



(12) 发明专利

(10) 授权公告号 CN 1943179 B

(45) 授权公告日 2011.05.04

(21) 申请号 200580011224.1

(22) 申请日 2005.01.21

(30) 优先权数据

10/779,083 2004.02.13 US

(85) PCT申请进入国家阶段日

2006.10.13

(86) PCT申请的申请数据

PCT/US2005/001602 2005.01.21

(87) PCT申请的公布数据

W02005/083946 EN 2005.09.09

(73) 专利权人 英特尔公司

地址 美国加利福尼亚州

(72) 发明人 索尔·雷维德斯 阿杰伊·加尔格

(74) 专利代理机构 上海专利商标事务所有限公司 31100

代理人 钱慰民

(51) Int. Cl.

H04L 12/46 (2006.01)

(56) 对比文件

GB 2389023 A, 2003.11.26, 说明书第7页第4-11行, 第8页第13-32行, 第9页第1-9行, 第10页第1-2行, 第12页第20-24行, 第22页第19-24行, 第28页第8-15行, 第36页第1-5行, 22-26行, 第40页第5-12行, 及附图3-6, 8.

WO 0167694 A, 2001.09.13, 说明书全文.

US 20020097730 A1, 2002.07.25, 说明书第1页第9段第1-4行, 第10段第1-7行, 12-17行, 第12段第1-8行, 第3页第39段第10-20行, 第2页第25-27段, 附图1-4.

审查员 平彧

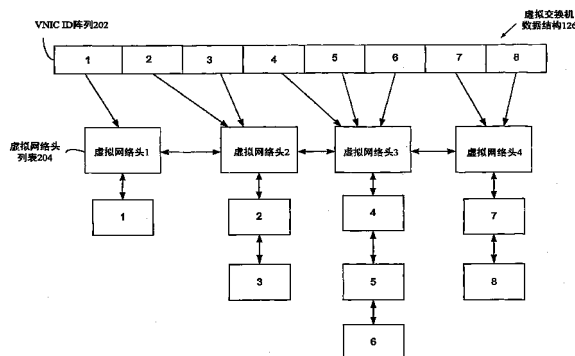
权利要求书 4 页 说明书 6 页 附图 9 页

(54) 发明名称

用于动态地可扩展的虚拟交换机的装置和方法

(57) 摘要

用于动态地可扩展的虚拟交换机的装置和方法。装置或虚拟交换机包括至少一个路由器和数据结构。所述路由器利用所述数据结构来组织一个或多个虚拟网络接口卡 (VNIC) 之间的连接以形成虚拟网络。所述虚拟交换机还用 VNIC 的独有的标识符来识别数据帧的 VNIC 节点, 利用所述独有的标识符来索引元素集合以取得虚拟网络头的指针, 并且向与所述虚拟网络头相关联的 VNIC 节点列表中的除涉及所述独有的标识符的 VNIC 节点之外的所有 VNIC 节点转发所述数据帧。



1. 一种用于动态地可扩展的虚拟交换机的装置,包括:

至少一个路由器,

其中所述至少一个路由器利用数据结构来组织一个或更多个虚拟网络接口卡之间的连接以形成虚拟网络,其中所述一个或更多个虚拟网络接口卡中的每一个位于虚拟机中,

其中所述一个或更多个虚拟网络接口卡中的每一个具有独有的标识符,并且所述数据结构包括:

元素集合,所述元素集合由所述一个或更多个虚拟网络接口卡的所述独有的标识符来索引;以及

虚拟网络头列表,其中在所述元素集合中的每个元素与所述虚拟网络头列表中的虚拟网络头相关联,其中所述虚拟网络头列表中的所述虚拟网络头中的每一个与虚拟网络接口卡节点列表相关联,并且其中所述虚拟网络接口卡节点列表中的每个虚拟网络接口卡节点包含所述一个或更多个虚拟网络接口卡的所述独有的标识符中的一个。

2. 如权利要求 1 所述的装置,其中所述虚拟机对于所述一个或更多个虚拟网络接口卡中的每一个是相同的。

3. 如权利要求 1 所述的装置,其中所述虚拟机对于所述一个或更多个虚拟网络接口卡中的每一个是不同的。

4. 如权利要求 1 所述的装置,其中所述元素集合是阵列。

5. 如权利要求 1 所述的装置,其中所述虚拟网络接口卡节点列表是虚拟网络接口卡节点双向链表。

6. 一种用于动态地可扩展的虚拟交换机的方法,包括:

通过至少一个路由器和数据结构形成虚拟网络,其中所述至少一个路由器利用所述数据结构来组织一个或更多个虚拟网络接口卡之间的连接,其中所述一个或更多个虚拟网络接口卡中的每一个位于虚拟机中,

其中所述一个或更多个虚拟网络接口卡中的每个具有独有的标识符并且所述数据结构包括由所述一个或更多个虚拟网络接口卡的所述独有的标识符来索引的元素集合,以及虚拟网络头列表,其中所述元素集合中的每个元素与所述虚拟网络头列表中的虚拟网络头相关联,其中所述虚拟网络头列表中的所述虚拟网络头的每一个与虚拟网络接口卡节点列表相关联,并且其中所述虚拟网络接口卡节点列表中的每个虚拟网络接口卡节点包含所述一个或更多个虚拟网络接口卡的所述独有的标识符中的一个。

7. 如权利要求 6 所述的方法,还包括创建新的虚拟网络,其中创建新的虚拟网络的操作包括:

创建新的虚拟网络头;以及

将所述新的虚拟网络头添加到所述虚拟网络头列表。

8. 一种用于动态地可扩展的虚拟交换机的方法,包括:

通过至少两个路由器和数据结构形成至少两个虚拟网络,其中所述至少两个路由器利用所述数据结构来组织一个或更多个虚拟网络接口卡之间的连接,其中所述一个或更多个虚拟网络接口卡中的每一个位于虚拟机中,

其中所述一个或更多个虚拟网络接口卡中的每个具有独有的标识符,其中所述数据结构包括由所述一个或更多个虚拟网络接口卡的所述独有的标识符来索引的元素集合,以及

虚拟网络头列表,其中所述元素集合中的每个元素与所述虚拟网络头列表中的虚拟网络头相关联,其中所述虚拟网络头列表中的所述虚拟网络头中的每一个与虚拟网络接口卡节点列表相关联,并且其中所述虚拟网络接口卡节点列表中的每个虚拟网络接口卡节点包含所述一个或更多个虚拟网络接口卡的所述独有的标识符中的一个。

9. 如权利要求 8 所述的方法,还包括删除所述至少两个虚拟网络中的一个,其中删除所述虚拟网络的操作包括:

- 确定所述虚拟网络头列表中表示所述要被删除的虚拟网络的所述虚拟网络头;
- 将所述元素集合中与所述确定的虚拟网络头相关联的每个元素解除关联;
- 将与所述确定的虚拟网络头相关联的所述虚拟网络接口卡节点列表解除关联;以及
- 从所述虚拟网络头列表删除所述确定的虚拟网络头。

10. 如权利要求 8 所述的方法,还包括从所述至少两个虚拟网络中的一个中删除虚拟网络接口卡,其中删除所述虚拟网络接口卡的操作包括:

- 确定所述虚拟网络头列表中表示与所述要被删除的虚拟网络接口卡相关联的所述虚拟网络的所述虚拟网络头;

- 将所述元素集合中由所述要被删除的虚拟网络接口卡的所述独有的标识符来索引的元素从所述确定的虚拟网络头解除关联;以及

- 在与所述确定的虚拟网络头相关联的所述虚拟网络接口卡节点列表中删除包含所述要被删除的虚拟网络接口卡的所述独有标识符的所述虚拟网络接口卡节点。

11. 如权利要求 8 所述的方法,还包括将虚拟网络接口卡添加到所述至少两个虚拟网络中的一个,其中添加所述虚拟网络接口卡的操作包括:

- 确定所述虚拟网络头列表中表示与所述要被添加的虚拟网络接口卡相关联的所述虚拟网络的所述虚拟网络头;

- 将所述元素集合中由所述要被添加的虚拟网络接口卡的所述独有的标识符来索引的元素关联到所述确定的虚拟网络头;以及

- 在与所述确定的虚拟网络头相关联的所述虚拟网络接口卡节点列表中添加包含所述要被添加的虚拟网络接口卡的所述独有的标识符的虚拟网络接口卡节点。

12. 一种用于动态地可扩展的虚拟交换机的装置,包括:

- 用于通过至少一个路由器和数据结构形成虚拟网络的装置,其中所述至少一个路由器利用所述数据结构来组织一个或更多个虚拟网络接口卡之间的连接,其中所述一个或更多个虚拟网络接口卡中的每一个位于虚拟机中,

- 其中所述一个或更多个虚拟网络接口卡中的每个具有独有的标识符,其中所述数据结构包括由所述一个或更多个虚拟网络接口卡的所述独有的标识符来索引的元素集合,以及虚拟网络头列表,其中所述元素集合中的每个元素与所述虚拟网络头列表中的虚拟网络头相关联,其中所述虚拟网络头列表中的所述虚拟网络头中的每一个与虚拟网络接口卡节点列表相关联,并且其中所述虚拟网络接口卡节点列表中的每个虚拟网络接口卡节点包含所述一个或更多个虚拟网络接口卡的所述独有的标识符中的一个。

13. 如权利要求 12 所述的装置,还包括用于创建新的虚拟网络的装置,其中用于创建新的虚拟网络的装置包括:

- 用于创建新的虚拟网络头的装置;以及

用于将所述新的虚拟网络头添加到所述虚拟网络头列表的装置。

14. 一种用于动态地可扩展的虚拟交换机的装置,包括:

用于通过至少两个路由器和数据结构形成至少两个虚拟网络的装置,其中所述至少两个路由器利用所述数据结构来组织一个或多个虚拟网络接口卡之间的连接,其中所述一个或多个虚拟网络接口卡中的每一个位于虚拟机中,

其中所述一个或多个虚拟网络接口卡中的每个具有独有的标识符,其中所述数据结构包括由所述一个或多个虚拟网络接口卡的所述独有的标识符来索引的元素集合,以及虚拟网络头列表,其中所述元素集合中的每个元素与所述虚拟网络头列表中的虚拟网络头相关联,其中所述虚拟网络头列表中的所述虚拟网络头中的每一个与虚拟网络接口卡节点列表相关联,并且其中所述虚拟网络接口卡节点列表中的每个虚拟网络接口卡节点包含所述一个或多个虚拟网络接口卡的所述独有的标识符中的一个。

15. 如权利要求 14 所述的装置,还包括用于删除所述至少两个虚拟网络中的一个的装置,其中用于删除所述虚拟网络的装置包括:

用于确定所述虚拟网络头列表中表示所述要被删除的虚拟网络的所述虚拟网络头的装置;

用于将所述元素集合中与所述确定的虚拟网络头相关联的每个元素解除关联的装置;

用于将与所述确定的虚拟网络头相关联的所述虚拟网络接口卡节点列表解除关联的装置;以及

用于从所述虚拟网络头列表删除所述确定的虚拟网络头的装置。

16. 如权利要求 14 所述的装置,还包括用于从所述至少两个虚拟网络中的一个删除虚拟网络接口卡的装置,其中用于删除所述虚拟网络接口卡的装置包括:

用于确定所述虚拟网络头列表中表示与所述要被删除的虚拟网络接口卡相关联的所述虚拟网络的所述虚拟网络头的装置;

用于将所述元素集合中由所述要被删除的虚拟网络接口卡的所述独有的标识符来索引的元素从所述确定的虚拟网络头解除关联的装置;以及

用于在与所述确定的虚拟网络头相关联的所述虚拟网络接口卡节点列表中删除包含所述要被删除的虚拟网络接口卡的所述独有标识符的所述虚拟网络接口卡节点的装置。

17. 如权利要求 14 所述的装置,还包括用于将虚拟网络接口卡添加到所述至少两个虚拟网络中的一个的装置,其中用于添加所述虚拟网络接口卡的装置包括:

用于确定所述虚拟网络头列表中表示与所述要被添加的虚拟网络接口卡相关联的所述虚拟网络的所述虚拟网络头的装置;

用于将所述元素集合中由所述要被添加的虚拟网络接口卡的所述独有的标识符来索引的元素关联到所述确定的虚拟网络头的装置;以及

用于在与所述确定的虚拟网络头相关联的所述虚拟网络接口卡节点列表中添加包含所述要被添加的虚拟网络接口卡的所述独有的标识符的虚拟网络接口卡节点的装置。

18. 一种用于动态地可扩展的虚拟交换机的方法,包括:

用源虚拟网络接口卡的独有的标识符来识别数据帧的源虚拟网络接口卡节点;

利用所述独有的标识符来索引元素集合以取得指向虚拟网络头的指针;以及

向与所述虚拟网络头相关联的虚拟网络接口卡节点列表中除涉及所述独有的标识符的虚拟网络接口卡节点之外的所有虚拟网络接口卡节点转发所述数据帧，

其中所述一个或多个虚拟网络接口卡中的每一个位于虚拟机中。

19. 如权利要求 18 所述的方法，其中所述元素集合是阵列，并且其中所述虚拟网络接口卡节点列表是虚拟网络接口卡节点双向链表。

20. 一种用于动态地可扩展的虚拟交换机的装置，包括虚拟交换机，所述虚拟交换机用源虚拟网络接口卡的独有的标识符来识别数据帧的源虚拟网络接口卡节点，利用所述独有的标识符来索引元素集合以取得指向虚拟网络头的指针，并且向与所述虚拟网络头相关联的虚拟网络接口卡节点列表中除涉及所述独有的标识符的虚拟网络接口卡节点之外的所有虚拟网络接口卡节点转发所述数据帧，其中所述一个或多个虚拟网络接口卡中的每一个位于虚拟机中。

21. 如权利要求 20 所述的装置，其中所述元素集合是阵列，并且其中所述虚拟网络接口卡节点列表是虚拟网络接口卡节点双向链表。

22. 一种用于动态地可扩展的虚拟交换机的装置，包括：

用于用源虚拟网络接口卡的独有的标识符来识别数据帧的源虚拟网络接口卡节点的装置；

用于利用所述独有的标识符来索引元素集合以取得指向虚拟网络头的指针的装置；以及

用于向与所述虚拟网络头相关联的虚拟网络接口卡节点列表中除涉及所述独有的标识符的虚拟网络接口卡节点之外的所有虚拟网络接口卡节点转发所述数据帧的装置，

其中所述一个或多个虚拟网络接口卡中的每一个位于虚拟机中。

23. 如权利要求 22 所述的装置，其中所述元素集合是阵列，并且其中所述虚拟网络接口卡节点列表是虚拟网络接口卡节点的双向链表。

用于动态地可扩展的虚拟交换机的装置和方法

[0001] 背景

[0002] 单个物理平台可以被分成多个虚拟网络。在这里,物理平台包括至少一个虚拟机监视器(VMM)。常规的VMM通常运行在计算机上并且向其他软件呈现一个或更多个虚拟机(VM)的抽象。每个VM可以作为运行其自身的“客户操作系统”(即以VMM为宿主的操作系统(OS))和其他软件的自主式平台,所述“客户操作系统”和其他软件被共同地称为客户软件。

[0003] 运行在VM内的进程被提供一些硬件资源的抽象并且可以不了解系统内的其他VM。每个VM认为它对分派给它的硬件资源具有完全的控制。VMM是负责适当地在VM之间管理和仲裁系统资源的实体,所述系统资源包括,但不限于,处理器、输入/输出(I/O)设备和存储器。

[0004] 网络接口卡(NIC)虚拟化是用于向VM提供一个或多个物理NIC的抽象的技术。通过虚拟化,同一物理NIC可以被多个VM共享。此外,NIC虚拟化允许向VM呈现同一物理NIC的多个实例。例如,系统可以具有单个物理NIC,但VM可以看见多个虚拟NIC(VNIC),所述多个VNIC中的每一个与物理平台内的不同网络和/或物理NIC所附接到的外部网络接口。事实上,实际的物理NIC甚至不是必须存在以使系统内的VM间通信能进行。呈现给VM的VNIC可以完全不同于实际的物理NIC,由此使得向VM展现在实际的物理硬件中可以不存在的特征成为可能。

[0005] 存在对单个物理平台中虚拟网络的总数和每个虚拟网络中VM的最大数的限制。此外,虚拟网络在VMM的运行时间不能被重新配置,因为它涉及将一个或更多个VM从一个虚拟网络移动到另一个,而这要求受影响的VM和整个VMM的重新启动。

[0006] 可以利用虚拟交换机来提供用于在多个虚拟网络之间路由信息或数据帧的交换功能。所述虚拟交换机通常通过数据帧曾从其被发送的VNIC的MAC地址,或通过被储存在数据帧本身中的MAC地址来识别数据帧的源VNIC节点。任何恶意软件(例如在VM中运行的客户OS)可能欺骗(spoof)MAC地址并因此导致数据帧的接收节点相信数据帧曾来自于与它曾来自的实际源VNIC节点不同的源VNIC节点。MAC地址的欺骗损害虚拟网络的完整性。

[0007] 附图简要说明

[0008] 通过参照以下描述和用来图示说明本发明的实施方案的附图,本发明可以得到最好地理解。在附图中:

[0009] 图1图示虚拟机环境的一个实施方案,其中本发明的一些实施方案可以在所述虚拟机环境中操作;

[0010] 图2根据本发明的一个实施方案,图示虚拟交换机数据结构;

[0011] 图3是用于创建新的虚拟网络的过程的一个实施方案的流程图;

[0012] 图4是用于删除现有虚拟网络的过程的一个实施方案的流程图;

[0013] 图5根据本发明的一个实施方案,图示删除现有虚拟网络的操作;

[0014] 图6是用于将VNIC添加到特定虚拟网络的过程的一个实施方案的流程图;

[0015] 图 7 图示将 VNIC 添加到特定虚拟网络的操作；

[0016] 图 8 是用于从特定虚拟网络删除 VNIC 的过程的一个实施方案的流程图；以及

[0017] 图 9 是用于数据帧路由方案的过程的一个实施方案的流程图，所述数据帧路由方案避免欺骗的可能性。

[0018] 实施方案描述

[0019] 描述了用于动态地可扩展的虚拟交换机的装置和方法。本发明的虚拟交换机提供在运行时间创建、删除或重新配置虚拟网络而无需重新启动任何 VM 或 VMM 的能力。此外，所述虚拟交换机对可以连接到任何虚拟网络的 VM 的数量没有施加固定的限制。所述虚拟交换机对虚拟网络的总数也没有施加限制。此外，所述虚拟交换机确保没有恶意软件可以损害由 VMM 管理者配置的虚拟网络的隔离。在以下的描述中，出于解释的目的阐述了各种具体细节。然而，本领域技术人员将清楚，无需这些具体的细节就可以实践本发明的实施方案。

[0020] 本发明的实施方案可以以软件、固件、硬件或由各种技术的任何组合来实现。例如，在一些实施方案中，本发明可以被提供为可以包括其上存有指令的机器或计算机可读介质的计算机编程产品或软件，所述指令可以被用来编程计算机（或其他电子设备），以根据本发明来完成过程。在其他实施方案中，本发明的步骤可以由包含用于完成所述步骤的硬连线逻辑的具体硬件组件来完成，或由编程的计算机组件和硬件组件的任何组合来完成。

[0021] 因此，机器可读介质可以包括用于以机器（例如计算机）可读形式储存和传输信息的任何机制。这些机制包括，但不限于，软盘、光盘、致密盘只读存储器（CD-ROM）、磁光盘、只读存储器（ROM）、随机访问存储器（RAM）、可擦除可编程只读存储器（EPROM）、电可擦除可编程只读存储器（EEPROM）、磁或光卡、闪存存储器，通过因特网、电、光、声或其他形式的传播信号（例如载波、红外信号、数字信号等）的传输等等。随着新的机制类型被开发出来并且根据本发明的特定应用，其他类型的机制可以被增加或被用来替换那些被描述的机制。

[0022] 下面的详细描述的一些部分是根据对计算机系统的寄存器或存储器中数据位的操作的算法和符号表示来介绍的。这些算法描述和表示是数据处理领域技术人员用来将他们工作的实质最有效地传达给本领域其他技术人员的工具。算法在这里，并且一般地，被认为是导致期望结果的自我一致的（self-consistent）操作序列。所述操作是要求物理量的物理操纵的那些操作。虽然不是必须的，这些量通常采取能够被储存、传输、组合、比较和以其他方式操纵的电或磁信号的形式。主要出于通用的原因，将这些信号称为位、值、元素、符号、字符、项、数等已常常证明是方便的。

[0023] 然而，应该记住所有这些和类似的术语都是与适当的物理量相关联的，并且仅仅是应用于这些量的简便标记。除非另外特别指出，正如从以下讨论中可以清楚认识到，利用例如“处理”、“计算”、“运算”、“确定”等术语的讨论可以是指计算机系统或者类似的电子计算设备的动作和过程，所述动作和过程将表示为计算机系统的寄存器和存储器内的物理量（例如，电子量）的数据操纵和转换为类似地表示为计算机系统存储器或寄存器或其他这样的信息储存、传输或显示设备内的物理量的其他数据。

[0024] 在整篇说明书中提及“一个实施方案”或“实施方案”意味着结合该实施方案描述的特定特征、结构或特性被包括在本发明的至少一个实施方案中。因此，短语“在一个实施

方案中”或“在实施方案中”在说明书中不同地方的出现不一定全是指同一实施方案。此外,特定特征、结构或特性可以以任何适当的方式被组合在一个或更多个实施方案。

[0025] 在以下的实施方案的详细描述中,参照了以图示说明的方式示出的可以在其中实践本发明的具体实施方案的附图。在附图中,同样的数字在所有几个视图中描述本质上类似的组件。这些实施方案以足够的细节来描述以使得本领域技术人员能够实践本发明。可以利用其他的实施方案并且可以做出结构、逻辑和电气改变而不背离本发明的范围。

[0026] 图 1 图示用于动态地可扩展的虚拟交换机的环境的一个实施方案,其中本发明的一些实施方案可以在所述环境中操作。图 1 中示出的具体组件表示可能适合于本发明的配置的一个实施例,并且不想要限制本发明。

[0027] 参照图 1,用于动态地可扩展的虚拟交换机的环境 100 包括一个或更多个 VM 102 到 112、VMM 114 和平台硬件 116,但不必要限于此。尽管在图 1 中示出了六个 VM,应该理解任何数量的 VM 可以存在于环境 100 中。下面更详细地描述了这些组件中的每一个。

[0028] VM 102 到 112 每个包括一个或更多个 VNIC。例如,VM 102 包括 VNIC 1 和 VNIC 2,而 VM 106 只包括 VNIC 5。图 1 中的每个 VNIC 具有独有的 ID。仅仅出于图示说明的目的,假定 VNIC 1 具有 ID 1, VNIC 2 具有 ID 2, VNIC 3 具有 ID 3,等等。

[0029] NIC 虚拟化是用于向 VM 提供物理 NIC 的抽象的技术。通过虚拟化,同一物理 NIC 可以被多个 VM 共享。每个 VM 认为它拥有所述物理 NIC。此外, NIC 虚拟化允许向 VM 呈现同一物理 NIC 的多个实例。例如,系统可以具有单个物理 NIC,但 VM 可以看见多个 VNIC,所述多个 VNIC 中的每一个与物理机器内的不同网络和 / 或物理 NIC 所附接到的 外部网络接口。在本发明的实施方案中, VNIC 1 到 8 是平台硬件 116 中的网络接口卡 (NIC) 128 的抽象。

[0030] VMM 114 包括虚拟交换机 118。虚拟交换机 118 包括路由器 120、122 和 124 的集合以及数据结构 126。尽管在图 1 中示出了三个路由器,应该理解任何数量的路由器可以存在于虚拟交换机 118 中。如以下将详细描述的,路由器 120、122 和 124 使用数据结构 126 来组织 VM 102 到 112 中的 VNIC 之间的连接以形成各种虚拟网络。在本发明的实施方案中,虚拟交换机 118 对被 VMM 114 使用的虚拟化模型 (例如,管理程序 (hypervisor) 式的,基于宿主 (host-based) 式的,混合式的,等等) 并不可知。随着新型的虚拟化模型被开发出来并且根据本发明的特定应用,其他类型的虚拟化模型可以被增加或被用来替换那些被描述的虚拟化模型。

[0031] 平台硬件 116 包括 NIC 128。NIC 128 被连接到物理网络 130。平台硬件 116 可以是个人计算机 (PC)、大型机、手持式设备、便携式计算机、机顶盒或任何其他计算系统。平台硬件 116 可以包括一个或更多个处理器和存储器 (在图 1 中未示出)。此外,平台硬件 116 可以包括存储器和各种其他的输入 / 输出设备 (在图 1 中也未示出)。

[0032] 平台硬件 116 中的处理器可以是任何类型的能够执行软件的处理器,所述处理器例如超线程、SMP、多核心、微处理器、数字信号处理器、微控制器等,或它们的任何组合。随着新型的处理器被开发出来并且根据用于环境 100 的特定应用,其他类型的处理器可以被增加或被用来替换那些被描述的处理器。所述处理器可以包括用于完成针对本发明的方法的实施方案的执行的微码、宏代码、软件、可编程逻辑、硬编码逻辑等,但不必要限于此。

[0033] 平台硬件 116 的存储器可以是任何类型的可记录 / 非可记录介质 (例如随机访

问存储器 (RAM)、只读存储器 (ROM)、磁盘储存介质、光储存介质、闪存存储器设备等), 以及电、光、声或其他形式的传播信号 (例如载波、红外信号、数字信号等)、以上设备的任何组合, 或任何其他类型的处理器可读机器介质。随着新型的可记录 / 非可记录介质被开发出来并且根据用于本发明的特定应用, 其他类型的可记录 / 非可记录介质可以被增加或被用来替换那些被描述的可记录 / 非可记录介质。存储器可以储存用于完成本发明的方法实施方案的执行的指令。

[0034] 在环境 100 中, 平台硬件 116 包括计算平台, 所述计算平台能够例如执行标准操作系统 (OS) 或例如 VMM 114 的虚拟机监视器 (VMM)。尽管 VMM 114 通常以软件实现, VMM 114 可以向更高层软件仿真并导出裸机接口。这样的更高层软件可以包括标准或实时 OS, 可以是具有有限的操作系统功能性的高度精简 (stripped down) 的操作环境, 或可以不包括传统的 OS 设施 (facility)。可替换地, 例如, VMM 114 可以在另一个 VMM 的内部或在另一个 VMM 之上运行。VMM 以及它们的典型特征和功能性的为本领域内技术人员所熟知, 并且可以例如以软件、固件、硬件或各种技术的组合来实现。

[0035] 在本发明的实施方案中, 虚拟交换机 118 中的每个路由器, 连同它相关联的 VNIC 表示一个虚拟网络。因此, 路由器 120、VM 102 中的 VNIC 2 和 VM 104 中的 VNIC 3 表示一个虚拟网络; 路由器 122、VM 104 中的 VNIC 4、VM 106 中的 VNIC 5 和 VM 108 中的 VNIC 6 表示第二虚拟网络; 并且路由器 124、VM 110 中的 VNIC 7 和 VM 112 中的 VNIC 8 表示第三虚拟网络。VM 102 中的 VNIC 1 是图 1 中示出的唯一的 (经由 NIC 128) 连接到物理网络 130 的 VNIC。尽管在图 1 中示出了四个虚拟网络, 可以理解, 任何数量的虚拟网络可以存在于环境 100 中。

[0036] 下面参照图 2 进一步描述虚拟交换机数据结构 126。如上面描述的, 路由器 120、122 和 124 使用数据结构 126 来组织 VM 102 到 112 中的 VNIC 之间的连接以在环境 100 中形成各种虚拟网络。数据结构 126 包括 VNIC ID 阵列 202 和虚拟网络头列表 204。下面更详细的描述这些中的每一个。

[0037] 在本发明的实施方案中, 在图 1 中的每个 VNIC 具有独有的 ID。仅仅出于图示说明的目的, 假定 VNIC 1 具有 ID 1, VNIC 2 具有 ID 2, VNIC 3 具有 ID 3, 等等。每一独有的 VNIC ID 用作对 VNIC ID 阵列 202 的元素的索引。VNIC ID 阵列 202 的元素包含指向虚拟网络头的指针和实际的 VNIC 数据结构 (在图 2 中未示出)。每个虚拟网络头包含指向 VNIC 节点双向链表 (doubly linked list) 的指针。每个 VNIC 节点包含 VNIC 的 ID (即对 VNIC ID 阵列 202 的元素的索引)。

[0038] 在本发明的实施方案中, 如上面参照图 1 描述的, VM 102 中的 VNIC 1 是唯一的 (经由 NIC 128) 连接到物理网络 130 的 VNIC。在图 2 中, 通过 VNIC ID 阵列 202 中包含指向 (虚拟网络头列表 204 的) 虚拟网络头 1 的指针的元素 1, 并且通过包含指向 VNIC 节点双向链表的指针的虚拟网络头 1, 所述 VNIC 节点双向链表包含 (表示 VNIC 1 的) ID 1, 示出了该虚拟网络。

[0039] 同样地, 如上面参照图 1 描述的, 路由器 120、VM 102 中的 VNIC 2 和 VM 104 中的 VNIC 3 创建另一个虚拟网络。在图 2 中, 通过 VNIC ID 阵列 202 中的元素 2 和 3, 所述元素每一个包含指向虚拟网络头 2 的指针, 并且通过包含指向 VNIC 节点双向链表的指针的虚拟网络头 2, 所述 VNIC 节点双向链表包含 (表示 VNIC 2 和 3 的) ID 2 和 3, 示出了该虚拟网

络。针对图 1 的其他虚拟网络,在图 2 中示出了类似的排列。

[0040] 在本发明的实施方案,虚拟交换机数据结构 126 用于动态地完成虚拟网络上的操作而无需重新启动 VM 102 到 112 和 / 或 VMM 114。这些操作包括,但不限于,(1) 创建新的虚拟网络;(2) 删除现有虚拟网络;(3) 将 VNIC 添加到特定虚拟网络;以及(4) 从特定虚拟网络删除 VNIC。下面参照图 3-8 描述这些操作的实施方案。

[0041] 图 3 是用于创建新的虚拟网络的过程的一个实施方案的流程图。参照图 3,过程在处理框 302 开始,其中在处理框 302 中创建新的虚拟网络头。在处理框 304,新创建的虚拟网络头被添加到虚拟网络头列表 204。图 3 的过程在这里结束。

[0042] 图 4 是用于删除现有虚拟网络的过程的一个实施方案的流程图。仅仅出于图示说明的目的,假定虚拟网络 3 要从图 2 的虚拟交换机数据结构 126 中被删除。在图 5 中示出了该图示说明。

[0043] 参照图 4,过程在处理框 402 开始,其中在处理框 402 中 VNIC ID 阵列 202 中的所有 VNIC ID 元素与要被删除的虚拟网络头解除关联 (disassociate)。在我们的图示说明中,这涉及从 VNIC ID 阵列 202 的元素 4、5 和 6 删除指针 (如图 5 中所示)。

[0044] 在处理框 404,要被删除的虚拟网络头的对应 VNIC 节点双向链表被删除。在我们的图示说明中,这涉及从虚拟网络头 3 删除 VNIC 节点双向链表 (如图 5 中所示)。

[0045] 在处理框 406,虚拟网络头从虚拟网络头列表 204 中被删除。在我们的图示说明中,这涉及从虚拟网络头列表 204 删除虚拟网络头 3 (如图 5 中所示)。图 4 中的过程在这里结束。

[0046] 图 6 是用于将 VNIC 添加到特定虚拟网络的过程的一个实施方案的流程图。出于图示说明的目的,假定 VNIC 4 (VNIC ID 阵列 202 中的元素 4) 要被添加到图 2 的虚拟网络 2。在图 7 中示出了该图示说明。

[0047] 参照图 6,过程在处理框 602 开始,其中在处理框 602 中 VNIC ID 阵列 202 中的 VNIC ID 元素与它的旧虚拟网络头解除关联。在我们的图示说明中,元素 4 目前与虚拟网络头 3 相关联。因此,从元素 4 指向虚拟网络头 3 的指针被删除 (如图 7 中所示)。

[0048] 在处理框 604, VNIC ID 阵列 202 中的 VNIC ID 元素与它的新虚拟网络头相关联。在我们的图示说明中,从元素 4 指向虚拟网络头 2 的指针被创建 (如图 7 中所示)。

[0049] 在处理框 606,针对旧的和新的虚拟网络头两者的 VNIC 节点双向链表被更新。在我们的图示说明中, VNIC 4 节点从虚拟网络头 3 的双向链表中被删除,并且被添加到虚拟网络头 2 的双向链表 (如图 7 中所示)。图 6 中的过程在这里结束。

[0050] 图 8 是用于从特定虚拟网络删除 VNIC 的过程的一个实施方案的流程图。参照图 8,过程在处理框 802 开始,其中在处理框 802 中 VNIC ID 阵列 202 中的 VNIC ID 元素与它的虚拟网络头解除关联。

[0051] 在处理框 804,虚拟网络头的 VNIC 节点双向链表被更新以移去对应于被删除 VNIC 的 VNIC 节点。图 8 中的过程在这里结束。

[0052] 如上面描述的,虚拟交换机可以被用来提供用于在多个虚拟网络之间路由信息或数据帧的交换功能。现有虚拟交换机通常通过数据帧曾从其被发送的 VNIC 的 MAC 地址,或通过被储存在数据帧本身中的 MAC 地址来识别数据帧的源 VNIC 节点。任何恶意的软件 (例如在 VM 中运行的客户 OS) 可能欺骗 MAC 地址并因此导致数据帧的接收节点相信所述数据

帧曾来自于与它所曾来自的实际源 VNIC 节点不同的源 VNIC 节点。MAC 地址的欺骗损害虚拟网络的完整性。

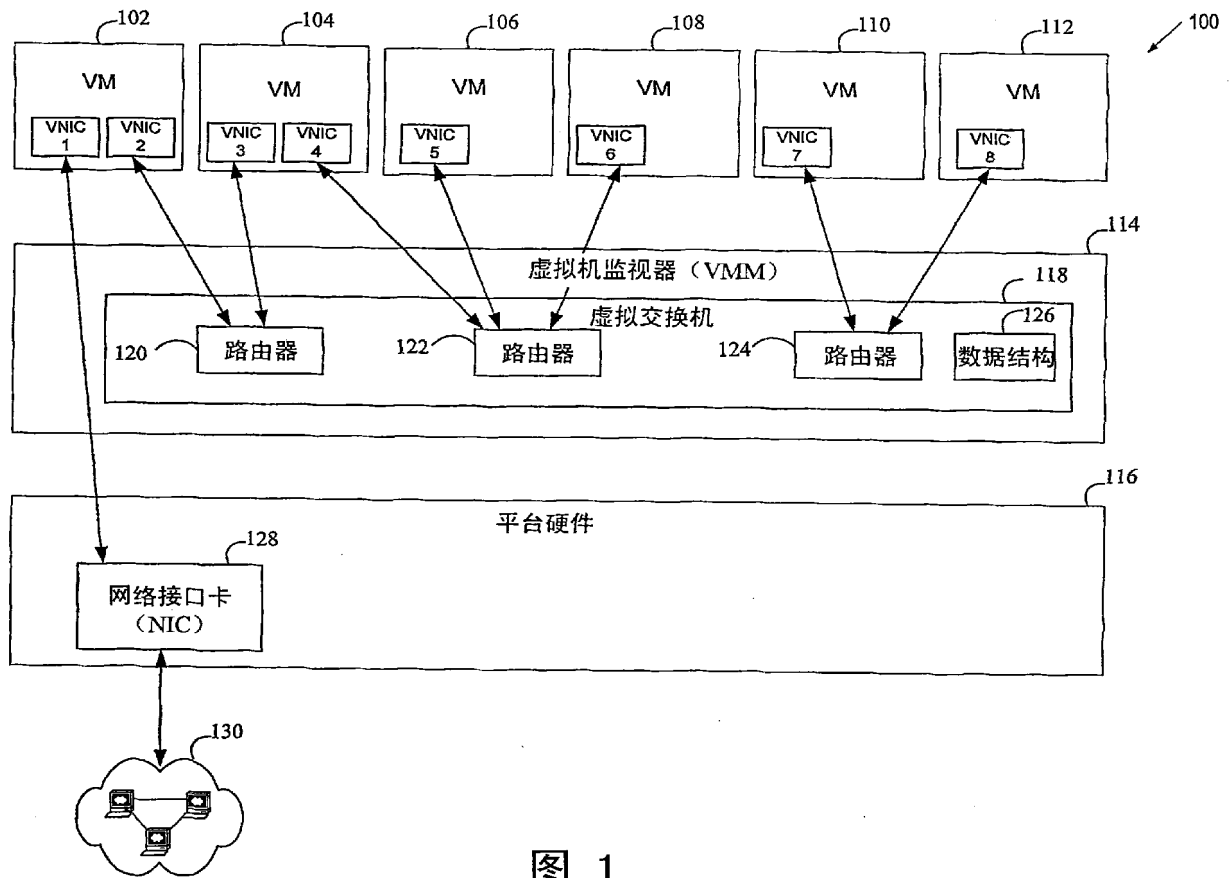
[0053] 在本发明的实施方案中,虚拟交换机 118 通过用源 VNIC 独有的 VNIC ID 而不是用 MAC 地址识别数据帧的源 VNIC 来克服其中恶意软件可以欺骗 MAC 地址的情况。通过指派和使用对在 VM 中运行的进程来说不可见的 VNIC ID,虚拟交换机 118 确保只是仅仅基于 MAC 地址来路由数据帧的恶意软件可能损害虚拟网络的完整性。

[0054] 图 9 是用于数据帧路由方案的过程的一个实施方案的流程图,所述数据帧路由方案避免欺骗的可能性。参照图 9,过程在处理框 902 开始,其中在处理框 902 中虚拟交换机 118 用源 VNIC 节点的独有的 ID 来识别数据帧的源 VNIC 节点。

[0055] 在处理框 904,虚拟交换机 118 利用 VNIC ID 来索引到 VNIC ID 阵列 202,以取得指向源 VNIC 节点所属的虚拟网络头的指针。

[0056] 在处理框 906,虚拟交换机 118 向 VNIC 节点双向链表中除源 VNIC 节点之外的每一个 VNIC 节点转发数据帧。图 9 的过程在这里结束。

[0057] 已经描述了用于动态地可扩展的虚拟交换机的装置和方法。应该理解,上面的描述旨在图示说明性的,而不是限制性的。一旦阅读和理解上面的描述,本领域技术人员将清楚许多其他实施方案。因此,本发明的范围应该参照所附权利要求书连同赋予权利要求书的整个等同物范围一起来确定。



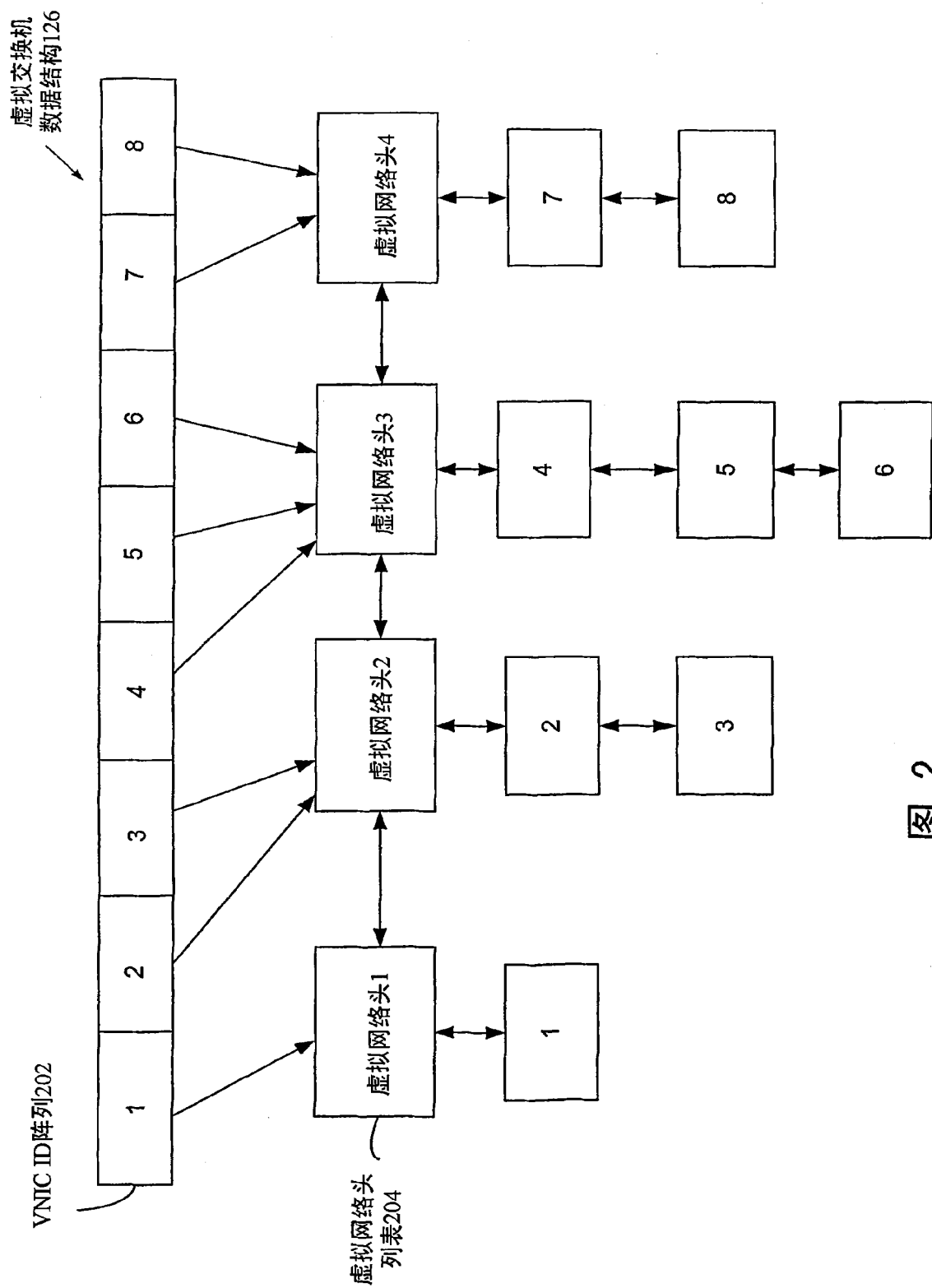


图 2

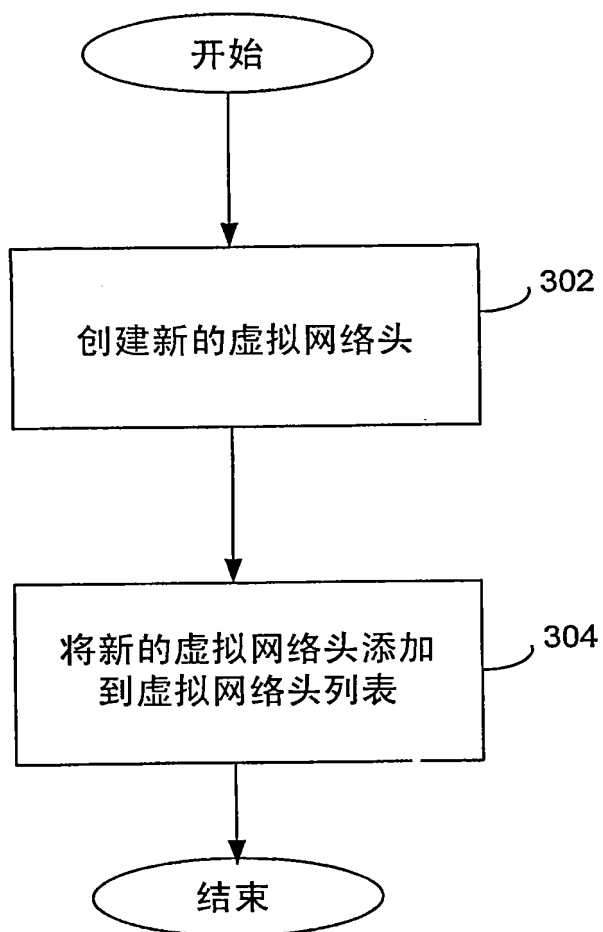


图 3

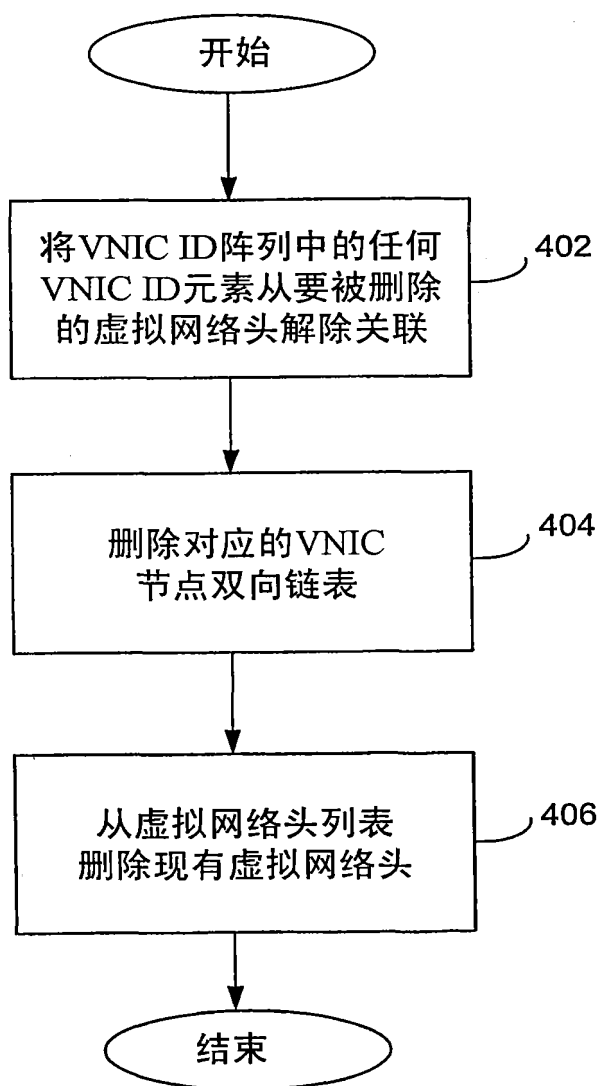


图 4

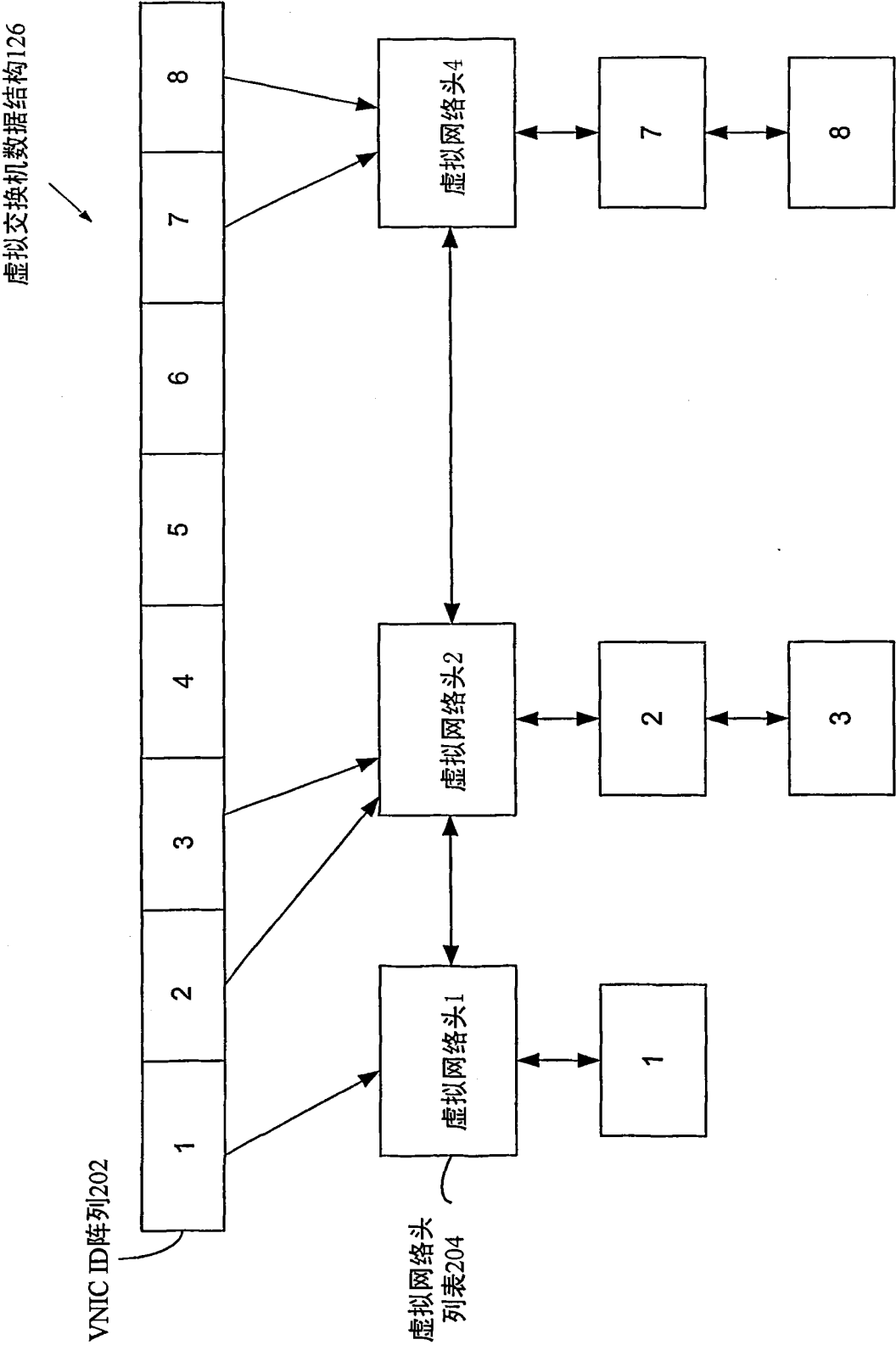


图 5

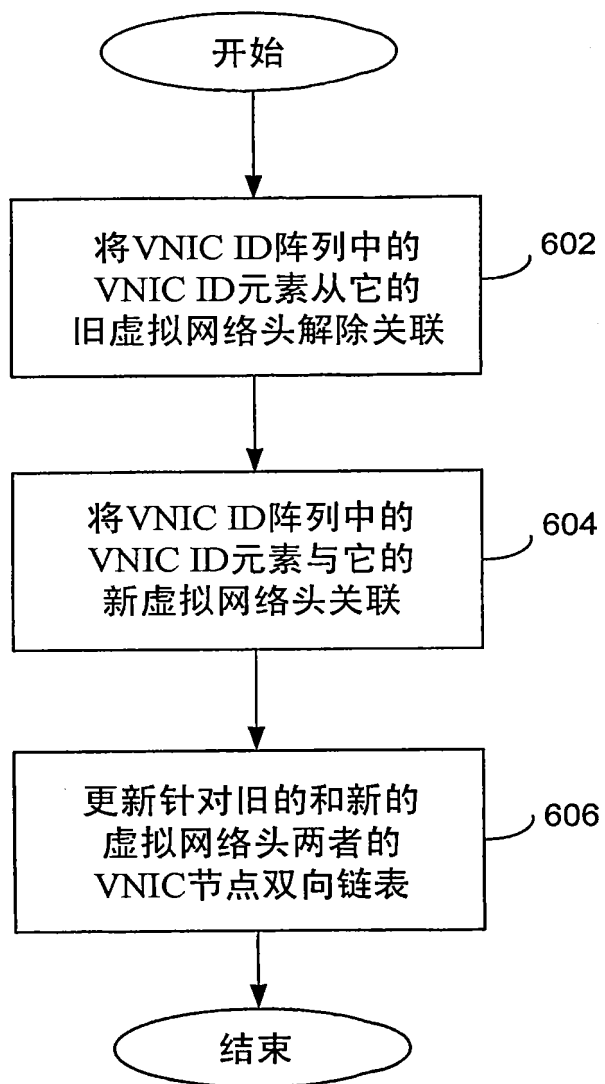


图 6

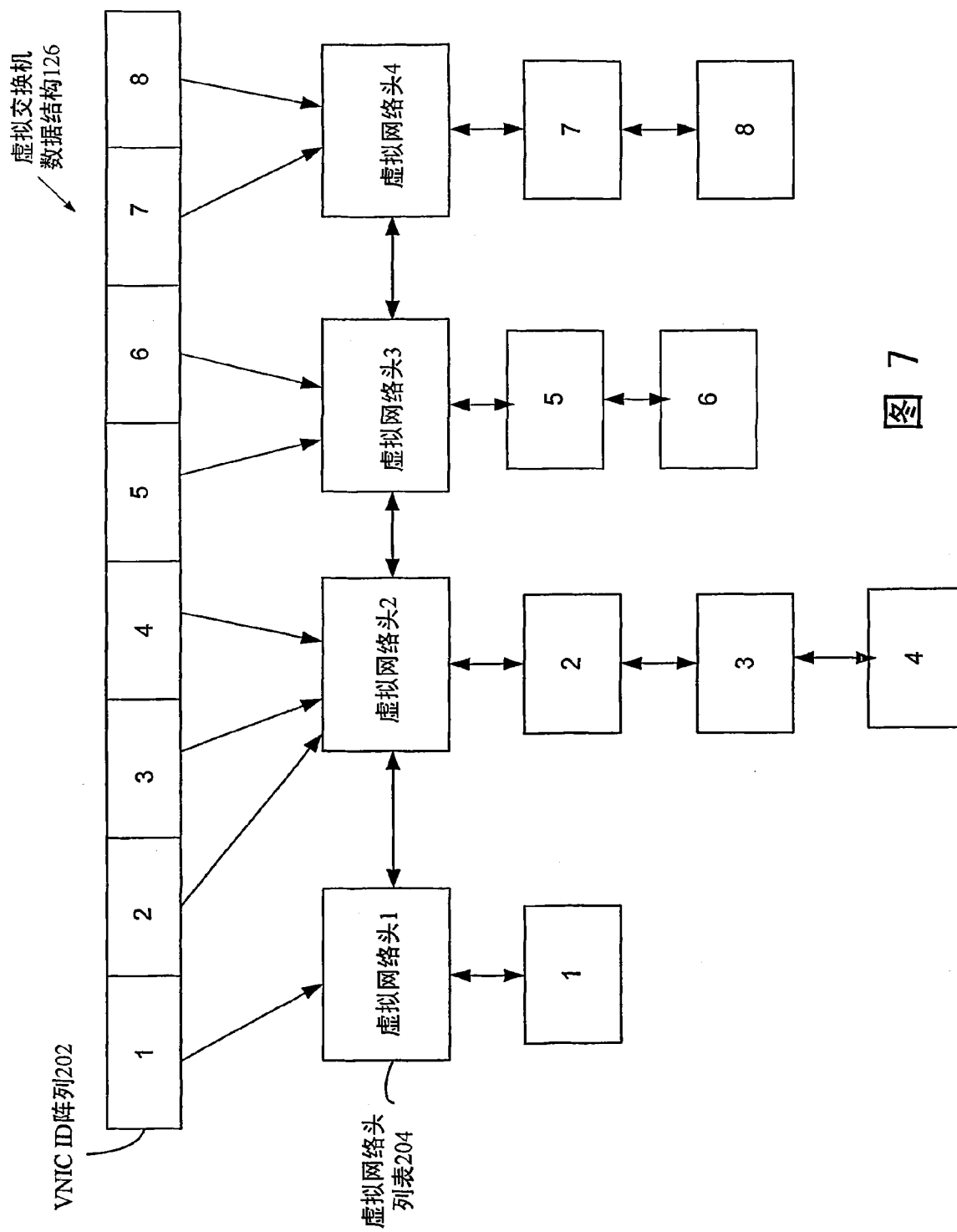


图 7

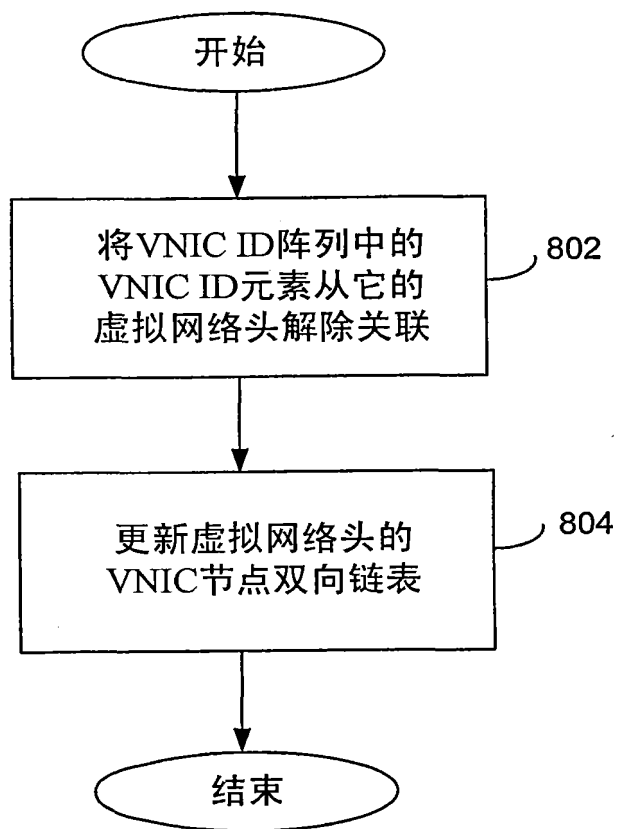


图 8

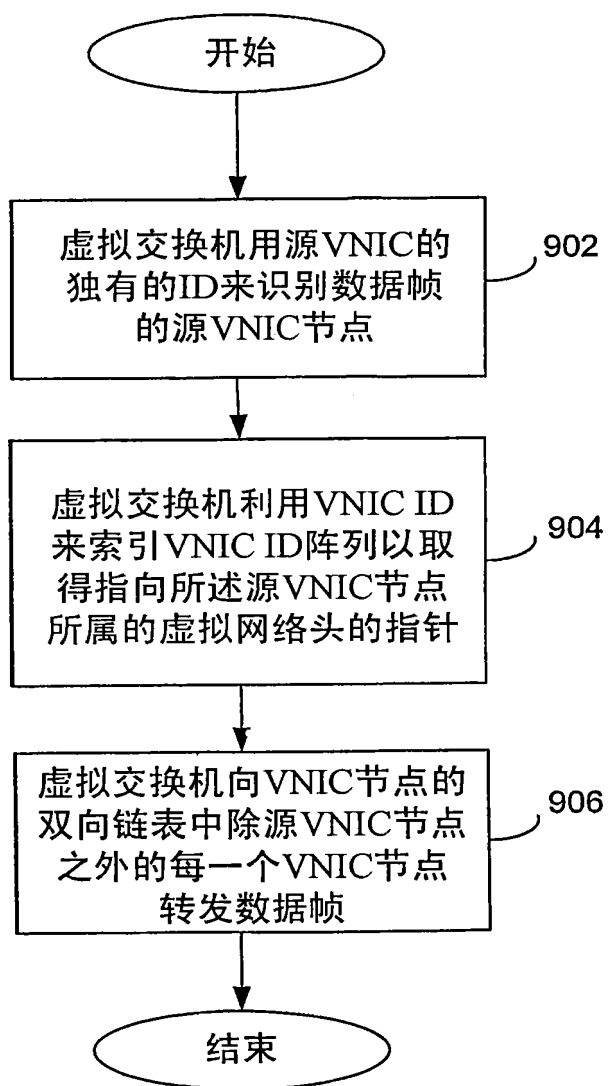


图 9