集合相关联。在一个实施例中,平局打破变换是掩码,该掩码用于通过对掩码和每个系统ID执行XOR操作来变换路径ID内的系统ID。在框404,节点402接收网络中的其它多个节点中的每个节点的第三和第四系统ID。第三和第四系统ID分别与第一和第二业务ID集合相关联。在本发明的一个实施例中,为网络中的其它多个节点中的每个节点接收多于两个系统ID,并且节点的每个系统ID与不同的业务ID集合相关联。在框406,该节点在网络的第一节点和第二节点之间构造多个不同的等成本最短路径。

在一个实施例中,当最短路径涵盖相同数量的链路时,它们视为是等成本。在一个实施例中,利用链路使用统计来确定最短路径。在框408,该节点通过以下步骤为每个业务ID选择所构造的等成本最短路径之一:1)利用与该业务ID相关联的平局打破变换以及包括系统ID中与该业务ID相关联的那些系统ID的系统ID集合来构造每个等成本最短路径的路径ID,并且其中系统ID集合中的每个系统ID与该等成本最短路径的不同节点相关联;以及2)对路径ID应用选择算法以选择等成本最短路径之一。在框410,响应于影响一个或多个所选择的等成本最短路径上的多个业务ID的业务故障,由于网络中的系统ID和业务ID的配置,该节点将与那些业务ID相关联的业务分配在多个其它等成本最短路径上。在本发明的一个实施例中,网络中的每个节点的所述多个系统ID、业务ID的选择以及业务的变换允许将受故障影响的业务分配在该业务的所有其它残存等成本最短路径上。

[0069] 图5示出根据本发明一个实施例,用于说明最短路径和业务标识符的示例性网络拓扑,并结合图6呈现,图6示出根据本发明一个实施例利用分离的平局打破器的最短路径的逻辑计算。图5描绘骨干网络,它包括六个节点,即,两个BEB节点B 502和A 500以及四个BCB节点W 510、X 512、Y 514和Z 516。每个节点包括两个系统ID;但是,为了易于说明,只描绘了BCB节点的系统ID。该网络包括四个等成本最短路径:A-W-B、A-X-B、A-Y-B和A-Z-B。图5中的圆圈指示,已经确定即将在特定路径上转发与业务ID相关联的业务。例如,将经由BCB节点W 510在BEB节点A 500与BEB节点B 502之间转发与业务ID 1和6相关联的业务。

[0070] 图6呈现如何选择与业务ID 1-8相关联的业务以在这四个物理等成本最短路径上行进的概念图示。如上文参考图1和图2所详述,每个节点保存表示网络拓扑的链路状态数据库。对于网络中的一对BEB节点(节点A 500和节点B 502),每个节点为在经过BEB节点之间的业务中所使用的每个业务ID计算BEB节点之间的等成本最短路径。当节点计算多于一个等成本最短路径时,该节点利用分离的平局打破器为具有业务ID的业务选择一个路径。

[0071] 图6包括用于理解分离的平局打破的第一表602A和第二表602B。每个表包括属于这四个等成本最短路径中的每个路径中的中间节点的系统ID的标题行(606、618)。每一行用与每个业务ID相关联的掩码604的第一列标引。第一表602A示出使用来自等成本最短路径中的每个中间节点的第一系统ID集合606。在该示例中,第一系统ID集合中的每个系统ID与从1到4范围内的业务ID集合相关联。另一方面,第二表602B示出使用来自等成本最短路径中的每个中间节点的第二系统ID集合618。该第二系统ID与从5到8范围内的业务ID集合相关联。当为业务ID 1-4中的每个业务ID确定等成本最短路径时,使用第一表602A和第一系统ID集合606。类似地,当为业务ID 5-8中的每个业务ID确定等成本最短路径时,使用第二表602B和第二系统ID集合618。

[0072] 当在源节点和目的地节点之间存在多个等成本最短路径时,例如在图5中,为每个路径构造路径ID。对于特定业务ID,利用与该业务ID相关联的变换来变换每个路径内的每个系统ID。将变换后的每个系统ID按字典顺序排序并进行连结以创建路径ID,并选择最低路径ID作为该业务ID的路径。图6(以及图8,将在本文稍后进行论述)示出该过程的相关部分。但是,为了易于理解,这些图中没有示出完整的路径ID。具体来说,没有示出传送BEB节点(在图5中为节点A 500)和接收BEB节点(在图5中为节点B 502)的系统ID和变换后的系统ID。由于在所构造的每个路径ID中都将存在这些值,所以这些值在比较时彼此抵消。例如,当比较A-W-B和A-X-B这两个路径时,每个路径的唯一区分方面是每个路径的第二跳,在此

情况下为节点W和节点X。为了易于理解,为清楚起见,省略这些多余值(例如,节点A和B的系统ID);但是,对于包括这些值的完整示例,请参见图1。

[0073] 第一表602A示出根据本发明一个实施例为业务ID 1-4选择路径。第一行610示出变换业务ID 1的中间节点系统ID 606并选择其中之一。与业务ID 1相关联的变换是掩码0000b,将它与系统ID集合中节点W 510的第一系统ID或0000b进行XOR操作,得到结果0000b。类似地,同样对节点X 512、Y 514和Z 516中的每个节点的第一系统ID应用掩码0000b。行610中的每个结果表示变换后的系统ID。在本发明的一个实施例中,以此方式变换每个路径中的每个节点的每个第一系统ID,并将结果按字典顺序排序并进行连结以创建路径ID,选择其中的最低路径ID作为路径。对于第一行610,选择(用圆圈表示)节点W 510的变换的系统ID,因为它在行610中具有最低的二进制变换的系统ID。在本发明的其它实施例中,可以根据诸如选择最高变换路径ID的另一规则或算法来选择所选择的路径ID。因此,根据表示节点A 500和节点B 502之间的这四个等成本最短路径的第一行610中的变换的系统ID,将利用路径A-W-B来转发业务ID 1的业务。尽管通过网络中的每个节点来执行该过程,但是在这一点上,节点A 500、W 510和B 502将在它们的相应的转发信息库236中配置转发信息以根据所确定的该路径识别和转发业务。

[0074] 将对于第二行612、第三行614和第四行616中的每个业务ID 2-4类似地执行该过程。根据该过程,业务ID 2的业务将经过路径A-X-B,因为节点X 512的变换系统ID 0000b是第二行612的最低值。类似地,业务ID 3的业务将经过路径A-Y-B,并且业务ID 4的业务将经过路径A-Z-B。

[0075] 但是,如第二表602B所示,业务ID 5-8的路径选择将包括变换与路径中的每个节点相关联的系统ID集合的第二系统ID。例如,第二表602B中的节点W 510由它的第二系统ID 0001b表示,而第一表602A中的节点W 510则由它的第一系统ID 0000b表示。在第二表602B的第一行620中,示出业务ID 5的路径选择。业务ID 5的变换是掩码0000b,将它应用于中间节点的每个第二系统ID。例如,节点W 510的第二系统ID是0001b,通过对掩码0000b应用XOR操作将其进行变换,结果是0001b。对于节点X 512,将第二系统ID 0011b与掩码0000b进行XOR操作,结果是0011b。对于节点Y 514,将第二系统ID 0000b与掩码0000b进行XOR操作,结果是0000b。对于节点Z,将第二系统ID 0010b与掩码0000b进行XOR操作,结果是0010b。在第二表602B的第一行620中变换每个系统ID之后,选择最低的变换系统ID。这个最低值是0000b,它指示包括节点Y 514的路径(A-Y-B)是利用业务ID 5的业务的业务的路径。类似地,对于业务ID 6-8中的每个业务ID重复该过程,并再次使用每个中间节点的系统ID集合的第二系统ID。图5表示在结束路径选择时根据业务ID的业务流。业务ID 1和6的业务经过路径A-W-B,业务ID 2和8的业务经过路径A-X-B,业务ID 3和5的业务经过路径A-Y-B,而业务ID 4和7的业务经过路径A-Z-B。

[0076] 图7示出根据本发明一个实施例在故障之后图5中示出的示例性网络拓扑,并结合图8呈现,图8示出根据本发明一个实施例利用分离的平局打破器的最短路径的逻辑计算。与图5和图6类似,图7和图8没有示出节点A 700和B 702的系统ID,而只是显示每个路径中的中间节点的系统ID和变换的系统ID,而没有描绘整个路径ID。

[0077] 图7包括如图5中所呈现的相同节点。但是,图7示出在发生影响业务ID 1和6的业务的故障718之后的网络。该故障可以因多种原因而发生,包括但不限于节点A 700和W 710

之间的链路的物理故障、节点A 700和B 710中的一个或两个节点的故障、或节点A 700和W 710中的一个或两个节点中的端口或网络接口卡的问题。在该示例网络中，节点A 700和W 710中的一个或两个节点检测到故障718，并将指示该故障718的链路状态消息250传送到网络的其它节点。当每个节点都知道故障718时，该节点将更新260它的链路状态数据库226，并将为业务ID 1和6计算新路径。

[0078] 图8呈现如何选择与业务ID 1-8相关联的业务以在图7的三个残存等成本最短路径上行进的概念图示。由于故障718，每个节点在计算节点A 500和节点B 502之间的路径时不会计算通过节点W 510的最短路径。因此，用“X”标记第一表802A和第二表802B中的节点W 510的列，并且不考虑它们。如同在图6中一样，第一表802A和第二表802B中的每个表包括与业务ID 1-8中的每个业务ID相关联的掩码804的第一列。类似地，每个表包括系统ID的行：第一表802A包括包含每个中间节点的系统ID集合的第一系统ID的行806，而第二表802B包括包含系统ID集合的第二系统ID的行818。

[0079] 如第一表802A中所描绘，通过网络中的每个节点计算节点A 500和节点B 502之间的三个等成本路径。如之前图6中所描述地那样进行业务ID 1-4中的每个业务ID的路径选择过程，但存在现在少一个路径的可能性。在第一行810中，与业务ID 1相关联的变换是掩码0000b，将它与每个中间节点的系统ID集合806的第一系统ID进行XOR操作。该变换导致节点X 512的变换系统ID 0001b、节点Y 514的变换系统ID 0010b和节点Z 516的变换系统ID 0011b。在这三个可能性中，节点X 512的0001b是最低值，因此为业务ID 1选择的路径现在是A-X-B。因此，节点A 500、X 512和B 502中的每个节点将配置它们的相应的转发信息库236以便允许根据该确定的路径转发与业务ID 1相关联的业务。对于第二行812、第三行814和第四行816中的业务ID 2-4继续进行该过程。

[0080] 第二表802B中表示业务ID 5-8中的每个业务ID的路径选择，它利用每个中间节点的系统ID集合818的第二系统ID。例如，通过首先用与业务ID 6相关联的变换（它是掩码0001b）来变换系统ID 818中的每个系统ID而在第二行822中确定业务ID 6的新路径。通过对系统ID 818中的每个系统ID应用掩码，节点X 512的变换系统ID是0010b，节点Y 514的变换系统ID是0001b，而节点Z的变换系统ID是0011b。由于节点Y 514的变换系统ID 0001b是行822的最低值，所以将利用节点Y 514来转发业务ID 6的业务。相应地，节点A 500、Y 514和B 502中的每个节点将配置它们的相应的转发信息库236以便允许根据该确定路径转发与业务ID 6相关联的业务。以类似方式，还对820、824和826中的每个业务ID 5、7和8执行平局打破。如上所述，除了变换中间节点的系统ID集合的第二系统ID以外，在一些实施例中，还变换节点A 500和节点B 502的系统ID集合的第二系统ID，并将所有三个变换的系统ID按字典顺序排序且连结到路径ID。在那些实施例中，比较每个路径ID以便执行平局打破并为业务ID选择路径。

[0081] 由于每个节点的多个系统ID、指派给业务的业务ID以及与每个业务ID相关联的变换的正确配置，分离的平局打破由此允许用于经过A-W-B的网络业务在发生影响该路径的故障之后经过两个不同路径：A-X-B和A-Y-B。这使得网络能够通过对于因故障而偏移的业务利用多个故障转移路径通过网络来更容易且可预测地适应故障场景。并非将所有受影响的业务作为一个块切换到一个其它路径，而是将业务分配在多个路径中，从而减轻故障的影响。

[0082] 图9示出根据本发明一个实施例在示例性网络上的多播业务的最短路径。为清楚起见,该图示描绘只利用两个业务ID(1和6)的业务;通常,此类网络将利用多于两个业务ID。在该实施例中,节点A 900将多播业务传送到节点B 902、节点C 904、节点D 906和节点E 908。在节点A 900和接收方节点902-908之间存在中间节点W 901、中间节点X 912、中间节点Y 914和中间节点Z 916。在节点之间的跳数指示路径成本的本发明的一个实施例中,确定在节点A 900与这四个接收方节点902-908中的每个接收方节点之间存在四个物理等成本最短路径。以图5和图6的相同方式,每个路径中的每个节点的第一系统ID与业务ID 1-4相关联,并且每个路径中的每个节点的第二系统ID与业务ID 5-8相关联。同样地,为了说明清楚起见,没有示出节点A 900和接收方节点902-908中的每个节点的系统ID。在为业务ID 1选择路径时,利用与业务ID 1相关联的变换(例如,掩码0000b,未示出)来变换中间节点 910-916中的每个节点的第一系统ID。这导致业务ID 1的业务经过节点W 910,并确定路径是A-W-B、A-W-C、A-W-D和A-W-E。类似地,为业务ID 6选择路径涉及利用与业务ID 6相关联的变换(例如,掩码0001b,未示出)来变换中间节点910-916的系统ID集合的第二系统ID。这导致业务ID 6的业务也经过节点W 910,并确定路径是A-W-B、A-W-C、A-W-D和A-W-E。

[0083] 图10示出根据本发明一个实施例在故障之后在图9的示例性网络上的多播业务的不同最短路径。在该图中,在中间节点W 910与接收方节点B 902之间发生影响业务ID 1和6的故障1018。在该场景中,只有去往节点B 902的业务ID 1和6的业务受影响,而去往节点C 904、节点D 906和节点E 908的业务ID 1和6的业务将继续通过节点W 901转发1020。假设业务ID 1利用掩码0000b作为变换,那么对照节点X 912、节点Y 914和节点Z 916中的每个节点的第一系统ID对该掩码进行XOR操作。结果,对于节点X 912,最低的变换系统ID将是0001b。类似地,假设业务ID 6利用掩码0001b,那么将该掩码应用于节点X 912、节点Y 914和节点Z 916的每个第二系统ID。结果,对于节点Y 914,最低的变换系统ID将是0001b。因此,从节点A 900去往节点B 902的业务ID 1的业务将经过路径A-X-B,而从节点A 900去往节点B 902的业务ID 6的业务将经过路径A-X-C。

[0084] 尽管图中的流程图示出通过本发明的某些实施例执行的操作的特定顺序,但应了解,该顺序是示例性的(例如,备选实施例可以按照不同的顺序执行操作,组合某些操作,重叠某些操作等)。

[0085] 尽管就若干实施例描述了本发明,但是本领域技术人员将意识到,本发明不限于所描述的实施例,而是可以用随附权利要求书的精神和范围内的修改和改变来加以实践。因此,本描述视为是说明性的而不是限制性的。

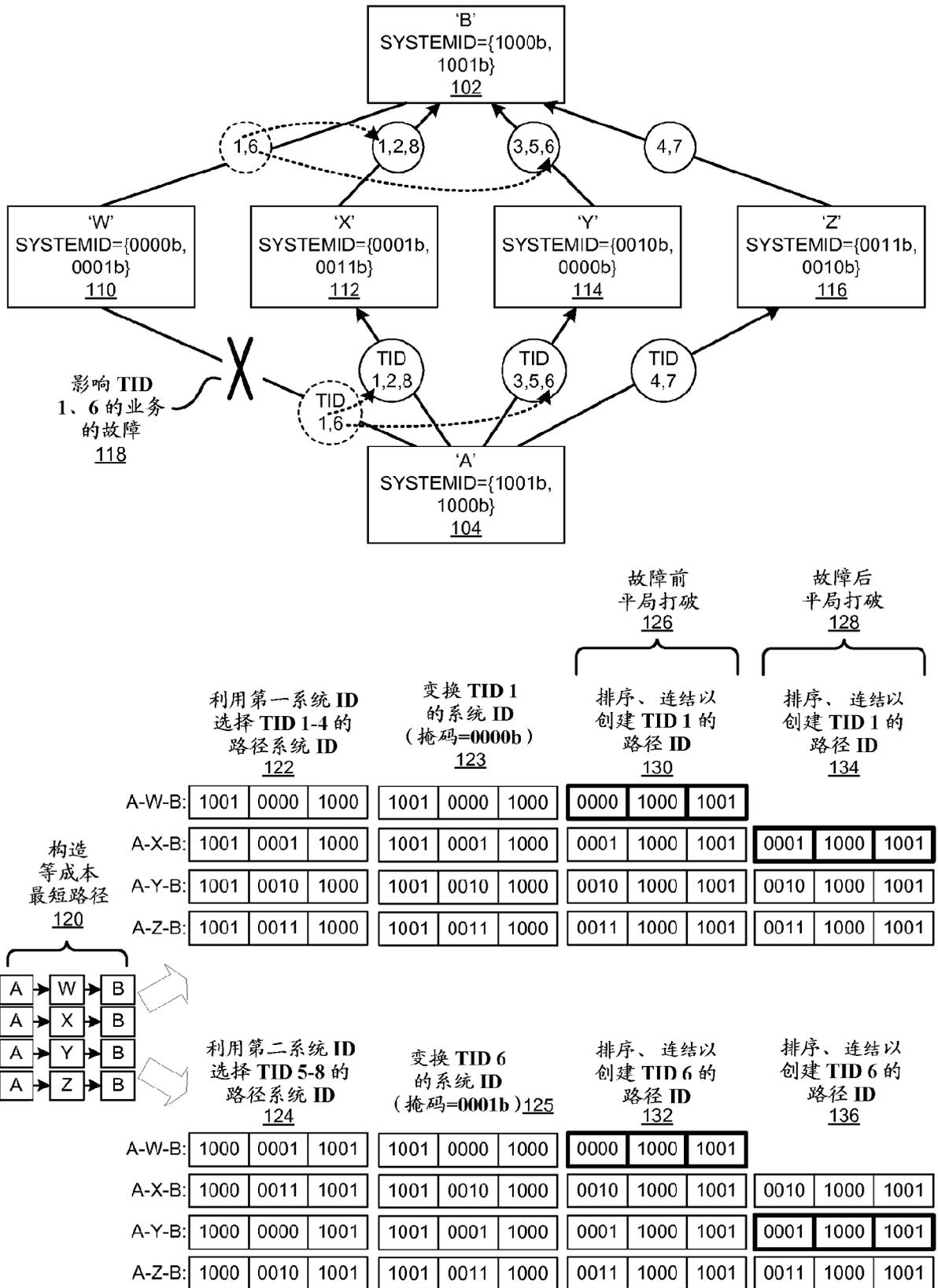


图 1

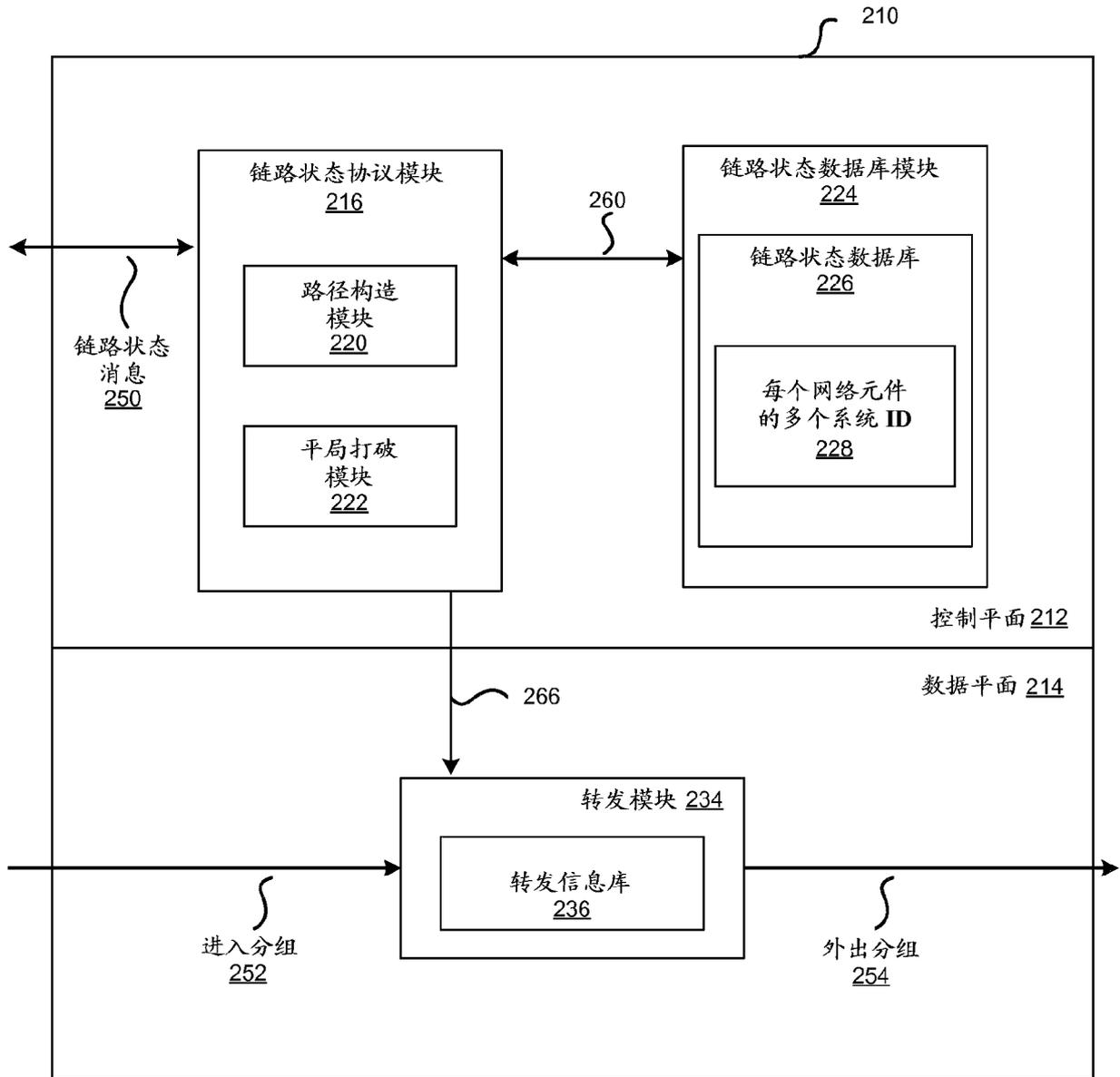


图 2

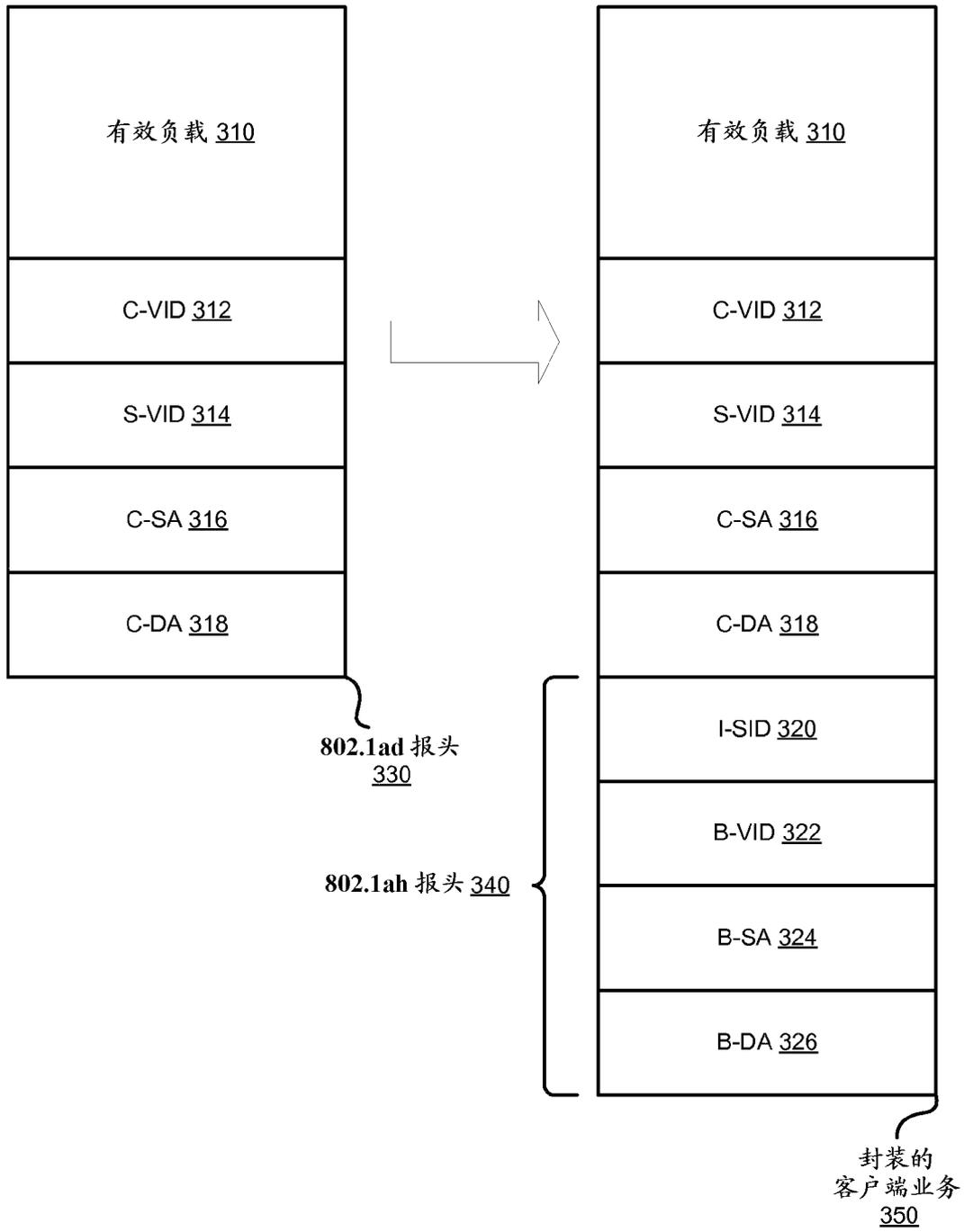


图 3

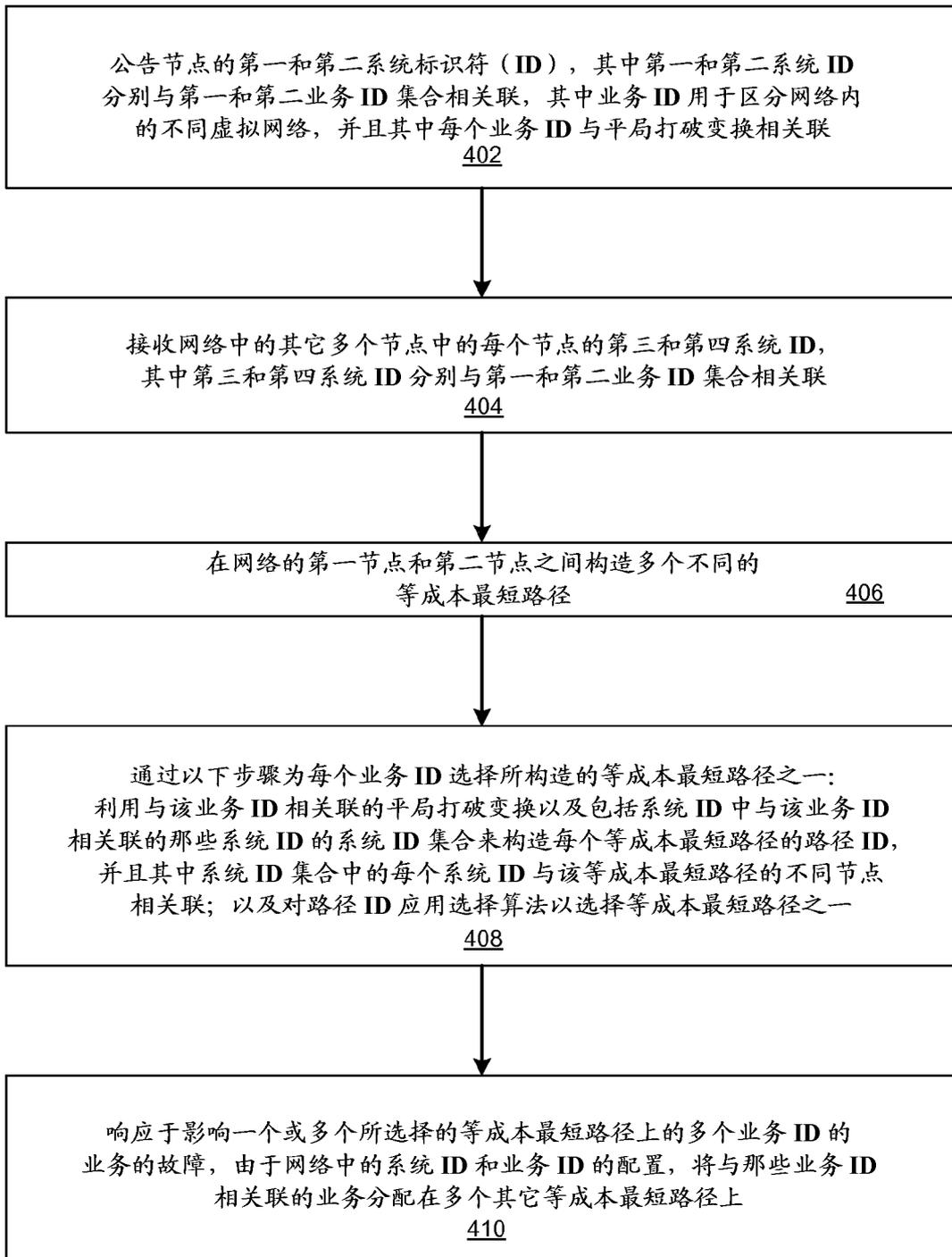


图 4

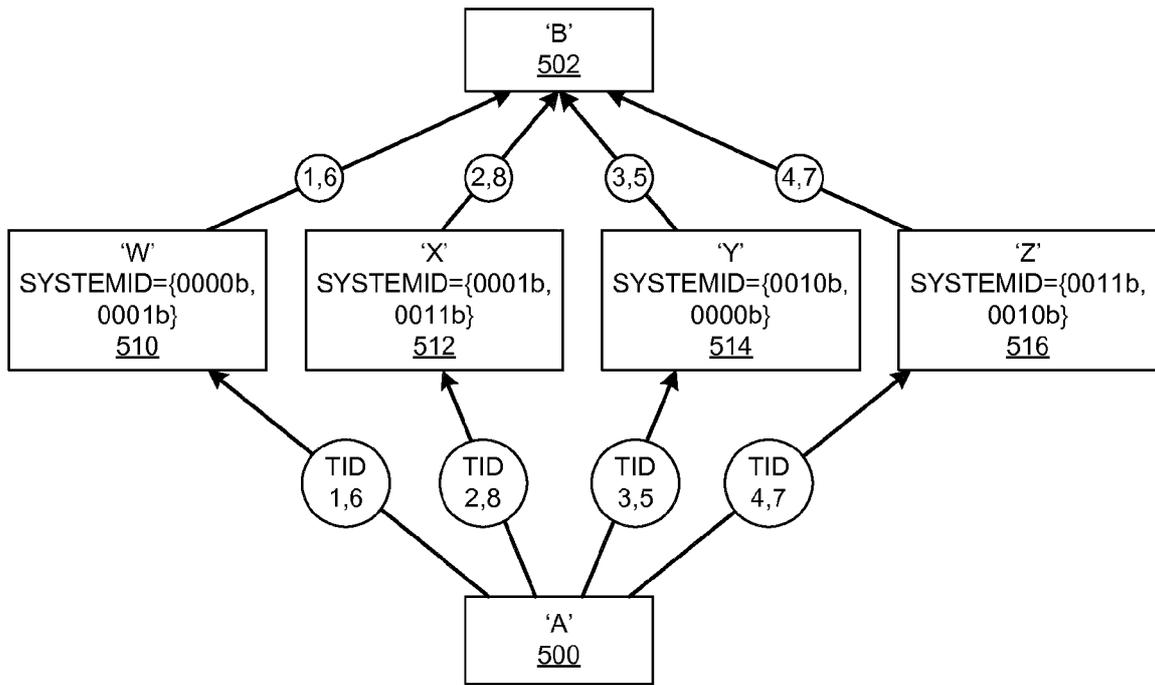


图 5

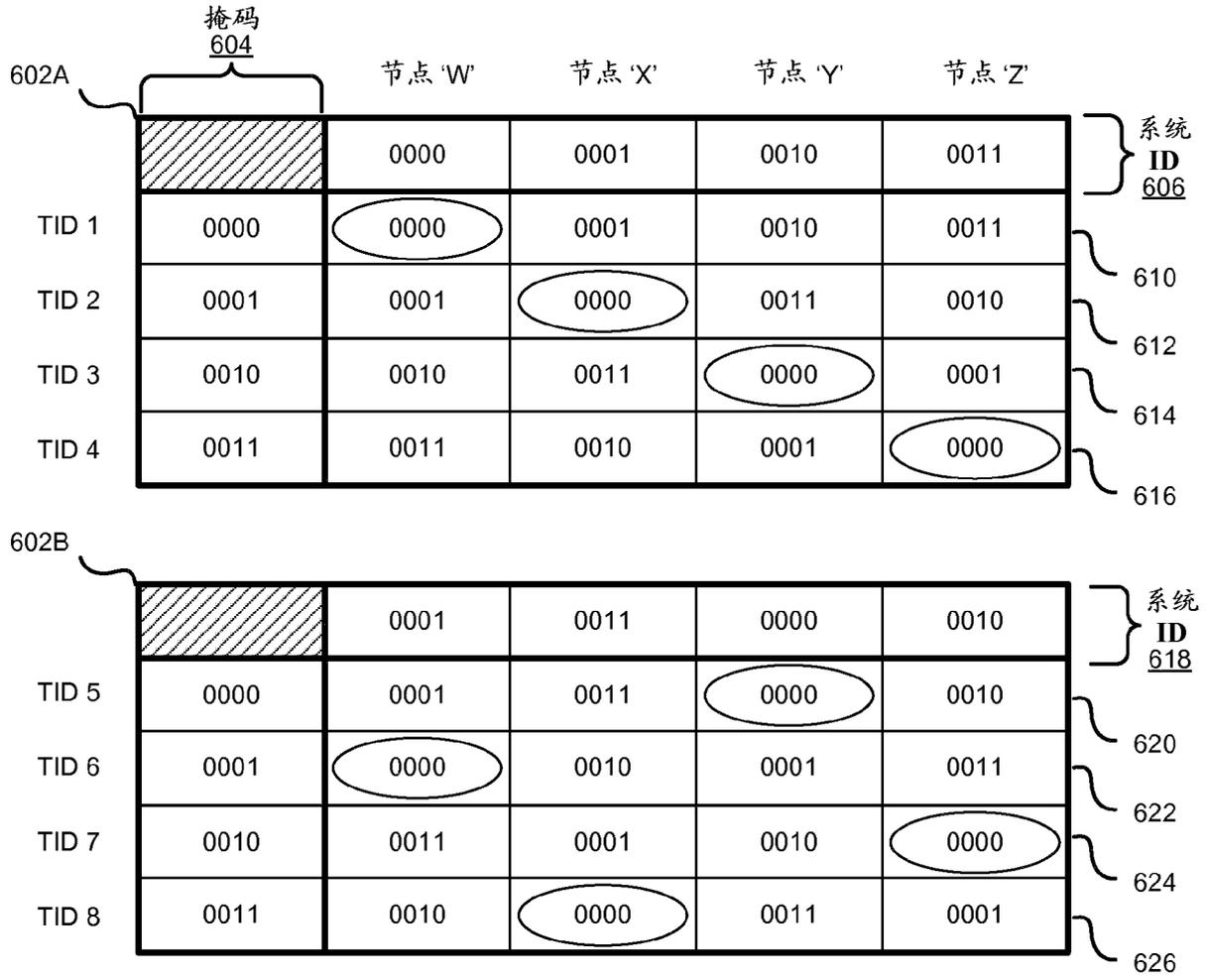


图 6

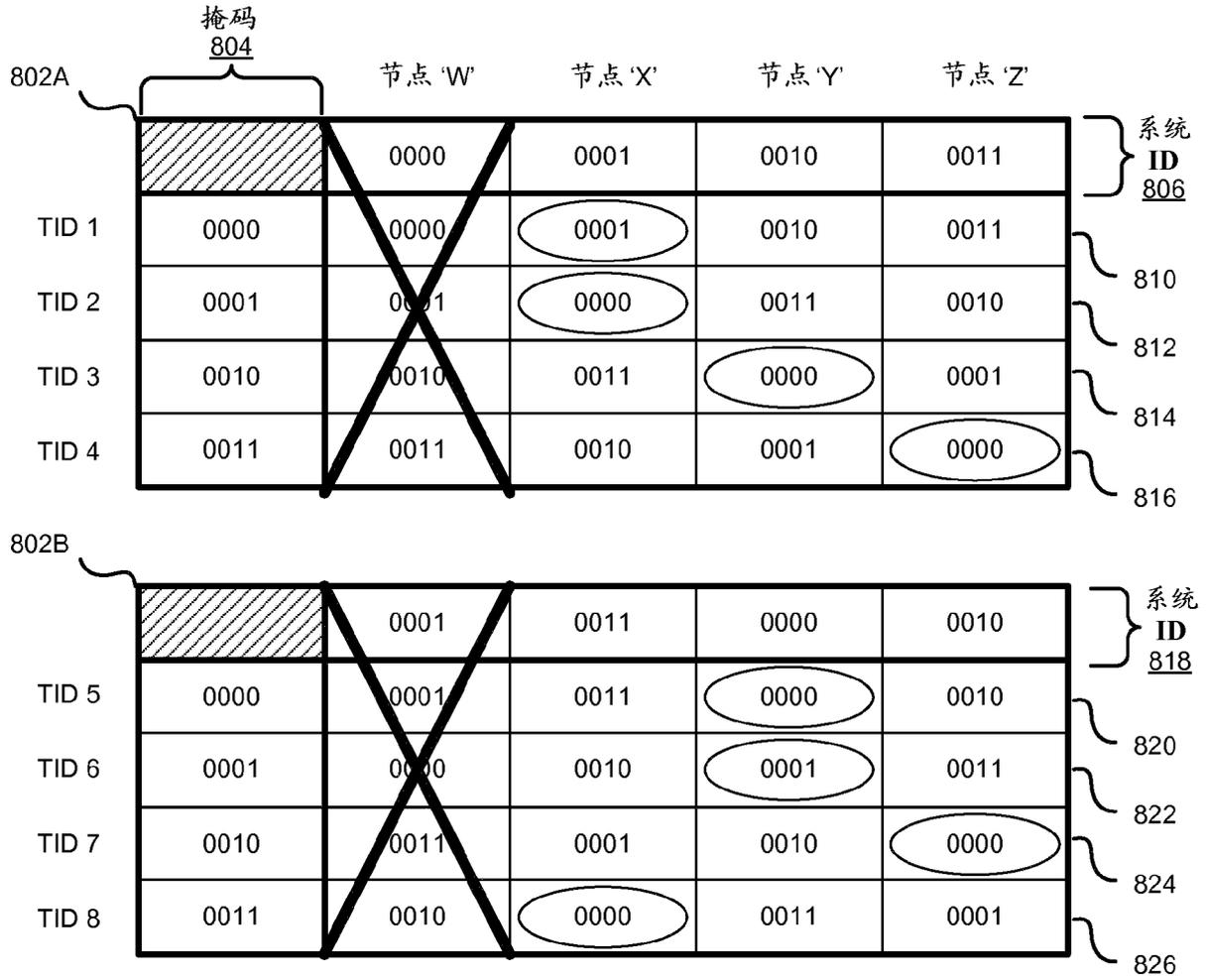


图 8

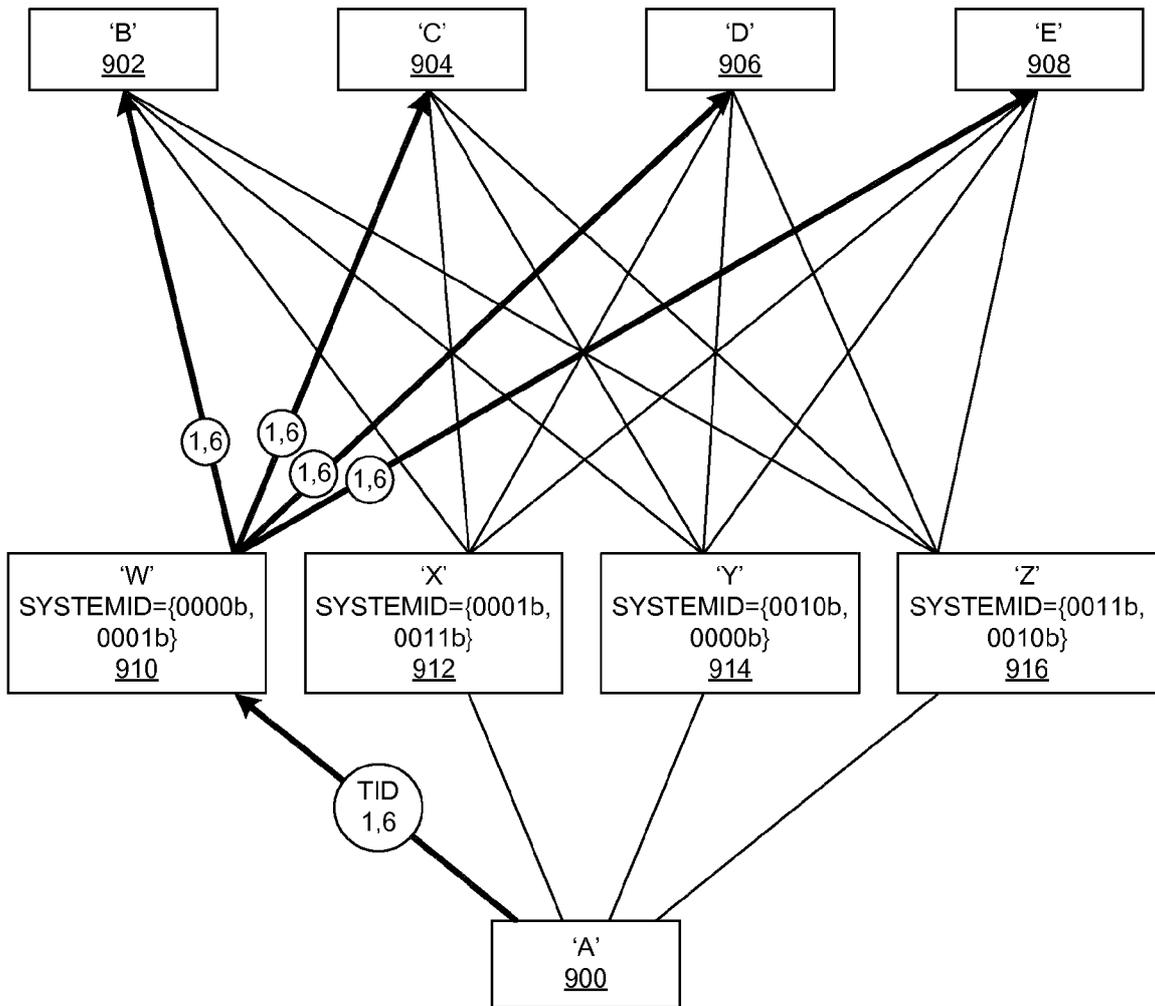


图 9

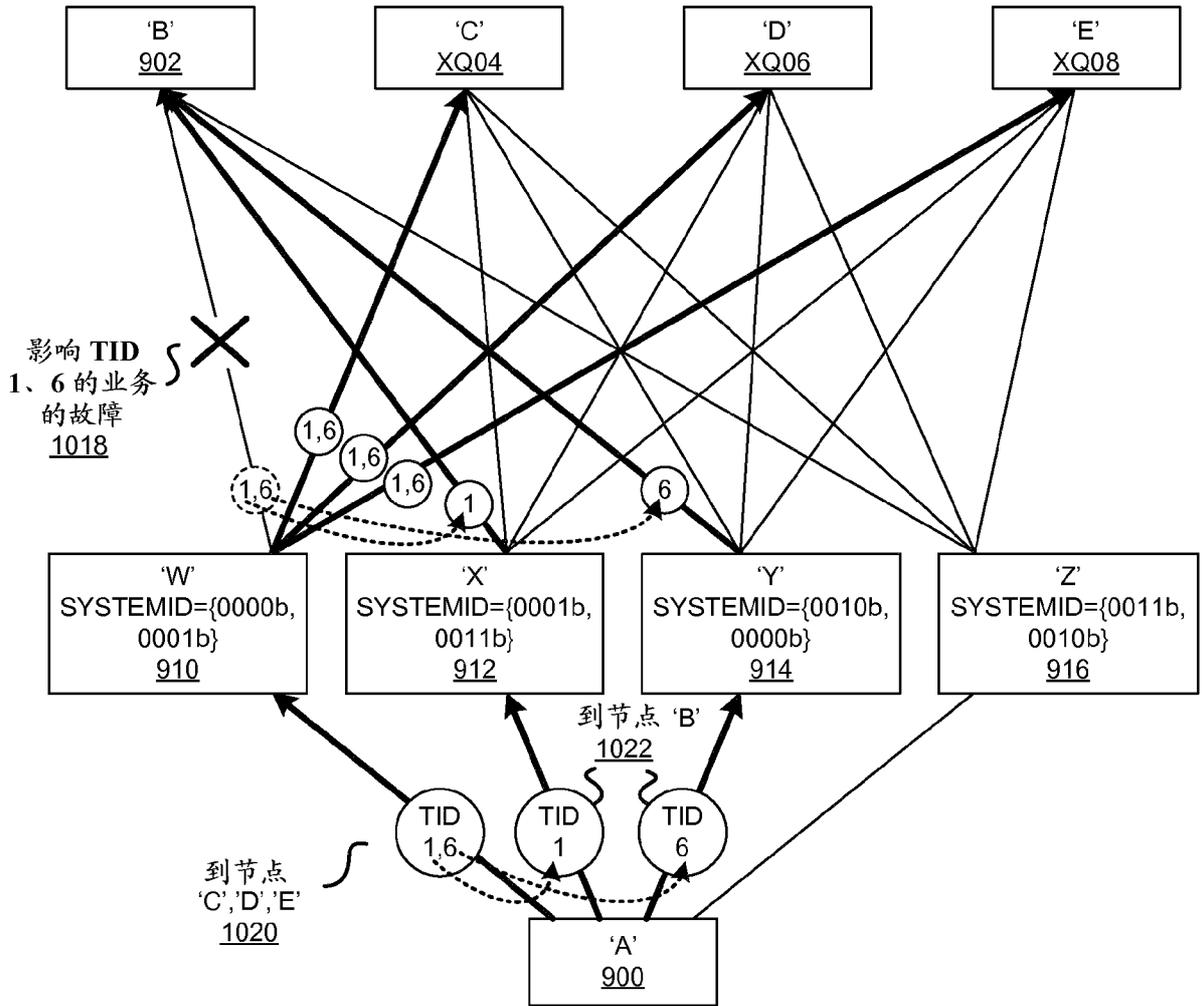


图 10