

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2016 年 12 月 8 日 (08.12.2016)



(10) 国际公布号
WO 2016/192214 A1

- (51) 国际专利分类号:
H04L 12/24 (2006.01) H04L 29/12 (2006.01)
- (21) 国际申请号: PCT/CN2015/087999
- (22) 国际申请日: 2015 年 8 月 25 日 (25.08.2015)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201510295684.7 2015 年 6 月 2 日 (02.06.2015) CN
- (71) 申请人: 清华大学 (TSINGHUA UNIVERSITY) [CN/CN]; 中国北京市海淀区清华园, Beijing 100084 (CN)。
- (72) 发明人: 尹浩 (YIN, Hao); 中国北京市海淀区清华大学 FIT 楼 3-406, Beijing 100084 (CN)。 张旭 (ZHANG, Xu); 中国北京市海淀区清华大学 FIT 楼 3-429, Beijing 100084 (CN)。 刘洪强 (LIU, Hongqiang); 中国北京市海淀区清华大学 FIT 楼 3-429, Beijing 100084 (CN)。
- (74) 代理人: 北京集佳知识产权代理有限公司 (UNITALEN ATTORNEYS AT LAW); 中国北京市朝阳区建国门外大街 22 号赛特广场 7 层, Beijing 100004 (CN)。

- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告(条约第 21 条(3))。

(54) Title: END SERVER DEPLOYMENT METHOD AND DEVICE

(54) 发明名称: 一种端服务器部署方法及装置

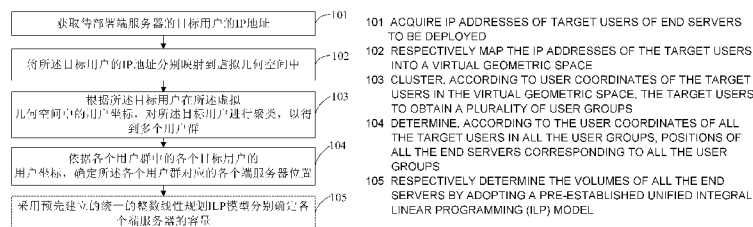


图 1

(57) Abstract: The present application provides an end server deployment method and device. The method comprises: acquiring IP addresses of target users of end servers to be deployed; respectively mapping the IP addresses of the target users into a virtual geometric space; clustering, according to user coordinates of the target users in the virtual geometric space, the target users to obtain a plurality of user groups; and determining, according to the user coordinates of all the target users in all the user groups, positions of all the end servers corresponding to all the user groups.

(57) 摘要: 本申请提供了一种端服务器部署方法及装置, 所述方法包括: 获取待部署端服务器的目标用户的 IP 地址; 将所述目标用户的 IP 地址分别映射到虚拟几何空间中; 根据所述目标用户在所述虚拟几何空间中的用户坐标, 对所述目标用户进行聚类, 以得到多个用户群; 依据各个用户群中的各个目标用户的用户坐标, 确定所述各个用户群对应的各个端服务器位置。

WO 2016/192214 A1

一种端服务器部署方法及装置

相关申请的交叉引用

本申请要求 2015 年 6 月 2 日提交中国专利局、申请号为 201510295684.7、发明名称为“一种端服务器部署方法及装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及互联网技术领域，特别涉及一种端服务器部署方法及装置。

背景技术

10 目前，随着互联网的发展，越来越多的用户对网络服务的质量要求越来越高，而许多网络服务（如搜索引擎，在线视频服务，社交网络，端计算服务（Edge Computing 等））都需要利用广泛分布于各个不同的地理和网络区域的端服务器来保证服务质量。因为网络服务质量高就要求服务器对其所服务的用户的网络延迟要足够小，而为了实现这一点，网络服务提供商必须要在离用户近的地方部署足够的端服务器资源。

15 但是发明人在研究过程中发现，目前已知的端服务器部署方法存在以下问题：仅仅在给定的候选服务器站点上部署端服务器，以及，仅仅针对单一目标进行优化，例如，仅仅考虑了服务器数量最小这一个方面。

发明内容

20 现有技术中，首先，如果仅仅在给定的候选服务器站点上部署端服务器，就缺乏发现未知的更好的端服务器站点的方法，这会导致网络服务提供商错过大量的提高网络延时、节约成本和改善用户性能的机会；其次，现有方法仅仅是对一个单一目标进行优化，但是却没有综合考虑

各种实际因素，提供一个权衡曲线来让网络服务提供商根据自己的现有情况（例如成本上限和用户性能要求等）来自行决定在哪里部署端服务器。

5 基于此，本申请所要解决的是上述问题中的一个或多个。本申请的一些实施例提供一种端服务器部署方法，用以尽量避免现有技术中仅仅在给定的候选服务器站点上以及仅仅考虑一个目标来部署端服务器导致的问题，从而提高端服务器对用户的服务质量，在改善用户性能的同时，还能节约部署端服务器的成本。

10 本申请还提供了一种端服务器部署装置，用以保证上述方法在实际中的实现及应用。

为了解决上述问题，本申请实施例公开了一种端服务器部署方法，包括：

获取待部署端服务器的目标用户的 IP 地址；

将所述目标用户的 IP 地址分别映射到虚拟几何空间中；

15 根据所述目标用户在所述虚拟几何空间中的用户坐标，对所述目标用户进行聚类，以得到多个用户群；

依据各个用户群中的各个目标用户的用户坐标，确定所述各个用户群对应的各个端服务器位置。

本申请实施例还公开了一种端服务器部署装置，包括：

20 获取模块，用于获取待部署端服务器的目标用户的 IP 地址；

映射模块，用于将所述目标用户的 IP 地址分别映射到虚拟几何空间中；

聚类模块，用于根据所述目标用户在所述虚拟几何空间中的用户坐标，对所述目标用户进行聚类，以得到多个用户群；

25 确定位置模块，用于依据各个用户群中的各个目标用户的用户坐标，确定所述各个用户群对应的各个端服务器位置。

本申请实施例还公开了一种端服务器的部署装置，包括：

存储器，和

处理器，

其中所述存储器存储有程序指令，所述程序指令当由所述处理器执行时使所述装置执行下述步骤：

获取待部署端服务器的目标用户的 IP 地址；

5 将所述目标用户的 IP 地址分别映射到虚拟几何空间中；

根据所述目标用户在所述虚拟几何空间中的用户坐标，对所述目标用户进行聚类，以得到多个用户群；

依据各个用户群中的各个目标用户的用户坐标，确定所述各个用户群对应的各个端服务器位置。

10 本申请实施例还公开了一种计算机可读存储介质，所述存储介质包含程序指令，所述程序指令当由处理器执行时，使处理器执行本申请中的方法。

与现有技术相比，本申请实施例包括以下优点：

在本申请实施例中，由于将所有待部署端服务器的目标用户的 IP 地址先映射到一个虚拟的几何空间中，进而参考用户在几何空间中的坐标来对用户进行聚类，进而根据聚类后的用户坐标来确定各个用户群对应的各个端服务器位置。可见本申请实施例考虑到了所有用户的地理位置特点，即，根据网络服务中用户的地理和/或网络位置来确定端服务器的位置，从而可以发现未知的可能部署位置更好的端服务器，使得端服务器部署的性能更好。

20

当然，实施本申请的任一产品并不一定需要同时达到以上所述的所有优点。

附图说明

为了更清楚地说明本申请实施例中的技术方案，下面将对实施例描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

25

图 1 是本申请的端服务器部署方法实施例的流程图；

图 2 是本申请的端服务器部署装置实施例的结构框图。

具体实施方式

下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本申请一部分实
5 施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术
人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本
申请保护的范围。

本申请可用于众多通用或专用的计算装置环境或配置中。例如：个
人计算机、服务器计算机、手持设备或便携式设备、平板型设备、多处
10 理器装置、包括以上任何装置或设备的分布式计算环境等等。

本申请可以在由计算机执行的计算机可执行指令的一般上下文中描
述，例如程序模块。一般地，程序模块包括执行特定任务或实现特定抽
象数据类型的例程、程序、对象、组件、数据结构等等。也可以在分布
式计算环境中实践本申请，在这些分布式计算环境中，由通过通信网络
15 而被连接的远程处理设备来执行任务。在分布式计算环境中，程序模块
可以位于包括存储设备在内的本地和远程计算机存储介质中。

本申请的一个实施例提供一种端服务器定位方法，即端服务器部署
方法。首先，从用户在地理和网络上的位置分布来反推适合的甚至是最
优化的端服务器的位置。随后，围绕着这些被计算出来的理想的端服务
器位置，网络服务提供商去搜索事先未知的可以部署端服务器的位置以
20 及方法。其中，部署方法可以例如，租用第三方服务器，自己部署服务
器机房，利用私有网络光纤连接数据中心，或者，利用虚拟网络功能服
务（Network Function Virtualization, NFV），等等。然后网络服务提供商
就可以得到一个候选端服务器部署节点的集合，其中包括了事先知道的
25 端服务器，以及事先未知并在反推过程中根据用户分布而发现的端服
务器。

下面结合图 1 对本申请实施例的端服务器部署方法进行描述。图 1
示出了根据本申请的示例实施例的一种端服务器部署方法的流程图。

步骤 101: 获取待部署端服务器的目标用户的 IP 地址。

在本步骤中, 为了从用户的位置信息来反推出端服务器部署的合理位置, 需要知道所有用户到所有端服务器的网络距离。但是, 网络距离这种信息又是不可能完整得到的, 因为在实际中无法控制数以百万计或者更多的用户来主动与端服务器进行网络延迟测量, 而且在完成端服务器的部署和发现之前, 用户本身没有办法测量与未知的端服务器之间的网络距离。因此, 为了解决这个问题, 步骤 101 先获取到所有待部署端服务器的目标用户的 IP 地址, 然后在下面要描述的步骤 102 中将所有的用户的 IP 地址映射到一个虚拟的几何空间里, 所有的网络节点 (包括用户和端服务器) 都是这个几何空间的一个有确定坐标的点, 那么任意两个网络节点之间的网络距离就可以由它们在虚拟网络空间中相应的坐标计算获得。

可以通过例如分析用户访问的日志来确定目标用户, 或者指定待服务区域, 从各区域进行随机抽样得到目标用户。

步骤 102: 将所述目标用户的 IP 地址分别映射到虚拟几何空间中。

步骤 102 将在步骤 101 中获取到的目标用户的 IP 地址分别映射到虚拟的几何空间 S 中。虚拟几何空间 S 例如可以是网络技术中经常使用的地理位置空间和网络空间。这两种虚拟几何空间分别采用网络坐标 (Network Coordinate) 和地理位置坐标, 其中, 地理坐标可以为经纬度坐标 (Geographical Coordinate)。

其中, 如虚拟的几何空间为地理位置空间, 即, 虚拟几何空间的虚拟坐标为地理位置坐标, 则步骤 102 具体可以包括步骤 A1~步骤 A2:

步骤 A1: 从地理位置数据库中查找各个目标用户的 IP 地址对应的地理位置信息, 所述地理位置数据库中保存有 IP 地址和地理位置信息之间的对应关系。

例如, 可以维护一个地理位置数据库, 在该地理位置数据库中保存有 IP 地址与地理位置信息之间的对应关系 (映射)。该对应关系可以从例如 IP 地址的地理位置数据库获得。那么在获取到各个目标用户的 IP 地址

之后，直接从该地理位置数据库中查找对应的地理位置信息即可。例如，用户的 IP 地址为：202.112.14.128，那么其映射出的地理位置信息可以包括该 IP 地址对应的区域，如“北京市”，也可以包括该 IP 地址对应的经纬度坐标，如（56.8，23.9），等等。

- 5 步骤 A2：根据所述地理位置信息确定所述目标用户在所述虚拟几何空间中的用户坐标。

在本步骤中例如直接将步骤 A1 中查找到的地理位置信息确定为目标用户在虚拟几何空间中的用户坐标。

- 10 其中，如虚拟的几何空间为网络空间，即，虚拟几何空间的虚拟坐标为网络虚拟坐标，那么步骤 102 具体可以包括：

步骤 B1：采用网络空间构建 GNP 算法将所述目标用户的 IP 地址分别映射为网络虚拟坐标。

- 15 例如，利用具体的网络空间构建算法（GNP）实现从 IP 地址到网络虚拟坐标的映射。具体的映射过程为：假设一共有 N 个已知其相互间的网络延迟的用户（即标记节点），N 为大于零的整数，则首先通过这 N 个标记节点相互间的网络延迟测量的结果，计算出这 N 个标记节点的网络空间的坐标；然后通过这 N 个标记节点向所有的待测量节点（即目标用户）发起网络延迟测量，并通过该测量的结果和已知的 N 个标记节点的网络坐标计算出所有待测量节点的网络坐标。

- 20 步骤 B2：将所述网络虚拟坐标确定为所述目标用户在所述虚拟几何空间中的用户坐标。

在本步骤中就将步骤 B1 映射得到的网络虚拟坐标确定为目标用户在网络空间中的用户坐标。

- 25 在得到虚拟几何空间中的用户坐标之后，接着进入步骤 103：根据所述目标用户在所述虚拟几何空间中的用户坐标，对所述目标用户进行聚类，以得到多个用户群。

在本步骤中，会按照目标用户在虚拟几何空间 S 中的用户坐标，即，用户之间的相对距离来对目标用户进行聚类，从而得到多个用户群。在

进行用户聚类之后，每个目标用户最后会归入一个用户群，由于一个用户群中的用户已经相互离得足够近，这样就可以更方便的找出一个或者几个端服务器的位置同时距离某一个用户群的所有用户都足够近。

5 步骤 103 可以根据所述虚拟几何空间中目标用户之间的网络距离与实际距离的一致性，和，最远用户优先原则，对所述目标用户进行聚类，以得到多个用户群。

需要说明的是，在本实施例中，由于虚拟几何空间 S 是一个对现实网络空间的近似估计，直接利用几何空间 S 中的虚拟坐标进行实际距离的估算或者节点位置的估算都是有误差的，因此，在本实施例中仅仅利用几何空间 S 提供的网络距离的相对大小信息来进行用户聚类，而不是直接利用几何空间 S 提供的网络距离的绝对值进行用户聚类。因此，在实际应用中判断几何空间 S 的优劣标准就不是几何空间 S 对实际距离的估算的绝对值有多准确，而是几何空间 S 有多好的一致性。其中，可以理解的是，用户之间的网络距离可以采用用户之间的网络延迟时间来表示。例如，用户 A 向用户 B 发送了一个数据包，而用户 B 在接收到数据包的时间与用户 A 发送数据包的时间差即为网络延迟时间，也即用户 A 15 和用户 B 之间的网络距离。

因此，网络距离与实际距离的一致性可以为：任意两个目标用户之间的网络距离大于其他两个目标用户之间的网络距离时，该任意两个目标用户的实际距离也大于其他两个目标用户之间的实际距离。例如，目标用户 A 和 B 之间的网络距离大于目标用户 C 和 D 之间的网络距离，而目标用户 A 和 B 之间的实际距离也同样大于目标用户 C 和 D 之间的实际距离，或者，目标用户 E 和 F 之间的网络距离小于目标用户 G 和 H 之间的网络距离，而目标用户 E 和 F 之间的实际距离也小于目标用户 G 和 H 25 之间的实际距离，那么这种情况就称为网络距离与实际距离的一致性。可见，几何空间 S 的一致性是所有网络节点中一致性成立的前提，一个虚拟网络空间 S 的一致性越好，进行用户聚类的性能就越好。

其中，最远用户优先原则可以为：在任一个用户群中，网络距离最

远的两个目标用户之间的网络距离小于预先设置的聚类距离阈值，也即，一个用户群里面的任何两个目标用户之间的距离都不大于一个预先设置的聚类距离阈值 $D_{\max}$ 。

具体的，步骤 103 进行用户聚类可以例如：假设一共有 A, B, C, D 四个用户，它们之间的距离分别是 $d[A,B] = d[A,C] = d[B,C] = 2$, $d[A,D] = 10$, $d[B,D] = 12$, $d[C,D] = 12$ ，假设 $D_{\max}=5$ 。首先可以随机选择一个用户，例如 A。首先找到离 A 最远的用户 D，由于 $d[A,D] > D_{\max}$ ，所以可以将 D 单独放在一个新的类里面，而由于用户 B 和 C 到用户 A 的距离都小于 $D_{\max}$ ，所以用户 B 和 C 将与用户 A 归为一类。最后的聚类结果是：{A, B, C} 一类，{D} 一类，即得到两个用户群。

接着进入步骤 104：依据各个用户群中的各个目标用户的用户坐标，确定所述各个用户群对应的各个端服务器位置。

然后在得到各个用户群之后，根据各个用户群中各个目标用户的用户坐标，来确定各个用户群对应的各个端服务器位置。

其中，假设几何空间 S 为地理位置空间，那么本步骤可以包括步骤 C1~步骤 C2：

步骤 C1：获取各个用户群中的目标用户在虚拟几何空间中的地理位置信息。

针对各个用户群，先获取到各个用户群中的目标用户在几何空间 S 中的地理位置信息。例如，用户 A 的地理位置信息为北京市，用户 B 的地理位置信息为保定市等。

步骤 C2：依据目标用户的地理位置信息，确定各个用户群对应的各个端服务器位置。

然后依据目标用户的地理位置信息，确定出各个用户群对应的各个端服务器位置。可以将出现的次数最多的位置，确定为用户群对应的端服务器位置，而如果多个位置出现的次数一样多，则将该多个位置都确定为用户群对应的端服务器位置。例如，假设一个用户群有 10 个目标用户，其中，5 个目标用户都在北京市，3 个目标用户在天津市，2 个目标

用户在保定市，那么对应的端服务器位置就是北京市。再例如，假设一个用户群中有 5 个目标用户，其中，2 个目标用户的地理位置信息为天津市，2 个目标用户的地理位置信息为保定市，1 个目标用户的地理位置信息为北京市，那么可以确定该用户群对应两个端服务器位置，分别为天津市和保定市。

其中，假设几何空间 S 为网络空间，那么本步骤可以包括步骤 D1~步骤 D4:

步骤 D1: 确定各个用户群的各个中心用户，以及各个中心用户对应的各个邻近用户集，所述邻近用户集包括: 一个或多个与所述中心用户的网络距离小于预设距离阈值的用户。

在用户聚类以后，本步骤会找出每个用户群的中心用户。在一个实施例中，中心用户指用户群中到其它用户的最大距离最小的用户，在另一实施例中，中心用户指用户群中到其它用户的平均距离最小的用户。找到中心用户后再在用户群中选择 N (其中， N 为大于零的整数，具体数值可以由本领域技术人员自主设置) 个离中心用户最近的用户，形成每个中心用户对对应的邻近用户集。其中，邻近用户集可以包括: 一个或多个与所述中心用户的网络距离小于预设距离阈值的用户，其中，该预设距离阈值也可以由本领域技术人员自主设置。

步骤 D2: 确定各个邻近用户集中各个邻近用户的地理位置。

将各个邻近用户的网络坐标反映射到真实的地理空间，从而得到各个邻近用户集中各个邻近用户的地理位置。例如，确定各个邻近用户在哪个市区，当然，也可以采用经纬度信息表示。

步骤 D3: 将每个邻近用户集中出现频率最多的地理位置，确定为对应的中心用户的地理位置。

在本步骤中，针对每一个邻近用户集，看在该邻近用户集中哪个地理位置出现的最多，然后将出现最多的地理位置确定为该邻近用户集对应的中心用户的地理位置。例如，针对一个邻近用户集，北京市出现的次数最多，那么就将北京市确定为该邻近用户集对应的中心用户的地理

位置，而如果多个城市出现的次数一样多，则将该多个城市都确定为
中心用户的地理位置。

步骤 D4：根据预先设置的部署判断原则，判断所述各个中心用户的
地理位置是否适合部署端服务器，如果是，则将适合部署端服务器的地
5 理位置分别确定为对应的用户群的候选端服务器位置。

在找到了中心节点的地理位置之后，可以会依次在这些地理位置寻
找端服务器部署的可行性。例如，参考预先设置的部署判断原则，来判
断各个中心用户的地理位置是否适合部署端服务器，如果是，则将适合
部署端服务器的地理位置分别确定为对应的用户群的候选端服务器位
10 置，最终每个用户群会得到一个对应的候选端服务器的位置集合。可以
理解的是，在某个用户群的端服务器集合中，包含了所有离这个用户群
的用户足够近的、新发现的和事先知道的端服务器。其中，在判断某个
地理位置是否适合部署端服务器时，可以判断在这个地理位置有没有第
三方机房，有没有云计算服务器，以及当地政策和网络光纤条件是否允
15 许等等。具体判断原则本领域技术人员可以根据实际需求自主设置。

在确定端服务器的位置之后，可以理解的是，还可以对每个端服务
器的容量进行规划，则还在步骤 104 之后还可以包括：

步骤 105：采用预先建立的统一的整数线性规划 ILP 模型分别确定各
个端服务器的容量。

20 在本实施例中，端服务器的容量规划主要如何决定每个端服务器节
点所服务的用户群。即，将用户对端服务器资源的需求，端服务器与用
户的相对位置以及端服务器节点自身的承载能力进行综合建模，利用整
数线性规划（ILP）来得到最后的端服务器容量。

在实际应用中，由于每个端服务器节点都有实际容量限制，并且收
25 费价格也不尽相同，而且由于一些地区的法律或者政策限制，一些端服
务器只能服务于一些特定的用户。因此，在确定各个用户群的候选端服
务器节点集合之后，网络服务商需要找到一个服务器容量规划的方案，
来决定如何使用各个端服务器节点，包括每个端服务器节点有多大的容

量（即能够服务多少用户）。而该规划的目标通常是让端服务器节点能够服务于更多的用户，并且降低端服务器资源的使用成本。

在本实施例中，整数线性规划 ILP 模型的建模过程如表 1 所示：

表 1

输入参数 Inputs	
$G=\{g\}$	用户组的集合
N_g	用户组 g 中的用户个数
S_g	用户组 g 中的各个用户节点
M_s	端服务器节点 s 的最大容量
p_s	端服务器节点 s 的单位资源价格
c_s	开放端服务器节点 s 的一次成本
输出参数 Outputs	
$n_{g,s}$	被端服务器节点 s 提供服务的用户个数
$r_{g,s}$	用户组 g 在端服务器节点 s 上使用的总容量
y_s	为 1 时表示端服务器节点关闭，为 0 时表示开放

5

具体的，在建模过程中，将所有具有相同候选端服务器集合的用户都归为一个用户组中，并用 S_g 来表示这个用户组， N_g 表示用户组 g 中的用户个数，那么建模过程就需要最终计算出每个用户组中有多少用户被端服务器节点 s 服务器。

10

首先，所有端服务器服务的用户数目的总和应该不多于一个用户组中的用户总数，可以表示为如下所示的公式（一）：

$$\forall g \in G: \sum_s n_{g,s} \leq N_g \quad (一)$$

接着，用 $r_{g,s}$ 表示用户组 g 在端服务器节点 s 上使用的总容量，它要大于 g 在 s 上的实际用户数，则有如下所示的公式（二）：

15

$$\forall g \in G: r_{g,s} \geq n_{g,s} \quad (二)$$

而对于端服务器节点 s 所能承受的最大用户并发数量，有如下所示的公式（三）：

$$\forall s: \sum_g r_{g,s} \leq M_s \quad (三)$$

接着针对 y_s 表示端服务器节点 s 是不是被使用了, y_s 为 1 意味着 s 被使用了, 为 0 则意味着没有被使用, 则有如下所示的公式 (四) 和 (五):

$$\forall s: \frac{\sum_g r_{g,s}}{N} \leq y_s \leq \sum_g r_{g,s} \quad y_s \in \{0,1\} \quad (四)$$

5

$$\forall s: y_s \in \{0,1\} \quad (五)$$

而对于端服务器节点, 总体的部署成本是使用端服务器的一次性成本和资源使用成本之和, 则有如下所示的公式 (六):

$$\sum_s (p_s \sum_g r_{g,s} + c_s y_s) \quad (六)$$

其中, p_s 为服务节点 s 的资源价格, c_s 为服务节点 s 的起步使用价格。

10

最后, 对于端服务器部署来说, 目标就是最大化其能服务的用户数量, 并且最小化部署成本, 因此, 有如下所示的格式 (七):

$$\max \sum_{g,s} n_{g,s} - \eta \sum_s (p_s \sum_g r_{g,s} + c_s y_s) \quad (七)$$

其中, η 大于零, 可取一个常数, 来权衡用户数量和部署成本之间的关系。

15

在计算得到公式一~七的解之后, $\sum_g r_{g,s}$ 的值即为端服务器节点 s 的容量, 即, 端服务器节点可以服务多少个用户。

20

可见, 在本申请实施例, 由于将所有待部署端服务器的目标用户的 IP 地址先映射到一个虚拟的几何空间中, 进而参考用户在几何空间中的坐标来对用户进行聚类, 进而根据聚类后的用户坐标来确定各个用户群对应的各个端服务器位置。可见本申请实施例考虑到了所有用户的地理位置特点, 即, 根据网络服务中用户的地理和网络位置来确定端服务器的位置, 从而可以发现未知的可能部署位置更好的端服务器, 使得端服务器部署的性能更好; 进一步的, 还能够综合考虑部署端服务器的成本和用户性能需求, 来综合考虑进而部署端服务器, 同时保证了低成本和用户性能。

25

对于前述的方法实施例, 为了简单描述, 故将其都表述为一系列的动作组合, 但是本领域技术人员应该知悉, 本申请并不受所描述的动作

顺序的限制，因为依据本申请，某些步骤可以采用其他顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作和模块并不一定是本申请所必须的。

与上述本申请一种端服务器部署方法实施例所提供的方法相对应，
5 参见图 2，本申请还提供了一种端服务器部署装置实施例，在本实施例中，该装置可以包括：

获取模块 201，用于获取待部署端服务器的目标用户的 IP 地址。

映射模块 202，用于将所述目标用户的 IP 地址分别映射到虚拟几何空间中。

10 其中，所述几何空间可以为地理位置空间，则所述映射模块 202 可以包括：

查找子模块，用于从地理位置数据库中查找各个目标用户的 IP 地址对应的地理位置信息，所述地理位置数据库中保存有 IP 地址和地理位置信息之间的对应关系；和，第一确定子模块，用于将所述地理位置信息
15 确定为所述目标用户在所述虚拟几何空间中的用户坐标。

其中，所述虚拟几何空间可以为网络空间，所述映射模块 202 可以包括：

映射子模块，用于采用网络空间构建 GNP 算法将所述目标用户的 IP 地址分别映射为网络虚拟坐标；和，第三确定子模块，用于将所述网络
20 虚拟坐标确定为所述目标用户在所述虚拟几何空间中的用户坐标。

聚类模块 203，用于根据所述目标用户在所述虚拟几何空间中的用户坐标，对所述目标用户进行聚类，以得到多个用户群；

其中，所述聚类模块 203，具体可以用于：

根据所述虚拟几何空间中目标用户之间的网络距离与实际距离的一致性，和，最远用户优先原则，对所述目标用户进行聚类，以得到多个
25 用户群；其中，所述网络距离与实际距离的一致性为：任意两个目标用户之间的网络距离大于其他两个目标用户之间的网络距离时，该任意两个目标用户的实际距离也大于其他两个目标用户之间的实际距离；所述

最远用户优先原则为：在任一个用户群中，网络距离最远的两个目标用户之间的网络距离小于预先设置的距离阈值。

确定位置模块 204，用于依据各个用户群中的各个目标用户的用户坐标，确定所述各个用户群对应的各个端服务器位置。

5 其中，所述确定位置模块可以包括：

获取子模块，用于获取各个用户群中的目标用户在虚拟几何空间中的地理位置信息；和，第二确定子模块，用于依据目标用户的地理位置信息，确定各个用户群对应的各个端服务器位置。

其中，确定位置模块 204 还可以包括：

10 第四确定子模块，用于确定各个用户群的各个中心用户，以及各个中心用户对应的各个邻近用户集，所述邻近用户集包括：一个或多个与所述中心用户的网络距离小于预设距离阈值的用户；第五确定子模块，用于确定各个邻近用户集中各个邻近用户的地理位置；第六确定子模块，用于将每个邻近用户集中出现频率最多的地理位置，确定为对应的中心
15 用户的地理位置；判断子模块，用于根据预先设置的部署判断原则，判断所述各个中心用户的地理位置是否适合部署端服务器；和，第七确定子模块，用于在所述判断子模块的结果为是的情况下，将适合部署端服务器的地理位置分别确定为对应的用户群的候选端服务器位置。

其中，该装置还可以包括：

20 确定容量模块，用于采用预先建立的统一的整数线性规划 ILP 模型分别确定各个端服务器的容量。

在本申请实施例中，由于将所有待部署端服务器的目标用户的 IP 地址先映射到一个虚拟的几何空间中，进而参考用户在几何空间中的坐标来对用户进行聚类，进而根据聚类后的用户坐标来确定各个用户群对应的各个端服务器位置。可见本申请实施例考虑到了所有用户的地理位置
25 特点，即，根据网络服务中用户的地理和网络位置来确定端服务器的位置，从而可以发现未知的可能部署位置更好的端服务器，使得端服务器部署的性能更好；进一步的，还能够综合考虑部署端服务器的成本和用

户性能需求，来综合考虑进而部署端服务器，同时保证了低成本和用户性能。

需要说明的是，本说明书中的各个实施例均采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间
5 相同相似的部分互相参见即可。对于装置类实施例而言，由于其与方法实施例基本相似，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

最后，还需要说明的是，在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不
10 一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的
15 过程、方法、物品或者设备中还存在另外的相同要素。

以上对本申请所提供的端服务器部署方法及装置进行了详细介绍，本文中应用了具体个例对本申请的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本申请的方法及其核心思想；同时，对于本
20 领域的一般技术人员，依据本申请的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本申请的限制。

权 利 要 求

1、一种端服务器的部署方法，该方法包括：

获取待部署端服务器的目标用户的 IP 地址；

将所述目标用户的 IP 地址分别映射到虚拟几何空间中；

5 根据所述目标用户在所述虚拟几何空间中的用户坐标，对所述目标用户进行聚类，以得到多个用户群；

依据各个用户群中的各个目标用户的用户坐标，确定所述各个用户群对应的各个端服务器位置。

2、根据权利要求 1 所述的方法，其中，所述虚拟几何空间为地理位置空间，所述将所述目标用户的 IP 地址分别映射到虚拟几何空间中，包括：

从地理位置数据库中查找各个目标用户的 IP 地址对应的地理位置信息，所述地理位置数据库中保存有 IP 地址和地理位置信息之间的对应关系；

15 将所述地理位置信息确定为所述目标用户在所述虚拟几何空间中的用户坐标。

3、根据权利要求 2 所述的方法，其中，依据各个用户群中的各个目标用户的用户坐标，确定所述各个用户群对应的各个端服务器位置，包括：

20 获取各个用户群中的目标用户在虚拟几何空间中的地理位置信息；

依据目标用户的地理位置信息，确定各个用户群对应的各个端服务器位置。

4、根据权利要求 1 所述的方法，其中，所述虚拟几何空间为网络空间，所述将所述目标用户的 IP 地址分别映射到虚拟几何空间中，包括：

25 采用网络空间构建 GNP 算法将所述目标用户的 IP 地址分别映射为网络虚拟坐标；

将所述网络虚拟坐标确定为所述目标用户在所述虚拟几何空间中的用户坐标。

5、根据权利要求 4 所述的方法，其中，所述依据各个用户群的各个目标用户的用户坐标，确定所述各个用户群对应的各个端服务器位置，包括：

确定各个用户群的各个中心用户，以及各个中心用户对应的各个邻近用户集，所述邻近用户集包括：一个或多个与所述中心用户的网络距离小于预设距离阈值的用户；

确定各个邻近用户集中各个邻近用户的地理位置；

将每个邻近用户集中出现频率最多的地理位置，确定为对应的中心用户的地理位置；

10 根据预先设置的部署判断原则，判断所述各个中心用户的地理位置是否适合部署端服务器，如果是，则将适合部署端服务器的地理位置分别确定为对应的用户群的候选端服务器位置。

6、根据权利要求 4 所述的方法，其中，所述根据所述目标用户在所述虚拟几何空间中的用户坐标对所述目标用户进行聚类，以得到多个用户群，包括：

根据所述虚拟几何空间中目标用户之间的网络距离与实际距离的一致性，和，最远用户优先原则，对所述目标用户进行聚类，以得到多个用户群；

其中，所述网络距离与实际距离的一致性为：任意两个目标用户之间的网络距离大于其他两个目标用户之间的网络距离时，该任意两个目标用户的实际距离也大于其他两个目标用户之间的实际距离；所述最远用户优先原则为：在任一个用户群中，网络距离最远的两个目标用户之间的网络距离小于预先设置的距离阈值。

7、根据权利要求 1 所述的方法，还包括：

25 采用预先建立的统一的整数线性规划 ILP 模型分别确定各个端服务器的容量。

8、一种端服务器部署装置，包括：

获取模块，用于获取待部署端服务器的目标用户的 IP 地址；

映射模块，用于将所述目标用户的 IP 地址分别映射到虚拟几何空间中；

聚类模块，用于根据所述目标用户在所述虚拟几何空间中的用户坐标，对所述目标用户进行聚类，以得到多个用户群；

5 确定位置模块，用于依据各个用户群中的各个目标用户的用户坐标，确定所述各个用户群对应的各个端服务器位置。

9、根据权利要求 8 所述的装置，其中，所述虚拟几何空间为地理位置空间，所述映射模块，包括：

10 查找子模块，用于从地理位置数据库中查找各个目标用户的 IP 地址对应的地理位置信息，所述地理位置数据库中保存有 IP 地址和地理位置信息之间的对应关系；

第一确定子模块，用于将所述地理位置信息确定为所述目标用户在所述虚拟几何空间中的用户坐标。

10、根据权利要求 9 所述的装置，其中，所述确定位置模块包括：

15 获取子模块，用于获取各个用户群中的目标用户在虚拟几何空间中的地理位置信息；

第二确定子模块，用于依据目标用户的地理位置信息，确定各个用户群对应的各个端服务器位置。

20 11、根据权利要求 8 所述的装置，其中，所述虚拟几何空间为网络空间，所述映射模块，包括：

映射子模块，用于采用网络空间构建 GNP 算法将所述目标用户的 IP 地址分别映射为网络虚拟坐标；

第三确定子模块，用于将所述网络虚拟坐标确定为所述目标用户在所述虚拟几何空间中的用户坐标。

25 12、根据权利要求 11 所述的装置，其中，所述确定位置模块，包括：

第四确定子模块，用于确定各个用户群的各个中心用户，以及各个中心用户对应的各个邻近用户集，所述邻近用户集包括：一个或多个与所述中心用户的网络距离小于预设距离阈值的用户；

第五确定子模块，用于确定各个邻近用户集中各个邻近用户的地理位置；

第六确定子模块，用于将每个邻近用户集中出现频率最多的地理位置，确定为对应的中心用户的地理位置；

5 判断子模块，用于根据预先设置的部署判断原则，判断所述各个中心用户的地理位置是否适合部署端服务器；

第七确定子模块，用于在所述判断子模块的结果为是的情况下，将适合部署端服务器的地理位置分别确定为对应的用户群的候选端服务器位置。

10 13、根据权利要求 11 所述的装置，其中，所述聚类模块，具体用于：
根据所述虚拟几何空间中目标用户之间的网络距离与实际距离的一致性，和，最远用户优先原则，对所述目标用户进行聚类，以得到多个用户群；

15 其中，所述网络距离与实际距离的一致性为：任意两个目标用户之间的网络距离大于其他两个目标用户之间的网络距离时，该任意两个目标用户的实际距离也大于其他两个目标用户之间的实际距离；所述最远用户优先原则为：在任一个用户群中，网络距离最远的两个目标用户之间的网络距离小于预先设置的距离阈值。

14、根据权利要求 8 所述的装置，其中，还包括：

20 确定容量模块，用于采用预先建立的统一的整数线性规划 ILP 模型分别确定各个端服务器的容量。

15、一种端服务器的部署装置，包括：

存储器，和

处理器，

25 其中所述存储器存储有程序指令，所述程序指令当由所述处理器执行时使所述装置执行下述步骤：

获取待部署端服务器的目标用户的 IP 地址；

将所述目标用户的 IP 地址分别映射到虚拟几何空间中；

根据所述目标用户在所述虚拟几何空间中的用户坐标，对所述目标用户进行聚类，以得到多个用户群；

依据各个用户群中的各个目标用户的用户坐标，确定所述各个用户群对应的各个端服务器位置。

5 16、根据权利要求 15 所述的装置，其中所述程序指令当由所述处理器执行时还使所述装置执行权利要求 2-7 中任一项所述的步骤。

17、根据权利要求 15 所述的装置，其中所述装置包括个人计算机、服务器计算机、手持设备或便携式设备、平板型设备或多处理器装置。

10 18、一种计算机可读存储介质，所述存储介质包含程序指令，所述程序指令当由处理器执行时，使处理器执行权利要求 1-7 中任一项所述的方法。

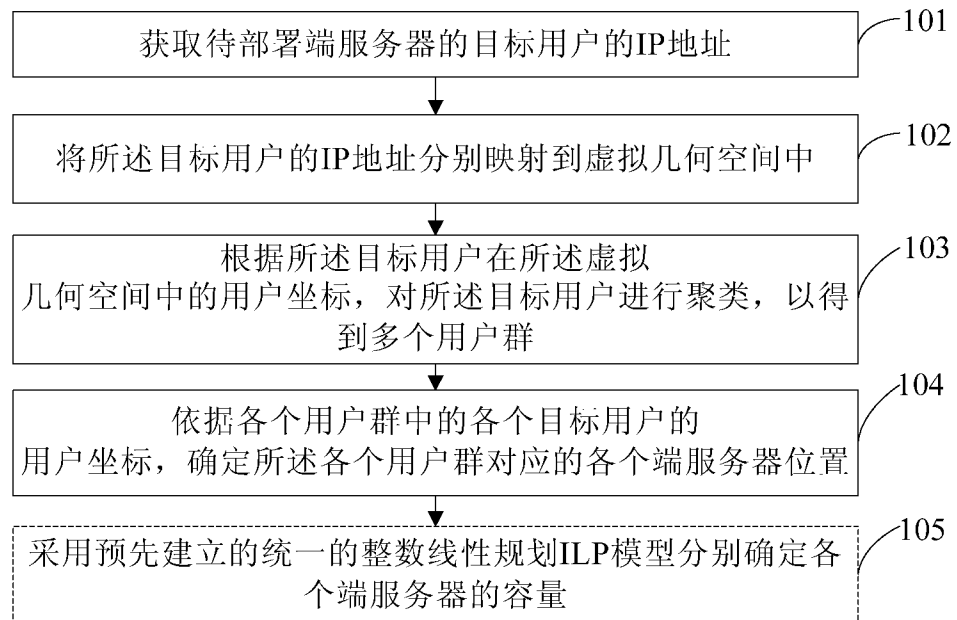


图 1

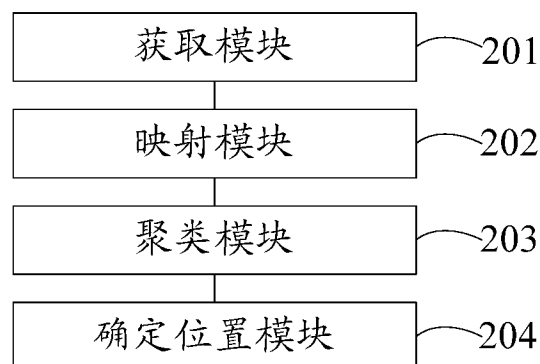


图 2

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2015/087999

A. CLASSIFICATION OF SUBJECT MATTER

H04L 12/24 (2006.01) i; H04L 29/12 (2006.01) i
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04L 12/-; H04L 29/-

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CPRSABS; CNABS; VEN; CNTXT; CNKI: address, location, server, distribut+, user, IP

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2014080304 A2 (ERICSSON TELEFON AB L M) 30 May 2014 (30.05.2014) description, page 13, line 20 to page 15, line 25	1-4, 7-11, 14-18
PX	CN 104935462 A (UNIV QINGHUA) 23 September 2015 (23.09.2015) description, paragraph [0028] to [0067]	1-18
A	US 7366108 B2 (ERICSSON TELEFON AB L M) 29 April 2008 (29.04.2008) the whole document	1-18
A	US 2008180446 A1 (GALLIVAN DAN) 31 July 2008 (31.07.2008) the whole document	1-18

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
“A” document defining the general state of the art which is not considered to be of particular relevance	“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
“E” earlier application or patent but published on or after the international filing date	“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	“&” document member of the same patent family
“O” document referring to an oral disclosure, use, exhibition or other means	
“P” document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 06 February 2016	Date of mailing of the international search report 19 February 2016
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer MA, Yingying Telephone No. (86-10) 62411362

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2015/087999

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
WO 2014080304 A2	30 May 2014	US 2014143407 A1	22 May 2014
		WO 2014080304 A3	24 July 2014
		US 9083757 B2	14 July 2015
		EP 2923459 A2	30 September 2015
		CN 104995870 A	21 October 2015
		IN 201502941 P1	18 September 2015
CN 104935462 A	23 September 2015	None	
US 7366108 B2	29 April 2008	CN 1480003 A	03 March 2004
		GB 0024679 D0	22 November 2000
		US 2004017783 A1	29 January 2004
		GB 2367970 A	17 April 2002
		WO 0235871 A3	10 October 2002
		GB 2367970 B	21 January 2004
		WO 0235871 A2	02 May 2002
		EP 1325650 A2	09 July 2003
		CN 1254994 C	03 May 2006
		AU 1398902 A	06 May 2002
		US 7948491 B2	24 May 2011
		US 8610719 B2	17 December 2013
US 2008180446 A1	31 July 2008	US 7196705 B2	27 March 2007
		US 2005179690 A1	18 August 2005
		US 2011221774 A1	15 September 2011
		US 6888548 B1	03 May 2005
		US 2014176556 A1	26 June 2014
		US 2007182742 A1	09 August 2007
		US 7352371 B2	01 April 2008

国际检索报告

国际申请号

PCT/CN2015/087999

A. 主题的分类

H04L 12/24(2006.01)i; H04L 29/12(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

H04L12/-; H04L29/-

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CPRSABS;CNABS;VEN;CNTXT;CNKI:地址, 位置, 服务器, 部署, 用户, 分布, 放置, IP, location, address, server, user, distribut+

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	WO 2014080304 A2 (ERICSSON TELEFON AB L M) 2014年 5月 30日 (2014 - 05 - 30) 说明书第13页第20行至第15页第25行	1-4, 7-11, 14-18
PX	CN 104935462 A (清华大学) 2015年 9月 23日 (2015 - 09 - 23) 说明书第[0028]段至第[0067]段	1-18
A	US 7366108 B2 (ERICSSON TELEFON AB L M) 2008年 4月 29日 (2008 - 04 - 29) 全文	1-18
A	US 2008180446 A1 (GALLIVAN DAN) 2008年 7月 31日 (2008 - 07 - 31) 全文	1-18

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2016年 2月 6日

国际检索报告邮寄日期

2016年 2月 19日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

授权官员

马莹莹

电话号码 (86-10)62411362

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2015/087999

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
WO	2014080304	A2	2014年 5月 30日	US	2014143407	A1	2014年 5月 22日
				WO	2014080304	A3	2014年 7月 24日
				US	9083757	B2	2015年 7月 14日
				EP	2923459	A2	2015年 9月 30日
				CN	104995870	A	2015年 10月 21日
				IN	201502941	P1	2015年 9月 18日
				无			
CN	104935462	A	2015年 9月 23日				
US	7366108	B2	2008年 4月 29日	CN	1480003	A	2004年 3月 3日
				GB	0024679	D0	2000年 11月 22日
				US	2004017783	A1	2004年 1月 29日
				GB	2367970	A	2002年 4月 17日
				WO	0235871	A3	2002年 10月 10日
				GB	2367970	B	2004年 1月 21日
				WO	0235871	A2	2002年 5月 2日
				EP	1325650	A2	2003年 7月 9日
				CN	1254994	C	2006年 5月 3日
				AU	1398902	A	2002年 5月 6日
US	2008180446	A1	2008年 7月 31日	US	7948491	B2	2011年 5月 24日
				US	8610719	B2	2013年 12月 17日
				US	7196705	B2	2007年 3月 27日
				US	2005179690	A1	2005年 8月 18日
				US	2011221774	A1	2011年 9月 15日
				US	6888548	B1	2005年 5月 3日
				US	2014176556	A1	2014年 6月 26日
				US	2007182742	A1	2007年 8月 9日
				US	7352371	B2	2008年 4月 1日

表 PCT/ISA/210 (同族专利附件) (2009年7月)