



(12)发明专利申请

(10)申请公布号 CN 107005597 A

(43)申请公布日 2017.08.01

(21)申请号 201580067641.1

(22)申请日 2015.10.13

(30)优先权数据

62/063,022 2014.10.13 US

(85)PCT国际申请进入国家阶段日

2017.06.12

(86)PCT国际申请的申请数据

PCT/US2015/055393 2015.10.13

(87)PCT国际申请的公布数据

W02016/061143 EN 2016.04.21

(71)申请人 七网络有限责任公司

地址 美国德克萨斯州马歇尔东区大道2660号

(72)发明人 阿里·贝克汉姆

(74)专利代理机构 杭州千克知识产权代理有限公司 33246

代理人 裴金华

(51)Int.Cl.

H04L 29/08(2006.01)

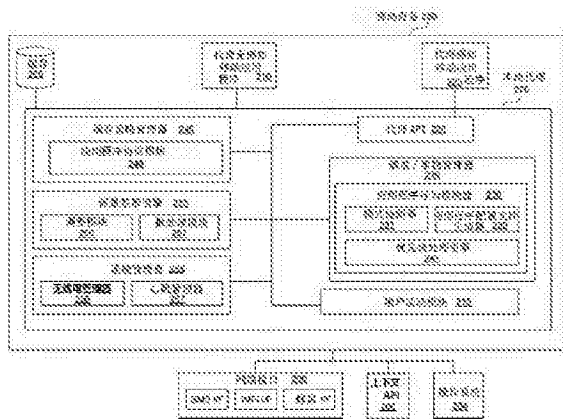
权利要求书3页 说明书60页 附图25页

(54)发明名称

基于用户特征在移动设备缓存的无线流量管理系统

(57)摘要

本发明提供一种通过缓存保存网络带宽的方法。该方法包括在移动设备接收关于用户的信息或特征;确定缓存信息,基于关于用户的信息或特征;在移动设备缓存所述确定的信息。



1. 一种通过缓存保存网络带宽的方法,包括:
在移动设备接收关于用户的信息或特征;
确定缓存信息,基于关于用户的信息或特征;
在移动设备缓存所述确定的信息。
2. 根据权利要求1所述方法,还包括在移动设备预缓存所述确定的信息。
3. 根据权利要求1所述方法,其中,关于用户的信息或特征包括至少下述之一:地理位置、使用网络、操作在移动设备上的应用程序、合并使用模式、在一用户的移动设备和一不同用户的移动设备之间的相似使用模式、用户人口统计资料、时间。
4. 根据权利要求1所述方法,其中,信息或特征在网络边缘由缓存服务器收集。
5. 根据权利要求1所述方法,其中,所述信息或特征被数据分析服务器收集。
6. 根据权利要求1所述方法,还包括延伸缓存周期超过基于信息或特征的缓存周期。
7. 根据权利要求1所述方法,其中,所述缓存遵循一个或多个规则。
8. 根据权利要求7所述方法,其中,一个或多个规则由用户控制。
9. 根据权利要求8所述方法,其中,一个或多个规则包括至少下述之一:选择缓存的应用程序或场所,确定当日加速时间,以及缓存用户代表的内容。
10. 根据权利要求7所述方法,其中,一个或多个规则由于移动设备相关联的移动网络操作器确定。
11. 根据权利要求10所述方法,其中,一个或多个规则包括缓存用户内容,通过调整给用户的内容价格。
12. 根据权利要求1所述方法,还包括根据压缩算法压缩信息。
13. 根据权利要求12所述方法,其中,所述压缩算法使用从网络部件接收到的信息,除了来自移动设备的信息。
14. 根据权利要求13所述方法,其中,网络部件除了移动设备包括至少下述之一:内容输送网络(CDN)、移动网络操作器、本地WIFI操作器、以及事业网络操作器。
15. 根据权利要求12所述方法,其中,所述信息在代理服务器被压缩。
16. 根据权利要求12所述方法,还包括接收输入,通过移动设备用于控制来自用户的压缩算法。
17. 根据权利要求16所述方法,其中,用于控制压缩算法的输入被代理服务器接收。
18. 根据权利要求1所述方法,还包括收集网络事务数据。
19. 根据权利要求18所述方法,其中,所述数据包括至少下述之一:经验指标质量、吞吐量、传输控制协议(TCP)连接延迟、以及超文本传输协议(HTTP)响应代码。
20. 根据权利要求19所述的方法,还包括使用数据来确定到指定内容源的特定路径的网络性能。
21. 一种通过缓存保存网络带宽的移动设备,包括:
连接管理器,用于在移动设备接收关于用户的信息或特征;
请求/事务管理器,用于基于关于用户的信息或特征确定缓存信息;以及
缓存策略管理器,用于在移动设备缓存上述确定信息。
22. 根据权利要求21所述的移动设备,其中,移动设备还被配置来与缓存所述确定信息。

23. 根据权利要求21所述的移动设备,其中,关于用户的信息或特征包括至少下述之一:地理位置、使用网络、操作在移动设备上的应用程序、合并使用模式、在一用户的移动设备和一不同用户的移动设备之间的相似使用模式、用户人口统计资料、时间。

24. 根据权利要求21所述的移动设备,其中,所述连接管理器被配置来接收在网络边缘由缓存服务器收集的信息或特征。

25. 根据权利要求21所述的移动设备,其中,所述连接管理器被配置来接收由数据分析服务器收集的信息或特征。

26. 根据权利要求21所述的移动设备,其中,缓存策略管理器被配置来延伸缓存周期超过基于信息或特征缓存的周期。

27. 根据权利要求21所述的移动设备,其中,所述缓存遵循一条或多条规则。

28. 根据权利要求27所述的移动设备,其中,一个或多个规则被用户控制。

29. 根据权利要求28所述的移动设备,其中,一个或多个规则包括至少下述之一:选择缓存的应用程序或场所,确定当日加速时间,以及缓存用户代表的内容。

30. 根据权利要求27所述的移动设备,其中,一个或多个规则由与移动设备相关联的移动网络操作器确定。

31. 根据权利要求30所述的移动设备,其中,一个或多个规则包括缓存用户内容,通过调整给用户的内容价格。

32. 根据权利要求21所述的移动设备,其中,所述移动设备还被配置来根据压缩算法压缩信息。

33. 根据权利要求32所述的移动设备,其中,所述压缩算法使用从网络部件接收到的信息,除了来自移动设备的信息。

34. 根据权利要求33所述的移动设备,其中,网络部件除了移动设备包括至少下述之一:内容输送网络(CDN)、移动网络操作器、本地WIFI操作器、以及事业网络操作器。

35. 根据权利要求32所述的移动设备,其中,所述移动设备还被配置来接收在代理服务器压缩的信息。

36. 根据权利要求32所述的移动设备,其中,所述移动设备还被配置接收输入,用于控制来自用户的压缩算法。

37. 根据权利要求36所述的移动设备,其中,用于控制压缩算法的输入被代理服务器接收。

38. 根据权利要求21所述的移动设备,其中,所述移动设备还被配置用来收集网络事务数据。

39. 根据权利要求38所述的移动设备,其中,所述数据包括至少下述之一:经验指标质量、吞吐量、传输控制协议(TCP)连接延迟、以及超文本传输协议(HTTP)响应代码。

40. 根据权利要求39所述的移动设备,还包括使用数据来确定到指定内容源的特定路径的网络性能。

41. 一种通过缓存保存网络带宽的计算机程序产品,所述计算机程序产品包括:

计算机可读介质,使用嵌入其内的计算机可读编程代码,计算机可读编程代码包括被配置于如下的计算机可读编程代码:

在移动设备接收关于用户的信息或特征;

确定缓存信息,基于关于用户的信息或特征;
在移动设备缓存所述确定的信息。

基于用户特征在移动设备缓存的无线流量管理系统

[0001] 相关申请的交叉引用

本申请请求美国临时申请号为62/063,022,名称为“基于用户特征在移动设备缓存的无线流量管理系统”,(代理人案号为455/195 PROV),申请提交日为2014年10月13日的美国临时专利的优先权及权益;本申请参考引用上述所申请的内容。

背景技术

[0002] 应用程序正在改变如运营商和手机制造业的产业,并寻求与应用程序和内容提供商合作而使他们的产品和服务不同。对于网络,这仅意味着更多发送信号和宽带的挑战,是当今未解决的问题中最重要的。4G/LTE网络扩展被期望能帮助增加的宽带,但它可能导致用户使用更多能抵消这个益处的数据。对于发送信号,4G/LTE对于诸多原因包括平面网络设计来说将不是困境,事实上是仅在4G网络上的设备是智能手机。

[0003] 无线运营商/提供商对LTE的益处很兴奋。这些包括可用光谱的更好利用、更快数据率、较低每千兆字节传输成本以及更简单的网络结构。LTE打开了惊奇的新应用程序和较丰富用户体验的大门,但是否解决网络阻塞问题或使得问题更糟糕?那发送信号呢?虽然LTE去除了网络中的一些瓶颈,但最直接的益处可能因其他挑战而被遮掩,如遍布全球的无线网络的信令风暴。

[0004] LTE通常伴随有简单的IP网络,平面化所述无线网络结构,并消除在无线网络层次结构中传统等效于无线网络控制器(RNC)的节点。这依次增加了在其他部件上的信令负载,这将不得不相应地变大。其次,今天的3G网络具有智能手机和功能手机的混合。然而LTE,将有接近100%智能手机,并因此,所述信令负载因为总是在智能手机应用程序的实质上将马上变高。进一步,运营商将采用移动到LTE,并因此产生更多数据流量。运营商/提供商将需要重新平衡网络资源,该资源被传统平衡来携带声音和信息发送且仅有小部分数据。据报道,信令负载增加伴随着LTE,但是总体上更多地与流量、应用程序和服务的本质相关,由LTE启用而不是LTE技术本身。

[0005] 随着大规模迁移从有线世界至工业和网络第一线的一切移动,流量优化继续成为保持网络性能和用户满意度高的重要部分。

附图说明

[0006] 附图1A示出了一个系统的示例图,在这个系统中,主机服务器有利于流量管理、内容缓存,和/或在移动设备(如无线设备)之间进行资源保护,还包含为了进行资源保护的应用程序服务器或内容提供商,或者其他服务器,例如无线网络(或宽带网络)中的广告服务器、促销内容服务器,或者电子优惠券服务器。所述系统能进一步优化缓存管理使用HTTP报头。

[0007] 附图1B示出了一个示例图,在该图中包含一个代理和缓存系统,该系统分布在主机和设备之间,该系统能促进各个设备之间的网络流量管理,还包含为了进行资源保护和内容缓存的应用程序服务器或内容提供商,或者其他服务器,例如无线网络(或宽带网络)

中的广告服务器、促销内容服务器,或者电子优惠券服务器。所述代理系统分布在所述主机服务器之间,且所述设备能进一步优化缓存管理使用HTTP报头。

[0008] 附图2A描绘了一个框图,说明在分布式代理和缓存系统中客户端组件的一个实施例,该种分布式代理和缓存系统驻留在一个移动设备(如无线设备)中,该移动设备可以对无线网络(或宽带网络)基于资源保护而进行数据传输管理,内容缓存,和/或数据管理。客户端代理(或本地代理)可以进一步分类移动流量和/或实施传送策略,基于应用程序的行为,内容的优先级、用户活动,和/或用户预期

附图2B描绘了附图2A中显示的缓存系统中的部件的一个进一步的实施例,能够缓存,也能够为移动应用程序的行为和/或网络条件来调整适应缓存策略,图中也阐述了部件能够检测长轮询请求和管理缓存的功能。

[0009] 附图2C描绘了应用程序行为检测器和图2A中显示的缓存系统缓存策略管理器的额外部件,进一步能够检测缓存失败并执行由标识符处理的内容缓存,这些标识符用来处理缓存。

[0010] 附图2D描绘了图2A中显示的本地缓存中的额外部件,进一步能够执行移动流量分类和基于应用程序行为和/或用户活动的策略实施。

[0011] 附图3A描绘了在分布式代理和缓存系统中的服务器端部件,该种分布式代理和缓存系统能基于保护资源,内容缓存,和/或数据流量管理而在无线网络(或宽带网络)中对数据流量进行管理。服务器端代理(或代理服务器)能进一步分类移动数据流量和/或实施传送策略,基于应用程序的行为,内容的优先级、用户活动,和/或用户的期望。

[0012] 附图3B描绘了在附图3A中显示的缓存系统中的缓存策略管理器部件,其能够缓存和适应缓存策略基于移动应用程序的行为和/或网络条件。图中也阐述了部件能够检测长轮询请求和管理缓存的功能。

[0013] 附图3C描绘了在附图3A中显示的代理系统部件的另一个实施例。这是进一步能够管理和检测缓存机制和监控内容源。

[0014] 附图3D描绘了在附图3A中显示代理服务器的额外部件,这是进一步能够在履行实施移动流量分类和基于应用行为和/或数据传输的优先级的情况下的策略实施。

[0015] 附图4A描绘了在分布式代理和缓存系统中客户端部件的另一实施例,该系统进一步包括响应报头管理器。

[0016] 附图4B描绘了在图4A中显示的所述响应报头管理器的额外部件。

[0017] 附图5A描绘了在分布式代理和缓存系统中服务器端部件的实施例,该系统进一步包括响应报头管理器。

[0018] 附图5B描绘了在图5A中显示的所述响应报头管理器的额外部件。

[0019] 附图6A描绘了一流程图,展示了在移动设备(例如任何无线设备)和远程代理之间的分布式内容缓存,以及内容缓存分布式管理的示例过程。

[0020] 附图6B描绘了一时序图,展示了来自于移动设备(如任何无线设备)的数据请求至一个无线网络(或宽带网络)中的应用程序服务器/内容提供商,是如何可以被一个分布式代理系统在某种方式下协调,所述方式为通过使用所述分布式代理系统实现的内容缓存和监管来保存网络和电池资源。

[0021] 附图7描绘了一份表格,这份表格记录了不同的流量或应用程序分类类型,这些在

网络接入执行过程中和内容传输策略中会被使用到。

[0022] 附图8描绘了一份表格,这份表格记录了不同内容分类类型,这些在网络接入执行过程中和内容传输策略中会被使用到。

[0023] 附图9描绘了一交互图,展示了轮询如何通过无线网络(或宽带网络)获取从移动设备(例如任何无线设备)至应用程序服务器/内容提供商的数据请求,其可以被缓存在本地代理上并且由分布式缓存系统管理。

[0024] 附图10描绘了一交互图,展示了如何通过无线网络(或宽带网络)轮询来自应用程序服务器/内容提供商的内容,其可以被检测并本地缓存,应用程序服务器/内容提供商采用标识符中的缓存击败机制(例如,旨在击败缓存的标识符)。

[0025] 附图11描绘了一流程图,展示了收集关于请求的信息以及相关的识别可缓存性的响应并且缓存该响应的示例过程。

[0026] 附图12描绘了一流程图,展示了判断请求的响应是否被缓存的决定流程的示例过程。

[0027] 附图13描绘了一流程图,展示了基本请求周期性和/或响应重复性判断潜在的可缓存性的示例过程。

[0028] 附图14描绘了一流程图,展示了动态调节给定请求或客户端的缓存参数的示例过程。

[0029] 附图15描绘了一流程图,展示了当对网络接入和内容传输策略实施的期望和用户活动考虑在内时,应用程序和/或流量(数据)分类的示例过程。

[0030] 附图16A描绘了一流程图,展示了流量处理的示例过程,所述流量因应用程序/流量分类,将被至少短暂抑制。

[0031] 图16B描绘了一流程图,展示了网络结构选择的示例过程,用来发送流量,基于应用程序和/或流量(数据)分类。

[0032] 图16C描绘了一流程图,展示了实现网络接入和内容传输策略的示例过程,其基于应用程序和/或流量(数据)分类。

[0033] 图17描绘了一流程图,展示了基于移动用户活动或用户期望的网络选择的示例过程。

[0034] 图18描绘了一数据时序图,展示了可适用于缓存的周期请求的检测示例。

[0035] 图19描绘了一数据时序图,展示了请求间隔中的变化的检测以及响应中的服务器轮询速率的更新的示例。

[0036] 图20描绘了一数据时序图,展示了服务具有缓存条目的前台请求的示例。

[0037] 图21描绘了一数据时序图,展示了缓存失效的可能效果示例,所述缓存失效发生在过时内容再次服务于请求内容应用程序之后。

[0038] 图22描绘了一数据时序图,展示了为缓存条目考虑生存时间(TTL)组合的缓存管理和响应的示例。

[0039] 图23描绘了一流程图,展示了通过网页内容缓存优化的无线流量减少的示例过程。

[0040] 图24描绘了一机器的图像化表现,以包含指令集的计算机系统为例,执行该指令,使得机器运行本申请所探讨的一种或多种方法。

[0041] 发明详述

以下详细说明和附图具有说明性,而并不解释为限制性。本申请描述了许多具体的细节,已完全公开本发明内容。然而,某些实例中,为避免说明内容的模糊化,一些惯用或常规细节并未描述。本发明公开的参照“一个实施例”或“某一实施例”可以,但不必然指参照同一实施例,并且参照指至少参照其中一个实施例。

[0042] 本发明公开的实施例包括选择性数据压缩来减少移动数据和发送信号流量的系统和方法。

[0043] 有多种因素有助于数据的增殖:最终用户、移动设备、无线设备、移动应用程序和网络。随着移动设备的发展,这样与它们相关的各种元素-可用性、应用程序、用户行为、位置,从而改变网络与设备和应用程序进行交互的方式。

[0044] 公开的技术提供了全面的和端到端的解决方案,其通过利用移动内容具有可定义的或相关的“新鲜度”值的前提能够为运营商和设备制造商处理每个元素以同时支持移动或无线设备中的转移和数据的激增。移动内容的“新鲜度”可以被确定,无论是肯定的,或有些具有在被增强的用户体验范围内的容限的试探,或不受负面影响地,或受到负面影响但不被用户察觉地或在可容忍的阈值水平内的。

[0045] 公开的创新通过监测、分析、并应用规则(其可能是试探性地确定)应用程序(例如,移动应用程序)和伙伴(相应的服务器或其他客户端)之间的交易(请求/响应)显而易见地确定了这样的“新鲜度”。此外,该技术还能够有效地缓存内容,其可能被它的始发/主机服务器标记为“不可缓存的”和标识一些“新鲜度”值,其可以稍后被用于执行应用程序特定的缓存。在一般情况下,“新鲜度”值有一个近似的最小值,其通常使用应用程序和对应的服务器/主机之间的更新间隔(例如,被发送请求时间间隔)进行确定。

[0046] 本发明所公开技术的一实施例包含,通过设备和应用程序活动的完整视图,优化有线和无线网络和设备各方面连接的系统,该活动包含:设备所需加载或当前应用程序,控制访问类型(推送vs.拉式或混合式)、访问位置、访问单一区域的用户集中度、访问时间、用户与应用程序、内容或设备互动的频率,并利用如上信息定型合作型客户/服务器的流量或在缺乏合作型客户时,同时定型各移动设备的流量。因为公开的服务器未与特定的网络提供商进行绑定,服务器对所有服务提供商的网络性能具有可视性。服务器进行的优化能适用所有设备,不论是运营商还是服务提供商,在漫游时,能够提升用户体验、管理网络利用率。如今带宽被视为无线网络的一大难题,对增加额外带宽,以解决访问需要,已进行了大量的研究——许多提高性能的解决方案和新一代标准,例如:那些通常被称为3.5G、LTE、4G和WiMAX,关注于提供加大的带宽。虽然由标准进行部分地处理,关键问题仍然是信令通道的带宽比数据通道更不足并且标准不能很好的解决电池寿命。

[0047] 本发明所公开技术的实施例包含,例如,调节多个应用程序的请求,尽量减少需要的数个轮询请求;利用特定的内容类型,确定如何代理/管理连接/内容;以及应用与设备相联系的特定试探式、用户行为模式(用户与设备/应用程序互动的频率)和/或网络参数。

[0048] 现有技术的实施例还包括:通过各控件或RSS阅读器执行,移动反复的HTTP轮询至远程网络节点(例如,网络操作中心(NOC)),从而大大降低了设备消耗的电池/功率、无线信道信令和使用的带宽。此外,程序卸载能透明地执行,因此无需改变现有的应用程序。

[0049] 一些实施例中,利用移动设备(例如,任何无线设备)的本地代理,自动检测符合特

定规则(如时间间隔为15分钟)相同内容的反复请求(RSS资源,控件数据集),能实现这一功能。本地代理将轮询指派给服务器时,能自动缓存移动设备上内容(例如,代理服务器作为通信网络元件之一)。若内容变化,服务器则通知移动/客户端代理;若内容未变化(变化不充分,或改变方式或数额不明显),移动代理将其缓存的最新内容提供给用户(根本无需使用无线)。利用此方式,如果请求的内容被监控,且未标记新/更改,移动或无线设备(例如:移动电话,智能手机、M2M模块/调制解调器或其它任何无线设备等)无需打开(例如,因此打开无线电)或使用数据连接。

[0050] 自动添加被监控的内容源/应用程序服务器(如包括网址/内容)的逻辑还能检查各种因素:内容一致的频率,提出相同请求的频率(是否存在固定的时间间隔/模式?),请求数据的应用程序等。本地代理和/或服务器也可实施和执行类似规则,确定是使用缓存还是向原始资源请求数据。

[0051] 例如,请求在未预定/非预期时间提出(用户启动的检查),或连续几次在缓存已提出响应后,亦或应用程序处于后台运行或前台的频繁互动模式。随着越来越多移动应用程序或无线启用应用程序的特征以其在网络中的可用资源为基础,这也显得愈发重要。此外,所公开的技术可以消除不必要的网络对话,使试图优化无线频谱应用的运营商从中受益。

[0052] 流量分类和策略

一些实施例中,公开的代理系统是能够为选择流量(数据、内容、信息、更新等)建立策略以缓存和/或整形。此外,通过结合来自观察作出网络请求的应用程序的信息,获得来自应用程序中的明确信息,或获知应用程序到达的网络目的地,所公开的技术可确定或推断所发送的流量属于哪个类别。

[0053] 例如,在一个实施例中,移动或无线流量可以分为:(a1)交互式流量或(a2)后台流量。区别在于(a1)中的用户正在积极等待响应,而(a2)的用户并不期待响应。这种分类可以配合或代替第二类型的流量分类使用:(b1)即时;(b2)低优先级;(b3)如果发出请求的应用程序在前台且活跃的,则为即时。

[0054] 例如,(b1)类别中的新的更新、消息或电子邮件为立即交付,但它仍然是(a2)后台流量-用户并不积极地等待。当它们来自活跃的聊天会话之外时,类似的分类适用于即时消息。在活跃的聊天会话期间,用户期待更快速的响应。这种用户的期望当在进行流量分类和策略执行中优化网络的使用和设备资源时被确定或推断并被考虑到。

[0055] 所述分类方案的应用程序的一些例子包括以下内容:(a1)交互流量可以被归类为(b1)即时-但(a2)后台流量也可以是(b2)或(b3)。低优先级传输例子是电子邮件或消息维护事件,例如,删除电子邮件或其他消息或将电子邮件在邮件或应用程序服务器中标记为已读。这样的传输通常发生在早于(a)时序器超过一个超时值(例如,2分钟)以及(b)发送的数据用于其他用途。

[0056] (b3)的例子是IM状态更新、股票代码更新、气象更新、状态更新、新闻提要。当应用程序的用户界面在前台和/或是活跃的(例如,由设备/电话的背光源的点亮或从其他传感器的状态所确定或推断出的表示)更新可以被认为立即通知服务器有东西要推送至设备。当应用程序没有在前台或不活跃时,这样的更新可被抑制直到应用程序到前台以及是活跃的。

[0057] 一些实施例,网络可以被选择或同时优化(a1)互动流量和(a2)后台流量。

[0058] 一些实施例中,由于无线设备或移动设备代理(单独或连同服务器代理)是能够分类为(例如)(a1)交互式流量或(a2)后台流量,其可以应用不同的策略至不同类型的流量。这意味着,它可以在内部不同地操作(a1)和(a2)流量(例如,通过允许交互式流量经过全部或部分的网络,并应用严格的流量控制至后台流量;或设备端只允许激活无线电的请求,如果其已接收到来自服务器的主机的内容已经更新,等等的信息)。

[0059] 当请求需要通过无线网络访问,所公开的技术可以请求无线层适用不同网络配置至不同的流量。根据不同的流量和网络,这可通过不同的手段实现:

(1) 为(a1)使用3G/4G以及为(a2)使用2G/2.5G;

(2) 明确为数据集FACH(前向接入信道)与DCH(专用信道)的不同数据集的使用指定网络配置,或以其他方式为后台流量要求较低/更多的网络高效数据率;或者

(3) 为不同的数据集利用不同的网络接入点(接入点将被配置为使用类似上述(1)和(2)的不同的网络资源)。

[0060] 此外,3GPP快速休眠电话的改进以使得应用程序、操作系统或移动设备将意识到流量类型在未来更有效。具有流量分类的知识并能适当地利用快速休眠的公开系统的实施例可解决在快速休眠中确定的问题。通过这种方式,移动或宽带网络并不需要配置一个妥协的配置,该配置产生电池消耗和网络信令的不利影响。

[0061] 轮询时间表

检测(或确定)轮询时间表允许代理服务器(分布式缓存系统的服务器端)尽可能使其轮询接近应用程序轮询。许多应用程序使用预定的时间间隔轮询(例如,每4小时或每30秒,在其他时间间隔)。客户端代理可以基于时间测量检测自动轮询并且为应用程序创建一个自动轮询配置文件。作为一个例子,本地代理尝试检测请求之间的时间间隔和在2、3、4或更多的轮询之后的间隔,如果所有时间间隔各自都在1秒(或相对接近的另一测量)内,则确定一自动的速率。如果没有的话,客户端可从更大数量的轮询事件(例如,10-12个轮询)中收集数据并且应用统计分析以确定、计算或估计所使用的平均时间间隔的值。轮询配置文件传递到其被使用的服务器。如果它是一个频繁地手动请求,本地代理可以为这个应用程序用默认的时间间隔替代它,这个应用程序来自非关键性应用程序的配置文件。

[0062] 在一些实施方案中,本地代理(例如,设备端代理)可能保持监控应用程序/客户端轮询并更新轮询间隔。如果其从当前值变化了30%以上(或另一个预定/动态/条件值),则它被传递到代理服务器(例如,服务器端代理)。这种方法可以称为“失去兴趣”的情况。在某些情况下,本地代理只能识别时间表之外作出的请求,认为他们是“手动的”,并相应地处理它们。

[0063] 应用程序类别/缓存模式

在一些实施例中,应用程序可以被分成三组或缓存模式。取决于一个或多个条件,每个移动客户端/应用程序可以被分类为以其中的一种模式进行处理,或使用多个模式进行处理。

[0064] A) 完全缓存-本地代理仅在代理服务器告诉本地代理更新时进行更新(例如,通过网络直接发送应用程序请求以由应用程序服务器/内容主机服务)。在这种模式中,本地代理可以忽略手动请求并且代理服务器使用检测到的自动配置文件(例如,体育比分的小程序、Facebook,每10、15、30或更多轮询)轮询应用程序服务器/内容提供商。

[0065] B) 部分缓存-本地代理为自动请求或其他预定的请求使用本地或内部缓存(例如,应用程序自动刷新),但通过一些手工请求(例如,电子邮件下载、EBay或一些Facebook请求);以及

C) 从不缓存(例如,实时股票行情、体育比分/状态,然而,在某些情况下,15分钟延迟引用可以安全地放置在30秒时间表-B乃至A)。

[0066] 实际应用程序或缓存模式分类可基于内容变化率和数据关键性特征被确定。未分类的应用程序默认情况下可以设置为C类。

[0067] 背光灯和活动应用程序

在一些实施例中,本地代理通过检测设备背光灯状态来启动。如果具有同一签名的请求被向代理服务器进行注册,显示屏灯“关闭”的请求可以被允许使用本地缓存,其轮询请求定向至的始发主机服务器/内容服务器。如果显示屏灯是“打开”,进一步的检测可以被作出以确定其是否是后台应用程序或者对于其它指示器,本地缓存条目可以或不可以被用于满足请求。当被识别时,本地条目可以被用于的请求可被相同地处理为显示屏灯关闭的状态。当缓存的数据可安全地用于处理请求时,前台请求可以使用上述应用程序分类以进行评估。

[0068] 图1A为系统示例图,其中主机服务器100促进流量管理、内容缓存和/或在移动设备(例如,无线设备150)和应用程序服务器或内容提供商110或其他服务器如广告服务器120A、促销内容服务器120B、或电子优惠券服务器120C之间的资源保存,在无线网络(或宽带网络)中用于资源保存。该系统还使用HTTP报头优化缓存管理。

[0069] 客户端设备150可以为能够创建连接,包括与其他设备、服务器和/或其他系统,如主机服务器100和/或应用程序服务器/内容提供商110的有线、无线和蜂窝连接,的任意系统和/或设备,和/或设备/系统的任意组合。客户端设备150通常包括一个显示屏和/或其他输出功能,以显示设备150和/或主机服务器100和/或应用程序服务器/内容提供商110所交换的信息和数据。

[0070] 例如,客户端设备150包含移动、手持或便携式、无线设备或非便携式设备和任意,但不限于:台式服务器、台式电脑、计算机集群或便携式设备,包括笔记本电脑,手提电脑,便携电脑,掌上电脑,移动电话,手机,智能手机,PDA,黑莓设备,PalM设备,手持平板电脑(如iPad或其它任何平板电脑),电子书阅读器,电子阅读器,上端安放的显示器,瘦客户端,手持控制台,手持游戏设备或控制台,任何超级手机如iPhone,和/或任何其它便携式、移动、手持设备或固定无线接口如M2M设备,等等。一个实施例中,客户端设备150、主机服务器100和应用程序服务器110通过网络106和/或网络108,耦合在一起。一些实施例中,设备150和主机服务器100直接相连。

[0071] 客户端设备150的输入机制包含触摸屏键盘(包括单点触摸,多点触摸,平面或三维手势感应等)、物理键盘、鼠标、指针、触模板,运动检测器(如包括1轴,2轴,3轴加速度计等)、光传感器、电容传感器、电阻传感器、温度传感器、接近性传感器,压电器件,设备方向探测器(例如:电子罗盘、倾斜传感器、旋转传感器、陀螺仪、加速度计)或上述装置的任意组合。

[0072] 过如上所述的一个或多个输入机制或其他接收或检测信号,表明客户端设备150的用户活动,能用于公开的技术获取客户端设备150的上下文感知。客户端设备150的上下

文感知一般包括,列举但不限于:客户端设备150运行或状态的确认、管理、用户活动/行为/互动感知、检测、传感、跟踪、趋势和/或应用程序类型(如移动应用程序)、行为、活动或运行状态等。

[0073] 本申请所公开的上下文感知还包含网络端上下文数据的知识检测和检测,并且包括网络信息,例如:网络容量、带宽、流量、网络/连接类型,和/或任何其他操作的状态数据。网络端上下文数据可以从网络106和/或网络108(例如:主机服务器和/或设备150)的网络服务提供商处(例如:手机提供商112和/或互联网服务提供商)接收和/或查询。除从客户端150确定的应用程序上下文感知外,应用程序上下文感知也可从各自的应用程序/服务提供商处110接收或获得/查询(通过主机100和/或客户端设备150)。

[0074] 主机服务器100可以使用,例如,客户端设备150、网络106/108、应用程序(如移动应用程序)、应用程序服务器/提供商110或上述任意组合所获取的上下文信息,管理系统流量,满足客户端设备150的数据要求(例如,满足应用程序或其他请求,包括HTTP请求)。一个实施例中,主机服务器100管理流量,满足所作出的数据请求,以响应显名或隐名用户103的请求和/或设备/应用程序的维护任务。管理流量能节约网络消耗,例如蜂窝网络,保证高效的带宽利用率。此外,主机服务器100能在优化性能和用户体验的同时,管理和协调系统流量,秉承节约资源的总思路,优化使用设备端150的资源(例如:包括但不限于:电池消耗电量、使用的无线、使用的处理器/内存)。

[0075] 例如,在节约电池方面,设备150能观察用户活动(例如,通过一个或多个输入机制观察用户的按键、背光状态或其他信号等),并且更改设备150的行为。基于用户活动或行为,设备150还可请求主机服务器100改变行为,降低消耗的网络资源。

[0076] 一个实施例中,利用主机服务器100和客户端设备150之间的分布式系统执行流量管理,以节约资源。分布式系统能包含服务器端100和设备/客户端的代理服务器和缓存组件,如服务器端100的缓存服务器135和客户端150的本地缓存185所示。

[0077] 本申请所公开上下文感知流量管理的功能与技术,节约网络(例如:网络106和/或108)和设备150的资源,设于分布式代理和缓存系统。代理和缓存系统能部分或全部分布,并设于给定的客户端设备150,和/或主机服务器100。分布式代理和缓存系统,将在如图1B所示的示例图作进一步描述。客户端设备150、主机服务器100和本申请中所述的相关组件中的代理和缓存部件执行该功能与技术,进一步详细描述分别参考示例图2-3。

[0078] 一个实施例中,客户端设备150与主机服务器100和/或应用程序服务器110通过网络106,可以为蜂窝网络和/或宽带网络,进行通信。为促进设备150和各应用程序服务器/内容提供商110的总体流量管理,以执行网络(带宽利用率)和设备资源(如电池消耗),主机服务器100可以与应用程序服务器/提供商110通过网络108,包含互联网(例如,宽带网络),进行通信。

[0079] 一般来说,客户端设备150与主机服务器100和/或应用程序服务器110通过网络106和/或108进行通信,网络可以为蜂窝网络、宽带网络、电话网络、开放网络,如互联网、专用网络,如内联网和/或外联网、或其任意组合。例如,互联网可以通过任何已知或方便协议提供文件传输、远程登录、电子邮件、新闻、RSS、云计算服务、即时消息、可视语音邮件、推送邮件、VoIP和其他服务,协议的列举包括,但并不限于:TCP/IP协议、UDP、HTTP、DNS、文件传输协议(FTP)、UPnP、NSF、ISDN、PDH、RS232、SDH、SONET等。

[0080] 网络106和/或108可以是部分或全部运营网络的任意组合,结合后向客户端设备150和主机服务器100提供连接,对于服务的系统和设备,作为一个或多个网络。一个实施例中,开放网络,如互联网、专用网络,宽带网络,如内联网和/或外联网都能实现客户端设备150通信的发送和接收。一个实施例中,安全通信协议,如安全套接层协议层(SSL)或安全传输层协议(TLS),也能实现此类通信。

[0081] 此外,一个或多个网络可以实现通信,包括但不限于:一个或多个WiMax、局域网(LAN)、无线局域网(WLAN)、个人区域网(PAN)、校园区域网(CAN)、城域网(MAN)、广域网(WAN)、无线广域网(WWAN)、或任何宽带网络并进一步由以下技术启用,例如:全球移动通信系统(GSM)、个人通信服务(PCS)、蓝牙、WiFi、固定无线数据、2G、2.5G、3G、4G,高级国际移动通信IMT-Advanced、预-4G、高级长期演进技术(LTEAdvanced)、移动WiMax、WiMax2、WirelessMAN高级网络,增强型数据速率GSM演进技术、通用分组无线服务技术(GPRS)、增强型通用无线分组业务、iBurst、UMTS、HSPDA、HSUPA、HSPA、UMTS-TDD,1xRTT, EV-DO、消息传送协议,例如:TCP/IP、SMS、MMS、可扩展消息与存在协议(XMPP)、实时消息传送协议(RTMP)、即时信息和空间协议(IMPP)、即时消息、USSD、IRC或其他无线数据网络、宽带网络或消息传送协议。

[0082] 图1B为代理和缓存系统示例图,分布在主机服务器100和设备150之间,该设备促进管理设备150与应用程序服务器或内容提供商110,或其他服务器,如广告服务器120A、促销内容服务器120B,或电子优惠券服务器120C之间的网络流量,以节约资源和内容缓存。分布于主机服务器100和设备150间的代理系统还可使用HTTP报头优化缓存管理。

[0083] 分布式代理和缓存系统可包括,例如,服务器端组件:代理服务器125(如远程代理)和服务器缓存135。如图所述,服务器端代理服务器125和缓存135能设在主机服务器100内部。此外,服务器端代理服务器125和缓存135可部分或全部设在主机服务器100外部,通过一个或多个网络106和108进行通信。例如,代理服务器125设在主机服务器外部,而服务器缓存135保持在主机服务器100内部。亦或,代理服务器125设在主机服务器内部,而服务器缓存135在主机服务器100外部。此外,代理服务器125和缓存135都可能部分设在主机服务器100内部,部分设于主机服务器100外部。应用程序服务器/内容提供商110可通过任何服务器,包括第三方服务器或服务/内容提供商,还包括广告、促销内容、公开或电子优惠券服务器或服务。类似地,将广告服务器120A、促销内容服务器120B、和/或电子优惠券服务器120C,按应用程序服务器或内容提供商分离,将以此示例的方式示出。

[0084] 分布式系统也能包括,一个实施例中,客户端组件,包括列举但不限于:本地代理175(如移动设备的移动用户)和/或本地缓存185,如图所示,能设在设备150(如移动设备)内部。

[0085] 此外,客户端代理175和本地缓存185能部分或全部设于设备150外部,并通过一个或多个网络106和108进行通信。例如,将本地代理175置于设备150外部,而本地缓存185保持在设备150内部。亦或,将本地代理175置于设备150内部,而本地缓存185则置于设备150外部。此外,每个代理175和缓存185都可能部分置于主机服务器100内部,部分置于主机服务器100外部。

[0086] 一个实施例中,分布式系统可以包括一个可选的缓存代理服务器199。缓存代理服务器199为一组件,由应用程序服务器/内容提供商110、主机服务器100、或网络服务提供商

112,或以上任意组合,为促进网络流量管理而执行,目的在于节约网络和设备资源。代理服务器199可用于,例如,为设备150提供从一个或多个应用程序服务器/提供商110、主机服务器100,和/或网络服务提供商112的内容缓存。远程代理125也能全部或部分地执行内容缓存,以满足应用程序请求或设备150的其他数据请求。

[0087] 在上下文感知流量管理和优化中,为节约网络资源(如蜂窝或其他无线网络),本地代理175跟踪移动设备150(例如,任何无线设备)的用户活动/行为和/或应用程序行为特征,并通过网络106与主机服务器100的代理服务器125组件进行通信,例如连接元数据。代理服务器125,依次与应用程序服务器/提供商110相耦合,提供内容和数据,以满足设备150提出的请求。

[0088] 此外,本地代理175能够识别和检索移动设备的属性,包含一个或多个电池电量、设备的注册网络、无线状态、移动设备是否在使用(例如,由用户互动)。一些实例中,本地代理175在传输数据给代理服务器前,一定情况下,能够延迟、加快(预取)和/或修改数据,具体情形参见示例图2和图3的相关说明。

[0089] 本地数据库185能包含或耦合至本地代理175,用于本地存储数据的查询,在数据请求被转发到代理服务器125前,回应该数据请求。本地代理175通过检索存储在缓存存储器185的缓存内容,若该内容仍然有效,使用本地的缓存响应,以满足移动设备150某些应用程序的请求。

[0090] 同样地,在本地代理服务器传输数据给内容源(例如:应用程序服务器/内容提供商110)之前,主机服务器100的代理服务器125可也延迟、加快或修改数据。此外,代理服务器125利用设备属性和连接元数据,生成满足移动设备150应用程序请求的规则。代理服务器125能够收集有关应用程序请求的实时流量信息,后续用于优化与移动设备150或其他移动设备的类似连接。

[0091] 一般来说,本地代理175和代理服务器125清楚移动设备上执行的多个应用程序。本地代理175通常清楚移动设备的操作系统或平台,并且可能或可能不明确设备的制造商。一些实例中,本地代理175可以选择自定义部分或全部具体设备。一些实施例中,本地代理175可能被捆绑进无线模型,防火墙,和/或路由器。

[0092] 一个实施例中,主机服务器100在一些实例中,利用短消息服务中心(SMSC)112的存储和转发功能,由网络服务提供商所提供,与设备150进行互动,实现网络流量管理。请注意112也可以使用任何其他类型的替代信道,包括USSD或其他网络控制机制。如图3示例所作的详细描述,主机服务器100将内容或HTTP响应转发给SMSC112,若设备150可用则自动转发,若设备150当前不可用,则后续转发。

[0093] 一般来说,公开的分布式代理和缓存系统能优化网络利用率,例如,通过回应本地缓存185的请求,本地代理175能减少需要通过网络106满足的请求数量。另外,本地代理175和代理服务器125能从通信数据中过滤无关数据。此外,为避免发送单个数据片段的协议消耗,本地代理175和代理服务器125也能累积低优先级数据,将其批量发送。本地代理175和代理服务器125还能将流量压缩或转码,减少通过网络106和/或108发送的数据量。由于网络的使用频率降低,网络流量在各应用程序间能实现同步,网络106和/或108的信令流量则会减少。

[0094] 至于移动设备150的电池寿命,通过回应本地缓存185的应用程序或内容请求,本

地代理175能减少无线模块的上电次数。当无线上电时,本地代理和代理服务器125能相互结合,累积低优先级的数据批量发送,以减少发送次数和/总时间。通过同时执行所有连接的批量数据传输,本地代理175能同步网络的使用。

[0095] 图2A为一框图,展示了设于移动设备250(例如,无线设备)的分布式代理系统客户端组件,管理无线网络(或宽带网络)的流量,以节约资源、内容缓存和/或流量管理。客户端代理(或本地代理275)还能分类移动流量和/或执行传送策略,基于应用程序行为、内容优先级、用户活动和/或用户期望。

[0096] 设备250,可以为便携式或移动设备(例如,任何无线设备),如手提电话,通常包括:例如,网络接口208、操作系统204、上下文API206,和感知代理210或未感知代理的移动应用程序220。需注意的是设备250指如图2具体描述的移动设备,但不限于于此,该设备250可能是能够通过网络,包括有线或无线网络(例如:WIFI,蜂窝网络,蓝牙等),接收、传输信号的任意便携式/移动或非便携式设备,以满足数据请求。

[0097] 网络接口208可以是网络模块,通过主机和外部实体支持的任何已知和/或方便协议,使设备250调解外部主机服务器250网络中实体的数据。网络接口208能包括:一个或多个网络适配卡、无线网络接口卡(例如:短信接口、WiFi接口、移动通信各代标准的接口,包括但不限于2G、3G、3.5G、4G、LTE等)、蓝牙、或者连接是否经由路由器、接入点、无线路由器、交换机、多层交换机、协议转换器、网关、网桥、桥接路由器、集线器、数字媒体接收器和/或中继器。

[0098] 设备250还能包含分布式代理和缓存系统的客户端组件,其中包括本地代理275(如移动设备的移动客户端)和缓存285。一个实施例中,本地代理275包括用户活动模块215,代理API225,请求/事务管理器235,具有应用程序协议模块248的缓存策略管理器245,流量定型引擎255和/或连接管理器265。流量整形引擎255可以更包含校准模块256和/或批处理模块257,连接管理器265可能会包括无线电控制器266。请求/事务管理器235还能包含应用程序行为检测器236和/或优先级引擎241,其中应用程序行为检测器236还能包含模式检测器237和/或应用程序配置文件生成器239。本地代理275和所述的任一组件均可另外增加或减少组件/模块/引擎。

[0099] 本申请所使用的“模块”、“管理器”、“处理程序”、“检测器”、“接口”、“控制器”、“规范化器”、“生成器”、“验证器”或“引擎”包含一般用途,专用或共享处理器,和通常由处理器执行的固件或软件模块。基于具体实施或其他因素的考量,该模块、管理器、处理器、检测器、接口、控制器、规范化器、验证器或引擎,可集中或依功能分布实施。该模块、管理器、处理程序或引擎包含嵌入在计算机可读(存储)介质,具有一般或特殊用途的硬件,固件或软件,由处理器予以执行。

[0100] 本申请所使用的计算机可读介质或计算机可读存储介质,包含所有法定介质(如根据美国专利法第101条),并明确排除所有非法定的介质,本质上说排除的必要范围,使得包含计算机可读(存储)介质的权利要求能成立。已知的法定计算机可读介质包含硬件(例如:寄存器,随机存储器(RAM)、非易失(NV)存储器,仅举几例),但可能或可能不限于硬件。

[0101] 一个实施例中,用于管理网络流量的部分分布式代理和缓存系统放置在设备250内或与其进行通信,包含本地代理275(移动客户端)和/或缓存285。本地代理275能为设备250提供接口,供用户访问设备的应用程序和服务,包含电子邮件、即时消息、语音信箱、可

视语音邮件、源订阅、互联网、游戏、效能工具或其他应用程序等。

[0102] 本地代理275通常独立于应用程序,能被应用程序(例如,代理感知和未感知应用程序210和220或移动应用程序)用于打开至远程服务器的TCP连接(例如:图1A-1B示例中的服务器100和/或图1B和图3A示例中的代理服务器125/325)。一些实例中,本地代理275包含代理API225,可以选择用于感知代理服务器应用程序220(或移动设备(例如任何无线设备)上的应用程序(移动应用程序))的接口。

[0103] 应用程序210和220通常包含任何用户应用程序,控件,软件,基于HTTP的应用程序,web浏览器,视频或其他多媒体播放器或下载应用程序,视频游戏,社交网络应用程序,电子邮件客户端,RSS管理应用程序,应用商店,文档管理应用程序,生产力促进应用程序等。设备制造商或网络服务提供商能够向应用程序提供设备的操作系统,该操作程序可由用户下载或其他人提供。

[0104] 本地代理275的一个实施例包含,或耦合到上下文API206。上下文API206可能为操作系统204或设备平台的一部分,亦或独立于操作系统204,如图示。操作系统204可包含任意操作系统,但不限于:以前、目前和/或未来版本/版本的Windows Mobile,iOS,Android,Symbian,Palm OS,Brew MP,Java2Micro Edition(J2ME),黑莓等。

[0105] 上下文API206可能为操作系统204的插件或设备250上特定客户端应用程序。上下文API206能够检测用户的指示信号或设备活动,例如检测位置,手势,设备位置的变化,设备背光,按键,点击,触摸屏被激活,鼠标点击或检测其他指针设备。上下文API206可以耦合至设备250的输入设备或传感器来识别这些信号。此信号通常包含接收的输入信息,以响应用户在设备250输入设备/机制的明确输入,和/或收集到的环境信号/上下文线索,在设备250内或毗邻处检测得到(如光,动作,压电等)。

[0106] 一个实施例中,用户活动模块215与上下文API206互动,用于识别、确定、推断、检测、计算、预测和/或预计设备250上用户活动特征。用户活动模块215汇总上下文API206收集的各输入值,生成用户活动特征的概况。这一概况由用户活动模块215利用各临时特征生成。例如,通过提供用户在给定时间所做之事或未做之事,瞬间生成实时的用户活动概况(例如由时间窗口定义最后一分钟或最后30秒等);用户活动概况也能由应用程序或网页定义的“时域”生成,“时域”描述了设备250上进行的特定任务,或在特定时间内用户行为的特征(例如前2小时或前5小时)。

[0107] 此外,用户活动模块215通过描绘用户活动和行为的历史趋势(例如1周,1个月,2个月等),生成上述特征概况。这些历史文件还可用于推断用户行为的趋势,例如:不同时间的访问频数、每周各天访问频率的趋势(周末或工作日)、基于位置数据的用户活动趋势(例如IP地址,GPS或基站协调数据)或位置数据的变化(例如基于用户位置的用户活动,或基于用户在忙碌中或在外旅行的用户活动等),以获取用户活动特征。

[0108] 一个实施例中,用户活动模块215能检测和跟踪设备250上应用程序、文档、文件、窗口、图标和文件夹的用户活动。例如,当一个应用程序或窗口(如Web浏览器或任何其它类型的应用程序)被退出,关闭,最小化,最大化,打开,移至前台,或后台,播放多媒体内容时,用户活动模块215都能检测到。

[0109] 一个实施例中,设备250上的用户活动特征可以用于本地调整设备行为(如移动设备或任何无线设备),优化消耗资源,如电池/功率消耗和更普遍地,包含内存、存储和处理

能源的其他设备资源的消耗。一个实施例中,基于耦合到用户活动模块215的用户行为特征,可调整设备上的无线电使用(如通过连接管理器265的无线电控制器266)。例如,根据设备250上的用户活动特征,无线电控制器266打开或关闭无线电。此外,无线电控制器266能依据用户活动特征,调整无线电的功耗模式(如高功耗模式或低功耗模式)。

[0110] 一个实施例中,设备250的用户活动特征也可用于引发其他与设备250存在通信(如通过蜂窝网络或其他网络)的设备(例如其他计算机、移动设备、无线设备或非便携式设备)或服务器(例如图1A-B和图3A示例的主机服务器100和300),修改其与设备250的通信频率。利用用户活动模块确定的用户行为特征信息,本地代理275能指示远程设备如何校准其通信频率(例如,若用户处于闲置状态,降低通信频率,如数据推送频率,或当新数据、更改数据或某重要性的数据可用时,请求远程设备通知设备250等)。

[0111] 在一个实施例中,用户活动模块215能确定用户行动的特征,作为响应确定用户行为特征,指示用户在闲置一段时间后的活跃,请求远程设备发送因之前降低通信频率的缓冲数据(例如图1A-B和图3A示例的主机服务器100和300)。

[0112] 此外,或本地代理275可另外选择将设备250的用户活动特征传输给远程设备(例如图1A-B和图3A示例的主机服务器100和300),并且远程设备确定改变与设备250的通信频率,以节约网络资源和设备250资源。

[0113] 实施例之一的本地代理275还包含请求/事务管理器235,能检测、识别、截获、处理和管理设备250提出的数据请求,例如通过应用程序210和/或220,和/或直接/间接的用户请求。基于事务特征,请求/事务管理器235能够确定处理给定的请求/事务或一组请求/事务的方式和时间。

[0114] 请求/事务管理器235能对设备250上应用程序和/或用户作出的请求或事务进行优先级排序,例如通过优先级引擎241。该请求/事务管理器235通过使用规则集,确定请求/事务的重要性或优先级,例如,根据事务的时间敏感性、事务内容的时间敏感性、事务的时间紧迫性,事务中传输数据的时间紧迫性,和/或应用程序发出请求的时间紧迫性/重要性。

[0115] 此外,事务特征也取决于该事务是否为用户互动或其他用户在设备上启动的操作(例如用户与应用程序(例如,移动应用程序)的互动)。一般来说,一项时间紧迫性事务包含一项由用户启动的数据传输事务,能被列为优先事务。事务特征还取决于根据所请求的事务中将传输或预计传输的数据量。例如,跟据需传输的数据量,连接管理器265调整无线电模式(例如通过无线电控制器266调整高功耗或低功耗模式)。

[0116] 此外,无线电控制器266/连接管理器265能依据事务的时间紧迫性/敏感性,调整无线电的功耗模式(高或低)。当启动或检测到一项时间紧迫性事务时,无线电控制器266触发使用高功耗无线电模式(例如:由用户启动的数据传输事务,前台运行的应用程序,或符合特定条件的其他事件)。

[0117] 一般来说,优先级通常为默认设置,例如,基于设备平台、设备制造商、操作系统等。优先级也能,或由特定的应用程序另外设置,例如,Facebook应用程序(例如,移动应用程序)为自己设置各项事务的优先级(例如,状态更新的优先级高于添加朋友请求或戳请求,发送消息请求的优先级高于删除消息请求),电子邮件客户端或IM聊天客户端有自己的优先级配置。优先级引擎241可能包含分配优先级的规则集。

[0118] 优先级引擎241还可以跟踪网络提供商对应用程序或事务优先级确定请求/事务

整体优先性状态的限制条件或详细说明。此外,优先级部分或全部,明确或隐含地取决于用户的喜好。用户一般能设置不同层次的优先级,例如为各时域、类型或应用程序设置具体的优先级(以浏览时域、游戏时域和IM聊天时域为例,用户可能设置游戏时域优先于IM聊天,而IM聊天的优先于浏览时域)。用户能设置应用程序的具体优先级(例如用户可设置Facebook的相关事务优先于LinkedIn的相关事务)、事务类型的具体优先级(例如所有应用程序的发送信息请求优先于删除信息请求,所有日程相关的事务具有优先级等),和/或文件夹的具体优先级。

[0119] 优先级引擎241跟踪和解决不同实体优先级设置的冲突。例如,用户具体的手动设置可能优先于设备操作系统设置,网络提供商参数/限制条件(例如,默认设置的网络服务领域,地理区域,具体时间或基于服务/费用类型的设置)可能会限制用户的具体设置和/或应用程序设置的优先级。一些实例中,用户的同步手动请求能优先于一些、大多数或全部的优先级设置,一经提出,该同步请求的执行不受单独分配的优先级或请求操作在所有优先级事务中的排序限制。

[0120] 优先级能以任何已知和/或简便方式在内部被指定并跟踪,包括但不限于:二进制表示形式,多值表示形式或分级表示形式,所有被认为的都在公开技术的范围内。

变化 (设备启动)	优先级	变化 (服务器启动)	优先级
发送邮件	高	接收邮件	高
删除邮件	低	编辑邮件	任意优先级
(未)接收邮件	低		同步(若可能,低)
移动邮件	低	删除草稿中新邮件	低
浏览更多消息	高		低
下载附件	高	(未)接收邮件	低
新日程事件	高	移动邮件	低
编辑/更改日程事件	高	任何日程变化	高
添加联系人	高	任何联系人变化	高
编辑联系人	高	删除/额定设备	高
查找通讯录	高	设置变化	高
更改设置	高	任何文件夹变化	高
手动发送/接收	高	连接重启	高(若不更改,则不发送信息)
即时消息状态改变	中	社交网络状态更新	中
竞标出牌或改变通知	高	初始天气警报	高
天气更新	低	新闻更新	低

[0121] 表一

如上述表1所示,为描述目的,展示了二进制表示方案中分配示例事务优先级的一些例子。针对另外的事件类型、请求或事务,存在其他分配方式,如上所述,优先级的分配可在更多或更少层级进行,例如,时域层或应用程序层等。

[0122] 如上述表格的例子所示,优先级较低请求/事务通常包含:更新信息状态为已读,未读、删除信息、删除联系人;优先级较高的请求/事务,一些实例包含:状态更新、新IM聊天消息、新电子邮件、日程事件的更改/取消/删除、手机游戏时域中的事件,其他娱乐相

关事件、网络购买或在线购物确认,加载或下载附加内容请求,通讯录相关事件,更改设备配置的事务,位置感知或基于位置的事件/事务,或任何由用户启动的事件/请求/事务,或已知、预计或猜想用户在等待响应等。

[0123] 收件箱整理事件(如电子邮件或其他类型的消息),通常被认为是低优先级和不含紧迫性的事件,一般不会触发使用设备250的无线电。具体来说,在预定的整理时间内,若没有无线电,移除旧邮件或其他内容的整理事件可作为其它通信的“背驮”事件。例如,如果用户喜好的设置为“信息保存7天”,当启动无线电时,再删除邮件,而非一旦邮件超过7天,就给设备的无线电上电以启动设备250的删除信息操作。如果已打开无线电,则在预定时间执行整理事件。

[0124] 请求/事务管理器235,可以利用请求的优先级(如通过优先级引擎241)管理设备250的输入流量,以优化资源(例如,为节约电池更有效地利用设备无线电)。例如,若无线电未启用,低于一定优先级的事务/请求可能不会触发使用设备250上连接管理器265控制的无线。与此相反,当检测到某一事务请求超过一定优先级,无线电控制器266则打开无线电,发送该请求。

[0125] 一个实施例中,利用优先级分配(如由本地代理275或其他设备/实体确定),引发远程设备修改其与移动设备或无线设备的通信频率。例如,远程设备被配置为当重要性较高的数据可以被发送到移动设备或无线设备时,发送通知至设备250。

[0126] 一个实施例中,事务优先级结合用户活动特征,能用于整形或管理流量,如通过流量整形引擎255。例如,检测到用户处于休眠或不活跃状态,作为响应,流量整形引擎255等待一段时间,再发送设备250的低优先级事务。此外,流量整形引擎255可以允许累积多个低优先级事务,实现设备250的批量传输(如通过批处理模块257)。一个实施例中,用户能设置、配置或调整优先级。例如,表1相同或类似表格中所述内容能在设备250的用户界面上获取,并且用户可用来调整或查看优先级。

[0127] 批处理模块257可以基于一定标准启动批量传输。例如,批量传输(例如,事件重复出现,其中一些事件发生在不同的实例)在检测到一定数量的低优先级事件,或在第一批低优先级事件启动后一定时间,可能启动。此外,当设备250启动或检测到较高优先级事件时,批处理模块257能启动所累积低优先级事件的批量传输。不然当其他原因触发使用无线电时,批量传输也能被传输(例如,从远程设备,如主机服务器100或300处接收数据)。一个实施例中,当发生批量传输时,可以执行即来的整理事件(整理收件箱)或其他低优先级的事件。

[0128] 一般来说,批处理功能可以在事件/事务层,应用程序层,或时域层被禁用或启用,根据以下任一或组合:用户配置、设备局限性/设置、制造商详细说明、网络提供商参数/限制性、平台具体限制条件/设置,设备操作系统设置等。一个实施例中,当应用程序/窗口/文件被关闭、退出或移至背景模式时,批量传输被启动;用户可以选择在启动批量传输前进行提示,用户还可以手动触发批量传输。

[0129] 在一个实施例中,本地代理275在缓存285中缓存数据,本地调整用于设备250上的无线电。若缓存285所存储的内容能满足设备250的请求或事务,无线电控制器266无需激活无线电,向远程实体发送请求(例如如图1A和图3A所示主机服务器100,300,或内容提供商/应用程序服务器,如图1A和1B示例中的服务器/提供商110)。正因如此,本地代理275可以使用

本地缓存285和缓存策略管理器245本地存储数据,满足数据请求,为节约网络资源和消耗的设备电池减少或降低设备无线电的使用。

[0130] 利用本地缓存,一旦请求/事务管理器225截获设备250应用程序的数据请求,本地仓库285被查询后确定是否具有本地存储的响应,并确定该响应是否有效。若本地缓存285存在一个有效响应,设备250可以将该响应提供给设备的应用程序,而无需访问蜂窝网络或无线宽带网络。

[0131] 若不存在有效响应,本地代理275可以查询远程代理服务器(如图3A所示的服务器代理325),以确定是否存在远程存储的有效响应。若存在,远程存储的响应(例如,存储在服务器缓存135或如图1B示例中的可选缓存服务器199)可以提供给移动设备,无需移动设备250访问蜂窝网络,从而缓解了网络资源的消耗。

[0132] 若不存在有效响应,或缓存的响应不适用被截获的数据请求,本地代理275,如缓存策略管理器245,可以将该数据请求发送至远程代理(如图3A所示代理服务器325),后者将数据请求转发内容源(如图1A所示应用程序服务器/内容提供商110),内容源通过远程代理提供响应,这将在图3示例主机服务器300的相关说明中作进一步详述。缓存策略管理器245可以管理或处理使用不同的协议请求,包括但不限于:HTTP、HTTPS、IMAP、POP、SMTP、XMPP和/或ActiveSync。缓存策略管理器245能为本地数据库285中的数据请求本地存储响应,作为缓存条目,以满足以后相同或类似的数据请求。

[0133] 缓存策略管理器245可以请求远程代理监控数据请求的响应,并且当检测到数据请求的意外响应时,远程代理通知设备250。在这种情况下,当被通知意外响应时(例如,新数据、更改数据、附加数据等),缓存策略管理器245能够清除或替换设备250上本地存储的响应。一个实施例中,缓存策略管理器245能够检测或识别具体请求所使用的协议,包括但不限于:HTTP、HTTPS、IMAP、POP、SMTP、XMPP和/或ActiveSync。一个实施例中,本地代理275应用程序的具体处理程序(例如通过缓存管理器245的应用程序协议模块246)允许优化任何端口映射到分布式代理服务器上处理程序的协议(例如图3A示例中代理服务器325的端口映射)。

[0134] 在一个实施例中,本地代理275通知远程代理,以便其监视从内容源接收的数据请求响应的更改结果,在将结果返回至移动设备之前更改结果,例如,当内容源的数据请求所产生的结果与返回到移动设备的结果一致。一般来说,本地代理275能使用本地缓存的内容,为设备250上应用程序模拟应用程序服务器的响应。这可以防止利用蜂窝网络传输不可获取的新/更改数据,从而释放网络资源,防止网络堵塞。

[0135] 一个实施例中,本地代理275包含应用程序行为检测器236,用于跟踪、检测、观察、监控可访问的应用程序(例如感知代理和/或未感知应用程序210和220)或设备250上安装的应用程序。本地代理275可使用应用程序行为,或设备250上可访问应用程序所检测行为模式(如通过模式检测器237),优化所需的无线网络流量,以满足这些应用程序的数据需要。

[0136] 例如,根据所检测的多个应用程序的行为,流量整形引擎255可以通过网络(无线网络)校准至少一些应用程序的内容请求(如通过校准模块256)。校准模块256能延迟或加快之前接收的请求以实现校准。当请求经校准后,流量整形引擎255能利用连接管理器进行网络轮询,以满足应用程序的数据请求。多个应用程序的内容请求能以行为模式或规则/设

置为基础进行校准,例如包含多个应用程序(音频,视频,文本等)请求的内容类型、设备(例如,移动设备)参数、和/或网络参数/流量情况、网络服务提供商限制/详细说明等。

[0137] 一个实施例中,模式检测器237可以检测由多个应用程序所提请求的重复,例如通过跟踪应用程序的行为模式。跟踪模式包含:检测某些应用程序,后台进程,定期轮询应用程序服务器,一天的某些时间,一周的某些天,以预定方式定期执行,某些频率,响应某些类型事件的某些频率,响应某种类型的用户查询,请求内容一致的频率,请求间歇,提出请求的应用程序,或任何上述组合。

[0138] 流量整形引擎255能利用这类重复,卸载从内容源(如图1A所示应用程序服务器/内容提供商110)轮询的内容,使得本该在移动设备或无线设备250执行的应用程序请求,转而由设备250的远程代理执行(如图1B所示代理服务器125或图3A所示代理服务器325)。当重复请求匹配某一规则时,流量匹配引擎255卸载查询内容。例如,相同资源的多个事件或请求具有完全相同的内容或返回的值或者基于请求和响应之间的可重复时间周期的检测,如在白天的特定时间请求的资源。卸载该轮询能降低移动设备250为重复内容轮询,创建与内容源的无线(蜂窝或其它无线宽带)连接所需的总带宽消耗。

[0139] 由于卸载查询内容,当未检测到内容源查询内容的更改时,存储在本地缓存285的缓存内容能被提供给设备250,满足数据请求。正因为如此,当数据没有更改时,应用程序的数据需求能得到满足,而无需使用无线电或占用蜂窝无线网络的带宽。当数据已更改和/或接收到新数据时,被卸载内容的远程实体能通知设备250。远程实体可以为如图3A示例中的主机服务器300。

[0140] 一个实施例中,本地代理275能降低所需/使用以维持TCP/IP连接的定期保活信息(心跳信息),其会消耗大量电能,从而不利于移动设备的电池寿命。本地代理(如心跳管理器267)的连接管理器265能检测、识别,并拦截任何或所有从应用程序发送的心跳(保活)信息。

[0141] 心跳管理器267可以阻止任何或所有通过蜂窝或其他网络发送的心跳消息,并改由分布式代理系统的服务器组件(如图1B所示)生成并发送的心跳消息,以维持与后台的连接(如图1A示例中的应用程序服务器/提供商110)。

[0142] 本地代理275通常代表所述单一管理器、模块、和/或引擎的任一或部分功能。本地代理275和设备250在不偏离本申请所公开新颖性范围时,可另外增加或减少组件;增多或减少功能;整体或部分。

[0143] 图2B为一框图,展示了如图2A所示缓存系统中的组件的进一步示例,其能够为移动应用程序行为和/或网络条件缓存和适配缓存策略。

[0144] 在一个实施例中,缓存策略管理器245包括元数据生成器203、缓存查找引擎205、缓存适当性决策引擎246、轮询时间表生成器247、应用程序协议模块248、缓存或连接选择引擎249和/或本地缓存无效器244。缓存适当性决策引擎246可进一步包括时序定时预测器246a、内容预测器246b、请求分析器246c和/或响应分析器246d,以及缓存选择引擎249包括:响应调度器249a。元数据生成器203和/或缓存查找引擎205耦合至缓冲285(或本地缓存)以用于缓存条目及其查询的修改或增补。

[0145] 缓存查找引擎205可以进一步包括ID或URI过滤器205a、本地缓存无效器244可进一步包括TTL管理器244a,以及轮询时间表生成器247可以进一步包括时间表更新引擎247a

和/或时间调整引擎247b。缓存策略管理器245的一个实施例包括应用程序缓存策略库243。在一个实施例中,应用程序行为检测器236包括模式检测器237、轮询间隔检测器238、应用程序配置文件生成器239和/或优先级要求241。轮询间隔检测器238可进一步包括具有响应/请求搜寻引擎238b的长轮询检测器238a。轮询间隔检测器238可进一步包括长轮询搜寻检测器238c。应用程序配置文件生成器239可以进一步包括响应延迟间隔跟踪器239a。

[0146] 模式检测器237、应用程序配置文件生成器239和优先级引擎241被与图2A中所示的模式检测器的说明结合描述。实施例中还包括应用程序配置文件存储库242,其可以由本地代理275使用以存储关于应用程序配置文件的信息或元数据(例如,行为、模式,HTTP请求的类型等等)。

[0147] 缓存适当性决策引擎246可以检测、评估或确定来自内容源(例如,图1B所示例子中的应用程序服务器/内容提供商110)的内容是否与移动设备250进行交互,以及是否具有可适合于缓存的内容。例如,决策引擎246可以使用与针对移动设备250启动的请求收到的请求和/或响应有关的信息以确定缓存能力、潜在缓存能力或者无缓存能力。在某些情况下,决策引擎246可以初步确认请求是否被定向到黑名单目的地或请求本身是否来源于黑名单的客户端或应用程序。如果是这样的话,额外的处理和分析可能不会被决策引擎246执行,并且请求可以被允许通过空中发送到服务器来满足该请求。黑名单目的地或应用程序/客户端(例如,移动应用)可被在本地代理(例如,在应用程序配置文件存储库242)中本地维持或远程(例如,在代理服务器325或其他实体中)维持。

[0148] 在一个实施例中,决策引擎246,例如,通过请求分析器246c,收集与在移动设备250产生的应用程序或客户请求有关的信息。请求信息包括请求特征信息,其包括,例如,请求方法。例如,请求方法可以表示移动应用程序或客户端产生的HTTP请求的类型。在一个实施例中,如果请求方法为GET请求或POST请求,响应一个请求可以被识别为可缓存的或潜在的可缓存的。其他类型的请求(例如,选项、头部、提交、删除、跟踪或连接)可能会或可能不会被缓存。在一般情况下,具有不可缓存请求方法的HTTP请求将不会被缓存。

[0149] 例如,请求特征信息可以进一步包括关于请求的大小的信息。主体大小超过一定尺寸的响应请求(如,HTTP请求)将不会被缓存。例如,如果有关请求的信息表明请求的请求主体大小没有超过一定的尺寸,缓存能力能够被确定。在某些情况下,最大可缓存请求体的大小可以设置为8092字节。例如,在其他情况下,取决于网络容量或网络运营商的特定设置,可能会使用不同的值。

[0150] 在一些情况下,基于一组标准,例如,标准指定内容的时间关键性,该内容正从内容源被请求,来自给定应用程序服务器/内容提供商(例如,图1B的服务器/内容提供商110)中的内容被确定为适合于缓存。在一个实施例中,本地代理(例如,图1B和图2A的本地代理175或275)应用一个选择标准以存储主机服务器中的内容,其被应用程序请求作为移动设备上的本地缓存中被缓存的元素以满足由应用程序作出的后续请求。

[0151] 进一步地基于检测到的请求和/和接收到的响应的模式,所述请求从移动设备250发送(例如,由移动应用程序或设备250上其它类型的客户端),缓存适当性决策引擎246可以检测请求和/或响应中的可预测性。例如,由决策要求246(例如,请求分析器246c)收集的请求特征信息,可进一步包括在请求和其它由移动设备上同一客户端生成的请求或其他定向到同一个主机(例如,具有相同或相似的标识参数)的请求之间的可预测信息。

[0152] 当由相同客户端产生的请求和其他请求,以固定速率或几乎固定速率或以具有一些可识别的或部分地或全部地可复制的变化模式的动态速率发生时,周期性可以被决策引擎246或请求分析器246c所检测。如果请求以一些可识别的模式被提出(例如,定期的时间间隔、具有可检测的模式的时间间隔或趋势(例如,增加、减少、恒定的,等等),定时预测器246a中可以确定由设备上给定应用程序作出的请求是可预测的并确定它是潜在的适于缓存,至少从定时的角度来看。

[0153] 可识别的模式或趋势通常可以包括任何应用程序或客户端行为,其可被本地模拟,例如,在移动设备上250的本地代理275上,或被远程模拟,例如,由主机300上的代理服务器325,或本地和远程模拟结合的模拟应用程序的行为。

[0154] 在一个实施例中,决策引擎246,例如,通过响应分析器246d,可以收集关于移动设备250产生的应用程序或客户端请求的响应的信息。响应通常是从应用程序的服务器或主机(例如,移动应用程序)或在移动设备250上发送请求的客户端处接收。在某些情况下,移动客户端或应用程序可以是应用程序的移动版本(例如,社交网络、搜索、旅游管理、语音邮件、联系人管理器、电子邮件)或通过网络浏览器或桌面客户端访问的网站。

[0155] 例如,响应特征信息可以包括是否传输编码或块传输用于发送响应的编码的指示。在某些情况下,具有传输编码或块传输编码的HTTP请求的响应未被缓存,因此也可从进一步分析中被删除。基本原理是分块响应通常是很大的和非最佳缓存的,因为这些事件的处理可能减慢整体性能。因此,在一个实施例中,当传输编码未用于发送响应中时,可缓存能力或潜在的可缓存能力可以被确定。

[0156] 此外,响应特征信息可以包括相关联的响应状态码,其可被响应分析器246d识别。在某些情况下,具有不可缓存状态码的HTTP响应通常不被缓存。响应分析器246d可以从响应中提取状态码,并确定它是否匹配缓存或不可缓存的状态码。一些可缓存的状态码以举例的方式包括:200-OK、301-重定向、302-发现、303-查看其他、304-未修改、307-暂时重定向或500-内部服务器错误。一些不可缓存状态码可以包括,例如,403-禁止或404-未找到。

[0157] 在一个实施例中,如果响应信息不表示不可缓存状态码或表示可缓存的状态码,则可缓存能力或潜在的可缓存能力可以被确定。如果响应分析器246d检测与给定响应相关的不可缓存的状态码,则具体的事件(请求/响应对)可能被从进一步的处理中淘汰并在临时性、半永久性或永久性的基础上被确定为不可缓存。如果状态码表示可缓存能力,事件(例如,请求和/或响应对)可能会受到进一步的处理和分析以确定可缓存能力,如图9-13的示例流程图中所示。

[0158] 响应特征信息还可以包括响应大小信息。在一般情况下,如果响应未超过一定的大小,响应可被本地缓存在移动设备250上。在某些情况下,默认最大被缓存的响应的大小被设置为115KB。在其他情况下,基于操作条件、网络条件、网络容量、用户偏好、网络运营商的要求,或其他应用程序特征、用户特征和/或设备特征原因,最大可缓存的响应的大小可能是不同的和/或动态调整的。在一个实施例中,响应分析器246d可以识别响应的大小,并且如果响应的大小不超过给定阈值或最大值,则可缓存能力或潜在的可缓存能力可被确定。

[0159] 此外,响应特征信息可以包括请求的响应的响应主体信息,以及其它请求的其它响应的响应主体信息,其它请求由在移动设备上的相同客户端生成或被定向至相同内容主

机或应用程序服务器。如果由缓存条目提供,响应和其它响应的响应主体信息可以被比较,例如,由响应分析器246d进行比较以防止动态内容的缓存(或频繁变化的以及不能有效配合于缓存条目的内容的响应,如财务数据、股票行情、新闻、实时的体育赛事活动等),如不再相关的或最新的内容。

[0160] 缓存适当性决策引擎246(例如,内容预测器246b)可以明确地识别重复性或识别重复性、潜在重复性或从内容源(例如,图1A-B的示例中所示的内容主机/应用程序服务器110)接收的响应中的可预测性的指示的。重复性可以通过,例如,跟踪从内容源接收的至少两个响应被检测,并确定两个响应是否相同。例如,如果响应的响应主体信息和由相同移动客户端发送的其它响应或或定向到同一主机/服务器其它响应的响应主体信息是相同的或基本相同的,则可缓存能力可以由响应分析器246d确定。两个响应可能是或可能不是在响应连续中被发送的响应。在一个实施例中,为了一般的应用程序和/或特定的请求,接收到的来自给定应用程序请求的响应的散列值被用来确定内容的可重复性(带或不带试探)。对于某些应用程序或在某些情况下,可能需要额外的相同的响应。

[0161] 接收内容中的重复性不需要100%确定。例如,如果一定数量或一定比例的响应是相同的或类似的,响应可以被确定为是可重复的。一定数量或一定比例的相同/相似的响应可以被跟踪一选定时间段,设置为默认或基于应用程序产生的请求被设置(例如,无论应用程序是具有不断更新的高动态或具有不频发更新的低动态)。任何指定的可预测性或可重复性,或者可能的可重复性,可以由缓存内容的分布式系统所利用以被提供至移动设备250上的请求应用程序或客户端。

[0162] 在一个实施例中,对于长轮询类型的请求,当前两个响应的响应延迟时间是相同的、基本相同的,或检测到在时间间隔中的增加时,本地代理175可以开始在第三请求上缓存响应。一般情况下,前两个响应的接收到的响应应该是相同的,以及基于验证第三个请求的接收到的第三响应是相同的(例如,如果 $R_0 = R_1 = R_2$),第三响应可以本地缓存在移动设备上。取决于应用程序的类型、数据的类型、内容的类型、用户偏好或运营商/网络运营商的规格,可能需要更少或更多的相同的响应以启动缓存。

[0163] 为长轮询增加相同响应的响应延迟可以指明搜索时期(例如,给定网络将允许一时段,在该时段中移动设备上的应用程序/客户端正在寻求的请求和响应之间的最长时间),搜索时期由应用程序行为检测器236的长轮询搜索检测器238c检测。

[0164] 使用 T_0 、 T_1 、 T_2 的例子将被描述如下,其中 T 表示当请求被发送和当连续请求的响应(例如,响应头部)被检测/接收之间的延迟时间:

$T_0 = \text{响应}0(T) - \text{请求}0(T) = 180 \text{秒} (+/- \text{公差})$

$T_1 = \text{响应}1(T) - \text{请求}1(T) = 240 \text{秒} (+/- \text{公差})$

$T_2 = \text{响应}2(T) - \text{请求}2(T) = 500 \text{秒} (+/- \text{公差})$

在上示的时序例子中, $T_0 < T_1 < T_2$,这可能表明当网络超时尚未达到或已超过时的长轮询搜索模式。此外,如果三个请求的接收到的响应 R_0 、 R_1 和 R_2 是相同的, R_2 可以被缓存。在这个例子中, R_2 是在长轮询搜索期间被缓存而无需等待处理长轮询,从而加快响应缓存(例如,这是可选的加速缓存行为,其可以实现所有的或选定的应用程序)。

[0165] 如果检测到响应延迟 $T_1 < T_2 < T_3 > T_4$ 的模式,如上述时序(例如,由应用程序行为检测器236的长轮询搜索检测器238c检测到的)所示,其可以判定 T_3 在长轮询搜索期间可能

超出网络时间。请求3中,在响应被发送或可获得之前,由于连接被网络、应用程序、服务器或其他原因终止,响应很可能不会被收到。请求4(在T4后),如果一个响应(例如,响应4)被检测或接收,本地代理275稍后可以使用响应用于缓存(如果内容重复性条件得到满足)。本地代理也可以使用T4作为为代理服务器设置的轮询时间表中轮询间隔以监测/轮询内容源。

[0166] 请注意上面的描述表明只要收到响应以及给定请求没有超时,则在长轮询在检测增加的响应延迟的事件中的搜索模式中时缓存可以开始。这可以被作为在长轮询期间的可选的加速缓存。缓存也可以在搜索模式(例如,在轮询请求确定在恒定或接近恒定的延迟值之后)完成之后开始。请注意,长轮询的搜索在搜索发生时可能会或可能不会发生,代理275通常可以检测到这个并决定是否在搜索期间开始缓存(增加相同响应的间隔),或等待直到搜索定到一个稳定的值。

[0167] 一个实施例中,缓存适当性决策引擎246的时序预测器246a可以跟踪来自应用程序(例如,移动应用程序)或客户端的出站请求接收到的响应的时序以检测到任何可识别的模式,该模式可以是部分地全部重现性以使得本地缓存响应可以以模拟内容源(例如,应用程序服务器/内容提供商110或310)行为的方式被提供给移动设备250上的请求客户端。例如,以何种方式(例如,从时序的角度来看)响应或内容将被传递到移动设备250上的请求应用程序/客户端。当应用程序或移动客户端请求的响应被从本地和/或远程缓存提供而不是被直接从内容源(例如,应用程序、内容提供商110或310)检索/接收时,这将确保用户体验的保存。

[0168] 在一个实施例中,决策引擎246或时序预测器246a确定给定的应用程序(例如,移动设备应用程序)或客户端的时序特征,该应用程序或客户端来自,例如,请求/响应跟踪引擎238b和/或应用程序配置文件生成器239(例如,响应延迟间隔跟踪器239a)。时序预测器246a使用时序特征确定接收到响应请求的内容是否适合或潜在适合缓存。例如,来自给定的应用程序的两个连续请求之间的轮询请求间隔可以用来确定请求间隔是否是可重复的(例如,恒定的、接近恒定的、随模式增加,随模式减少等)并可预测的,从而在至少一些时间精确地或近似地在公差范围水平内复制。

[0169] 在一些情况下,为特定应用程序、应用程序的多个请求或多个应用程序的给定请求的时序特征可以存储在应用程序配置文件存储库242中。应用程序配置文件存储库242通常可以存储任何类型的关于应用程序请求/响应特点,包括时序模式、时序重复性、内容重复性等的信息或元数据。

[0170] 应用程序配置文件存储库242也可以存储元数据,该元数据指示由给定应用程序使用的请求的类型(例如,长轮询,长期持有的HTTP请求,HTTP流,推送,COMET推送,等等)。应用程序配置文件由应用程序指示请求类型,可以用于当检测到随后的相同/相似的请求时或当从已被分类的应用程序检测到请求时。以这种方式,已被跟踪和/或分析的给定请求类型或特点应用程序的请求的时序特征不需要被重新分析。

[0171] 应用程序配置文件可以与生存时间(例如,或者默认的过期时间)相关联。应用程序配置文件或应用程序或请求的配置文件的各个方面的过期时间的使用可以按个别情况来使用。应用程序配置文件条目的生存时间或实际到期时间可以设置为默认值或单独确定或以其组合确定。应用程序配置文件也可以具体到无线网络、物理网络、网络运营商或特定

的运营商。

[0172] 实施例包括应用程序黑名单管理器201。黑名单管理器201可以被耦合到应用程序缓存策略存储库243并可以部分或全部内置于本地代理或缓存策略管理器245。同样,黑名单管理器201可以部分或全部内置于本地代理或应用程序行为检测器236。黑名单管理器201可以在永久或临时的基础上聚集、跟踪、更新、管理、调整或动态监测列入“黑名单”的、或认定为不缓存的服务器/主机的目的地列表。目的地的黑名单,当在请求中的确定时,有可能被使用以允许该请求被经由(蜂窝)网络发送用于服务。额外的处理请求可能不会被执行,因为它检测到引导至被列为黑名单的目的地。

[0173] 列入黑名单的目的地可在应用程序缓存策略存储库243中由地址标识符识别,该地址标识符包括标识符的特定的URI或模式,该标识符包括URI模式。一般情况下,被列入黑名单的目的地可以被任何一方设置或因任何原因被任何一方修改,该任何一方包括用户(所有者/移动设备250的使用者)、设备250的操作系统/移动平台、目的地本身、网络运营商(蜂窝网络)、互联网服务提供商、其他第三方,或可根据已知不能被缓存/不适于缓存的应用程序的目的地列表。一些在黑名单目的地中的条目可能包括基于由本地代理(例如,缓存适当性决策引擎246)执行的分析或处理的聚集的目的地。

[0174] 例如,在移动设备上的应用程序或移动客户端的已被确定为不适合用于缓存的响应可以加入到黑名单。其相应的主机/服务器可能会增加以补充或代替在移动设备250上的请求应用程序/客户端的标识。部分或全部的这样的由代理系统标识的客户端的可以被添加到黑名单中。例如,对于所有的暂时标识为不适于缓存的应用程序客户端或应用程序中,仅有那些具有某些检测到的特征(基于时序、周期性、响应内容变化的频率、内容的可预测性、大小等)可以被列为黑名单。

[0175] 被列入黑名单的条目可包括在移动设备(而不是目的地)上的请求应用程序或请求客户端以使得当从给定应用程序或给定客户端检测到请求时,它有可能通过网络发送至响应,因为列入黑名单的客户端/应用程序的响应在大多数情况下不被缓存。

[0176] 给定应用程序配置文件还可以被不同地对待或处理(例如,本地代理275和远程代理325的不同的行为),这取决于与移动设备相关的应用程序正在访问的移动帐户。例如,更高的付费帐户,或最初的帐户可能允许更频繁访问无线网络或更高的带宽分配,从而影响本地代理275和代理服务器325之间实施的缓存策略,该策略相比于节约资源,其强调更好的性能。给定的应用程序配置文件也可以根据不同的无线网络条件(例如,基于拥塞或网络中断等)被不同地等待或处理。

[0177] 需注意的是可以为移动设备上250上的多个客户端或应用程序确定、跟踪和管理缓存适当性。也可以为移动设备250上的给定客户端或应用程序发起的不同的请求或请求类型确定缓存适当性。缓存策略管理器245与时序预测器246a和/或试探性地确定或估计可预见性或潜在的可预见性的内容预测器246b,可以跟踪,管理和存储可缓存给定的应用程序的各种应用程序或各种请求的可缓存信息。可缓存信息可能还包括条件(例如,应用程序可以在一天的某些时候、一周的某几天被缓存,或者给定的应用程序的某些请求可被缓存,或所有具有给定目标地址的请求可以被缓存),在该条件下,缓存是适当的,其可由缓存适当性决策引擎246确定和/或跟踪以及当在适当的时候应用程序缓存策略存储库243耦合到缓存适当性决策引擎246时被存储和/或更新。

[0178] 当相同的请求被检测到时,应用程序缓存策略存储库243中的关于请求、应用程序的缓存性能和/或相关条件的信息以后可以稍后使用。以这种方式,决策引擎246和/或时序和内容预测器246a/b不需要跟踪并分析请求/响应时序和内容特征以作出关于缓存性能的评估。此外,缓存性能信息可在某些情况下与其他移动设备的本地代理通过直接通信或通过主机服务器(例如,主机服务器300的代理服务器325)的方式共享。

[0179] 例如,在各种移动设备是的本地代理275检测到的缓存性能信息可以被发送到主机服务器上的代理服务器325或远程主机服务器(例如,图3A的示例中所示的主服务器300或代理服务器325,图1A-B的示例中所示的主机100和代理服务器125)。然后远程主机或代理服务器分发信息,该信息关于应用程序-特定、请求-特定的缓存性能信息和/或各种移动设备或在无线网络中或跨越多个无线网络(相同的服务提供商或多个无线服务提供商)的本地代理的用于它们使用的任何相关条件。

[0180] 一般来说,缓存的选择标准可以进一步包括,通过举例的方式但不限制于,指示移动设备是否是有效或无效、网络条件和/或无线电覆盖范围统计信息的移动设备的状态。缓存适当性决策引擎246可在标准的任何一种或任何组合之中,并可以以任何顺序,识别缓存可能适合的源。

[0181] 一旦应用程序服务器/内容提供商具有可能适合于在移动设备250上本地缓存的被识别的或被检测的内容,缓存策略管理器245通过存储从内容源接收到的内容为在移动设备250上的本地缓存(例如,图1B和图2A的例子中分别所示的本地缓存185或285)中的缓存元素可以缓存从所识别的源接收的相关的内容。

[0182] 响应可以被存储在缓存285(例如,也称为本地缓存)作为缓存条目。除了请求的响应,缓存条目可包括具有关于响应缓存的附加信息的缓存元数据。元数据可以由元数据生成器203生成,并且可以包括,例如,时序数据,例如缓存条目的访问时间或者缓存条目的创建时间等等。元数据可以包括附加的信息,如适合于在确定作为缓存条目存储的响应是否用于满足随后的响应中使用的任何信息。例如,元数据信息可以进一步包括,请求时序历史(例如,包括请求时间、请求启动时间、请求结束时间),请求和/或响应的哈希值,间隔或在间隔中的变化等。

[0183] 缓存条目通常存储在与生存时间(TTL)相联的缓存285中,例如可能会由缓存无效器244的TTL管理器244a指定或确定。缓存条目的生存时间保持在缓存285中的条目时间的总量,不管响应是否依然有效或者与给定请求或在移动设备250上的客户端/应用程序相关。例如,如果设置的给定的缓存条目的生存时间为12个小时,即使包含在缓存条目的响应主体仍然是当前的并适用于相关请求的,缓存条目被清除、删除或以其他方式指明为超过生存时间。

[0184] 默认生存时间除非另有规定可以自动使用于所有的条目(例如,由TTL管理器244a),或每个缓存条目可以创建其个自的TTL(例如,基于各种动态或静态的标准由TTL管理器244a所确定)。注意,每个条目可以有与响应数据和任何相关联的元数据相关联的单个的生存时间。在某些情况下,相关联的元数据可具有不同于响应数据的生存时间(例如,较长的生存时间)。

[0185] 具有用于缓存的内容的内容源,除了或可替代的,被确定为代理服务器(例如,分别在图1B和图3A的例子中所示的代理服务器125或325)远离移动设备250并与移动设备250

无线通信以使得代理服务器可以为新的或更改的数据监测内容源(例如,应用程序服务器/内容提供商110)。类似地,本地代理(例如,分别在图1B和图2A的本地代理175或275)可以识别到代理服务器,从特定的应用程序服务器/内容提供商接收的内容作为缓存的元素285被存储在本地缓存中的。

[0186] 一旦内容已经在本地缓存,缓存策略管理器245,基于接收未来轮询请求以与应用程序服务器/内容主机(例如,110或310)联系,可以从本地缓存检索缓存的元素以响应在移动设备250作出的轮询请求以使得移动设备的无线电未被激活为服务轮询请求。例如,缓存查找引擎205可以查询缓存285以确定响应被提供至响应。响应可以从缓存中提供以响应识别匹配的缓存条目和使用任何在缓存条目中与响应一起存储的元数据。缓存条目可被缓存查找引擎通过使用请求的URI或另一种类型的标识符(例如,通过ID或URI过滤器205a)进行查询。缓存查找引擎205还可以使用与匹配的缓存条目一起存储的元数据(例如,提取任何时序信息或其他有关信息)以确定响应是否仍然适合在被提供至当前请求中使用。

[0187] 需要注意的是缓存查找可以由引擎205通过使用一个或多个不同的多种策略执行。在一个实施例中,可以按顺序存储在缓存285中的每个条目上执行多个查找策略,直到至少有一个策略识别匹配的缓存条目。进行缓存查找所采用的策略可以包括严格匹配标准或允许非匹配参数的匹配标准。

[0188] 例如,查找引擎205可以执行严格的匹配策略,该策略寻找在代理试图识别缓存条目的目前的请求中引用的标识符(例如,一个主机或资源的URI)以及与缓存条目一起存储的标识符之间的精确匹配。在标识符包括URI(s)或URL(s)的情况下,严格匹配的匹配算法将查找在URL匹配中的所有参数的缓存条目。例如:

例1

1、缓存包含条目http://test.com/products/

2、请求正在作出URI http://test.com/products/严格策略将发现匹配,因为这两个URI是相同的。

[0189] 例2

1、缓存包含条目http://test.com/products/?query=all

2、请求正在作出URI http://test.com/products/?query=sub

在上述严格的策略下,匹配将不会被发现,因为URI在查询参数中不同。

[0190] 在另一个示例性策略中,查找引擎205的用标识符查找缓存,该标识符部分匹配在代理试图识别匹配缓存条目的目前的请求中引用的标识符。例如,查找引擎205可查找具有与由查询参数识别的请求不同的标识符的缓存条目。在运用这一策略中,查找引擎205可以收集信息,该信息收集多个之前请求(例如,标识符中任意参数的列表)用于稍后检查在目前请求中的检测到的任意参数。例如,在缓存条目存储与URI或URL的标识符一起存储的情况下,查找引擎搜索具有URI的缓存条目,该URI随查询参数不同而不同。如果找到了,引擎205可以检查缓存条目在以前的请求(例如任意参数列表)中收集的信息并检查在目前URI/URL中检测到的或从目前的URI/URL中提取的任意参数是否属于任意参数列表。

[0191] 例1

1、缓存包含条目http://test.com/products/?query=all,其中查询被标记为任意。

[0192] 2、请求正在作出URI `http://text.com/products/?query=sub`匹配将被发现,因为查询参数被标记为任意。

[0193] 例2

1、缓存包含条目`http://test.com/products/?query=all`,其中查询被标记为任意。

[0194] 2、请求正在作出URI `http://test.com/products/?query=sub&sort=asc`匹配将不会被发现,因为目前的请求包含没有在缓存条目中标记为任意的分类参数。

[0195] 可采用附加策略检测缓存命中。这些策略可以单独使用或以其任意的组合实施。当这些策略中的任何一个决定匹配时,缓存命中可被确定。当查找引擎205确定请求的数据因任何原因不能从缓存285提供时,缓存未命中可被指示。例如,当没有缓存条目被识别为任何或所有的被使用的查找策略,缓存未命中被确定。

[0196] 当匹配的缓存条目存在但被确定为无效或当前请求无关时,缓存未命中也可以被确定。例如,查找引擎205还可以分析与匹配的缓存条目有关的元数据(例如,其可以包括缓存条目的时序数据)以确定它是否仍然是适合用于响应当前请求。

[0197] 当查找引擎205已经识别了缓存命中(例如,指示请求的数据可以从缓存中提供的事件),在匹配缓存中的存储的响应可以由缓存提供以满足应用程序/客户端的请求。

[0198] 通过使用存储在缓存285中的缓存条目服务请求,网络带宽和其他资源不需要被用于请求/接收轮询响应,该轮询响应未从已经在移动设备250接收的响应改变。这种服务和实现应用程序(例如,移动应用程序)通过在本地缓存285中的缓存条目本地请求,以允许更有效的资源和移动网络流量的利用和管理,因为不需要通过无线网络发送的请求进一步消耗带宽。在一般情况下,缓存285可在移动设备250电源的开/关之间保持,并在应用程序/客户端刷新和重启时保持。

[0199] 例如,基于收到的来自其移动设备250或移动设备250上的应用程序或其他类型客户端的出站请求,本地代理275可以拦截该请求并确定缓存响应是否在移动设备250的缓存285中可用。如果是这样,出站请求由本地代理275使用移动设备的缓存上的缓存的响应作出响应。因此,发出的请求可以被填充或满足,而不需要通过无线网络发送出站请求从而节省了网络资源和电池消耗。

[0200] 在一个实施例中,响应设备250上的请求的应用程序/客户端定时至对应到内容服务器将通过持续连接响应出站请求的方式(例如,通过持续连接或长期持有的HTTP连接,长轮询式连接,将由本地代理建立缺席拦截)。通过由本地缓存285提供存储的内容而不是从预期内容源接收到的新鲜内容(例如,图1A-B的内容主机/应用程序服务器110),响应时序可以由本地代理275仿真或模拟以保存应用程序行为,以使得最终用户体验不会受到影响,或受到最小影响。时序可以准确地复制或公差参数内被估计,其可能会被用户忽视或由应用程序类似处理以便不引起操作问题。

[0201] 例如,出站请求可以是用于内容服务器(例如,图1A-1B的例子的应用程序服务器/内容提供商)的持久连接请求。在内容源(服务器)的持久连接(例如,长轮询、COMET式推送或任何其他在异步HTTP请求中推送模拟、长期持有的HTTP请求、HTTP流媒体或其他)中,在请求被发送之后连接被保持一定的时间。连接通常可以在移动设备和服务器之间被保存,直到内容在服务器上被提供以发送到移动设备。因此,当长轮询请求被发送以及当响应

从内容源接收之间时间上通常有一些延迟。如果内容源没有为一定量的时间提供响应,如果响应未被发送,则可能由于网络原因(例如,套接字关闭)终止该连接。

[0202] 因此,模拟来自内容服务器的经由持久连接(例如,长轮询方式连接)发送的响应,内容服务器的响应的方式可以通过在响应具有缓存的响应的出站请求之前允许间隔消逝来模拟。例如,间隔的长度可以在请求-请求的基础上或者应用程序-应用程序(客户端-客户端的基础上)被确定。

[0203] 在一个实施例中,间隔时间确定基于产生出站请求的移动设备上的应用程序的请求特征(如时序特征)被确定。例如,轮询请求的间隔(例如,它可以被跟踪、检测以及由轮询间隔检测器238的长轮询检测器238a确定)可以用来确定在响应具有本地缓存的请求之前的等待的间隔,并可以由响应调度器249a进行管理。

[0204] 缓存策略管理器245的一个实施例包括为移动设备250上的一个或多个应用程序产生轮询时间表的轮询时间表生成器247。轮询时间表可以指定轮询间隔,该轮询间隔可由物理上不同于和/或独立于移动设备250的实体在监测代表移动设备的一个或多个应用程序(例如缓存的响应可以通过轮询请求被定向至的主机服务器(主机服务器110或310)来进行定期验证)的内容源中采用。这样的可以监测移动设备250的源的内容的外部实体的例子是代理服务器(例如,图1B和图3A-C的例子中的代理服务器125或325)。

[0205] 轮询时间表(例如,包含轮询率/频率)可以被确定,例如,基于定向至来自移动设备的内容源的轮询请求之间的间隔。轮询时间表或轮询速率可以在移动设备250(由本地代理)确定。在一个实施例中,为了确定由任何或所有应用程序(例如,移动应用程序)作出的轮询请求之间的间隔,应用程序行为检测236的轮询间隔检测器238可以监测定向至来自移动设备250的内容源的轮询请求。

[0206] 例如,轮询间隔检测器238可以跟踪设备250上的应用程序或客户端的请求和响应。在一个实施例中,连续请求在由相同的移动客户端或应用程序(例如,移动应用程序)作出在移动设备250上的应用程序(例如,移动应用程序)发起的出站请求的检测之前被跟踪。轮询率利用收集响应缓存请求的请求信息确定。在一个实施例中,轮询率由在由相同的产生请求的客户端产生的在先请求之间的间隔的平均值来确定。例如,第一间隔可能被在当前请求和前一请求之间被计算,第二间隔可以在前两个请求之间被计算。轮询速率可由第一和第二间隔的平均来设置并在设置缓存策略中发送到代理服务器。

[0207] 可选的间隔可在产生平均值中被计算,例如,除了前两个请求的多个先前的请求可用于,以及多于两个的间隔可用于计算平均值。一般情况下,在计算间隔中,给定请求不需要导致响应从主机服务器/内容源接收以便它用于间隔计算。换句话说,给定请求的时序特征可以在间隔计算中被使用,只要该请求已被检测到,即使请求在发送中失败或如果响应检索失败。

[0208] 轮询时间表生成器247的实施例包括时间表更新引擎247a和/或时间调整引擎247b。时间表更新引擎247a确定更新速率的需要或者或轮询来自先前设定值的给定应用程序服务器/内容主机具有的间隔,基于在移动设备250上的客户端或应用程序(如移动设备应用程序)产生的实际请求中检测到的间隔变化。

[0209] 例如,监测率确定的请求现在可以被从应用程序(例如,移动应用程序)或客户端以不同的请求间隔发送。预定的更新引擎247a可确定实际请求更新的轮询间隔,并生成一

个不同于先前设定速率的新的速率以轮询代表移动设备250的主机。更新的轮询率可经由蜂窝网络传播至远程代理(代理服务器325)用于远程代理监测给定主机。在某些情况下,更新的轮询速率可在远程代理或监测主机的远程实体被确定。

[0210] 在一个实施例中,时间调整引擎247b可以进一步优化的被生成用来监测应用程序服务器/内容源(110或310)的轮询时间表。例如,时间调整引擎247b可选地指定开始轮询到代理服务器的时间。例如,除了设置代理服务器监测应用程序的轮询间隔,服务器/内容主机也可以指定在移动客户端/应用程序产生的实际请求的时间。

[0211] 然而,在某些情况下,由于固有的传输延迟或增加的网络延迟或其他类型的恢复时间,远程代理服务器接收来自本地代理的具有一些延迟(例如,几分钟或几秒钟)的轮询设置。在请求由移动客户端/应用程序产生后,这在源具有检测响应改变的效果,在响应不再是最新的或有效的之后,缓存的响应一旦再次被供应到应用程序则会引起缓存的响应发生的无效。

[0212] 为了在无效之前,再次解决供应超时内容的这种非最优的结果,时间调整引擎247b可以指定轮询应该开始除速率之外的时间(t_0),其中指定的到代理服务器325的初始时间 t_0 作为一个时间是小于当请求被移动应用程序/客户端产生时的实际时间的。通过这种方式,在由移动客户端产生实际请求稍前服务器轮询资源使得在实际应用程序请求之前任何内容的变化可以被检测。这可以防止无效的或不相关的超时内容/响应在提供新鲜内容之前被再次发送。

[0213] 在一个实施例中,基于来自移动设备250上的相同应用程序或客户端的在先请求的时序特征,来自移动设备250的出站请求被检测到以用于持久连接(例如,长轮询、COMET式推送和长期持有的(HTTP)请求)。例如,请求和/或相应响应可以由轮询间隔检测238的长轮询检测器238a的请求/响应跟踪引擎的238b进行跟踪。

[0214] 连续请求的时序特征可以被确定以设置应用程序或客户端的轮询时间表。轮询时间表可用于监测内容源(内容源/应用程序服务器)的内容变化,以使得存储在移动设备250的本地缓存上的缓存内容可被适当管理(例如,更新或丢弃)。在一个实施例中,时序特征可包括,例如,响应延迟时间(D)和/或空闲时间(IT)。在一个实施例中,响应/请求跟踪引擎238b可以跟踪请求和响应以确定、计算和/或估计请求者或客户端的请求的时序图。

[0215] 例如,响应/请求跟踪引擎238b检测到移动设备上的客户端启动的第一请求(请求0)和在响应于第一请求的移动设备处接收响应之后在移动设备上的客户端启动的第二请求(请求1)。第二请求是在第一次请求之后。请求之间的关系可见于图17A-B的时序图。

[0216] 在一个实施例中,响应/请求跟踪引擎238b可以跟踪请求和响应以确定、计算、和/或估计请求者或客户端的请求的时序图。例如,响应/请求跟踪引擎238b可以检测移动设备上的客户端启动的第一请求和在响应于第一请求的移动设备处接收响应之后在移动设备上的客户端启动的第二请求。第二请求是在第一次请求之后。

[0217] 响应/请求跟踪引擎238b进一步确定在第一、第二请求以及接收到的响应于第一个请求的响应之间的相对时序。一般情况下,相对时序可被长轮询检测器238a用于确定由应用程序生成的请求是否是长轮询请求。

[0218] 需注意的是,在一般情况下,由响应/请求跟踪引擎238b在计算相对时序中使用的第一和第二请求被选择用于在长轮询搜索期间固定之后或在长轮询搜索不会发生时进行

使用。典型的长轮询搜索期间的时序特征被在图8的例子中说明并可以,例如,由长轮询搜索检测器238c检测。换言之,由响应/请求跟踪引擎238b跟踪的请求并被用于确定给定的请求是否是长轮询请求的请求发生在长轮询固定之后。

[0219] 在一个实施例中,通过识别增加的请求间隔(例如,增加的延迟),长轮询搜索检测器238c可以识别或检测搜索模式。长轮询搜索检测器238a也可以通过检测增加的跟随有没有响应的请求(例如,连接超时)的请求间隔来检测搜索模式,或通过检测增加的跟随有在间隔中的减少的请求间隔来检测搜索模式。此外,长轮询搜索检测器238c可以应用过滤器的值或阈值的请求-响应时间的延迟值(例如,一个绝对值)检测的延迟高于该值则可以被认为是一个长轮询请求响应延迟。该过滤器的值可以是任何适当长轮询和/或网络条件的值特征(例如,2秒、5秒、10秒、15秒、20秒,等等),并可以用作过滤器的值或阈值。

[0220] 响应延迟时间(D)是指在请求发送后开始接收响应的的时间以及空闲时间是指在响应接收之后发送后续请求的时间。在一个实施例中,出站请求被检测用于持久连接,其基于响应延迟时间相对(D)或(D)的平均值(例如在一段时间的任何平均值)与空闲时间(IT)的比较(例如,由跟踪引擎238b执行的),例如,长轮询检测器238a。所使用的平均的数量可以是固定的、动态调整的或在一段较长的时间内可改变的。例如,由客户端发起的请求被确定为长轮询请求,如果响应延迟间隔大于空闲间隔($D > IT$ 或 $D \gg IT$)。在一个实施例中,长轮询检测器的跟踪引擎238b计算、确定或估计响应延迟间隔作为在第一请求的时间和初始检测或完整接收响应之间所经历的时间量。

[0221] 在一个实施例中,当空闲时间(IT)短暂时,请求被检测为持久连接,因为持久连接被建立至长轮询请求或长轮询的HTTP请求的响应中,例如,在收到前一个请求的响应(例如, $IT \sim 0$)之后,也可以在检测随后请求的即时或接近即时的发布中被特征化。因此,空闲时间(IT)也可以被使用来检测这样的即时或接近即时的重新请求以识别长轮询请求。第一请求的响应被收到后,由跟踪引擎238b确定的绝对或相对时序被用于确定第二请求是否是即时或接近即时的重新请求。例如,一个请求可能被归类为一个长轮询请求,如果 $D+RT+IT \sim D+RT$ 因为IT较小是成立的。如果IT小于阈值,则它可能被确定为小。请注意,该阈值可以是固定的或在有限的时间段(一个会话、一天、一个月等)计算出的或在一个较长的时间周期(例如,数月或分析的寿命)中计算出的。例如,对于每一个请求,平均IT可被确定,并可使用该平均(例如,平均IT小于百分比可能被作为阈值)确定阈值。这可以允许阈值随时间自动适应网络条件和服务器性能、资源的可用性或服务响应中的变化。固定的阈值时可以采取任何值,包括但不限于(例如,1秒、2秒、3秒.....等等)。

[0222] 在一个实施例中,长轮询检测器238a可以比较相对时序(例如,由跟踪引擎238b所确定的)和其他应用程序的请求-响应时序特征以确定应用程序的请求是否是长轮询请求。例如,由客户端或应用程序启动的请求可以被确定为长轮询请求,如果响应延迟间隔(D)或平均响应延迟间隔(例如,平均x个请求或任何数目的延迟间隔时间平均x时间量)大于一个阈值。

[0223] 阈值可以使用由其他客户端,例如,请求/响应跟踪引擎238b和/或应用程序配置文件生成器239(例如,响应延迟间隔跟踪器239a),所产生的请求的响应延迟间隔时间所确定。其他客户端可能同一移动设备上,并通过在移动设备上的组件本地确定阈值。例如,所有网络的所有资源服务器的所有请求的该阈值可被确定。该阈值可以被设置为一个特定

的恒定值(例如,30秒)被用于所有的请求,或任何不具有可适用的阈值(例如,如果 $D > 30$ 秒,则检测到长轮询)的请求。

[0224] 在一些情况下,其他客户端位于不同的移动设备上并且阈值可以由代理服务器(例如,图3A-B的例子中所示的主机300的代理服务器325)确定,该代理服务器在移动设备的外部并能够通过无线网络与多个不同的移动设备进行通信,正如将由图3B进一步地描述。

[0225] 在一个实施例中,缓存策略管理器245发送轮询时间表到代理服务器(例如,图1B和图3A的例子中所示的代理服务器125或325),并被由代理服务器在监测内容源中使用,例如,用于改变的或新的内容(不同于缓存的响应的更新的响应,缓存的响应与请求或应用程序相关联)。发送到代理的轮询时间表可以包括多个时序参数,包括但不限于间隔(从请求1请求2的时间)或超时间隔(例如,在长轮询中使用的等待响应的时间)。参照图17A-B的例子中所示的请求/响应时序的时序图,间隔 R_I 、 D 、 R_T 和/或 I_T ,或上述值的一些统计操作(例如,平均值,标准偏差等)可全部或部分地被发送到代理服务器。

[0226] 例如,当本地代理275检测到到长轮询的情况下,请求/响应时序中的不同的间隔(例如, D 、 R_T 和/或 I_T)可以被发送到代理服务器325以轮询内容源(如应用程序服务器/内容主机110)中使用。本地代理275还可以识别至代理服务器325,被监测的给定的应用程序或请求是长轮询请求(例如,指示代理服务器设置“长轮询标志”)。此外,代理服务器使用不同的间隔以确定何时发送代表移动设备的保活指令。

[0227] 缓存策略管理器245的本地缓存无效器244可以无效缓存在本地缓存(例如,缓存185或285)中的缓存元素,当从应用程序服务器/内容源检测到给定请求的新的或更改的数据(例如,更新的响应)。基于从代理服务器(例如,代理325或主机服务器300)收到的一份通知,缓存的响应可为出站请求被确定为无效。供到响应至移动客户端的请求的源可以被监测以确定存储在移动设备250的缓存中的请求中的缓存的响应的相关性。例如,当缓存的响应对于给定的请求或给定的应用程序不再有效时,缓存无效器244可以进一步从移动设备的缓存中移除/删除缓存的响应。

[0228] 在一个实施例中,在缓存的响应再次被提供给应用程序后,缓存的响应被从缓存中移除,该应用程序在确定缓存的响应不再有效后生成出站请求。缓存的响应在无需等待间隔的情况下可以被再次提供,或在等待间隔(例如,确定为特定的模拟长轮询中的响应延迟的间隔)之后被再次提供。在一个实施例中,间隔是响应延迟“ D ”或响应延迟“ D ”的两个或多个值的平均值。

[0229] 新的或更改的数据可以,例如,由代理服务器(例如,图1B和图3A的例子中的代理服务器125或325)检测。当给定请求/轮询的缓存条目失效时,移动设备250上的无线电的使用可以启用(例如,由本地代理275或者缓存策略管理器245)以满足后续的轮询请求,参考图4B的交互图表将进一步说明。

[0230] 缓存策略管理器245的一个实施例包括缓存或连接选择引擎249,缓存或连接选择引擎249可以决定是否使用本地缓存的条目以满足轮询/内容在移动设备250处由应用程序或部件产生的请求。例如,本地代理275或缓存策略管理器245可以拦截移动设备上的应用程序(例如,移动应用程序)作出的轮询请求以联系应用程序服务器/内容提供商。选择引擎249可以确定拦截的请求接收到的内容是否已被本地存储为缓存元素以用于决定移动设备

的无线电是否需要被激活以满足由应用程序(例如,移动应用程序)作出的请求,也确定缓存的响应在使用缓存的响应响应于出站的请求之前是否仍然对出站请求是有效的。

[0231] 在一个实施例中,响应于确定相关的缓存的内容存在并仍然有效,本地代理275中可以从本地缓存中检索缓存的元素以提供响应给应用程序(例如,移动应用程序),该应用程序作出轮询请求以使得移动设备的无线电不被激活以提供响应至应用程序(例如,移动设备应用程序)。在一般情况下,每次收到出站请求本地代理275继续提供缓存的响应直到检测到不同于缓存的响应的更新的响应。

[0232] 当确定缓存的响应已不再有效时,对于给定的请求的新的请求被通过无线网络发送以获取更新的响应。该请求可以被发送到应用程序服务器/内容提供商(例如,服务器/主机110)或主机服务器上的代理服务器(例如,主机300上的代理325)的以获取新的和更新的响应。在一个实施例中,在缓存的响应从移动设备上的缓存移除之前,如果新的响应在间隔内未被接收,可以再次提供缓存的响应作为出站请求的响应。

[0233] 图2C是一框图,展示了图2A的例子中所示的分布式代理系统中的客户端上的本地代理275中的应用程序行为检测器236以及缓存策略管理器245中的另一组件。例如,图示的应用程序行为检测器236和缓存策略管理器245可以使本地代理275检测缓存失败并由旨在战胜缓存的标识符编址的缓存的内容进行执行。

[0234] 在一个实施例中,缓存策略管理器245包括缓存失败解析引擎221、标识符规化器211、缓存适当性决策引擎246、轮询时间表生成器247、应用程序协议模块248、具有缓存查询模块229的缓存或连接选择引擎249和/或本地缓存无效器244。缓存失败解析引擎221还可以包括模式提取模块222和/或缓存失败参数检测器223。缓存失败参数检测器223还可以包括随机参数检测器224和/或时间/日期参数检测器226。一个实施例中还包括耦合到决策引擎246的应用程序缓存策略存储库243。

[0235] 在一个实施例中,应用程序行为检测器236包括模式检测器237、轮询间隔检测器238、应用程序配置文件生成器239和/或优先级引擎241。模式检测器237还可以包括缓存失败参数检测器223,该缓存失败参数检测器223还具有,例如,随机参数检测器233和/或时间/日期参数检测器234。一个实施例中还包括耦合至应用程序配置文件生成器239的应用程序配置文件存储库242。应用程序配置文件生成器239,已被描述的优先级引擎241与图2A的例子中的应用程序行为检测器236的描述相关联。

[0236] 缓存失败解析引擎221能够检测、识别、跟踪、管理和/或监测内容或内容源(例如,服务器或主机),其采用标识符和/或由标识符(例如,资源标识符,如URL和/或URI)与一个或多个击败缓存或旨在击败缓存的机制进行编址。例如,缓存失败解析引擎221可以从由应用程序或客户端产生的给定的数据请求检测,该应用程序或客户端标识失败或潜在的失败缓存,其中数据请求以不同的方式编址来自可缓存的主机或服务器(例如,应用程序检测服务器/内容主机110或310)的内容或响应。

[0237] 在一个实施例中,缓存失败解析引擎221使用在移动设备250检测到的数据请求的标识符检测或识别由内容源(例如,应用程序服务器/内容主机110或310)所使用的缓存失败机制。缓存失败解析引擎221可以检测或识别表明缓存失败机制被使用的标识符中的参数。例如,参数的格式、语法或模式可用于标识缓存失败(例如,由模式提取模块222确定或提取的模式、格式或语法)。

[0238] 模式提取模块222可以解析标识符为多个参数或组件并在每个参数上执行匹配算法,以识别任何匹配一个或多个预定的格式(例如,日期和/或时间的格式。例如,来自标识符的匹配或解析出的参数的结果可用于(例如,由缓存失败参数检测器223)以确定包括一个或多个改变参数的缓存击败参数。

[0239] 在一个实施例中,缓存失败参数检测器223可以检测随机参数(例如,由随机参数检测器224)和/或时间和/或日期参数,其典型地用于缓存失败。缓存失败参数检测器223使用这些参数通常采用的格式并执行匹配算法和测试的模式可以检测随机参数(例如,图7中所示的参数752中所示出的)和/或时间/日期。

[0240] 除了检测模式、格式和/或语法,缓存失败参数检测器223进一步确定或确认给定参数是否击败缓存以及和被编址的内容是否可以被分布式缓存系统缓存。缓存失败参数检测器223可以通过分析收到的由给定的数据请求使用的标识符的响应检测到这种动作。在一般情况下,当响应对应的多个数据请求相同时,即使当多个数据请求使用具有变化的参数的标识符因多个数据请求的每个而不同,该标识符中的变化的参数被标识为指明缓存失败。例如,请求/响应对展示了接收到的响应相同的,即使资源标识符包括随每个请求而变化的参数。

[0241] 例如,至少有两个相同的响应可能被要求识别变化参数表明为指示缓存失败。在某些情况下,可能需要至少三个相同的响应。相同响应的数量的要求可以是应用程序特定、上下文相关和/或用户依赖的/用户指定的,或上述的组合,该相同响应需要确定具有在请求之间的变化值的给定参数是缓存击败。这样的要求也可以由分布式缓存系统静态地或动态地调整以满足某些性能阈值和/或显式/隐式关于用户体验(例如,用户或应用程序是否正在接收响应于请求的相关的/新鲜的内容)的反馈。如果应用程序由于响应缓存开始出现故障和/或如果用户表示不满(显式用户反馈)或系统检测到用户受挫(隐式用户提示),可能需要更多的相同的响应以确认缓存失败,或为系统处理给定的参数用于缓存失败。

[0242] 缓存适当性决策引擎246可以检测、评估或确定是否来自移动设备250进行交互的内容源(例如,图1B的例子中的应用程序服务器/内容提供商110)的内容是否具有可适合于缓存的内容。在某些情况下,基于一套标准(例如,指定被从内容源请求的内容的时间关键性的标准),来自给定应用程序服务器/内容提供商(例如,图1B的服务器/提供商110)的内容被确定为适合于缓存。在一个实施例中,本地代理(例如,图1B和图2A的本地代理175或275)应用选择标准存储来自主机服务器的内容,该主机服务器由应用程序请求作为在移动设备上的本地缓存中的缓存的元素以满足由应用程序作出的后续请求。

[0243] 选择标准也可以包括,通过举例的方式但不限于,指示移动设备是否是活跃或不活跃的移动设备的状态、网络条件和/或无线电覆盖范围的统计信息。缓存适当性决策引擎246可采用标准的任何一个或任何组合,并以任何顺序,以识别缓存可能适合的源。

[0244] 一旦应用程序服务器/内容提供商具有被识别的或被检测到的可能适合在移动设备250上本地缓存的内容,缓存策略管理器245通过存储从内容源接收到的内容为在移动设备250上的本地缓存中的缓存元素(例如,分别在图1B和图2A的例子中显示的本地缓存185或285)可以缓存从所识别的源接收到的相关联的内容。内容源也可以被识别至远离移动设备250并与移动设备250无线通信的代理服务器(例如,分别在图1B和图3A的例子中显示的代理服务器125或325)以使得代理服务器可以为新的或变更的数据监测内容源(例如,应用

程序服务器/内容提供商110)。类似地,本地代理(例如,分别在图1B和图2A中的本地代理175或275)可以识别到代理服务器,从特定应用程序服务器/内容提供商接收的内容被存储为本地缓存中的缓存的元素。

[0245] 在一个实施例中,缓存的元素被存储在本地缓冲285中作为与标识符的归一化版本相关联以用于标识符采用一个或多个只在击败缓存的参数。标识符可以由标识符规范化模块211标准化以及规范化过程可以包括,通过举例的方式,一个或多个:转化URI方案和主机为小写,大写在百分比转义序列中的字母,移除默认的端口并删除重复的斜线。

[0246] 在另一个实施例中,标识符取可以通过移除缓存失败的参数和/或通过更换参数为静态值进行标准化,该静态值可用于编址接收到的缓存的响应或与接收到的缓存的响应相关联,该接收到的缓存的响应由规范化器211或缓存失败参数处理器212利用标识符响应于请求。例如,存储在本地缓冲285(图2A中所示)中的缓存的元素可使用标识符的归一化版本或标识符的归一化版本的哈希值进行识别。该标准化的标识符或标识符的哈希值可由哈希引擎213产生。

[0247] 一旦内容已经在本地缓存,缓存策略管理器245可以,基于接收未来轮询请求以与内容服务器联系,检索来自本地缓存中的缓存的元素以响应在移动设备250作出的轮询请求以使得未激活的移动设备的无线电维护轮询请求。这样的通过本地缓存条目本地地维护和实现应用程序(例如,移动应用程序)请求允许更有效的资源和移动网络流量利用率和管理的,因为网络带宽和其他资源不需要被用于请求/接收轮询响应,该轮询响应可能没有从已经在移动设备250接收到的响应改变。

[0248] 缓存策略管理器245的一个实施例包括轮询时间表生成器247,该轮询时间表生成器247可以为在移动设备250上的一个或多个应用程序产生轮询时间表。轮询时间表可以指定由代理服务器(例如,图1B和图3A的例子中所示的代理服务器125或325)在监测一个或多个应用程序的内容源中采用的轮询间隔。轮询时间表可以被确定,例如,基于定向至来自移动设备的内容源的轮询请求之间的间隔。在一个实施例中,应用程序行为检测器的轮询间隔检测器238可以监测定向至来自移动设备250的内容源的轮询请求,以确定的任何或所有的应用程序(例如,移动应用程序)作出的轮询请求之间的间隔。

[0249] 在一个实施例中,缓存策略管理器245发送轮询时间表发送到代理服务器(例如,图1B和图3A的例子中所示的代理服务器125或325)并可被由代理服务器在监测的内容源中,例如,为改变的或新的内容所使用。当从应用程序服务器/内容源检测到给定请求的新的或变更的数据时,缓存策略管理器245的本地缓存无效器244可以无效在本地缓存(如缓存185或285)中缓存的元素。新的或更改的数据,例如,可以被代理服务器检测。当给定请求/轮询的缓存条目在无效后被失效和/或移除(例如,从缓存中删除)时,移动设备250上的无线电的使用可以被启用(例如,由本地代理或策略管理器245)以满足随后的轮询请求,参照图4B的交互图表将进一步描述。

[0250] 在另一个实施例中,代理服务器(例如,图1B和图3A的例子中所示的代理服务器125或325)使用在数据请求中使用的资源标识符的修改后的版本用于为新的或更改的数据监测给定的内容源(数据请求被编址至的图1A和图1B中的应用程序服务器/内容主机110)。例如,在该实施例中,在其中内容源或标识符被检测到采用缓存击败机制的情况下,修改(例如,标准化)后的标识符可以用来代替轮询的内容源。标识符的修改或归一化版本可以

通过缓存策略管理器245传送到代理服务器,或更具体地,标识符规化器211的缓存失败参数处理器212。

[0251] 由代理服务器所使用的用于轮询代表移动设备/应用程序(例如,移动应用程序)的内容源的修改后的标识符可能或可能不会与标准化的标识符相同。例如,标准化的标识符可以是具有被移除的改变缓存参数的原始标识符,然而修改后的标识符使用替代参数来替代被用于击败缓存的参数(例如,由本地代理和/或代理服务器已知的静态值或其他预定值替代改变的参数)。修改后的参数可以由本地代理275确定并传送到代理服务器。修改后的参数也可由代理服务器(例如,由图3C的例子中所示的标识符修改器模块353)产生。

[0252] 缓存策略管理器245的一个实施例包括缓存或连接选择引擎249,该引擎可以决定是否使用本地缓存条目以满足在移动设备250由应用程序或部件产生的轮询/内容请求。例如,本地代理275或缓存策略管理器245可拦截由移动设备上的应用程序(例如,移动应用程序)作出的轮询请求以与应用程序服务器/内容提供商联系。选择引擎249可以确定拦截请求的已接收到的内容是否被本地存储作为缓存元素以用于决定移动设备的无线电是否需要被激活来满足由应用程序(例如,移动应用程序)作出的请求。在一个实施例中,响应于确定相关的缓存的内容存在并仍然是有效的,本地代理275可以从本地缓存检索缓存的元素以提供到应用程序(例如,移动应用程序)的响应,该应用程序作出轮询请求以使得移动设备的无线电没有被激活以提供到应用程序(例如,移动应用程序)的响应。

[0253] 在一个实施例中,存储在本地缓存285(图2A中所示)中的缓存的元素可以使用标识符的归一化版本或标识符的归一化版本的哈希值被识别,例如,使用缓存查询模块229。缓存的元素可以与标准化的标识符一起存储,该标识符具有被移除的或以其他方式被替代的缓存击败参数以使得相关的缓存的元素可以在未来被识别并检索以满足采用相同类型的缓存击败的其他请求。例如,当在后续请求中使用的标识符被确定为采用相同的缓存击败参数时,该标识符的归一化版本可以被生成的并用于识别存储在移动设备缓存中的缓存的响应以满足数据请求。标识符或标准化的标识符的哈希值可由标识符规化器211的哈希引擎213产生。

[0254] 附图2D描绘了图2A中显示的本地代理275中的额外部件,进一步能够执行移动流量分类和基于应用程序行为和/或用户活动的策略实施。

[0255] 本地代理275的一个实施方式中,用户活动模块215,进一步包括一个或更多的用户活动检测器/追踪器215a、用户活动预测引擎215b和/或用户预期管理器215c。应用程序行为检测器236还可以包括优先级处理引擎241A、时间关键性检测引擎241B、应用程序状态分类器241C和/或应用程序流量分类器241D。本地代理275可以进一步包括背光检测器219和/或网络配置选择引擎251。所述网络配置选择引擎251进一步包括一个或多个无线生成标准选择器251a、一个数据速率符251b、接入信道选择引擎251c和/或一个访问点选择器。

[0256] 在一个实施例中,应用程序行为检测器236可检测、确定、识别或推断移动设备250上的一个应用程序的活动状态,流量源自或被导向(例如经由应用程序状态分类器241C和/或应用程序流量分类器241D)该移动设备。活动状态可以通过移动设备(通过应用程序状态分类器241C)上的应用程序在前景状态还是在背景状态来确定,因为前台应用程序的流量与后台应用程序的流量可通过不同的方式处理。

[0257] 在一个实施例中,活动状态可以按启发式的可靠性水平确定、检测、识别或推断,

其基于移动设备250的背光状态(例如通过背光检测器219)或者在移动设备上的其他软件代理或硬件传感器,包括但不限于电阻传感器、电容传感器、环境光传感器、运动传感器、触摸传感器等。在一般情况下,如果背光开启,流量可以被视为正在或被确定从活动的或处于前台中的应用程序产生,或者该流量是交互式的。此外,如果背光开启,流量可以被视为或被确定来自用户交互或用户活动的流量,或包含用户预期在某个时间段内数据的流量。

[0258] 在一个实施例中,活动状态被基于流量是否是交互式流量或维护流量来确定。交互式流量可以包括来自从用户的应用程序活动/交互直接产生之响应和请求的事务,并且可以包括用户正在等待或预期接收的内容或数据。维护流量可被用于支持不直接由用户检测到的应用程序的功能。维护流量还包括可能发生响应用户动作的动作或事件,但用户不是主动等待或预期响应。

[0259] 例如,在移动设备250上的邮件或消息删除操作生成一个在服务器上删除相应邮件或消息的请求,但用户通常不等待响应。因此,这样的请求可以被分类为维护流量或具有较低优先级(例如通过优先级处理引擎241a)的流量和/或非时间关键性(例如通过时间关键性检测引擎214b)的流量。

[0260] 与此相反,由用户在移动设备250上发起的邮件“读取”或消息“读取”请求可以被归类为“交互流量”,因为用户请求读取消息或邮件时通常在等待访问内容或数据。类似地,这样的请求可以被归类为具有较高优先级(例如通过优先级处理引擎241a)和/或时间关键/时间敏感(例如通过时间关键性检测引擎241b)。

[0261] 时间关键性检测引擎241B可大致确定、识别、推断包含在发送自移动设备250或发送至主机服务器(例如主机300)或应用程序服务器(例如应用程序服务器/内容源110)的移动设备流量中数据的时间灵敏度。例如,时间敏感数据可以包括状态更新、股票信息更新、即时通讯在线状态信息、电子邮件或其他消息、手机游戏应用程序产生的行动、网页请求、位置更新等。根据内容或请求性质,非时间敏感或时间关键性的数据可以包括删除消息请求、标记为已读或已编辑的操作、添加好友或删除好友请求等特定于应用程序的操作、某些类型的消息或或其他性质不经常变化的信息等等。在某些情况下,当数据不是时间关键性时,其允许流量通过的时机是基于何时需要从移动设备250发送附加数据而设定的。例如,流量整形引擎255可以把流量与一个或多个后续事务调整,使其被一起在移动设备无线(例如使用调整模块256和/或批处理模块257)的单次启动事件中发送。调整模块256也可以把被导向同一主机服务器的时间接近的轮询请求进行调整,因为这些请求可能用相同的数据响应。

[0262] 活动状态的替代或组合可以从评价、确定、评估、推断、识别移动设备250上的用户活动(例如通过用户活动模块215)确定。例如,用户活动可以使用用户活动跟踪器215a直接检测和跟踪。然后,由此产生的流量可以被适当归类,从而用于后续处理以确定处理策略。此外,用户活动可以由用户活动预测引擎215b来预测或预期。通过预测用户活动或预期用户活动,预测后由此而产生的流量可以被视为是用户活动产生的结果,并可被适当分类以确定传输策略。

[0263] 此外,用户活动模块215还可以管理用户预期(例如通过用户预期管理器215c和/或结合使用活动跟踪器215a和/或预测引擎215b),以确保流量被适当归类,从而大致满足用户预期。例如,用户启动操作应被分析(例如通过预期管理器215c)以确定或推断用户是

否将等待响应。如果是这样,这种流量应按能够使用户在接收这种响应或操作时不会经历令人不愉快延迟的策略来处理。

[0264] 在一个实施例中,选择新一代的无线标准网络,以供基于流量源自或被导向至的移动设备上应用程序的活动状态在无线网络中在移动设备和主机服务器之间发送流量。可以选择3G、3.5G、3G+、4G或LTE网络等高级技术标准,用于处理用户交互、用户活动所产生的流量,或含有用户预期或等待数据的流量。对于响应前台活动的移动设备,也可选择高级无线标准网络用于传输包含在移动设备流量中的数据。

[0265] 在分类流量和确定移动通信传输策略时,可以选择网络配置在(例如通过网络配置选择引擎251)移动设备250上使用,用于在移动设备和代理服务器(325)和/或应用程序服务器(例如应用程序服务器/主机110)之间发送流量。所选择的网络配置可以基于由应用程序行为模块236收集的应用程序活动状态(例如后台或前台的流量)、应用程序流量类别(例如交互或维护流量)、数据/内容的任何优先级、时间灵敏度/关键性方面的信息确定。

[0266] 所述网络配置选择引擎2510可选择或者指定一个或多个生成标准(例如通过无线生成标准选择器251a)、一个数据比率(例如通过数据比率区分符251b)、一个访问通道(例如访问通道选择引擎251c)和/或一个任何结合的访问点(例如通过访问点选择器251d)。

[0267] 例如,可选择或者指定更先进的一代(例如3G、LTE或者4G或更后)用于流量,当活跃状态与使用者互相影响或者在移动设备最显著的位置。相反,一种较旧的标准(例如2G、2.5G或者3G或者更老的)可指定用于流量,当检测到一种或多种情况,应用程序与使用者之间没有相互影响,应用程序位于移动设备的背景位置,或者流量中包含的数据不是时间关键的或者另外确定具有较低优先级。

[0268] 同样地,一个具有较慢数据速率的网络配置可被指定用于流量,当检测到一种或多种情况,应用程序与使用者之间没有相互影响,应用程序位于移动设备的背景位置,或者流量中包含的数据不是时间关键的。所述访问通道(例如提出接入信道或者专用信道)可被指定。

[0269] 图3A系一框图,展示了分布式代理和缓存系统的服务器端组件,放置于主机服务器300,管理无线网络的流量,以节约资源。

[0270] 主机服务器300一般包含,例如,网络接口308和/或一个或多个存储库312、314和316。注意:该服务器300可以为任何便携/移动或非便携设备、服务器、计算机集群和/或(例如图24示例中所示任何机器)能通过网络,包含任何有线或无线网络(例如,WiFi、蜂窝、蓝牙等),用于接收或传输信号满足数据请求的其他类型处理单元。

[0271] 网络接口308可以包含网络模块或设备,能使服务器通过任何由主机和外部实体所支持的已知和/或方便通信协议,与主机服务器300外部实体调解网络数据。具体而言,网络接口308允许服务器300与多台设备进行通信,含移动电话设备350,和/或一个或多个应用程序服务器/内容提供商310。

[0272] 主机服务器300可以将设备的连接信息(例如,网络特征、条件、连接类型等)存储至连接元数据存储库312。此外,第三方应用程序或内容提供商的相关信息也可以存储在存储库312中。主机服务器300可以存储设备信息(例如,硬件性能、属性、设备设置、设备语言、网络性能、制造商、设备模型、操作系统、操作系统版本等)至设备信息存储库314。此外,主机服务器300可以将网络提供商和各类网络服务区域的信息存储至网络服务提供商存储库

316。

[0273] 具备通信功能的网络接口308允许与设备350进行即时连接(如包含蜂窝连接)和/或与内容服务器/提供商310的连接(例如,包含有线/无线,HTTP,互联网连接,网卡,WIFI等),为最优化网络资源利用率和/或节约服务设备350消耗的电源(电池),管理设备350和内容提供商310的流量。主机服务器300能与由不同的网络服务提供商和/或在相同/不同的网络服务区域内提供服务的移动设备350进行通信。主机服务器300能操作,并且兼容各类型或各整体性能水平的设备350,包含如下列举但不限于:1G、2G、2G过渡(2.5G,2.75G)、3G(IMT-2000)、3G过渡(3.5G,3.75G,3.9G),4G(IMT-Advanced)等。

[0274] 一般来说,网络接口308可以包含一个或多个网络适配卡、无线网络接口卡(例如:SMS接口、WiFi接口、各代移动通信标准接口,包含但不限于:1G,2G,3G,3.5G,4G类型网络,例如LTE、WiMAX等)、蓝牙、WiFi或任何其他是否经由路由器,接入点、无线路由器、交换机、多层交换机、协议转换器、网关、网桥、桥接路由器、枢纽、数字媒体接收方,和/或中继器连接的网络。

[0275] 主机服务器300还包含,包含代理服务器325和服务器缓存335的分布式代理和缓存系统服务器端组件。一个实施例中,代理服务器325可以包含HTTP访问引擎345,缓存策略管理器355,代理控制器365,流量整形引擎375,新数据检测器347,和/或连接管理器395。

[0276] HTTP访问引擎345还能包含心跳管理器398;代理控制器365还能包含数据无效模块368;流量整形引擎375还能包含控制协议376和批处理模块377。代理服务器325和各所述组件能增加或减少包含的组件/模块/引擎。

[0277] 本申请所使用的“模块”,“管理器”,“处理程序”,“检测器”,“接口”,“控制器”、“规化器”、“生成器”、“无效器”或“引擎”包含一般用途,专用或共享处理器,和通常由处理器执行的固件或软件模块。基于具体实施或其他因素的考量,该模块、管理器、处理器、检测器、接口、控制器、规化器、生成器、无效器或引擎,可集中或依功能分布式实施。该模块、管理器、处理器、检测器、接口、控制器、规化器、生成器、无效器或引擎能包含嵌入在计算机可读(存储)介质,具有一般或特殊用途的硬件,固件或软件,由处理器予以执行。本申请所使用的计算机可读介质或计算机可读存储介质,包含所有法定介质(例如,美国专利法第101条),并明确排除所有非法定的介质,本质上说排除必要范围,使得包含计算机可读(存储)介质的权利要求能成立。已知的法定计算机可读介质包含硬件(例如:寄存器,随机存储器(RAM)、非易失(NV)存储器,仅举几例),但可能或可能不限于硬件。

[0278] 示例中设备(如移动设备350)向应用程序服务器或内容提供商310提出应用程序或内容请求,该请求可能被截获,并路由到代理服务器325,其耦合至设备350和应用程序服务器/内容提供商310。具体来说,代理服务器能够与移动设备350的本地代理进行通信(例如图1和图2示例中各自的代理服务器175和275),本地代理将数据请求转发到代理服务器325,某些实例中,如需要进一步处理,为响应数据请求将其发送至应用程序服务器/内容服务器310。

[0279] 在这样的配置下,主机300或主机服务器300的代理服务器325可以利用本地代理提供的智能信息,调整其与设备的通信方式,以最优化网络使用和设备资源。例如,代理服务器325可以识别设备350的用户活动特征,修改其通信频率。用户活动特征可以由,例如,代理控制器365的活动/行为感知模块366,通过设备350上本地代理所收集的信息进行确

定。

[0280] 一个实施例中,代理服务器325的连接管理器395可以控制通信频率,例如,调整设备350内容的推送或更新频率。例如,当用户活动特征表明用户处于不活跃状态时,连接管理器395可以降低推送频率。在一个实施例中,当用户活动特征表明用户处于闲置一段时间后的活跃状态,连接管理器395能调整与设备350的通信频率,将因降低通信频率的缓冲数据发送至设备350。

[0281] 此外,代理服务器325包含感知优先级的各项请求、事件、时域、应用程序,和/或特定事件。这种感知由设备350的本地代理确定,并提供给代理服务器325。代理服务器325的优先级感知模块367一般能评估各事件或应用程序的优先级(例如,包含时间的紧迫性,时间敏感性等);此外,优先级感知模块367能跟踪设备350的本地代理所确定的优先级。

[0282] 一个实施例中,连接管理器395通过优先级感知还能修改服务器300与设备350的通信频率(例如,使用或无线电由无线电控制器396控制)。例如,当需要发送满足一定重要性/优先级标准的数据或更新时,若未使用无线电,服务器300可以通知设备350,请求使用无线电。

[0283] 一个实施例中,代理服务器325能检测事件的重复(如从服务器/提供商310接收的事件、内容或数据),并允许事件累积,批量传输至设备350。基于模块367和/或模块366所跟踪的优先级感知和/或用户活动/应用程序行为感知,批量传输可以累积,并且能延迟传输事件。例如,当服务器300上检测到高优先级事件(满足阈值或标准)时,批处理模块377可以启动批量传输多个事件至设备350(低优先级)。此外,当服务器从设备350接收数据,表明已在使用无线电设备,这能触发服务器300的批量传输。一个实施例中,代理服务器325可以基于事件/事件的优先级对传输的批量信息/数据包进行排序,高优先级内容优先发送,防止连接中断或电池耗尽等。

[0284] 一个实施例中,服务器300缓存数据(例如,由缓存策略管理器355管理),能通过网络(例如,蜂窝网络)修改(如降低)其与设备350的通信频率。数据能缓存,例如至服务器缓存335,为后续检索或批量发送至设备350减少设备350打开无线电的需要。服务器缓存335可以部分或者全部放置于主机服务器300内,尽管在图3A示例中,其置于主机服务器300外部。某些实例中,服务器缓存335可能相同于和/或部分或全部统一于另一缓存,由另一实体进行管理(例如图1B所示的可选缓存代理服务器199),如被应用程序服务器/内容提供商310、网络服务提供者,或其他第三方进行管理。

[0285] 在一个实施例中,内容缓存由设备350在主机服务器300的支持下本地执行。例如,主机服务器300的代理服务器325向应用程序服务器/提供商310查询请求,监测响应的变化。当检测到更改或新响应时(如通过新数据检测器347),代理服务器325通知移动设备350,因此设备350上本地代理可以确定本地缓存存储作为响应的相关缓存条目无效(如标记为陈旧)。另外,数据无效模块368能基于从应用程序服务器/提供商310接收的响应,自动指示设备350的本地代理使某些缓存数据失效。当新的内容被从内容服务器310接收时,缓存的数据被标记为无效,也可以被替换或删除。

[0286] 应注意的是,检测器347能通过一种或多种方式检测到数据更改。例如,服务器/提供商310能将更改通知主机服务器300。为响应源服务器/提供商310的直接检索,主机服务器300也能检测到更改。某些实例中,代理服务器325可以另外,预加载设备350上本地缓存

的新/更新数据。当主机服务器300检测到移动设备上的无线电已在使用时,或服务器300有附加内容/数据发送至设备350,可以执行此操作。

[0287] 一或多个上述机制可以同时被执行或基于应用程序(例如不同服务器/提供商310的不同策略)进行调整/配置。某些实例中,源提供商/服务器310可通知主机300某些类型的事件(如满足优先级阈值级别的事件)。此外,提供商/服务器310被配置为不论事件优先级,均在特定间隔期通知主机300。

[0288] 一个实施例中,主机300的代理服务器325可以监测/跟踪从内容源接收的数据请求的响应,在将结果返回至移动设备之前更改结果,当内容源的数据请求所产生的结果与返回到移动设备的结果一致时,这种监测才适合,从而在特定的请求不存在新/更新时,防止使用网络/功耗。设备350的本地代理能指示代理服务器325执行这种监测或代理服务器325对于一特定请求,接收到几个相同的响应时,能自动启动这一过程(或在一段时间内几个相同的响应)。

[0289] 一个实例中,服务器300,例如,通过活动/行为感知模块366,能够在独立于移动设备350的设备中识别或检测用户活动。例如,模块366可以检测被访问的用户收件箱(例如电子邮件或收件箱类型)。这可以表明,用户利用设备,而非移动设备350与他/她的应用程序进行互动,可能根本无需频繁的更新。

[0290] 在本实例中,服务器300降低将新或更新后内容发送至移动设备350的频率,或只要检测到用户使用另一台设备访问,消除所有通信。降低具体应用程序的频率(例如,用户在另一台设备上与应用程序互动)或普遍降低与移动设备350的频率(例如,因为用户被检测到通过另一台设备与服务器或应用程序进行互动,他/她同时利用该设备访问其他服务)。

[0291] 在一个实施例中,主机服务器300能够代表设备350轮询内容源310,以节约电能或设备350的电池消耗。例如,移动设备350上的某些应用程序可以预定的方式反复轮询其各自服务器310。代理控制器365的活动/行为模块366能够跟踪其重复或其他类型的应用程序行为。主机服务器300从而可以轮询移动设备350上应用程序的内容源,要不然设备350通过无线(如包含蜂窝连接)也能执行。主机服务器通过HTTP访问引擎345,创建HTTP连接或利用无线控制器396通过蜂窝网络连接到源310,轮询源310新或更改数据。当检测到新或更改数据时,新数据检测器347通知设备350,该数据可用和/或将新/更改数据提供给设备350。

[0292] 在一个实施例中,连接管理器395确定移动设备350不可用(例如无线电已关闭),并利用SMS传输内容至设备350,如通过图1B所示的SMSC。SMS是用来传输无效消息、批次无效消息,甚至在内容足够小以适应到仅有几个(通常是一个或两个)SMS消息的情况下的内容。这就避免了访问的无线信道发送的开销信息的必要。主机服务器300可以将SMS用于某些优先级高于阈值或满足其他标准的事件或响应。主机服务器300还能将SMS作为带外抑制触发器维持或唤醒IP连接,这是维持IP持续连接的替代之法。

[0293] 一个实施例中,代理服务器325的连接管理器395(如心跳管理器398)能够生成和/或代表连接的设备350传输心跳信息,为设备350上运行的应用程序维持与提供商310的后台连接。

[0294] 例如,在分布式代理系统中,设备350的本地缓存可以防止通过蜂窝网络或其他网络发送任何或全部维持应用程序必要TCP/IP连接所需心跳信息,却反而依赖主机服务器

300的代理服务器325生成和/或发送心跳信息,以维持与后台的连接(例如图1A所示应用程序服务器/提供商110)。代理服务器能够独立于移动设备本地代理的操作,生成保活(心跳)信息。

[0295] 存储库312、314和/或316可以另外存储软件、说明性数据、图像、系统信息、驱动程序、和/或任何由主机服务器300的其他组件和/或任何其他服务器操作所使用的的数据项目。存储库可以由数据库管理系统(DBMS)管理,例如但不限于:Oracle、DB2、Microsoft Access、Microsoft SQL Server、PostgreSQL、MySQL、FileMaker等。

[0296] 存储库可以通过面向对象技术和/或文本文件实施,由分布式数据库管理系统,面向对象的数据库管理系统(OODBMS)(例如:概念库、FastDB主存数据库管理系统、JDOInstruments、ObjectDB等),对象-关系数据库管理系统(ORDBMS)(例如:Informix、OpenLink Virtuoso、VMDS等),文件系统,和/或任何其他方便或已知的数据库管理程序包进行管理。

[0297] 图3B系一框图,展示了如图3A的例子中所示的缓存系统的缓存策略管理器中的组件的进一步的示例,其能够为移动应用程序(例如,移动应用程序)行为和/或网络条件缓存并适配缓存策略。

[0298] 在一个实施例中,缓存策略管理器355可以进一步包括元数据生成器303、缓存查找引擎305、应用协议模块356、具有轮询时间表管理器358的内容源监测引擎357、响应分析器361和/或更新的或新的内容检测器359。在一个实施例中,轮询时间表管理器358还包括主机时序模拟器358a、长轮询请求检测器/管理器358b、时间表更新引擎358c和/或时间调节引擎358d。为修改或增补,元数据生成器303和/或缓存查找引擎305可以耦合到缓存335(或服务器缓存)以缓存条目或其查询。

[0299] 在一个实施例中,代理服务器(例如,图1B和图3A的例子的代理服务器125或325)可以通过监测引擎357监测新的或更改的数据的内容源。如图所示的代理服务器是图2A-B的移动设备250的外部的实体。内容源(例如,图1B的应用程序服务器/内容提供商110)可以是已被识别至代理服务器(例如,由本地代理)作为具有被本地缓存在移动设备(例如,移动设备150或250)上的内容的一种。例如,内容源可以由监测引擎357以一频率进行监测,该频率是基于移动设备的内容源的轮询频率。例如,轮询时间表可以由本地代理产生并发送到代理服务器。轮询频率可以被轮询时间表管理器358跟踪和/或管理。

[0300] 例如,代理服务器可以轮询代表移动设备的主机(例如,内容提供商/应用程序服务器)和通过主机定时模拟器358a模拟至主机的客户端行为。轮询行为可以被模拟以包括在主机持久连接(例如,通过长轮询请求检测器/管理器358b)中经历的长轮询请求-响应序列的特征。请注意,一旦轮询间隔/行为被设定,设备端上的本地代理275和/或服务器端上的代理服务器325可以验证应用程序和应用程序服务器/内容主机行为是否匹配或是否可以被这种预测模式所表示。在一般情况下,本地代理和/或代理服务器可以检测出偏差,并且在适当的时候,重新评估和计算、确定或估计另一项轮询间隔。

[0301] 在一个实施例中,分布式代理的服务器端的缓存策略管理器355可以配合或独立于移动设备上的代理服务器275识别或检测长轮询请求。例如,缓存策略管理器355可以确定阈值,该阈值被用于与在请求-响应序列中的应用程序请求的响应延迟间隔时间进行比较以识别或检测长轮询请求、可能的长轮询请求(例如,用于与客户端通信的主机的永久连

接的请求,包括但不限于,长期持有的HTTP请求、能够以COMET式推送的持久连接、HTTP流的请求,等等)或其他可以以不同方式被作为长轮询请求处理的请求。

[0302] 例如,该阈值可以由代理服务器325使用由移动设备的客户端/应用程序生成的请求的响应延迟间隔时间进行确定,该移动设备由多个不同的蜂窝或无线网络提供服务。由于代理325位于主机300上能够通过多个网络与多个移动设备通信,缓存策略管理器355在全球层面上访问应用程序/客户端信息,其可用于设置阈值以分类和检测长轮询中。

[0303] 通过经由不同或相同的网络跟踪设备的应用程序的响应延迟间隔时间,缓存策略管理器355可以设置一个或多个阈值以用于与长轮询检测的响应延迟间隔时间的比较。由代理服务器325设置的阈值可以是静态或动态的,并且可以与条件和/或生存时间(相对或绝对的过期时间/日期)相关联。

[0304] 另外,代理服务器325的缓存策略管理器355可以进一步确定该阈值,全部或部分地基于给定无线网络的网络延迟、由给定运营商(服务提供商)服务的网络,或多个无线网络。代理服务器325还可以基于一个或多个应用程序(例如,移动应用程序)或移动客户端请求定向至的应用程序服务器/内容提供商(例如,110)的延迟确定长轮询请求识别的阈值。

[0305] 代理服务器可以在被监测的内容源检测到新的或更改的数据并传输消息到移动设备通知其这样的改变以使得移动设备(或移动设备上的本地代理)可以采取适当的动作(例如,无效在本地缓存中的缓存的元素)。在某些情况下,代理服务器(例如,缓存策略管理器355)基于检测到新的或更改的数据也可以在其缓存(例如,分别在图1B和图3A的例子中的服务器缓存135或335)中存储新的或更改的数据。存储在服务器缓存335中的新的/更新的数据可以在某些情况下被使用以满足移动设备的内容请求,例如,它可以用于代理服务器通知移动设备新的/更改的内容之后以及本地缓存的内容已无效之后。

[0306] 类似于图2B的例子中所示的元数据生成器203的元数据生成器303可以在移动设备250生成为请求缓存的响应的元数据。元数据生成器303可以生成存储在服务器缓存335中的缓存条目的元数据。同样地,缓存查找引擎305可以包括相同或相似的用于描述图2B的例子中所示的缓存查找引擎205的功能。

[0307] 响应分析器361可以执行任何或所有的功能,该功能与分析为请求接收的响应相关,该请求在移动设备250以相同或类似的方式被生成至,图2B的例子中所示的本地代理的响应分析器246d。由于代理服务器325能够从定向至移动设备250的应用程序服务器/内容源310接收响应,代理服务器325(例如,响应分析器361)可以执行类似的响应分析步骤以确定如描述本地代理的响应分析器的缓存性能。响应分析程序的例子也结合图11-13的例子中所示的流程图被描述。响应可以被分析以补充或代替可以被在移动设备250上的本地代理275执行的分析。

[0308] 此外,基于如为本地代理275的时间表更新引擎描述的移动设备250的应用程序的请求间隔改变,时间表更新引擎358c可以更新给定应用程序服务器/内容主机的轮询间隔。时间调整引擎358d可以设置应用程序服务器/内容主机的轮询即将开始的初始时间,以防止在服务为在本地代理275中的时间表更新引擎所描述的新鲜内容之前再次地过时内容的服务。时间表更新和时间调整算法可连同或替代移动设备250上的本地代理275执行的类似的过程。

[0309] 图3C系一框图,展示了如图3A的例子所示的分布式代理系统中的服务器端组件上

的代理服务器375中的缓存策略管理器355中的组件的另一示例,其能够管理并检测缓存击败机制及监测内容源。

[0310] 在一个实施例中,缓存策略管理器355可以进一步包括:缓存击败源管理器352、具有轮询时间表管理器358的内容源监测引擎357,和/或更新或新的内容检测器359。缓存击败源管理器352可以进一步包括标识符修改器模块353和/或标识符模式跟踪模块354。

[0311] 在一个实施例中,代理服务器(例如,图1B和图3A的例子中的代理服务器125或325)可以通过监测引擎357的为新的或更改的数据检测内容源。内容源(如图1B的应用程序服务器/内容提供商110或图3A的310)可以是已被识别至代理服务器(例如,由本地代理)以具有在移动设备(例如,移动设备150或250)上被本地缓存的内容。例如,内容源310可以由监测引擎357以一基于移动设备的内容源的轮询频率监测。例如,轮询时间表可以由本地代理产生并发送到代理服务器325。轮询频率可以被轮询时间表管理器358跟踪和/或管理。

[0312] 在一个实施例中,代理服务器325在轮询内容源310中使用标准化的标识符或修改的标识符以检测新的或更改的数据(响应)。标准化标识符或更改的标识符也可由代理服务器325用于在服务器缓存335上存储响应。在一般情况下,标准化的或修改的标识符可以被当为缓存的内容采用击败机制时使用。缓存击败机制可以是在例如URI或者URL的标识符中的变化参数的形式,并且可以包括变化的时间/数据参数、随机变化的参数或其它类型的参数。

[0313] 标准化标识符或更改标识符删除或以其他方式取代变化参数为相关联的后续请求和相关响应的标识,也可以被用于轮询内容源。在一个实施例中,修改后的标识符由代理服务器325(分布式代理系统的服务器端组件)上的缓存策略管理器355的缓存击败源管理器352(例如,标识符修改器模块353)的生成。修改后的标识符可以利用替代参数(一般在一段时间是静止的)代替用来击败缓存的改变的参数。

[0314] 缓存击败源管理器352可选地包括标识符模式跟踪模块354以跟踪、存储和监测各种标识符的修改或为一个或多个内容源(如应用程序服务器/内容主机110或310)编址内容的标识符以不断验证由代理服务器325使用修改的标识符和/或标准化的标识符以轮询的作为预测或意图的内容源(例如,接收相同的响应或以其他方式与相比原始的、未修改的标识符相关的响应)。

[0315] 模式跟踪模块354检测导致内容源上的不稳定或不可预知行为(例如,将被发送的意想不到的响应)的标识符的修改和标准化,跟踪模块354可以登录修改并指示缓存击败源管理器352生成另一个修改/标准化或通知本地代理(例如,本地代理275)生成另一个修改/标准化以在轮询内容源中使用。在替代方案或并行方案中,来自移动设备(例如,移动设备250)的给定应用程序/客户端的请求可以暂时通过网络发送到内容源用于提供直接响应给移动设备和/或直到工作可以产生标识符的修改。

[0316] 在一个实施例中,当新的或更改的数据被检测用于已经存储在移动设备(例如,移动的移动设备250)上的本地缓存(例如,缓存285)中的响应时,响应被作为服务器缓存原始被存储在服务器缓存中。因此,移动设备或本地代理275可以连接到代理服务器325以检索请求的响应的新的或更改的数据,该请求是之前被本地地缓存在本地缓存285中(现在是无效的,过时的,或以其他方式确定为无关的)。

[0317] 代理服务器325可以在被监测的应用程序服务器/内容主机310检测新的或改变的

数据并传输消息至移动设备通知它这样的变化以使得移动设备(或在移动设备上的本地代理)可以采取适当的措施(例如,无效在本地缓存中的缓存的元素)。在某些情况下,代理服务器(例如,缓存策略管理器355),基于检测到新的或更改的数据,也可以存储新的或更改的数据在其缓存中(例如,分别在图1B和图3A的例子中的服务器缓存135或335)。存储在服务器缓存中的更新的/新的数据可以被使用,在某些情况下,以满足移动设备的内容请求,例如,它可以在代理服务器通知移动设备新的/更新的内容以及本地缓存的内容已经失效之后被使用。

[0318] 附图3D描绘了在附图3A中显示代理服务器325的额外部件,这是进一步能够在履行实施移动流量分类和基于应用行为和/或数据传输的优先级的情况下的策略实施。

[0319] 在代理服务器325的示例中,流量整形引擎375进一步耦合至流量分析器336,用于对策略定义分类移动流量,并用于指向一个或多个移动设备(例如图2A-2D中的移动设备250)或指向应用程序服务器/内容主机(例如,图1A-1B的110)的移动流量和事务的应用。通常,所述代理服务器325远离移动设备并远离主机服务器,如图1A-1B示例所示。代理服务器325或主机服务器300能监测多种移动设备的流量,并能够分类流量且设计不同移动设备的流量策略。

[0320] 另外,代理服务器325或主机服务器300可与多个运营商或网络操作者操作,并能实施与多种分类的流量策略应用和流量分类相关的运营商特定策略。例如,代理服务器325或主机服务器300的流量分析器336能包括优化引擎341a、时间关键性检测引擎341b、应用程序状态分类器341c、和/或应用程序流量分类器341d中的一个或多个。

[0321] 这些引擎或模块中的每个,可追踪不同基于无线运营商的优选顺序、临界时间、后台/前台、或交互/维护所考虑的标准。不同标准还存在用于不同移动设备类型(例如,设备模型、制造商、操作系统等)。在一些情况下,移动设备的用户能调整关于流量分类的设置或标准,以及所述代理服务器325能够追踪并实施这些用户调整的/配置的设置。

[0322] 在一实施例中,流量分析器336能检测、确定、识别或推断在一个或多个移动设备上的应用程序的活动状态(例如,移动设备150或250),其中流量来自于或指向该移动设备,例如通过应用程序状态分类器341c和/或流量分类器341d。活动状态可被确定,基于应用程序是否在一个或多个移动设备上的前台或后台状态(通过应用程序状态分类器341c),当前台应用程序流量与后台应用程序流量相比可被不同地处理,以优化网络使用。

[0323] 在交替或结合中,应用程序的活动状态可通过无线连接至移动设备而被确定(例如,通过在本地代理中的应用程序行为检测器)并传输给代理服务器325。例如,活动状态可以按启发式的可靠性水平确定、检测、识别或推断,其基于移动设备的背光状态(例如通过背光探测器)或者在移动设备上的其他软件代理或硬件传感器,包括但不限于电阻传感器、电容传感器、环境光传感器、运动传感器、触摸传感器等。在一般情况下,如果背光开启,流量可以被视为正在或被确定从活动的或处于前台中的应用程序产生,或者该流量是交互式的。此外,如果背光开启,流量可以被视为或被确定来自用户交互或用户活动的流量,或包含用户预期在某个时间段内数据的流量。

[0324] 活动状态从在移动设备250处评估、确定、重估、推断、识别用户活动来确定,并传输至代理服务器325。在一实施例中,活动状态被基于流量是否是交互式流量或维护流量来确定。交互式流量可以包括来自从用户的应用程序活动/交互直接产生之响应和请求的事

务,并且可以包括用户正在等待或预期接收的内容或数据。维护流量可被用于支持不直接由用户检测到的应用程序的功能。维护流量还包括可能发生响应用户动作的动作或事件,但用户不是主动等待或预期响应。

[0325] 时间关键性检测引擎341b可通常确定、识别、推断包含在发送自移动设备250或发送至主机服务器(例如主机300)或应用程序服务器(例如应用程序服务器/内容源110)的移动设备流量中数据的时间灵敏度。例如,时间敏感数据可以包括状态更新、股票信息更新、即时通讯在线状态信息、电子邮件或其他消息、手机游戏应用程序产生的行动、网页请求、位置更新等。

[0326] 根据内容或请求性质,非时间敏感或时间关键性的数据可以包括删除消息请求、标记为已读或已编辑的操作、添加好友或删除好友请求等特定于应用程序的操作、某些类型的消息或其他性质不经常变化的信息等等。在某些情况下,当数据不是时间关键性时,其允许流量通过的时机是基于何时需要从移动设备发送附加数据而设定的。例如,流量整形引擎375可以把流量与一个或多个后续事务调整,使其被一起在移动设备无线(例如使用调整模块378和/或批处理模块377)的单次启动事件中发送。调整模块378也可以把被导向同一主机服务器的时间接近的轮询请求进行调整,因为这些请求可能用相同的数据响应。

[0327] 通常,新的或变化的数据是否从主机服务器发送至移动设备,是基于移动设备上的应用程序是否与新的或变化的数据相关确定的,运行在前台(例如,通过应用程序状态分类器341c)或新的或变化的数据的优先级或时间关键性。所述代理服务器325可发送新的或变化的数据至移动设备,如果应用程序在移动设备上的前台,或如果应用程序在前台和与移动设备上的用户交互的活跃状态,和/或是否用户等待将被提供于新的或变化的数据的响应。所述代理服务器325(或流量整形引擎375)能发送具有高优先级或时间关键性的新的或变化的数据。

[0328] 相似地,代理服务器325(或流量整形引擎375)可抑制新的或变化的数据发送,如果应用程序在移动设备的后台中。代理服务器325还能抑制新的或变化的数据发送,如果用户不在等待提供于新的或变化的数据中的响应;其中所述抑制由代理服务器完成,代理服务器耦合至主机服务器并能够无线连接至移动设备。

[0329]

通常,如果数据包括低优先级或不是时间关键性的新的或变化的数据,代理服务器可等待传输数据直到一段时间后,或直到额外数据被发送(例如,通过调整模块378和/或批量模块377)。

[0330] 附图4A描绘了在分布式代理和缓存系统中客户端部件的另一实施例,该系统进一步包括响应报头管理器401。附图4B描绘了在图4A中显示的所述响应报头管理器的额外部件。

[0331] 在一实施例中,图4A的本地代理275包括具有报头跟踪引擎405的压缩管理器401,报头和缓存地图引擎406和/或报头寻回器。所述响应报头管理器401和每个所示部件可增加或减少部件/模块/引擎。

[0332] 报头跟踪引擎405能跟踪各种报头(例如,HTTP报头),所述报头与接收到在移动设备启动的轮询或客户请求的响应和/或缓存响应相关联。所述报头包括或示出过期参数和/或其他用于缓存有效的标签,示出响应部件过时允许客户/应用程序请求仅目前变化的部

件。一旦报头或标签被响应识别或被响应处提取,他们与移动设备上的本地缓存的缓存条目相关联(例如,通过地图引擎)。报头寻回器407能被用于寻回报头或标签,来使用确定基于网络服务器和/或标签设定的过期时间的缓存条目是否过期,用于确定内容是否实际上已变化。

[0333] 上述功能和特征的部分或全部可在服务器端全部或部分实施(例如,通过如图5A-5B所示的服务器端响应报头管理器501),在响应报头管理器501处由服务器端部件部分或全部实现。

[0334] 附图5A描绘了在分布式代理和缓存系统中服务器端部件的实施例,该系统进一步包括响应报头管理器501。附图5B描绘了在图5A中显示的所述响应报头管理器501的额外部件。

[0335] 在一实施例中,图5A代理服务器325包括报头生成器505、报头监管/重计算引擎506、报头返回引擎507、具有内容主体分析器516和/或内容主体变化检测引擎518的策略创建引擎515。

[0336] 报头生成器505可产生用于响应的报头或标签。报头生成器505还能产生有效器,该有效器用于比较响应来确定缓存响应的特定部件在网络服务器或第三方主机处是否变化或更新。报头返回引擎507能基于该有效器与之前存储在有效器的比较结果,返回合适响应给本地代理(例如,移动设备250的代理275)。如果响应尚未在网络服务器处修改,响应报头管理器501(例如,报头返回引擎507)可返回报头显示如(例如,“无修改”报头)。如果响应或响应部分被修改,管理器501可识别其他部件给移动设备250,还可识别其他未变化的部件。

[0337] 策略创建引擎515包括内容主体分析器516和/或变化检测引擎518,来确定响应部分已变化,这样仅变化的部分需要通过网络传输至移动设备250。策略创建引擎515能创建与基于内容主体和内容响应部件如何和/或何时变换的标签或有效器相关联的策略。响应报头管理器501可计算来自内容主体响应的更新标签(例如,实体标签)或更新有效器,该内容主体响应对先前版本比较以检测和识别那个部分已改变。更新的标签/有效器能使用内容的特定部分或所用功能来计算。

[0338] 响应报头管理器501和所示每个部件可增加或减少部件/模块/引擎。每个部件的功能参照图4A-4B示例的客户端响应报头管理器401的相同/相似部件被相似描述。

[0339] 图6A系一流程图,展示了在移动设备和代理服务器之间的分布式内容缓存以及内容缓存的分布式管理的示例过程。

[0340] 如图4的例子中的分布式交互图所示,所公开的技术是一种具有各方面缓存任务的分布式缓存模型,该缓存任务在客户端/移动设备端(例如,图4的例子中的移动设备450)以及服务器端(例如,服务器端470包括主机服务器485和/或可选的缓存代理475)之间拆分。

[0341] 一般而言,设备端的责任可包括确定特定请求的响应是否可以和/或应该被缓存的。代理的设备端可以基于从请求和响应中/在请求和响应期间收集的信息(例如,时序特征、检测模式、具有启发法的检测模式、可预测性或重复性的指示)做出这个决定并缓存这个决定(例如,把它存储在移动设备上的本地缓存中)。设备端也可以通知在分布式缓存系统中的服务器端的本地缓存事件,并通知其监测内容源(如图1A-B的应用程序服务器/内

容提供商110)。

[0342] 设备端可以进一步指示分布式代理的服务器端定期验证缓存响应(例如,通过轮询的方式或发送轮询请求到内容源)。设备端可以进一步决定特定缓存请求的响应是否应从本地缓存(例如,是否检测缓存击中)返回。设备端(例如,设备上的本地代理)可以通过使用信息作出决定,该信息从由内容源接收到的请求和/或响应中收集/在由内容源接收到的请求和/或响应的过程中收集。

[0343] 在一般情况下,服务器端的责任可包括验证缓存的响应的关联(例如,确定缓存的响应是否依然有效或有其相关的要求)。当缓存的响应被检测到不再有效或不再是相关的时候(例如,服务器无效给定的内容源),服务器端可发送移动设备无效的请求以通知设备。然后设备端可从本地缓存移除响应。

[0344] 图6A显示了为每个检测到的或在移动设备(例如,分布式代理的客户端)检测到的拦截的请求(例如,HTTP请求)所执行的缓存逻辑流程。在步骤602中,代理的客户端(例如,在图2A-B示出的本地代理275或图4的移动设备450)接收到一个请求(从应用程序(例如,移动应用程序)或移动客户端)。在步骤604中,URL被标准化,并在步骤606中客户端检查以确定该请求是否是可缓存的。如果在步骤612中该请求被确定为不可高速缓存,则该请求在步骤608中被发送到源(应用程序服务器/内容提供商),并在610中接收到响应以及交付给请求应用程序622,类似于无需由客户端代理的拦截的请求-响应序列。

[0345] 如果该请求被确定为是可缓存的,在步骤612中,客户端查找缓存以确定当前请求的缓存条目是否存在。如果存在,在步骤624中,客户端能够确定该条目是否是有效的,如果有效,客户端可以检查请求以查看是否在步骤615中包括一个验证器(例如,修改后的报头或实体标签)。例如,验证的概念是躲避至RFC2616的13.3节,其描述报头的可能类型(例如,电子标签、修改_何时、需要_重新生效,编译指示未_缓存)并形成有效响应632,如果是这样,则在步骤622中被传递到请求应用程序。如果请求不包含步骤615所确定的验证,在步骤630中从本地缓存中形成一个响应,并在步骤622中传送到请求应用程序。该验证步骤可用于通常被认为不能被缓存的内容。

[0346] 相反地,如果在步骤624中缓存条目被发现但被确定为不再有效或无效,则在步骤618中,代理的客户端发送请求616至内容源(应用程序服务器/内容主机),并直接从源接收响应。同样,如果在步骤612中,缓存条目在查找过程中被发现,该请求在步骤616中也被发送。一旦接收到响应,客户端检查响应以确定在步骤626中它是否是可缓存的。如果是这样,在步骤620中响应被缓存。然后客户端在步骤614中发送另一个轮询,然后在步骤622中传递响应至请求应用程序。

[0347] 图6B描述了时序图,示出了请求自移动设备450至应用程序服务器/内容提供商496的数据在无线网络中如何通过分布式代理系统460以一种方式协同,该方式是网络和电池资源被节约,通过使用内容缓存和由分布式代理系统460实现的监测。

[0348] 在满足移动设备450上无需分布式代理系统460的应用程序或客户端请求中,移动设备450或执行与设备450上的软件部件执行数据请求452(例如,HTTP GET, POST, 或其他请求)直接至应用程序服务器495并接收直接来自服务器/提供商495的响应404。如果数据已被更新,移动设备上的部件455能刷新自身反应所述更新并等待一小段时间,并触发对服务器/提供商495的另一数据请求。

[0349] 在一实施例中,请求设备450上的客户端或软件部件455能利用分布式代理系统460处理对服务器/提供商495的数据请求。通常,分布式代理系统460可包括本地代理465(通常被认为系统460的客户端组件并能附属在移动设备450上),缓存代理475(被认为系统的服务器端组件470且附属在主机服务器485上或全部或部分在主机服务器485的外部),以及主机服务器485。本地代理465可通过任何网络或网络结合被连接至缓存代理475和主机服务器485。

[0350] 当分布式代理系统460被用于数据/应用程序请求时,部件455可通过本地代理465执行数据请求456。本地代理465,可拦截由设备应用程序产生的请求,并能识别所述请求的连接类型(例如,HTTP get 请求或其他类型请求)。本地代理465之后能问询本地缓存,任何之前关于请求的信息)。如果本地存储响应不可利用或如果无效响应被存储,本地代理465能更新或存储关于请求的信息、产生的时间和任何其他数据在本地缓存内。所述信息可被更新用于潜在满足后续请求。

[0351] 本地缓存465能之后发送请求至主机服务器485,且主机服务器485能执行所述请求456并响应458返回所述结果。本地代理465能存储所述结果,且此外关于结果的信息并返回请求部件455的结果。

[0352] 在一实施例中,如果相同请求出现多次(在某一时期内)且经常趋向于相同结果,本地代理465可通知460服务器485,所述请求应该被监测(例如步骤462和464),对优先于返回结果至本地代理465或请求部件455的结果变化。

[0353] 在一实施例中,如果请求被标记用于监测,本地代理465目前能存储结果至本地缓存内。目前,当数据请求466,本地响应可利用的,由部件455产生时并在本地代理465处被截获,本地代理465能在不需要跨越无线网络建立连接通信下自本地缓存返回响应468。另外,服务器代理执行被标记未监测470的请求,以确定给定请求的响应472是否改变。通常,主机服务器485能执行独立于部件455或本地代理465操作的监测。无论何时未期待的响应472接收用于请求,服务器485可通知本地代理465响应已变化(例如步骤474中的无效通知)以及通知本地存储在客户端的响应应该被消除或被新的响应替代。

[0354] 在这个情况下,来自设备450的部件455的随后数据请求476产生返回自主机服务器485(例如通过缓存代理475)的数据并在步骤478中,所述请求满足缓存代理475。因此,通过利用分布式代理系统460,无线(蜂窝)网络被智能使用,当移动设备40上的用于部件或软件应用程序455的内容/数据通常被改变时。如此,所需流量检查应用程序数据变化在跨越无线(蜂窝)网络上未执行。这减少了产生网络流量的数量并缩短了总共时间,且无线电模块的次数在移动设备450上上电,因此减少电池消耗并,除此之外,释放网络带宽。

[0355] 附图7描绘了一份表格700,这份表格记录了不同的流量或应用程序分类类型,这些在网络接入执行过程中和内容传输策略中会被使用到。例如,;流量/应用程序分类可包括交互的或后台,用户是否在等待响应,前台/后台应用程序,以及背光灯是否开启或关闭。

[0356] 附图8描绘了一份表格800,这份表格记录了不同内容分类类型,这些在网络接入执行过程中和内容传输策略中会被使用到。例如,内容分类类型可包括高或低优先级内容,以及时间关键性或时间非关键性内容/数据。

[0357] 图9系一交互图,展示了具有通过无线网络从移动设备至应用程序服务器/内容提供商995的数据请求的应用程序(例如,移动应用程序)955轮询如何被缓存在本地代理上并

且由分布式缓存系统(包括本地代理965和主机服务器985(具有服务器缓存935或缓存代理服务器975))管理。

[0358] 在一个例子中,当移动应用程序/部件955轮询应用程序服务器/提供商932时,轮询可以在移动设备上由本地代理965进行本地拦截934。本地代理965可以检测缓存的内容是可用于请求中的轮询内容并因此可以从本地缓存中检索响应以满足被拦截的轮询936,而无需使用无线网络带宽或其他无线网络资源。移动应用程序/部件955可随后接收响应至来自缓存条目938的轮询。

[0359] 在另一个例子中,移动设备应用程序部件955轮询应用程序服务器/提供商940。轮询由本地代理965拦截942并检测缓存内容无法在本地缓存中可用并决定为缓存944设置被轮询的源。为了满足要求,轮询被转发至内容源946。应用程序服务器/提供商995接收来自应用程序的轮询请求并提供响应以满足当前的请求948。在950中,应用程序(例如,移动应用程序)/部件955接收来自应用程序服务器/提供商的响应以满足请求。

[0360] 为了设置内容缓存,结合本地代理965跟踪应用程序的轮询频率并可以设置轮询时间表以被发送到主机服务器952。本地代理发送缓存设置至主机服务器954。主机服务器985可以使用缓存设置,其包括,例如,应用程序服务器/提供商的识别将可选地被轮询时间表956轮询。主机服务器985现在可以轮询应用程序服务器/提供商995以监测代表移动设备的请求958的响应。主机服务器985确定被接收的相同的响应并根据特定的轮询时间表962轮询应用程序服务器995。应用程序服务器/内容提供商995接收轮询并做出相应的响应964。

[0361] 主机服务器985检测更改的或新的响应并通知本地代理965。主机服务器985可以额外存储在服务器中缓存或缓存代理968中的更改的或新的响应。本地代理965从主机服务器985收到新的或更改的数据现已可用并可受影响的缓存条目970无效的通知。下一次,应用程序(例如,移动应用程序)/部件955为同一服务器/内容提供商972产生了同样的请求,本地代理确定没有有效的缓存条目而检索来自服务器缓存974的响应,例如,通过HTTP连接。主机服务器985接收新响应的请求并发送响应回976至本地代理965。请求被从服务器缓存或缓存代理978满足,而不需要利用移动设备使用其无线电或消耗移动网络带宽,从而节省了网络资源。

[0362] 另外,当应用程序(例如,移动应用程序)在步骤980产生相同的请求,本地代理965,响应于确定没有有效的缓存条目可用,则在步骤982通过移动网络转发轮询到应用程序服务器/提供商。应用程序服务器/提供商995接收轮询并在步骤984中通过移动网络将响应返回到移动设备。因此,在步骤986中该请求使用移动网络从服务器/提供商被满足。

[0363] 图10系一交互图,展示了应用程序1055如何轮询来自应用程序服务器/内容提供商1095的内容,其通过无线网络采用标识符(例如,旨在击败缓存的标识符)中的缓存击败机制仍可以被检测并本地缓存。

[0364] 在一个示例中,当应用程序(例如,移动应用程序)/部件1055在步骤1032中轮询应用程序服务器/提供商时,在步骤1034中轮询可以在移动设备上由本地代理1065本地地拦截。在步骤1034中,在移动设备上的本地代理1065也可能确定(以某种程度的确定性和启发式)缓存击败机制由服务器提供商采用或可能采用。

[0365] 本地代理1065可以为请求中的轮询的内容检测缓存的内容是可用的,因此可以从

本地缓存中检索响应以满足被拦截的轮询1036,而无需使用无线网络带宽或其他无线网络资源。在步骤1038中,应用程序(例如,移动应用程序)/部件1055可随后接收响应至来自缓存条目(例如,在移动设备上本地存储的缓存条目)的轮询。

[0366] 在另一个例子中,在步骤1040中,应用程序(例如,移动应用程序)部件1055轮询应用程序服务器/提供商1095。该轮询在步骤1042中被由本地代理1065拦截,本地代理1065确定缓存击败机制由服务器/提供商1095采用。本地代理1065也可以检测该请求在本地缓存中的缓存的内容是不可用,并决定在步骤1044中为缓存设置轮询内容源。本地代理1065可以稍后提取请求的标识符的模式(例如,格式或语法)并在步骤1046中跟踪应用程序的轮询频率以设置主机服务器1085的轮询时间表。

[0367] 为了满足该请求,在步骤1048中,轮询请求被转发给内容提供商1095。应用程序服务器/提供商1095从应用程序中接收轮询请求并在步骤1050中提供响应以满足当前的请求。在步骤1052中,应用程序(例如,移动应用程序)/部件1055从应用程序服务器/提供商1095接收响应以满足请求。

[0368] 为了设置内容缓存,结合本地代理1065缓存响应并存储标识符(或标准化的标识符的哈希值)的归一化版本,该标识符与接收到的响应相关联以用于未来在步骤1054中的识别和检索。在步骤1056中,本地代理发送缓存设置到主机服务器1085。缓存设置包括,例如,标识符和/或标识符的归一化版本。在某些情况下,不同于标准化的标识符的修改后的标识符被发送到主机服务器1085。

[0369] 主机服务器1085可以使用缓存设置,其中包括,例如,将要轮询的应用程序服务器/提供商的识别和在步骤1058中可选的轮询时间表。主机服务器1085现在可以轮询应用程序服务器/提供商1095以监测在步骤1060中代表移动设备的请求的响应。在步骤1062中,应用程序服务器1095从主机服务器1085接收轮询响应。主机服务器1085确定相同的响应已被接收并轮询应用程序服务器1095,例如,根据指定的轮询时间表并在步骤1064中使用标准化的或修改过的标识符。在步骤1066中,应用程序服务器/内容提供商1095接收投票并做出相应的响应。

[0370] 这一次,在步骤1068中,主机服务器1085检测更改的或新的回应并通知本地代理1065。在步骤1070中,主机服务器1085可附加地在服务器缓存中1035或缓存代理1075中存储更改的或新的响应。在步骤1072中,本地代理1065从主机服务器1085接收新的或更改的数据现已可用并可以使受影响的缓存条目无效的通知。下一次,在步骤1074中,应用程序(例如,移动应用程序)/部件为同一服务器/内容提供商1095产生同样的请求,本地代理1065确定没有有效的缓存条目可用,而在步骤1076中,从服务器缓存中检索响应,例如,通过一个HTTP连接。在步骤1078中,主机服务器1085接收新的响应的请求并将响应发送回本地代理1065。在步骤1080中,请求被因此从服务器缓存或缓存代理满足,而不需要为移动设备利用其无线或消耗移动网络带宽,从而节省了网络资源。

[0371] 可替代地,当应用程序(例如,移动应用程序)1055产生了同样的要求,在步骤1084中,本地代理1065响应于确定没有有效的缓存条目,在步骤1082中,通过移动网络转发轮询到应用程序服务器提供商1095。应用程序服务器/提供商1095接收轮询并在步骤1086中将响应通过移动网络返回到移动设备。在步骤1088中,通过使用移动网络1086,该请求因此被从服务器/提供商满足。

[0372] 图11系一流程图,展示了收集关于请求以及相关的识别缓存性能的响应的信息并缓存该响应的示例过程。

[0373] 在过程1102中,关于请求的信息和关于请求的接收到的响应的信息被收集。在过程1104和1106中,关于在移动设备发起的请求和信息以及关于请求的接收到的响应的信息被合计或独立使用以在步骤1108确定缓存性能。使用请求和响应信息以用于评估缓存性能的步骤的细节被将在图12的例子中进一步描述的流程A进行说明。

[0374] 在步骤1108中,如果基于流程A它确定响应是不可缓存的,那么响应在步骤1110中不可被缓存,并且流程可选的在1102重新开始以收集关于请求或响应的信息以用于再次评估缓存性能。

[0375] 在步骤1108中,如果从流程A确定响应是可缓存的,则在1112响应可以作为缓存条目被存储在缓存中,缓存条目包括具有有关响应缓存的附加信息的元数据。缓存条目,除了响应,包括具有响应有关响应缓存的附加信息的元数据。元数据可以包括,例如,缓存条目的访问时间或缓存条目的创建时间。

[0376] 在响应被存储在缓存中之后,可出现并行的过程以确定存储在缓存中的响应是否需要过程1120中被更新。如果是,在过程1122中,存储在移动设备的缓存中的响应是无效的或从移动设备的缓存被移除。例如,响应的相关性或有效性可以通过定期轮询请求被定向至的代表移动设备的主机服务器而被定期验证。主机服务器可以通过使用被收集用于响应请求的请求信息以在移动设备上所确定的速率进行轮询。该速率由产生请求的相同客户端产生的之前的请求之间的间隔的平均所确定。

[0377] 验证可以由物理上不同于移动设备的实体进行执行。在一个实施例中,该实体是耦合到移动设备的代理服务器并能够与移动设备进行无线通信并且以及基于由产生请求的相同客户端产生的之前的请求之间的间隔,代理服务器以一移动设备所确定的速率轮询请求所定向的主机服务器。

[0378] 在过程1114中,相同的客户端或应用程序的后续请求被检测到。在过程1116中,本地缓存中的缓存查找被执行以确定的缓存条目被用于响应后续请求。在一个实施例中,元数据被用确定存储作为缓存条目的响应是否被使用以满足随后的响应。在过程1118中,可以从缓存中提供响应以满足后续请求。响应可以提供以响应为后续请求识别匹配的缓存条目,该后续请求至少部分地使用元数据来确定。

[0379] 图12系一流程图,展示了确定请求的响应是否能被缓存的决定流程的示例过程。

[0380] 过程1202确定如果请求被定向到被列为黑名单的目的地。如果是这样的话,响应不会在步骤1285中被缓存。如果被列为黑名单的目的地被检测到或如果请求本身与被列为黑名单的应用程序相关联,图中所示分析的其余部分可能无法执行。如果请求及其目的地不被列入黑名单,这个过程可以继续到步骤1204和1206中。

[0381] 在过程1204中,与该请求关联的请求特征信息被分析。在分析该请求中,在过程1208中,请求方法被识别,并且在步骤1214中,判断响应是否可以基于请求方法被缓存。如果检测到不可缓存的请求,该请求不被缓存并且过程可能会在1285终止。如果请求方法被确定为可缓存的,或不不可缓存的,那么在步骤1295响应可以被识别为可缓存的或潜在可缓存的(例如,缓存,但受其他测试和分析图中所示)。

[0382] 在过程1210中,该请求的大小是确定的。在过程1216中,确定请求的大小是否超过

可缓存的大小。如果是的话,响应不被缓存以及在分析可能在过程1285终止。如果在步骤1216中请求的大小不超过可缓存的大小,那么在步骤1295中响应可以被识别为可缓存的或潜在可缓存的(例如,可缓存但受限于图中所示的其他测试和分析)。

[0383] 在步骤1212中,相同的客户端产生的请求和其它请求之间的周期性的信息被确定。在步骤1218中,判断周期性是否已被确定。如果不是,响应不被缓存并分析可在1285终止。如果是,那么在步骤1295中响应可以被识别为可缓存的或潜在可缓存的(例如,可缓存但受限于图中所示的其他测试和分析)。

[0384] 与请求收到的响应相关的请求特征信息在过程1206中被分析。

[0385] 在过程1220中,状态代码被识别并确定在过程1228中状态代码是否表示可缓存的响应状态码。如果不可缓存的状态代码被检测到,请求不被缓存并且过程可能会在1285终止。如果响应状态代码表示缓存性能,或不可缓存的,那么在步骤1295中响应可以识别为可缓存的或潜在可缓存的(例如,可缓存但受限于图中所示的其他测试和分析)。

[0386] 在过程1222中,确定响应的大小。在过程1230中,判断响应的大小是否超过可缓存的大小。如果是的话,响应不被缓存以及分析可能在在1285终止。如果响应的大小不超过在步骤1230中的可缓存的大小,那么在步骤1295中响应可以识别为可缓存的或潜在可缓存的(例如,可缓存但受限于图中所示的其他测试和分析)。

[0387] 在过程1224中,响应主体被分析。在过程1232中,判断响应是否包含动态内容或高动态内容。由于数据的内在本质(例如,股票行情、体育比分、快节奏的体育赛事等)动态内容包括以高频率进行改变的和/或具有较短生存时间或较短相关时间的数据。如果是的话,响应是不被缓存以及分析可能在1285终止。如果不是,那么在步骤1295中响应可以识别为可缓存的或潜在可缓存的(例如,可缓存但受限于图中所示的其他测试和分析)。

[0388] 过程1226确定传输编码或块传输编码是否在响应中使用。如果是的话,响应是不被缓存以及分析可能在1285终止。如果不是,那么在步骤1295中响应可以识别为可缓存的或潜在可缓存的(例如,可缓存但受限于图中所示的其他测试和分析)。

[0389] 不是所有的上述测试需要被执行以确定响应是否被缓存。未示出的附加的测试也可以被执行。需要注意的是,测试1208、1210、1212、1220、1222、1224和1226中的任何一个都可以被单独或以任何组合执行以确定缓存性能。在某些情况下,所有的上述产测试都被执行。在某些情况下,所有被执行的测试(任意数量的上述被实际执行的测试)需要为确定为可缓存的响应确认缓存性能。换句话说,在某些情况下,如果上述的测试中任一项指示非不可缓存性能,则响应不被缓存,而不管其他测试的结果在其他情况下,基于请求特征和响应特征的组合,可以使用不同的标准以确定哪些测试或多少测试需要通过系统来决定缓存给定的响应。

[0390] 图13系一流程图,展示了基于请求周期性和/或响应重复性判断潜在的缓存性能的示例过程。

[0391] 在过程1302中,由客户端生成的请求被跟踪以检测请求的周期性。在过程1306中,判断在请求的时序中是否有可预测的模式。如果是,响应内容可在过程1395中被缓存。如果不是,在过程1308中,判断请求间隔是否落入公差范围内。如果是,响应内容可以在过程1395中被缓存。如果不是,响应在过程1385中不被缓存。

[0392] 在过程1304中,由客户端产生的请求的接收到的响应被跟踪以检测响应内容中的

重复性。在过程1310中,为客户端接收的响应的响应主体的哈希值被检查并在过程1312中检查与响应相关的的状态码。在过程1314中,通过使用哈希值和/或状态码判断是否在至少两个响应的内容中存在相似。如果是的话,响应在过程1395中可以被缓存。如果不是,响应在1385中不被缓存。

[0393] 图14系一流程图,展示了为给定请求或客户端动态调节缓存参数的示例过程。

[0394] 在步骤1402中,由客户端生成的或定向至主机的请求被在移动设备处跟踪以检测请求的周期性。过程1404确定两个或多个请求之间的请求间隔是否是相同的或大致相同的。在过程1406中,判断两个或多个请求的请求之间的请求间隔是否落入公差范围内。

[0395] 基于步骤1404和1406的结果,周期性被检测的请求的响应被在过程1408中接收。

[0396] 在过程1412中,响应被缓存作为移动设备的缓存中的缓存条目。在过程1414中,主机被以一速率进行检测以验证缓存条目的相关性或有效性,同时在过程1416中,响应可以从缓存中被提供以满足后续请求。

[0397] 在步骤1410中,监测主机的速率从请求的间隔确定,使用,例如,过程1404和/或1406的结果。在过程1420中,给定主机被监测所使用的速率被设置以验证请求的缓存条目的相关性或有效性。在过程1422中,由客户端生成的请求的请求间隔的变化被检测。在过程1424中,不同的速率基于请求的间隔的变化被技术。在过程1420中,给定主机被监测所使用的速率被更新以验证请求的缓存条目的相关性或有效性。

[0398] 附图15描绘了一流程图,展示了当对网络接入和内容传输策略实施的期望和用户活动考虑在内时,应用程序和/或流量(数据)分类的示例过程。

[0399] 在步骤1502中,系统或服务器检测可利用被发送至移动设备的新的或变化的数据。新的,变化的,或更新的数据能包括一个或多个IM存在更新、股票行情更新、天气更新、电邮、文本信息、新闻推送、好友推送、日志条目、文章、文件、任何多媒体内容(例如,图片、音像、相片、视频等),或任何其他可被经HTTP或无线带宽网络发送,或被用户用于位置终端设备或应用程序操作而消耗。

[0400] 在步骤1504中,新的或变化的数据指向的应用程序被识别。在步骤1506中,应用程序基于应用程序被分类。在步骤1508中,确定新的或变化的数据的优先级或关键性时间。在步骤1510中,数据被分类。基于由相关数据的应用程序和/或优先级/时间敏感度确定的信息,任何或所有系列的评估可被完成来分类流量和/或构建用于在移动设备无线上传输和/或上电的策略。

[0401] 例如,使用识别的应用程序的信息,在步骤1512中,来确定应用程序是否在活跃状态与用户在移动设备上交互。在步骤1514中,确定应用程序是否运行在移动设备的前台。

[0402] 如果步骤1512或1514测试答案为“是”,系统或服务器之后能在步骤1526中确定新的或变化的数据是否被发送至移动设备,并无延迟发送。另外,正如图31进一步示出,步骤在流程C处继续,此处时序伴随其他传输参数,如网络配置,可被选择。如果步骤1512或1514测试答案为“否”,其他测试按任意次序进行。只要步骤1512或1514的任意一个测试为“是”,具有数据系统和服务器之后能处理步骤1526和/或流程C。

[0403] 基于应用程序或应用程序特征的步骤1512和1514的测试答案为“否”,之后过程执行步骤1524。其中新的或变化的数据的发送被抑制,至少在临时基础上。过程能在A流程继续,例如还确定何时发送数据的时序以优化网络使用和/或设备供电消耗,如图29中所描述

的示例。

[0404] 相似地,在步骤1516中,确定应用程序是否运行在后台。如果是,过程继续执行步骤1524,此时新的或变化的数据发送被抑制。但是,即使应用程序在后台状态,任何保持测试可被执行。例如,即使应用程序在后台状态,如果是高优先级或关键性时间,则新的或变化的数据可能被发送。

[0405] 在步骤1518中,使用优先级或时间关键性信息,确定数据是否有高优先级1518。在步骤1520中,确定数据是否为关键性时间。在步骤1522中,确定用户是否等待将提供于可利用数据的响应。

[0406] 如果步骤1518,1520或1522的答案为“是”,系统或服务器之后能确定步骤1526中新的或变化的数据是否被发送至移动设备,并无延迟发送。另外,过程能继续流程C,此处时序伴随其他传输参数,如网络配置,可被选择,如图31示例进一步示出。如果步骤1518,1520或1522测试答案为“否”,其他测试按任意次序进行。只要步骤1512或1514的任意一个测试为“是”,具有数据系统和服务器之后能处理步骤1526和/或流程C。

[0407] 如果步骤1518,1520或1522的一个或多个答案为“否”,之后过程执行步骤1524。其中新的或变化的数据的发送被抑制,至少在临时基础上。过程能在A流程继续,例如还确定何时发送数据的时序以优化网络使用和/或设备供电消耗,过程能继续在有或没有其他执行测试的步骤1524,如果其中之一测试趋向于“否”响应。

[0408] 确定应用程序分类步骤1504能被用于与确定数据分类步骤1510所替代或相结合。例如,新的或变化的数据具有高优先级或关键性时间的能在步骤1526发送,即使应用程序在前台状态但与移动设备的用户不积极交互,或者如果应用程序不在前台或在后台。

[0409] 相似地,即使用户不在等待提供于新的或变化的数据的响应(步骤1522),数据可被发送至移动设备1526,如果应用程序在前台,或者如果数据有高优先级或保持关键性时间内容。

[0410] 一般,在内容源(例如,起始于新的或变化的数据的服务器/内容主机)或代理服务器处执行抑制。例如,代理服务器可来自远程的接收移动设备(例如,能够无线连接接收移动设备)。代理服务器还可来自远程的起始服务器/内容主机。尤其是,逻辑并智力确定数据是否被发送或抑制存在于相同的服务器或相同主体,如数据起源器被发送或部分或全部来自远程(例如,代理能够与内容起始服务器通信)。

[0411] 在一实施例中,等待传输数据由移动设备上的本地代理所管理,其能够与接收服务器无线通信(例如,移动应用程序或客户的主机服务器)。移动设备上的本地代理能控制移动设备上无线电使用,当时间周期消逝,或当额外数据被发送时,数据的传输被检测到。

[0412] 附图16A描绘了一流程图,展示了流量处理的示例过程,所述流量因应用程序/流量分类,将被至少短暂抑制。

[0413] 例如,步骤1602,时间周期在新的或变化的数据在步骤1606传输前消逝。这可被实现,如果数据有低优先级或无关键性时间,否则被确定发送抑制(例如,正如图15流程图所确定)。时间周期可被应用程序、用户、第三方、和/或默认值设置。时间周期还可在应用程序的特定类型或实时网络操作条件的条件下被更新。如果被发送的新的或变化的数据来源于移动设备,等待传输数据直到消逝的时间周期,能由移动设备上的本地代理管理,其与主机服务器通信。本地代理还能或允许移动设备无线电使用,用于当时间周期消逝时传输数据。

[0414] 在一些实例下,当有额外数据被发送在步骤1604中,新的或变化的数据在步骤1606传输。如果被发送的新的或变化的数据来源于移动设备,等待传输数据直到消逝的时间周期,能由移动设备上的本地代理管理,其与主机服务器通信。本地代理还能或允许移动设备无线电使用,用于当额外数据被发送时传输数据,这样设备资源被保存。注意额外数据可能源自于相同移动应用程序/客户或不同应用程序/客户。额外数据可包括较高优先级或关键性时间的内容。额外数据还可有相同或较低优先级。在一些实施例下,某些无优先级或无时间敏感时间可触发一个发送事件。

[0415] 如果被发送的新的或变化的数据起源于服务器(内容的代理服务器或主机服务器),等待传输数据直到消逝的时间周期或等待额外数据被发送,由能与移动设备无线通信的代理服务器管理。通常,代理服务器等待直到额外数据可利用于相同移动设备,在发送共同在单个事务下的数据,以减少设备电池电源开启次数并优化网络使用。

[0416] 图16B描绘了一流程图,展示了网络结构选择的示例过程,用来发送流量,基于应用程序和/或流量(数据)分类。

[0417] 在步骤1608中,移动设备上的应用程序的活动状态被检测指向流量或来源流量。并列的或代替的活动状态,在移动设备和主机服务器之间发送保持在流量中的数据时间关键性在步骤1610中确定。活动状态能部分或在等应用程序是否在移动设备的前台或后台状态时确定。活动状态还可在用户是否在于应用程序交互时确定。

[0418] 使用活动状态和/或数据特征,图15的步骤1612中当由数据确定被发送至移动设备。过程能继续步骤3006用于网络配置选择。

[0419] 例如,步骤1614中,一代无线标准被选择。被选择的无线标准的一代包括2G或2.5G,3G,3.5G,3G+,3GPP,LTE或4G,或任何其他未来代。例如,无线标准较慢或较老代由较少关键性事务或保留较少关键性数据的流量指定。例如,较老标准如2G,2.5G或3G能被选择规划流量当一个或多个下述被检测,应用程序不与用户交互,应用程序运行在移动设备后台,或保留在流量中的数据不是时间关键性的。较新代如被指定为较高优先级流量或事务。例如,较新代如3G,LTE或4G当活动状态与用户交互或在移动设备的前台时被指定为流量。

[0420] 在步骤1616中,访问信道类型被选择。例如,前向访问信道(FACH)或专用信道(DCH)能被指定。在步骤1618中,网络配置基于数据率或数据率性能被选择。例如,具有交底数据率的网络配置可被指定为流量,在下述一个或多个被检测时,应用程序不与用户交互,应用程序在移动设备后台运行,或保留于流量的数据不是非时间关键性的。

[0421] 在步骤1620中,网络配置被选择通过制定访问点。步骤1614,1616,1618或1620的任意一个或所有可被执行或任意结合用于网络配置。

[0422] 图16C描绘了一流程图,展示了实现网络接入和内容传输策略的示例过程,其基于应用程序和/或流量(数据)分类。

[0423] 在步骤1634中,移动设备上应用程序的活动状态被检测,流量来源于或被指向移动设备。例如,活动状态能通过应用程序是否在移动设备的前台或后台状态时被确定。活动状态还可被用户是否期待保留于指向移动设备的流量的数据。

[0424] 在步骤1636中,确定发送在移动设备和主机服务器之间的保留于流量的数据的时间关键性。

[0425] 应用程序状态和/或数据特征能被用于应用程序分类和/或数据分类,来确定来自

上述的流量是否被发送至移动设备或至少在发送前的临时基础上被抑制,如图15示例示出的流程图。

[0426] 在确定后继续流程C被用于发送流量,与流量如何和何时发送相关的参数能被确定。例如,在步骤1638中,允许流量通过的时序基于活动状态或时间关键性被确定。

[0427] 在步骤1640中,移动设备的无线电使用基于流量允许通过的时序而被控制。例如,启动于移动设备的流量,位于移动设备的本地代理能控制无线电是否被开启用于事务,如果是,当基于由应用程序状态或数据优先级/时间关键性确定的事务特征时,其将被打开。

[0428] 在步骤1642中,无线网络中的网络配置被选择用于传递流量至移动设备和/或传递来自于移动设备的流量。例如,更高容量或数据率网络(例如,3G,3G+,3.5G,LTE或4G网络)能被选择来通过流量传递,当应用程序活跃时或当保留在流量的数据是时间关键性的或具有较高优先级/重要性的。

[0429] 图17描绘了一流程图,展示了基于移动用户活动或用户期望的网络选择的示例过程。

[0430] 在步骤1702中,移动设备的背光灯状态被检测。背光灯状态能被用于确定或推断关于用户活动和/或用户期望的信息。例如,步骤1704中,与应用程序在移动设备上的用户交互被检测和/或在步骤1706中,如果确定在背光灯开启时,用户正在期待指向移动设备保留于流量的数据。

[0431] 用户交互1704和/或用户期望1706能确定或通过其他直接或非直接线索推断。例如,设备动作传感器、环境灯光、数据活动、无线电活动和模式检测、呼叫处理等,能单独使用,或结合起来使用,以进行关于用户活动、交互或期望的评估。

[0432] 在步骤1708中,移动设备上的应用程序的活动状态被确定,流量来自于移动设备或被指向移动设备。在一实施例中,应用程序的活动状态通过与在移动设备上的应用程序用户交互确定,和/或通过用户是否期待保留在指向移动设备的流量的数据。

[0433] 在步骤1710中,3G,4G或LTE网络被选择用于发送在无线网络中移动设备和主机服务器之间的流量。其他网络配置或技术也能被选择,包括但不限于2.5G GSM/GPRS网络,EDGE、EGPRS,3.5G,3G+,turbo 3G,HSDPA等。例如,当用户交互被检测有应用程序请求网络访问时,更高带宽或更高容量网络能被选择。相似地,如果被确定或在某些可能下被推断,用户可能期待保留于流量请求网络访问,较高容量或较高数据率网络也被选择。

[0434] 活动状态还可通过保留于指向移动设备的流量的数据响应应用程序中前台活动而被确定。对于在前台的应用程序,较高容量(例如,3.5G,4G或LTE)网络可被选择用于执行事务。

[0435] 移动设备可被确定,通过设备参数如移动设备的背光灯状态,或任意其他基于设备传感器,包括但不限于电阻性传感器、电容性传感器、光检测器、动作传感器、接近传感器、触摸屏传感器等的软件或硬件。被选择使用的网络配置还可基于保留于流量的数据的时间关键性和/或优先级,以在移动设备和主机服务器之间发送。

[0436] 图18系一时序图1800,展示了可能适于缓存的周期请求的检测的示例。

[0437] 在所示的例子中,来自移动设备上的客户端/应用程序的第一请求被时间1:00(t_1)检测到。此时,缓存条目可以被在步骤1802中创建。在时间2:00(t_2),第二请求被从同一客户端/应用程序检测到,并且被创建的缓存条目现在可被在步骤1804中的时间 t_2 与 t_1

之间的1小时的间隔更新。来自同一客户端的第三请求现在可以在时间 $t_3=3:00$ 检测到,并且它现在可以被确定周期性的请求在步骤1806中被检测到。本地代理现在可以缓存响应并发送指定到代理服务器的间隔(例如,在这种情况下,1小时)的启动轮询请求。

[0438] 时序图进一步示出2:54和3:06之间的时序图,这指示一个窗口的边界,如果第三请求在时间帧1810中被接收,则在该窗口内的周期性将被确定。2:54和3:06分之间的时序窗口1808对应于以前的间隔的20%并且是所示公差示例。其他公差也可以被使用,以及可以动态地确定或根据不同的情况(应用程序)被确定。

[0439] 图19系一数据时序图1900,展示了请求间隔中的变化的检测以及响应中的服务器轮询速率的更新的示例。

[0440] 在步骤1902中,代理确定周期性请求被检测到,本地代理缓存响应并设置轮询请求到代理服务器,并且,例如,间隔被在第三请求设置为1小时。在时间 $t_4=3:55$,请求在55分钟而不是1小时之后被检测到。55分钟的间隔仍适合窗口1904给定的20%的公差。然而,在步骤1906中,在五请求在时间 $t_5=4:50$ 被接收,其不再适合从第一和第二,以及1小时的第二和第三请求之间的间隔被确定的公差窗口设置。现在本地代理从代理服务器中检索资源或响应,并刷新本地缓存(例如,不用来提供第五请求的缓存条目)。本地代理也以一更新的间隔(例如,在本例中为55分钟)重发启动轮询请求至代理服务器并且窗口由公差定义,例如设置为20%,现在变成了11分钟,而不是12分钟。

[0441] 需注意的是,在一般情况下,当间隔的改变被检测到时和/或当新的速率已经被确定时,本地代理以一更新的轮询间隔通知代理服务器。然而,这通常只用于后台应用程序请求或自动/程序化的刷新(例如,不涉及用户交互的请求)。一般情况下,如果用户与应用程序在前台进行交互并导致检测到一段请求,指定至代理服务器的轮询或轮询间隔的速率通常不会被更新,如图20所示。图20描绘了数据时序图2000,其显示提供具有缓存条目的前台请求的一个例子。

[0442] 例如,时代 $t=3:00$ 和3:30之间,本地代理在 $t=3:10$ 和 $t=3:20$ 检测第一和第二前台请求。这些前台请求在为检测前台应用程序或自动应用程序请求的周期性之外。然而,为前台请求检索的响应数据可以被缓存和更新,前台请求的请求间隔在过程2008不被发送到服务器。

[0443] 如图所示,在 $t=4:00$ 从应用程序(例如,后台请求,程序性/自动刷新)检测到的下一次周期请求,响应是从缓存中提供,作为在 $t=5:00$ 的请求。

[0444] 图21系一数据时序图,展示了发生在过时的内容可能再次服务于请求应用程序之后的缓存失效的非最佳影响的示例。

[0445] 由于代理服务器轮询的间隔设置为大致相同的间隔,应用程序(例如,移动应用程序)以该间隔发送请求,有可能的情况是代理服务器通常在缓存的条目(现在已过时)已经被提供至请求之后检测改变的内容(例如,在 $t=5:02$) (例如,在 $t=5:00$ 至第五请求)。在所示例子中,资源在 $t=4:20$ 更新或更改以及之前的发生在 $t=4:02$ d服务器轮询不能够捕获这种改变直到在5:02的下一个轮询,并在2100发送缓存失效至本地代理。因此,在第五请求在时间 $t=5:00$ 已经被提供旧内容之后,本地缓存不能在某些时间无效缓存。新鲜的内容现在不会提供给请求应用程序直到第六请求在 $t=6:00$,在过程2106的1周期以后。

[0446] 要优化缓存性能并解决这个问题,本地代理可以通过指定请求的初始时间调节时

间设置,除了到代理服务器的轮询间隔。在此的请求的初始时间在请求实际发生之前设置为某些时间(例如,几分钟),以使得代理服务器轮询稍微发生在未来的实际应用请求之前。通过这种方式,代理可以及时获得响应中的任何改变以被提供至随后的应用程序请求。

[0447] 图22系一数据时序图,展示了缓存管理和考虑缓存条目的生存时间(TTL)设置的和响应的示例。

[0448] 在一个实施例中,在本地缓存中的缓存的响应数据指定时间量,缓存条目可以存储在本地缓存中直到它被删除或移除。

[0449] 在给定的缓存条目中的响应数据将被移除的时间可以使用下面的公式确定: $\langle \text{响应数据_缓存时间} \rangle + \langle \text{TTL} \rangle$,如图所示在 $t=3:00$,响应数据在TTL过期后由于在步骤2212中的缓存(例如,在这个例子中,在步骤2212中的缓存之后的24小时)被自动移除。在一般的情况下,生存时间(TTL)适用于整个缓存条目(例如,包括响应数据和包括有关周期性和用于计算周期的信息的信息的任何元数据)。在一个实施例中,缓存的响应数据TTL由默认值或某些其他值(例如,6小时、12小时、48小时等)被设定为24小时。TTL也可能是动态可调节的或在管理员/用户重新配置和/或基于不同情况、设备、应用程序、网络提供商,网络条件、运营商和/或特定用户而不同。

[0450] 图23描绘了一流程图,展示了通过网页内容缓存优化的无线流量减少的示例过程。

[0451] 来自网页服务器的网页内容按照步骤2302中缓存部件在移动设备本地缓存方式存储。步骤2304中,在移动设备处提出的用于网页服务器内容的请求被接收。请求可被任何移动应用程序或客户,包括但不限于,社会网络应用程序、任何基于浏览器的应用程序、游戏应用程序、地图应用程序、导航应用程序、规划应用程序、在线或移动市场应用程序、电子阅读应用程序/客户、日历、或电子邮件客户等提出。

[0452] 在步骤2306中,在本地缓存检索缓存部件。不管示出在网络内容的报头缓存过期,缓存部件能被检索到,这样至少部分能响应使用从本地缓存检索的一些缓存部件。该步骤还包括验证示出过期的缓存内容是否实际上发生变化,能进一步减轻并减少空气流量来满足移动设备应用程序/内容请求。

[0453] 例如,在一些示例中,缓存部件在本地缓存获取,并用于响应移动设备处的请求,即使超过报头过期时间。使用过期报头的缓存条目可通过使用用于内容报头的标签来管理,以确定网络内容的缓存部件是否在当前可利用于响应请求。尤其是,网页内容的报头标签被用来确定用于网页内容的缓存部件是否在当前可利用于响应请求,即使一些缓存部件在服务器处发生变化且一些没有变换。

[0454] 标签由代理服务器在无线流量管理系统中使用,其远离移动设备,来确定用于在本地代理的网页内容的缓存部件是否仍然有效。例如,在HTTP中,标签是一个用于网页缓存有效的HTTP报头实体标签,其允许客户发送传统请求,并且网页服务器并不需要去发送全部响应,仅内容部分发生变化/更新。

[0455] 在步骤2308中,计算用于随后请求的更新标签。这个更新标签由与移动设备通信的代理服务器计算。当代理服务器在移动设备的移动应用程序的内容服务器处发生内容变化时,代理服务器能周期性地询问内容服务器/应用程序主机,用于响应缓存于移动设备的请求。更新的标签能被计算来识别并确定哪个缓存部件是在移动设备且哪个不能(在网页

服务器处真实发生变化的)。在步骤2310中,更新的标签与识别有效响应随后请求的缓存部件的标签进行比较。

[0456] 在步骤2312中,验证器(例如,软验证器)由标签产生。在一些示例中,代理服务器能产生来自标签的验证器,与接收来自网页服务器的响应相关联,并示出哪个缓存部件被改变。验证器或软验证器能被使用,即使当失败机制被采用时。例如,即使在网页服务器或主机采用缓存击败的环境下,响应能仍然缓存在移动设备,且响应的部件能被验证使用标签或代理服务器的验证器的有效性。

[0457] 在步骤2314中,更新的验证器被重新计算来自网页内容的响应。更新的标签能被使用哈希公式计算(例如,抗碰撞散列函数)。在步骤2316中,更新的验证器与原始验证器比较缓存的内容。更新的验证器或更新的标签显示,在本地缓存的哪个缓存部件对于响应移动设备请求是有效的。

[0458] 在步骤2318中,确定哪个缓存部件对于响应移动设备请求是有效的。在一个实施例中,当更新的验证器或更新的标签分别与验证器或标签不同时,本地缓存的缓存部件被无效。在一个实施例中,报头(例如,未修改的报头)被返回到移动设备表明网页内容尚未被修改且缓存部件可能被使用。

[0459] 在步骤2320中,至少部分响应使用从本地缓存获得的一些缓存部件的请求。在响应移动设备处的应用程序或客户请求时,在网页服务器并未发生变化的那些缓存部件由本地缓存提供。已经改变/更新的响应部分能通过直播来自网页服务器或来自代理服务器获得。例如,如果请求能被缓存部件全部满足,移动设备的无线电不被激活来响应请求;并且无线电可在移动设备上满足已经发生变化的响应的部分。这个优化还缩小了需要通过网络(蜂窝网络)传输信息的数量。

[0460] 以往,内容传输网络或CDNsD 的内容加速通过将缓存带至网络边缘来实现,并且甚至在无用户活动下,可能主动地保持缓存更新。

[0461] 在本文公开的一个或多个实施例中,提供在终点(可能是一个或多个实施例中的移动设备)缓存网络另一端(可能是移动网络,可能具有较高传输成本特性,传输成本考虑带宽、延迟或用户直接成本)内容的方法。基于缓存的标准HTTP被RFC2616/章节13涵盖。这个标准缓存是基于用户的观察行为、并确保可被频繁使用的内容存储在移动设备的缓存内。本发明拓展了标准缓存通过下述:

(1) 利用来自其他用户的信息,该信息收集通过,例如,在网络边缘处的缓存服务器,或其他数据分析服务器,来预先取得由用户使用的特定数据集(或者延伸缓存周期超过缓存算法所使用的周期)。基于地理位置、使用网络(例如店内wifi、公共wifi、私人网络)、执行或安装在移动设备上的应用程序中的一个或多个识别的其他用户,整合使用类型、特定识别的相似使用流、人口统计资料、设备特征、所有要么通用时间或时间特定点(例如,当日时间、一周里的一天或动态的,与目前时刻或过去时刻相似的分类)。

[0462] (2) 利用预先定义的规则:

a. 规则,可至少部分被终点用户控制,选择应用程序或场所加速或使用,在一日中的何时进行加速,间接加速被观察在某时间(当日、一周里的一天等)使用或终点用户(例如企业IT管理器)代表使用的内容。

[0463] b. 规则,可至少部分被至少部分网络的管理器控制(例如CDN,移动网络操作器)-

例如通过内容所有者协议(或由内容所有者间接控制的规则),来加速用户内容。这可能伴随调整给终点用户的内容价格;例如零税率预先获取或所有的加速内容;或更昂贵的价格。

[0464] (3) 终点可能还包括压缩算法,该算法与由某方操作的网络部件一同工作,该方运行至少部分网络路径到内容所有者(例如,CDN,或移动网络操作器,本地wifi操作器,或事业网络操作器)。这似乎对于双方在终点和内容服务器的客户来说是透明代理。终点还可提供压缩算法控制给终点用户(例如,图像和影像质量的理想水平,或者压缩或数据保存的理想水平),还与客户端通信配置压缩参数给下游流量的参数(来自内容源设备的流量)。

[0465] (4) 终点可收集网络事务数据,包括那些通过一些网络部件参与加速、压缩或提供其他功能给数据(例如,为隐私移除第三方记录程序来修改数据)。它可收集与来自终点侧经验的质量相关的指标(例如,延迟、吞吐、TCP连接延迟、HTTP响应代码),以及标签如具有服务网络部件识别的信息(例如,位置、cellids、区域代码,服务事务的代理/CDN 服务器(例如,来自HTTP报头),或者IP地址或者请求的主机名(可识别最终目的地,或者服务请求的CND服务器))。这个信息可被用于测量来自终端用户/终点侧的网络性能,查明在特定路径中内容源的区别/异常/问题,并潜在地识别这个差异的根源。数据分析,如变量分析(ANOVA)可被用于识别显著偏差。

[0466] 根据一实施例,流量优化可包括内容加速。内容加速可动态预先获取选择内容至设备,用于提供较快访问至所述内容。动态获取可被用于与本文描述的边缘缓存服务相结合,如蜂窝和WLAN网络。动态获取的示例使用例可包括Pos WLAN,企业应用程序以及其他的内容加速还可包括SPDY压缩和HTTP内容加速。SPDY是一个开放网络协议,主要开发于Google,用于传输网页内容。SPDY操作HTTP流量,具有特定目标:减少网页负荷延迟并提高网页安全。SPDY通过压缩、复用和优化获得减少延迟,尽管这取决于网络和网页调度条件的混合。“SPDY”名字是Google商标而不是首字母缩写。回到流量优化,除了内容加速外,流量优化还可包括终端用户控制基于云图像和视频优化,以及改善在设备DNS上的缓存,该缓存有用于提供快速浏览性能的云帮助的届期。

[0467] 本发明因此提供一个或多个方法。例如,提供一种通过缓存保存网络带宽的方法。所述方法包括在移动设备接收信息或关于另一用户的特征,确定基于信息缓存的信息或关于另一用户的特征,以及缓存在移动设备确定的信息。

[0468] 根据一个或多个实施例,提供一种通过缓存保存网络带宽和资源的方法。该方法包括接收关于第一用户的信息或特征,确定基于信息缓存的信息或关于第一用户特征的信息,并预先缓存确定信息在第二用户的移动设备上。

[0469] 根据一个或多个实施例,关于用户的信息或特征包括至少一个:地理位置、使用网络、操作或安装在移动设备上的应用程序、合并使用模式、在第一用户移动设备和第二或另一用户的移动设备之间的相似使用流、用户人口统计资料;以及时间的特定点。

[0470] 根据一个或多个实施例,来自其他或第一用户的信息或特征由网络边缘的缓存服务器收集。

[0471] 根据一个或多个实施例,来自其他或第一用户的信息或特征由其他数据分析服务器收集。

[0472] 根据一个或多个实施例,该方法包括延伸缓存周期超过使用缓存算法的周期,否则基于信息或特征。

- [0473] 根据一个或多个实施例,缓存或预缓存遵守一个或多个规则。
- [0474] 根据一个或多个实施例,一个或多个规则被至少部分由终端用户控制。规则可由终端用户全部控制。
- [0475] 根据一个或多个实施例,一个或多个规则包括至少一:选择哪个应用程序或场所去加速,或一天的何时去加速使用,以及间接加速用户被观察用于特定时间或终端用户代表的内容。
- [0476] 根据一个或多个实施例,一个或多个规则被至少部分移动网络管理器控制,移动设备在该移动网络上操作。
- [0477] 根据一个或多个实施例,受规则控制的一个或多个操作器包括由于与内容所有者有协议建立的规则(或者直接由内容所有者控制的规则),来加速用户内容。该规则可通过调整给终端用户内容的价格执行,例如零税率预获取或所有加速内容,或者更贵价格。
- [0478] 根据一个或多个实施例,该方法包括依据压缩算法压缩信息。
- [0479] 根据一个或多个实施例,压缩算法基于与相关信息或从其他网络部件接收到的信息被确定。
- [0480] 根据一个或多个实施例,其他网络部件有CDN、移动网络操作器、本地WIFI操作器或企业网络操作器、
根据一个或多个实施例,信息压缩在代理服务器处压缩。
- [0481] 根据一个或多个实施例,该方法包括提供压缩算法控制给移动设备用户。
- [0482] 根据一个或多个实施例,控制算法通信给服务器。
- [0483] 根据一个或多个实施例,该方法包括收集网络事务数据。
- [0484] 根据一个或多个实施例,数据包括至少一个与来自终点侧经验的质量相关的指标,如延迟、吞吐、TCP连接延迟、HTTP响应代码;服务网络部件识别如位置、cell ids、区域代码,服务事务的代理/CDN 服务器,如来自HTTP报头或IP地址或可识别最终目的地的请求的主机名,或者服务请求的CDN服务器。
- [0485] 根据一个或多个实施例,数据被用于测量来自终端用户/终点侧的网络性能,以及查明不同/差异/问题在特定路径的内容源,并潜在识别这个差异的根源。
- [0486] 根据一个或多个实施例,提供一移动设备。移动设备具有存储其上的指令,当执行使得移动设备实现本文所述的任意方法。
- [0487] 根据一个或多个实施例,提供一系统。该系统包括移动设备、代理服务器、网络服务器,被配置用来是实现本文所描述的任意方法。
- [0488] 根据一个或多个实施例,通过缓存保存网络带宽的方法包括在移动设备处接收用户信息或特征,确定基于所述用户信息或特征的缓存信息,在移动设备处缓存基于一个或多个规则的确定信息。
- [0489] 根据一个或多个实施例,提供一通过缓存保存网络带宽和资源的方法。方法包括接收关于第一用户的信息或特征,确定基于第一用户信息或特征的缓存信息,并预获取基于一个或多个规则的第一用户移动设备上的确定信息。
- [0490] 根据一个或多个实施例,方法包括在服务器端或服务器代理处压缩信息,依据由移动设备用户确定的算法。
- [0491] 根据一个或多个实施例,压缩算法基于相关信息或从其他网络部件接收到的信息

来确定。

[0492] 根据一个或多个实施例,其他网络部件为CDN,移动网络操作器,本地WIFI操作器,或企业网络操作器。

[0493] 根据一个或多个实施例,信息在代理服务器处被压缩。

[0494] 根据一个或多个实施例,该方法还包括提供压缩算法控制给移动设备用户。

[0495] 根据一个或多个实施例,所述控制算法由移动设备传输至服务器。

[0496] 图27系一机器2700的图形表示形式,以包含指令集的计算机系统为例,执行该指令,使得机器运行本申请所探讨的一种或多种方法。

[0497] 其他实施例中,该机器能作为独立运行的设备,或可连接(如网络)到其它机器。在网络部署中,机器可能作为服务器或客户端机器在客户端服务器网络环境中运行,或作为对等(或分布式)网络环境中的同行机。

[0498] 该机器可能为服务器计算机、客户端计算机、个人电脑(PC)、用户设备、平板电脑、便携式电脑、机顶盒(STB)、个人数字助理(PDA)、移动电话、iPhone、iPad、黑莓、处理器、电话、web装置、网络路由器、交换机或网桥、控制台、手持式的控制台、(手持式)游戏设备、音乐播放器、笔记本电脑、移动,手持设备或能够执行该机器所采取具体操作指令集的机器(顺序或其它)。

[0499] 示范实施例中的机器可读介质或机器可读存储介质为单一介质,术语“机器可读介质”和“机器可读存储介质”应包含存储一个或多个指令集的单个或多个介质(例如:集中式或分布式数据库,和/或相关的缓存和服务器等)。术语“机器可读介质”和“机器可读存储介质”也应包含能够存储,编码或执行机器所带指令集,并能够使机器执行一种或多种所公开的技术和创新方法的介质。

[0500] 一般来说,执行例程可作为一部分操作系统或特定应用程序,组件,程序,对象,模块或称为“计算机程序”的指令序列来实施公开的实施例。计算机程序通常由不同时期计算机内各内存和存储设备中的一个或多个指令集构成,当计算机一个或多个处理单元或处理器读取和执行指令集时,导致计算机执行涉及所公开各方面元素的操作。

[0501] 此外,本申请描述了实施例计算机和计算机系统的全部功能,本领域技术人员能够领会各实施例能分布成各种形式的程序产品,且所公开内容对实际影响分布均适用,不论机器的特定类型或使用的计算机可读介质。

[0502] 机器可读存储介质,机器可读介质或计算机可读(存储)介质的更多实例包括但不限于:可刻录类型介质,如易失性和非易失性存储设备、软盘和其它可移动磁盘、硬盘驱动器、光盘(例如:只读光盘存储器(CDROMS)、数字多功能磁盘(DVD)等)及其他,和传输类型介质,如数字和模拟通信链路。

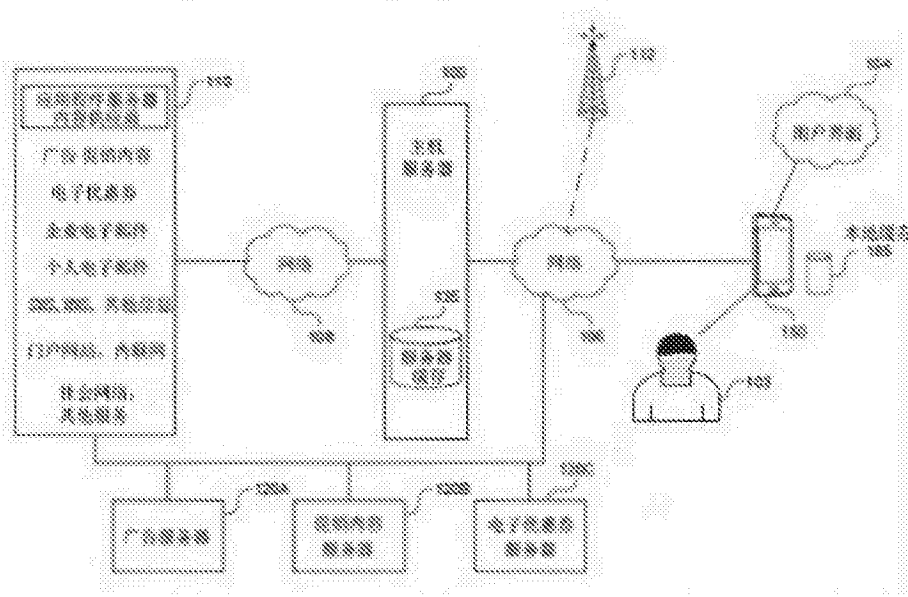


图1A

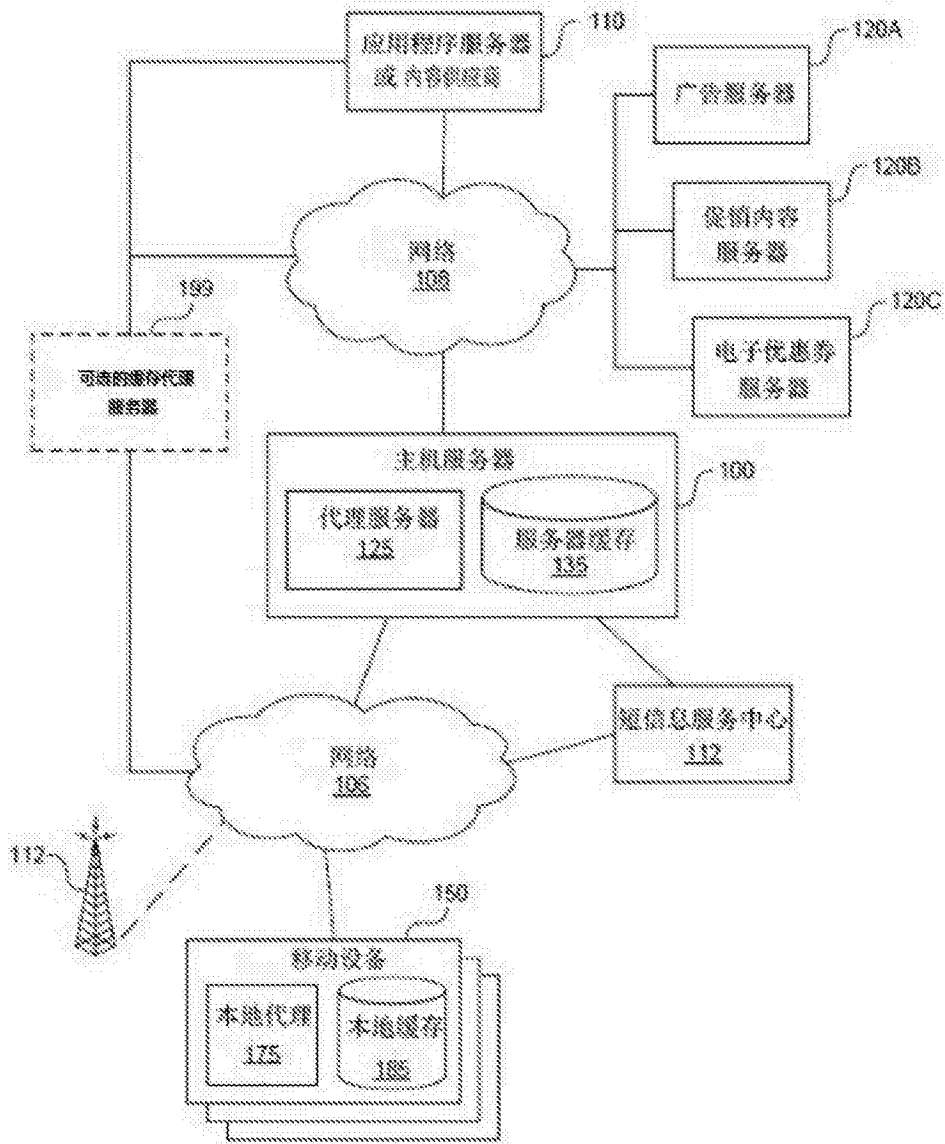


图1B

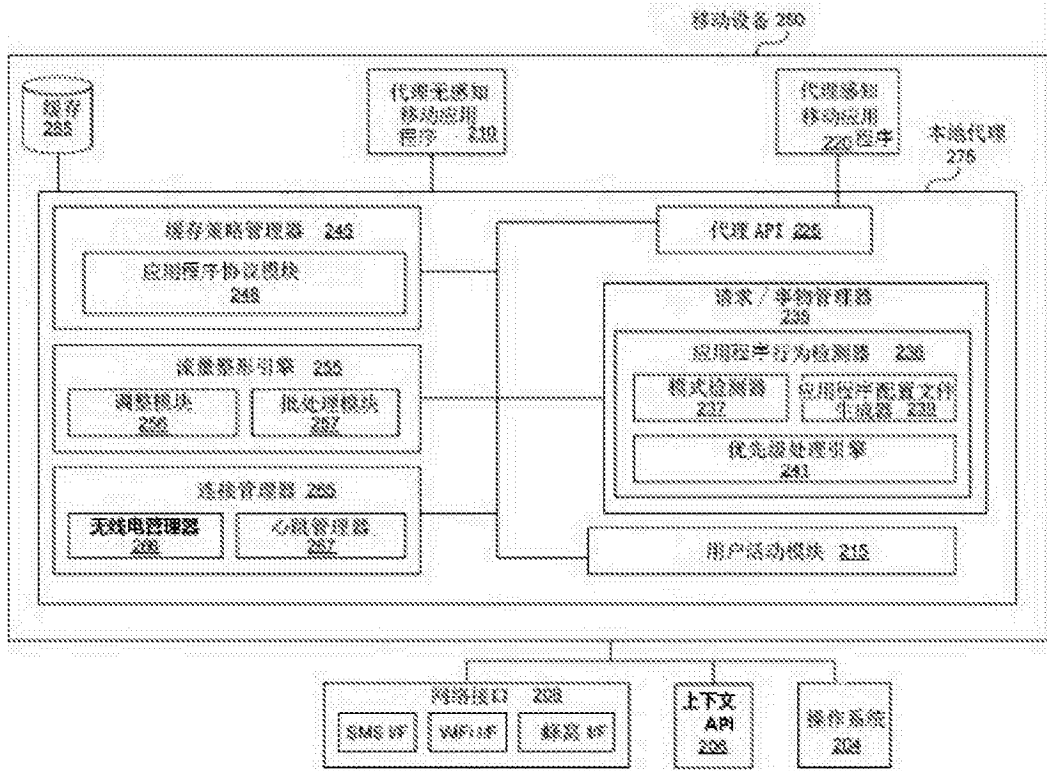


图2A

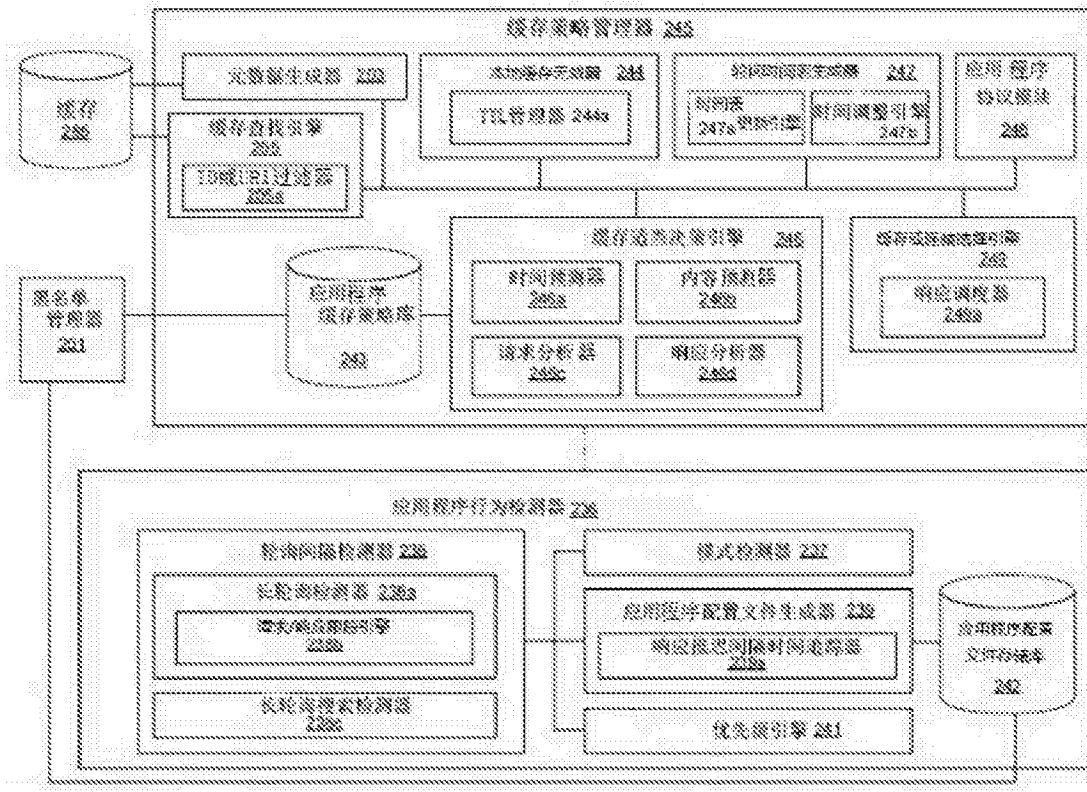


图2B

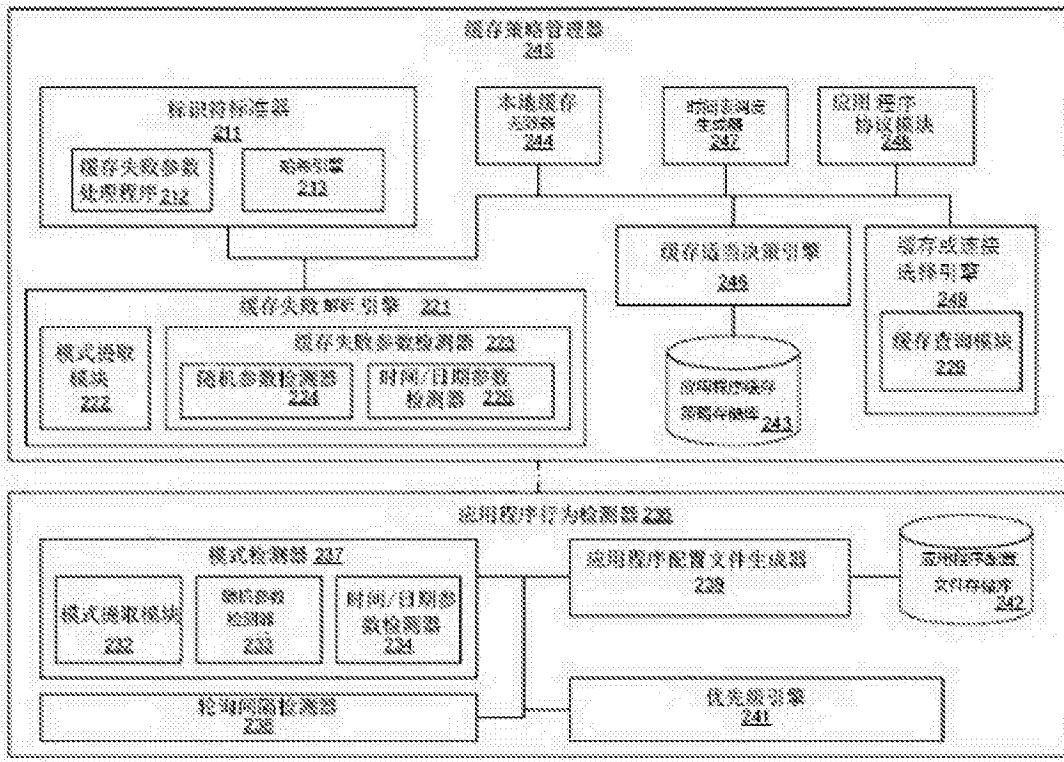


图2C

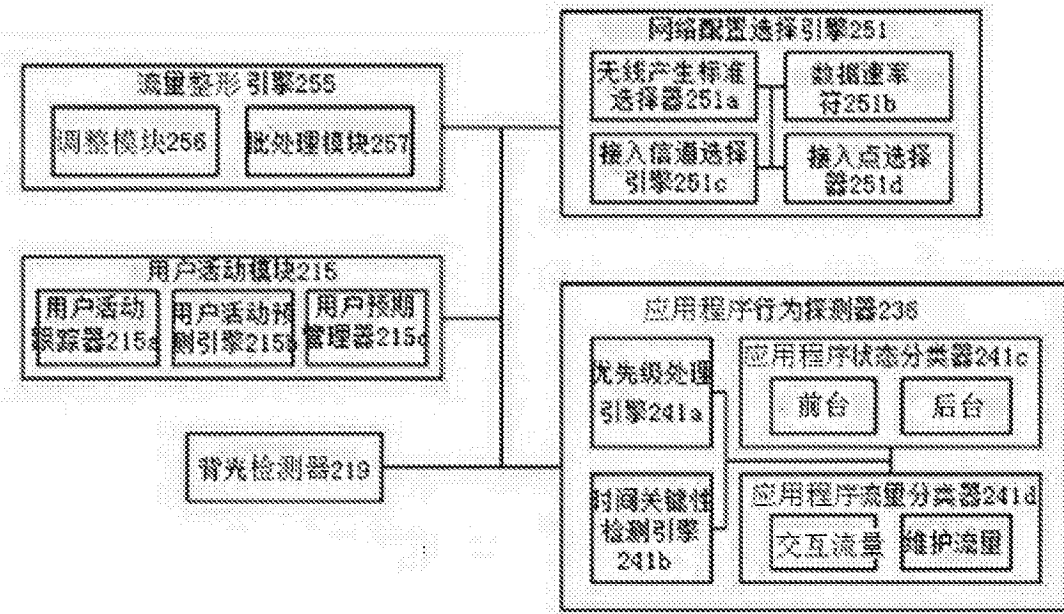


图2D

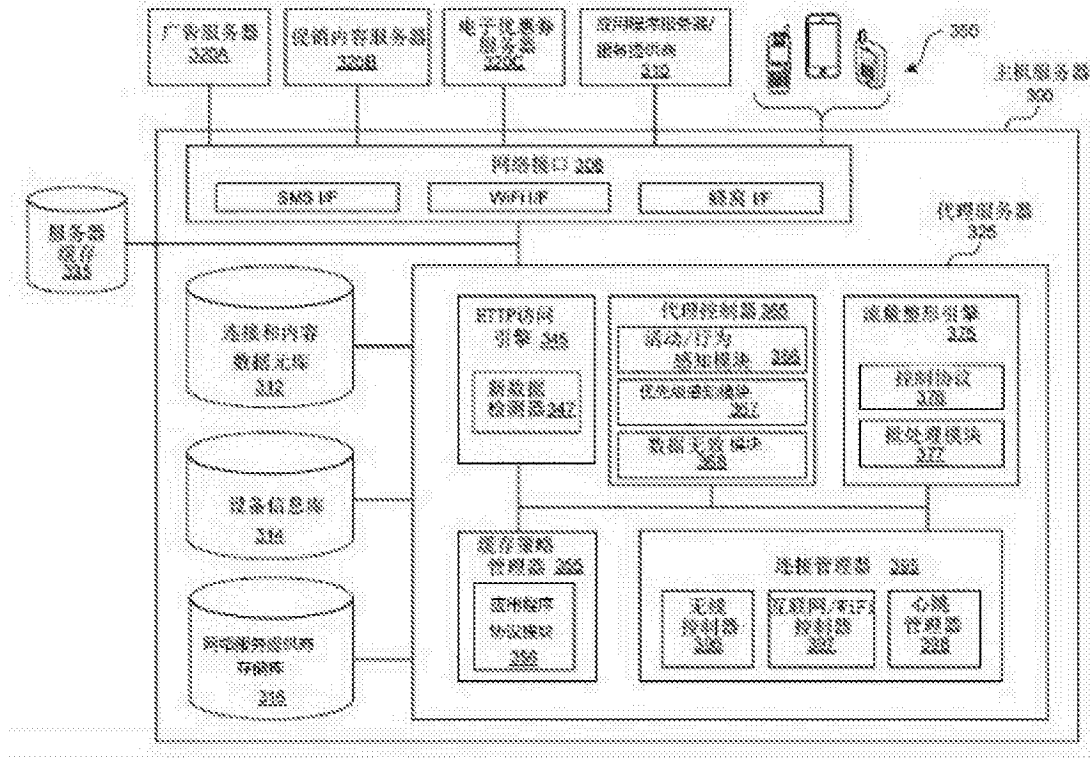


图3A

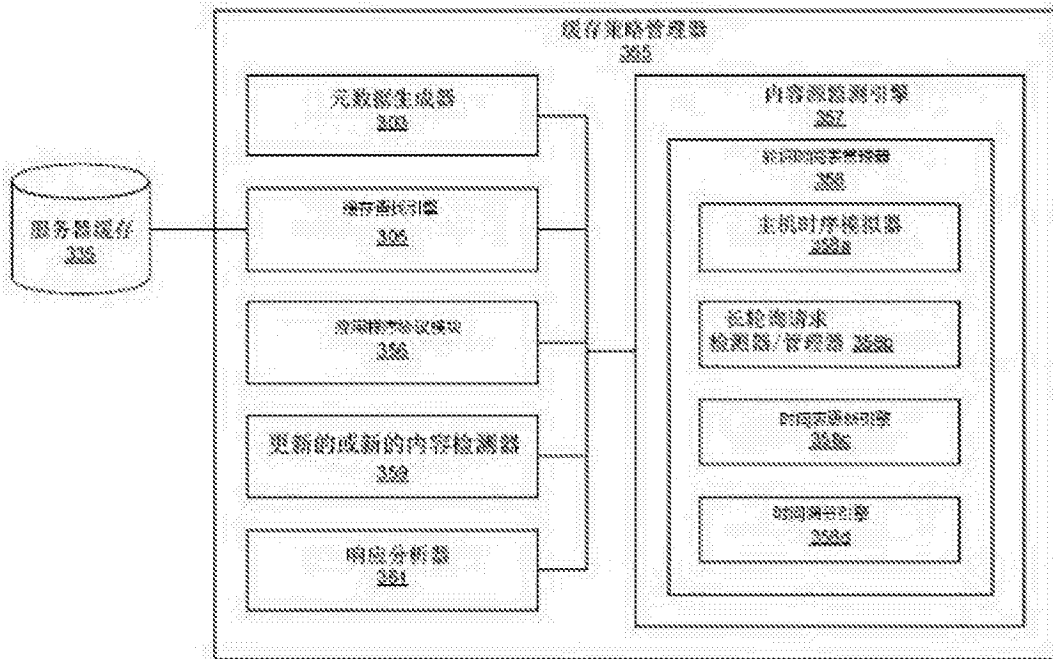


图3B

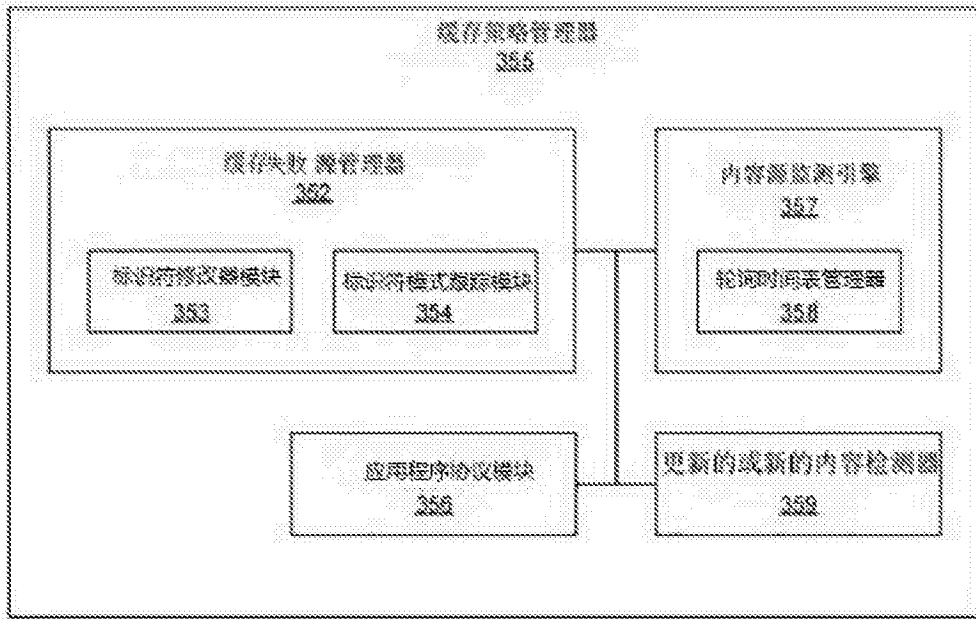


图3C

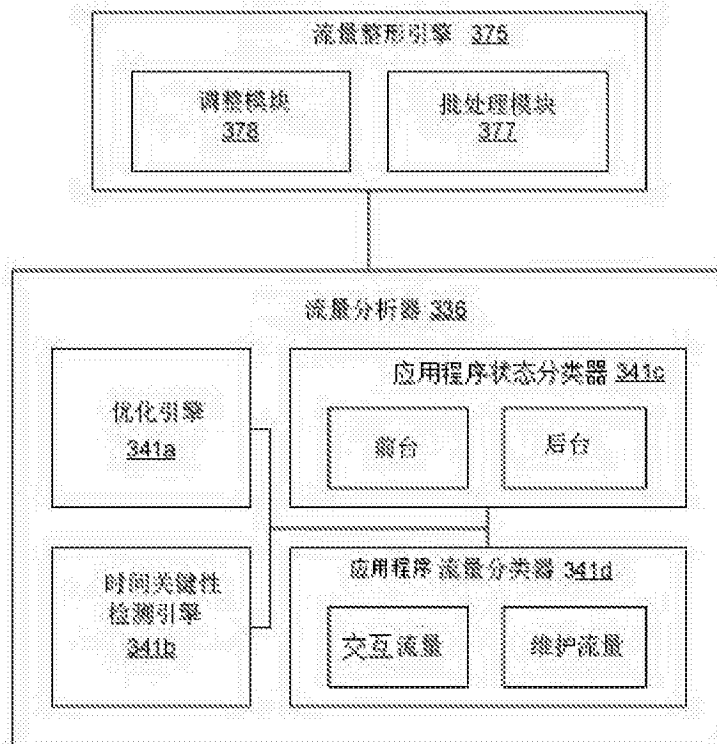


图3D

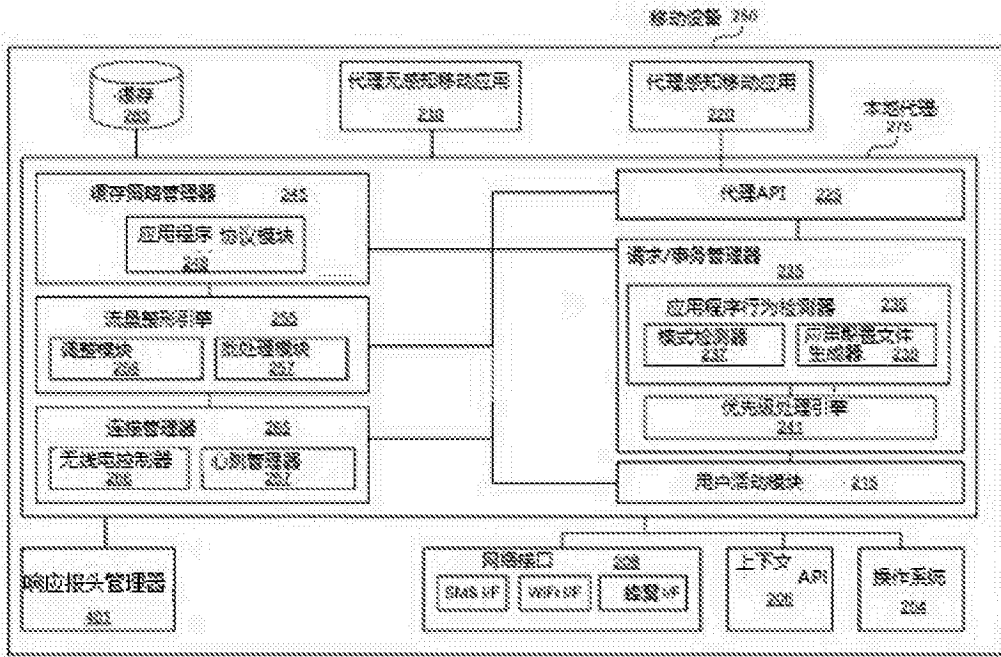


图4A

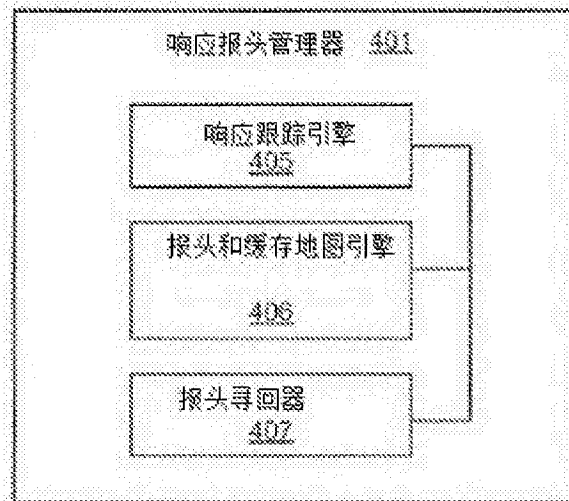


图4B

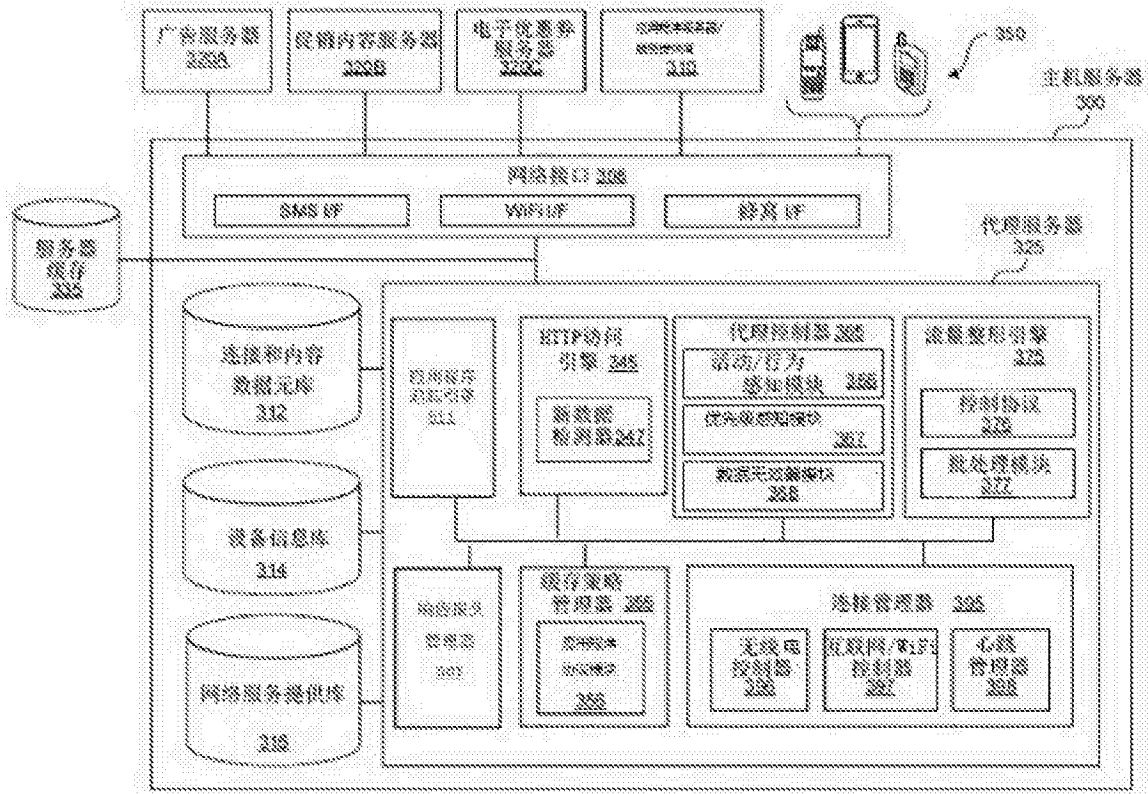


图5A

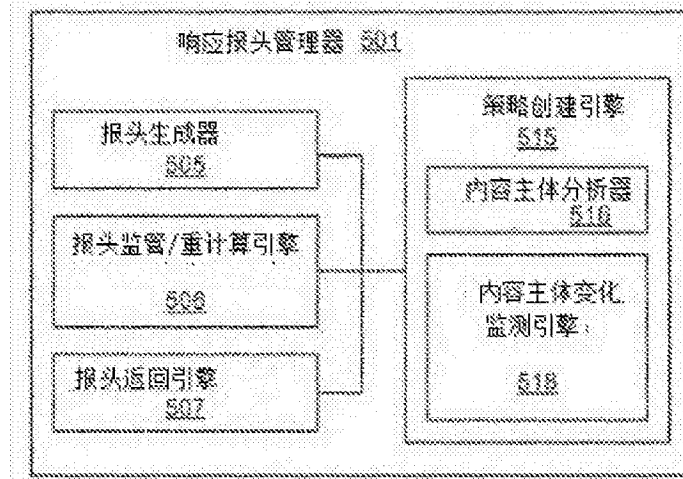


图5B

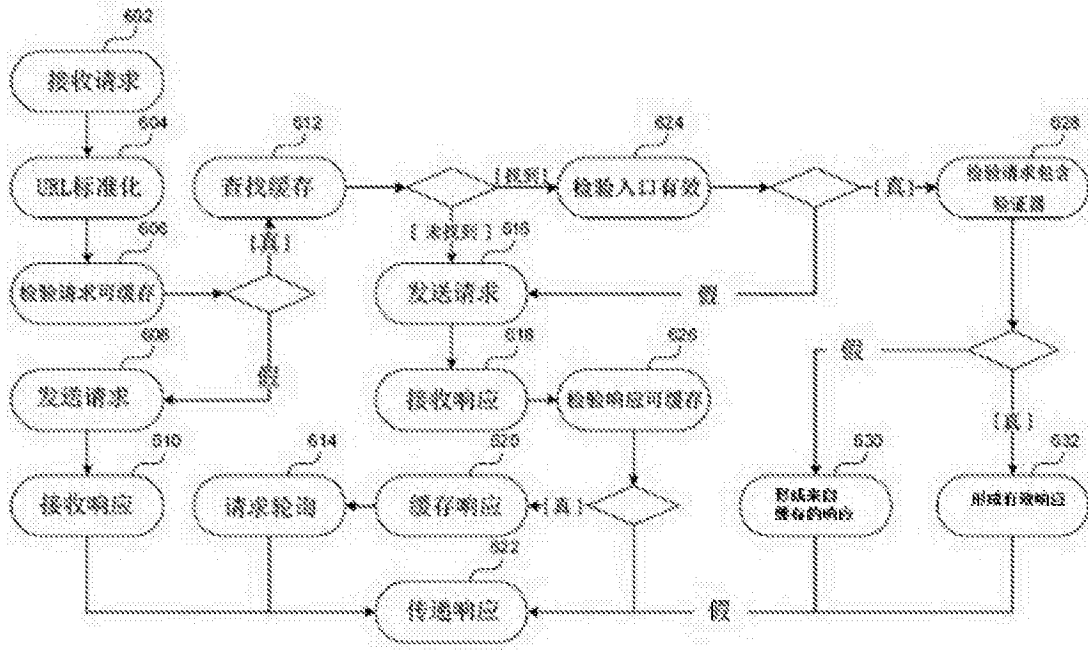


图6A

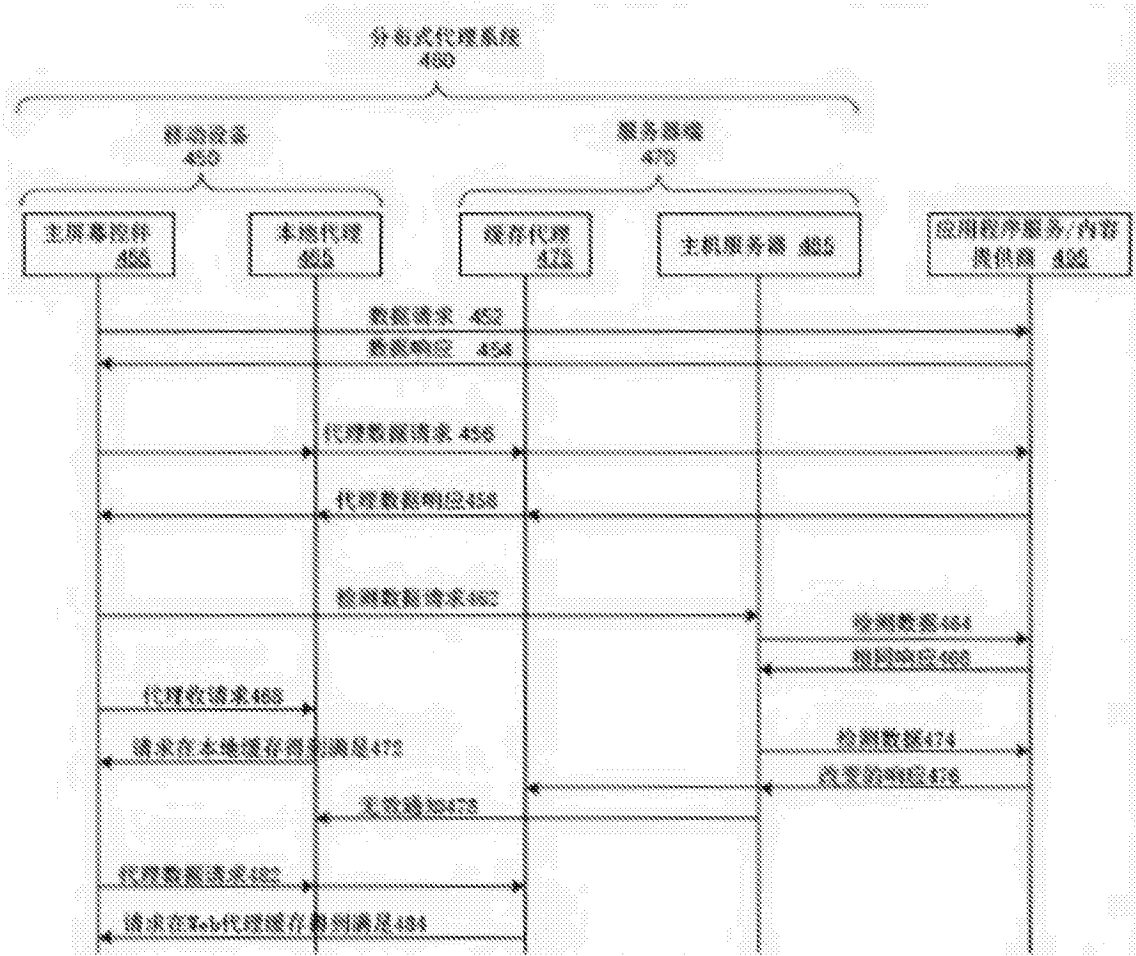


图6B

流量分类/应用程序分类类型700	
交互流量	后台流量
用户等待响应	用户未等待响应
前台应用程序	后台应用程序
背光灯开启	背光灯关闭

图7

内容分类类型800	
高优先级	低优先级
关键时间	非关键时间

图8

移动应用程序/部件 808	本地代理 809	主机服务器 806 服务器组 923或缓存代理923	应用程序服务器/ 内容提供商925
移动应用程序服务器/提供者922	接收缓存 924 代理检测缓存内容是否可用于满足中间 程序内容系统的需求或满足其他的需求926		
接收去缓存来自远程的响应925			
移动应用程序服务器/提供者922	接收缓存 922 代理检测缓存是否可用并 决定为缓存建立地址924		
	检测缓存是否可用 922		接收来自应用程序的缓存请求 并检测响应以确定是否请求926
接收来自应用程序服务器/提供者的 响应925, 以决定是否请求922	根据应用程序的缓存需求并检测是否请求 缓存或是否请求服务器922		
	发送缓存设置给主机服务器924	接收缓存建立请求以检测缓存的应用程序服 务器/提供者的缓存需求926	
		在应用程序服务器/提供者以检测 代理的缓存需求922	接收来自主机服务器的 缓存请求926
		检测缓存是否可用, 并检测是否请求缓存 或是否请求922	接收来自服务器的缓存 请求926
		检测变化的或新的响应, 通知本地代理926	
		存储或从缓存的响应或缓存 缓存或缓存代理922	
	接收新的或更新的响应或可利用的响应, 指示缓存是否可用或是否922		
移动应用程序服务器/提供者 922	确定无缓存是否作为可利用的并检测是否 请求缓存或缓存的响应924	接收响应的请求并发送响应给应 用系统925	
接收缓存或缓存请求或缓存代理923			
移动应用程序服务器/提供者 922	确定无缓存是否作为可利用的并 转发或向应用程序服务器/提供者922		接收缓存并检测响应924
接收来自应用程序服务器/提供者 响应925			

图9

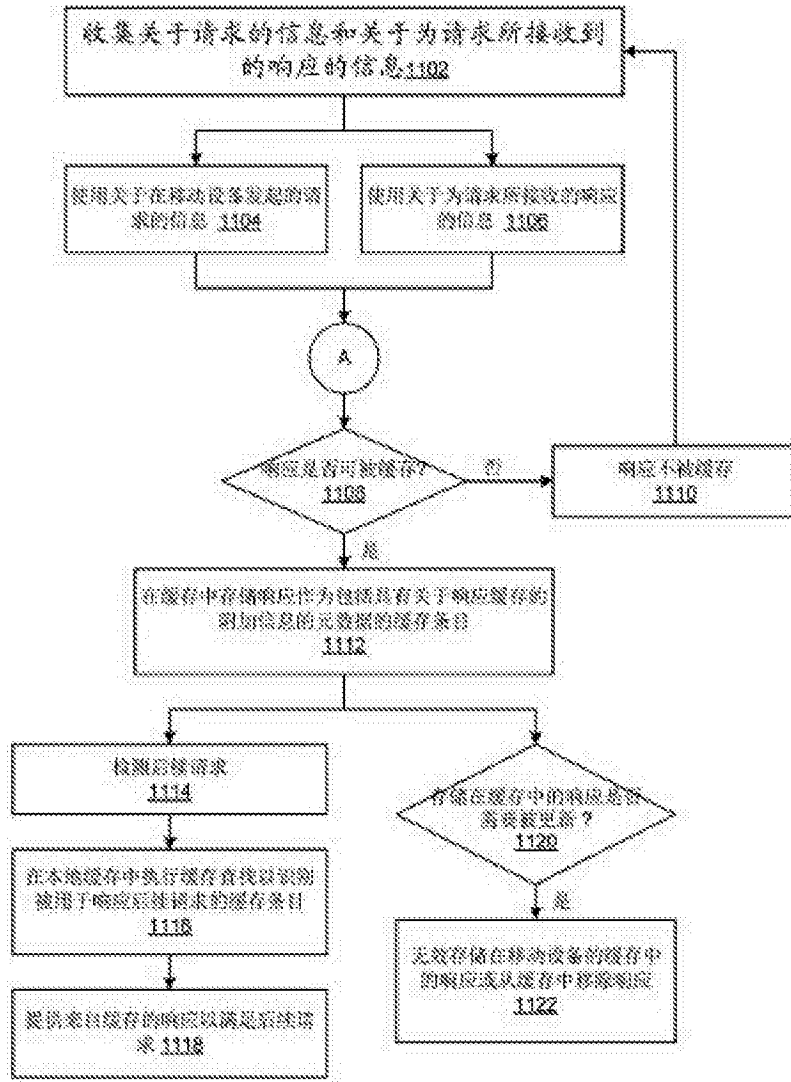


图11

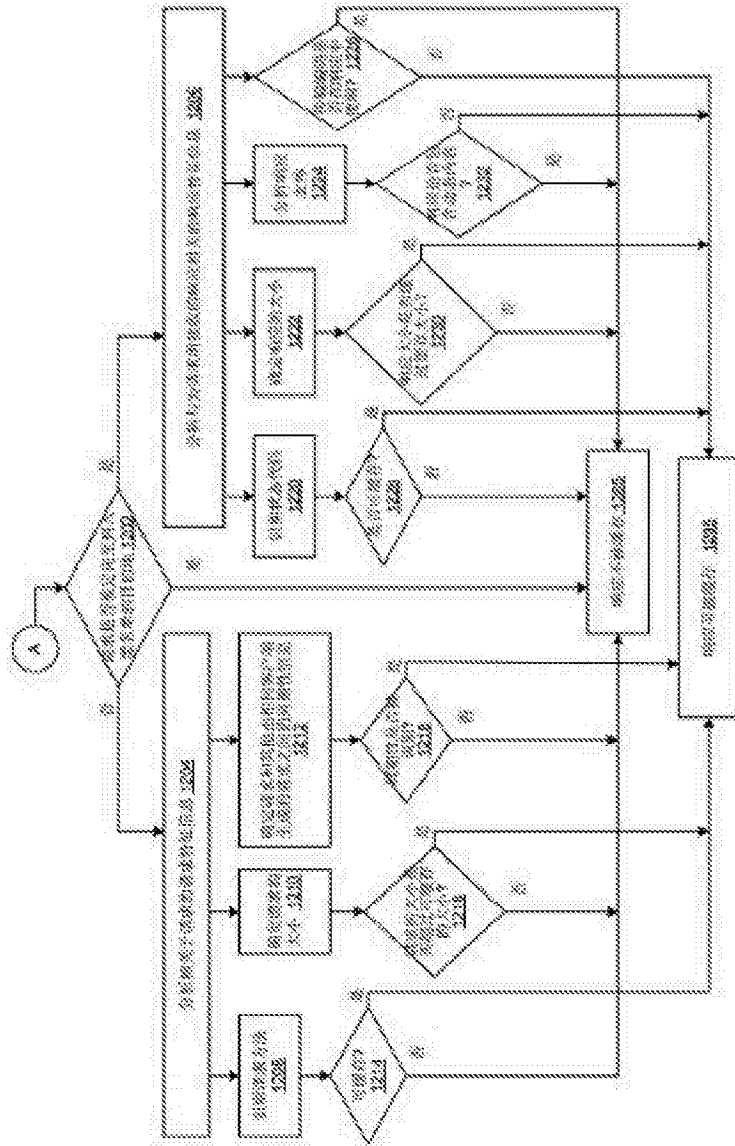


图12

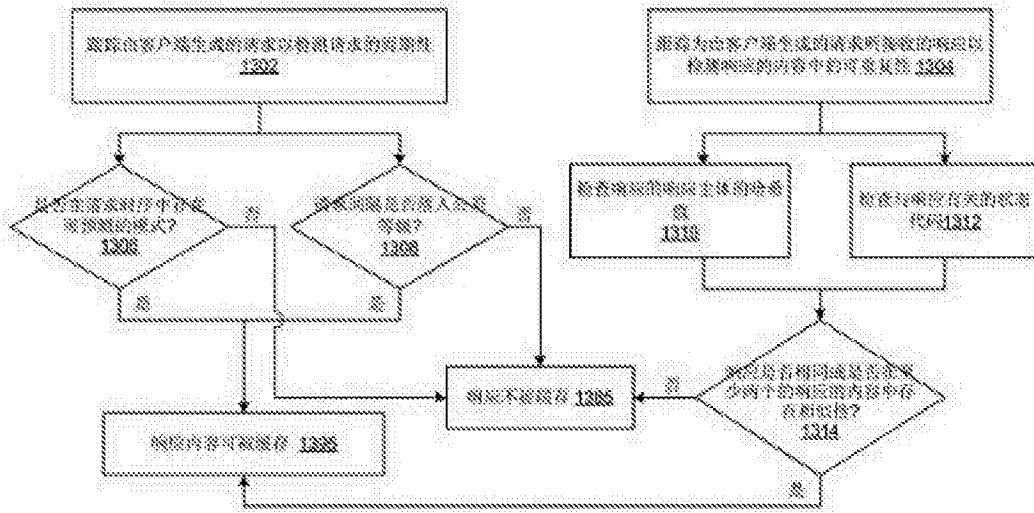


图13

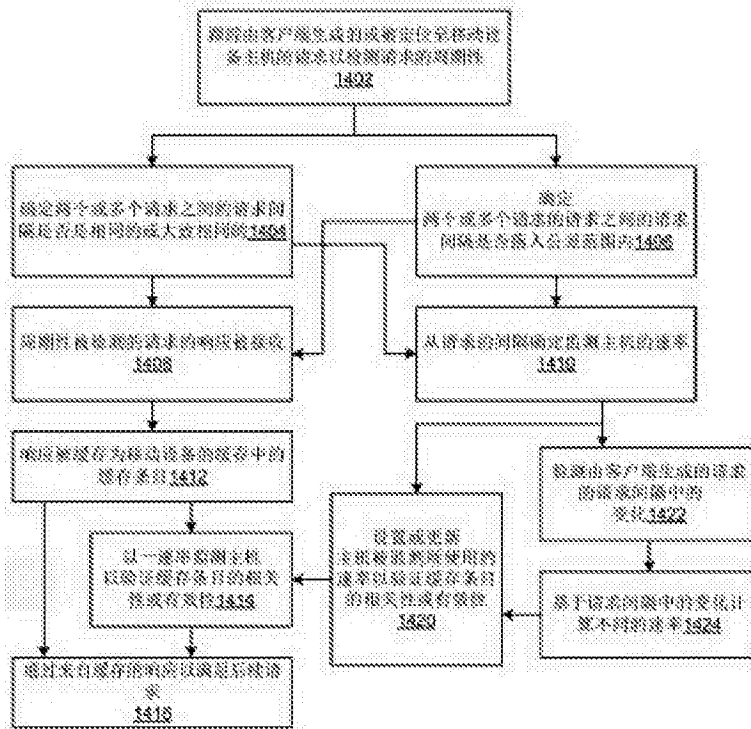


图14

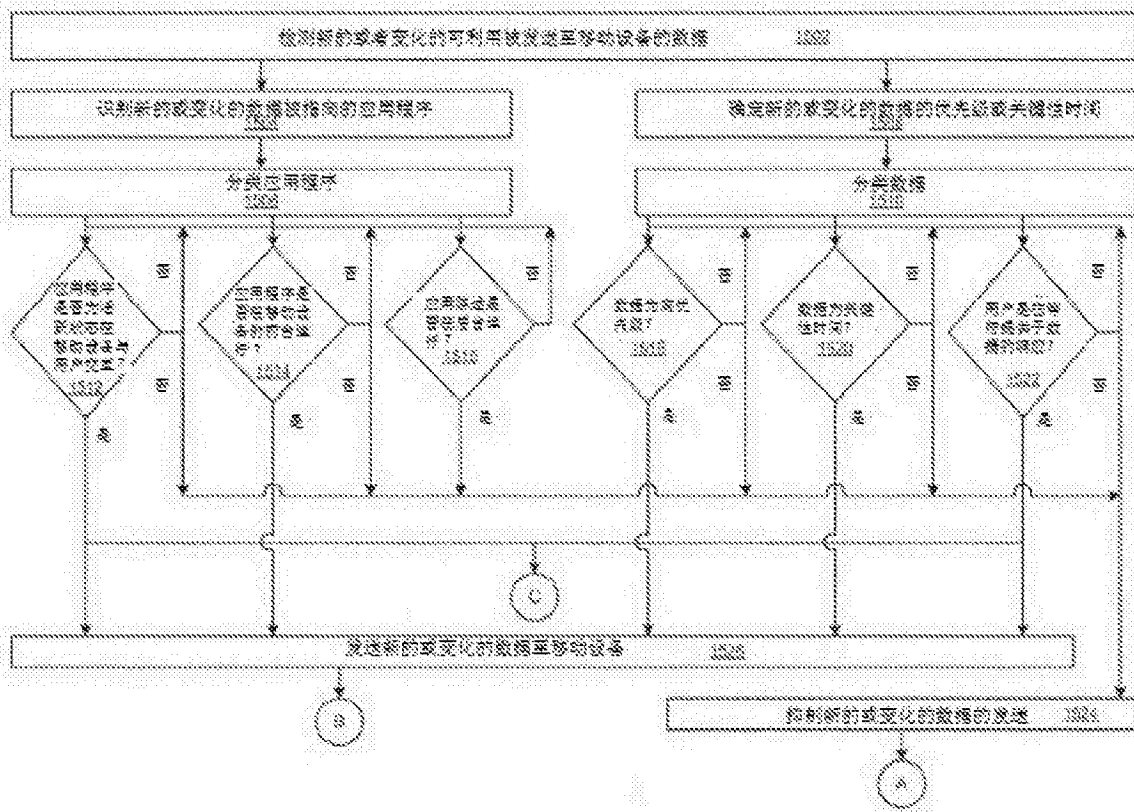


图15

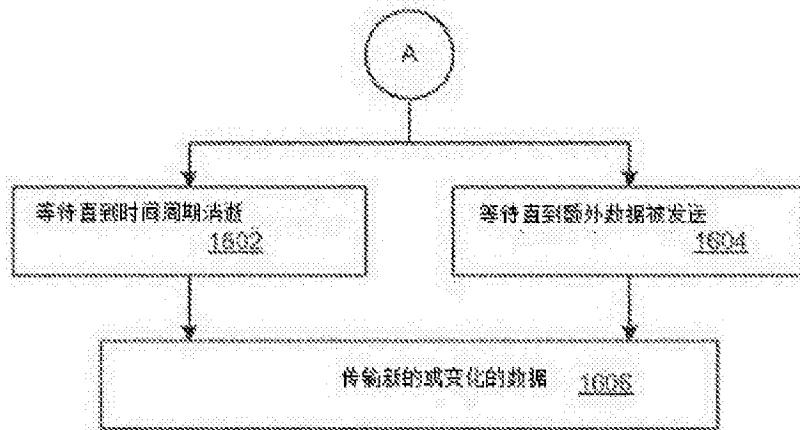


图16A

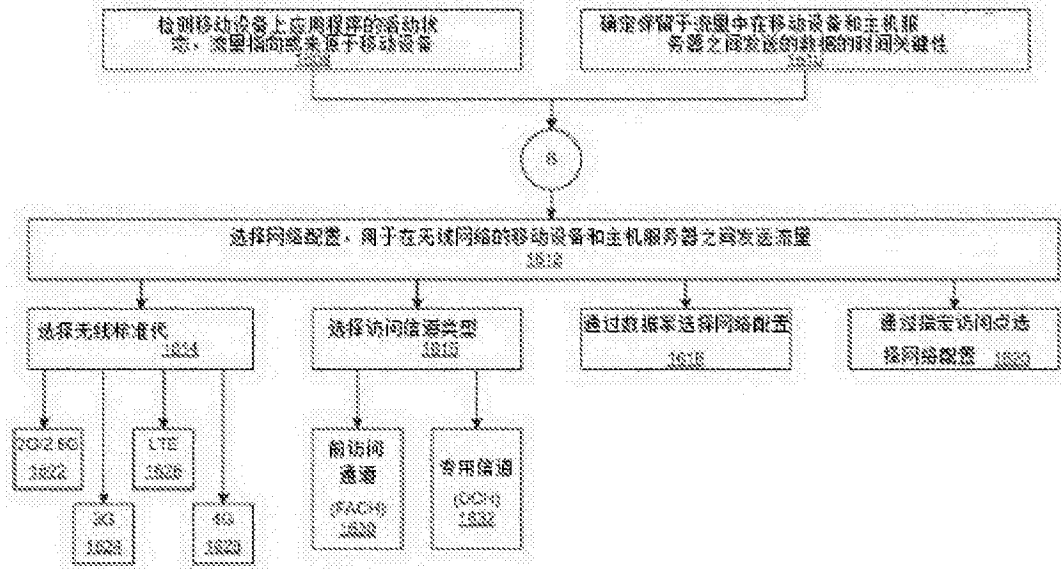


图16B

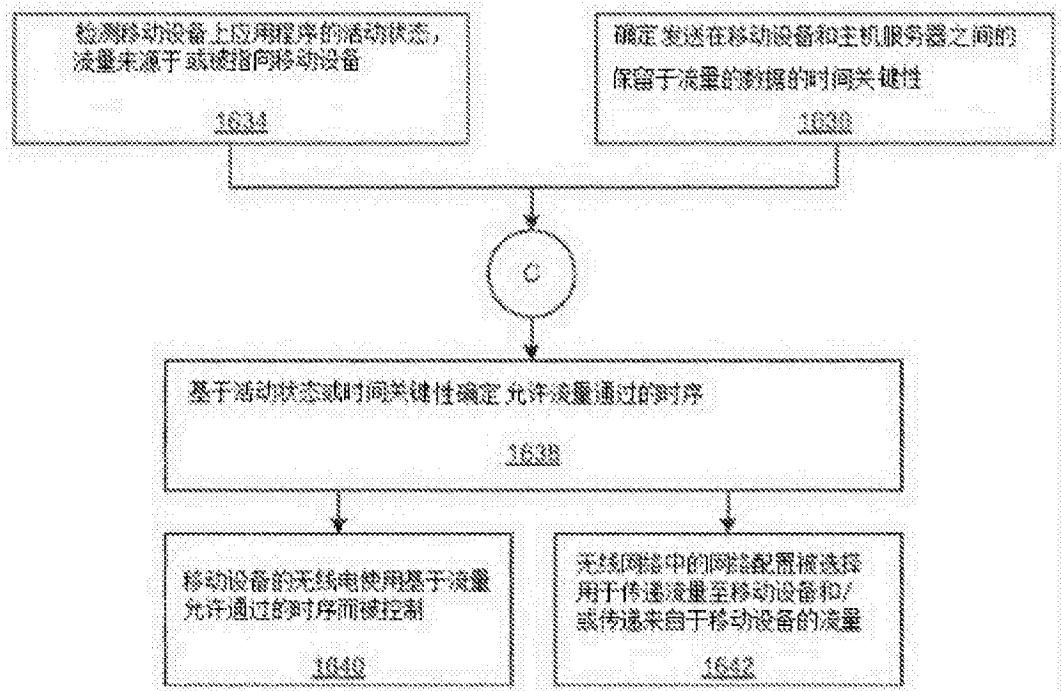


图16C

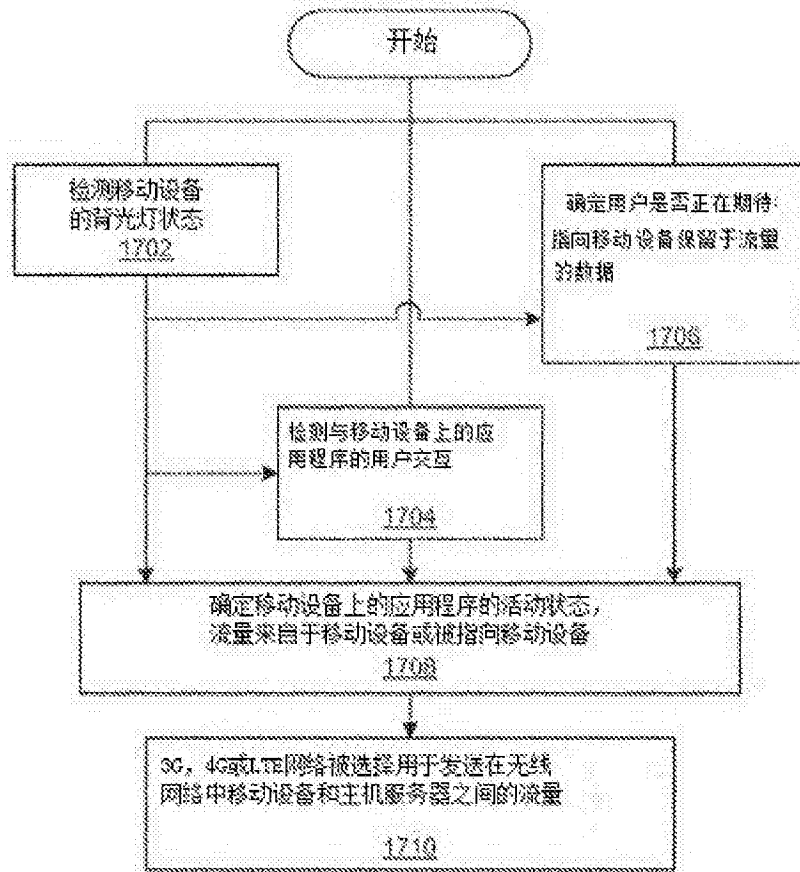


图17

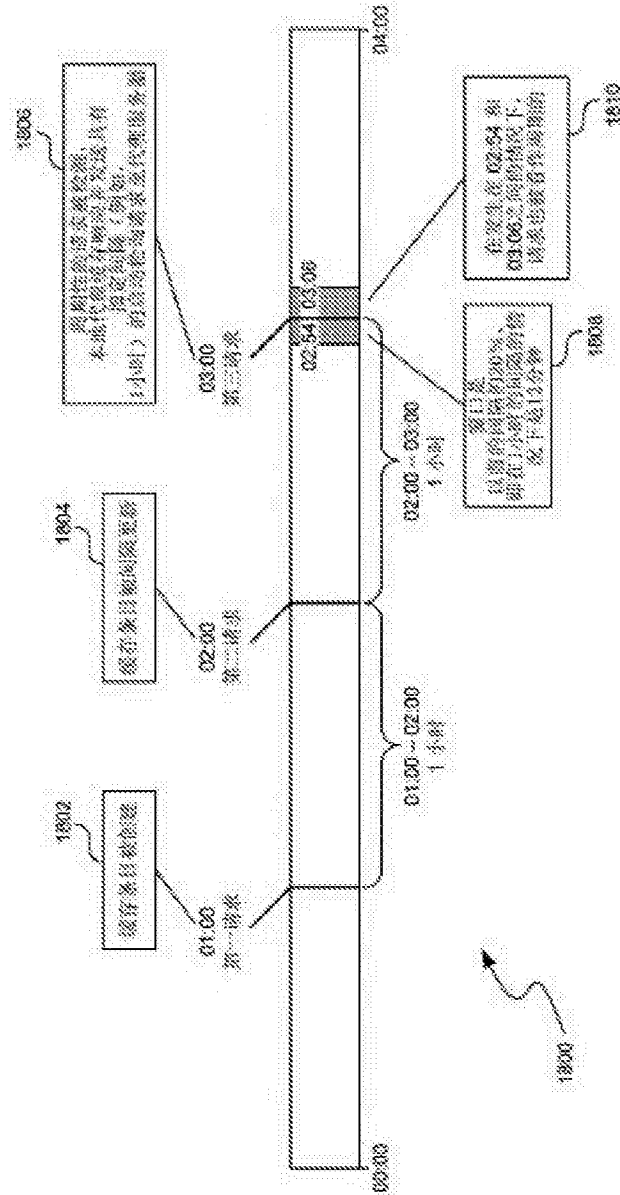


图18

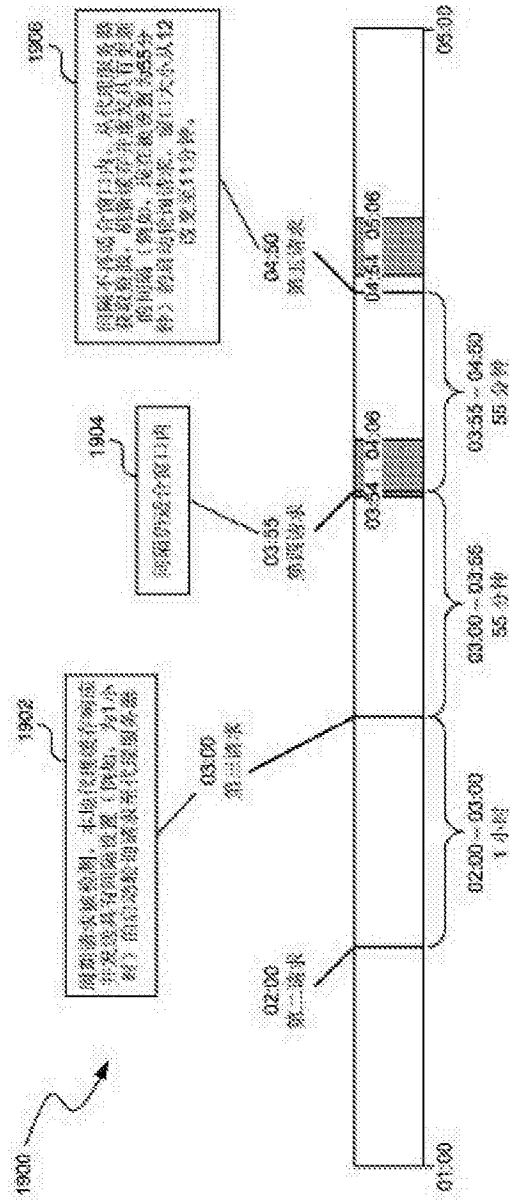


图19

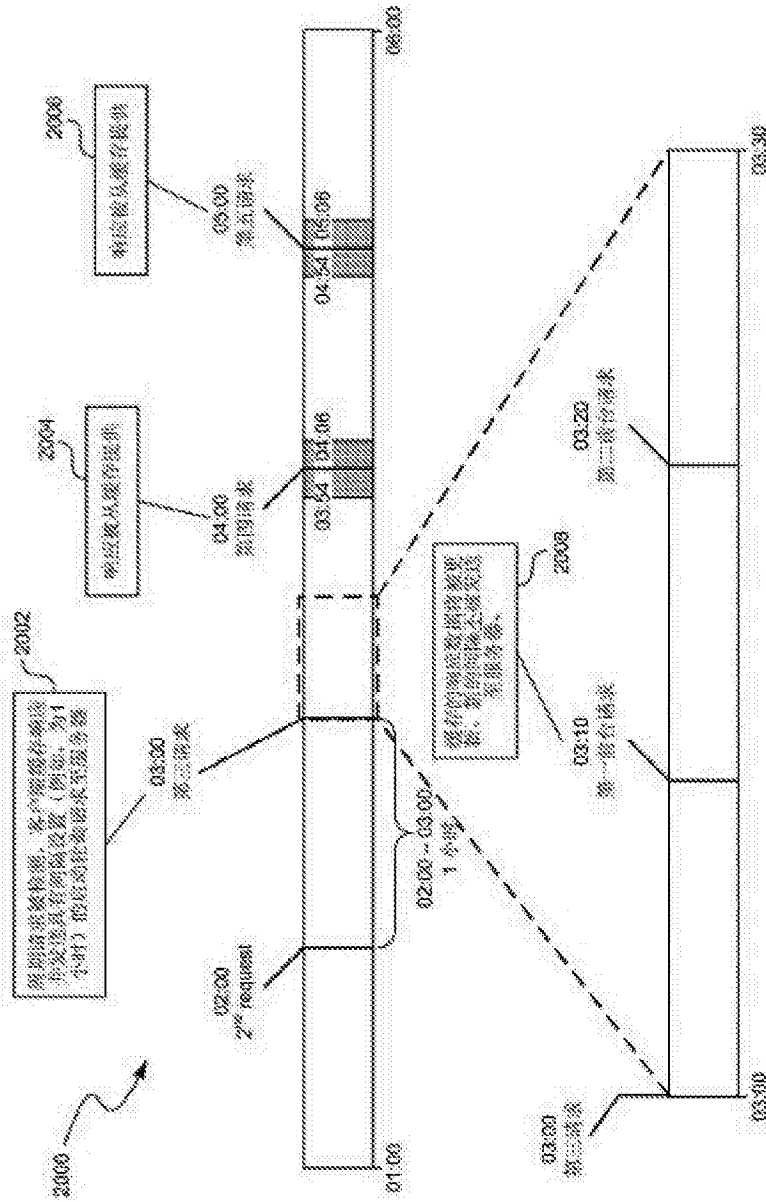


图20

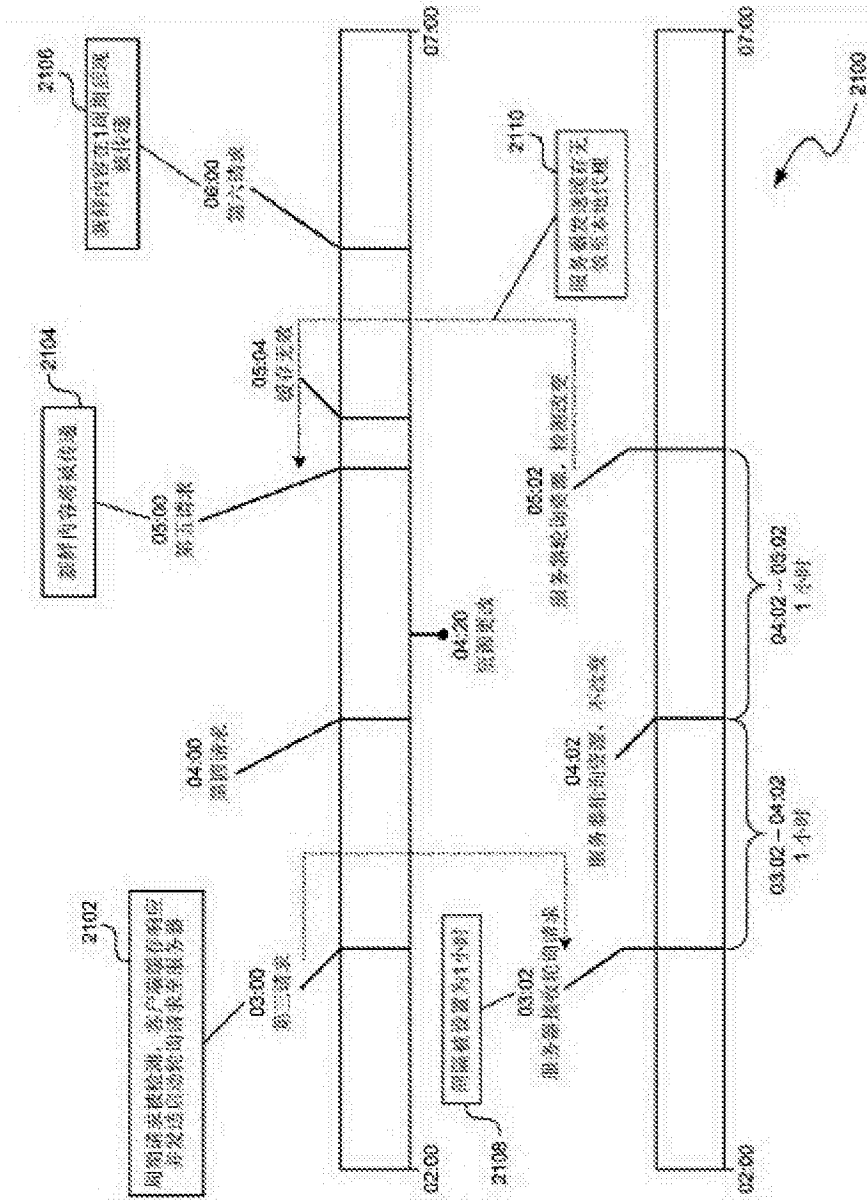


图21

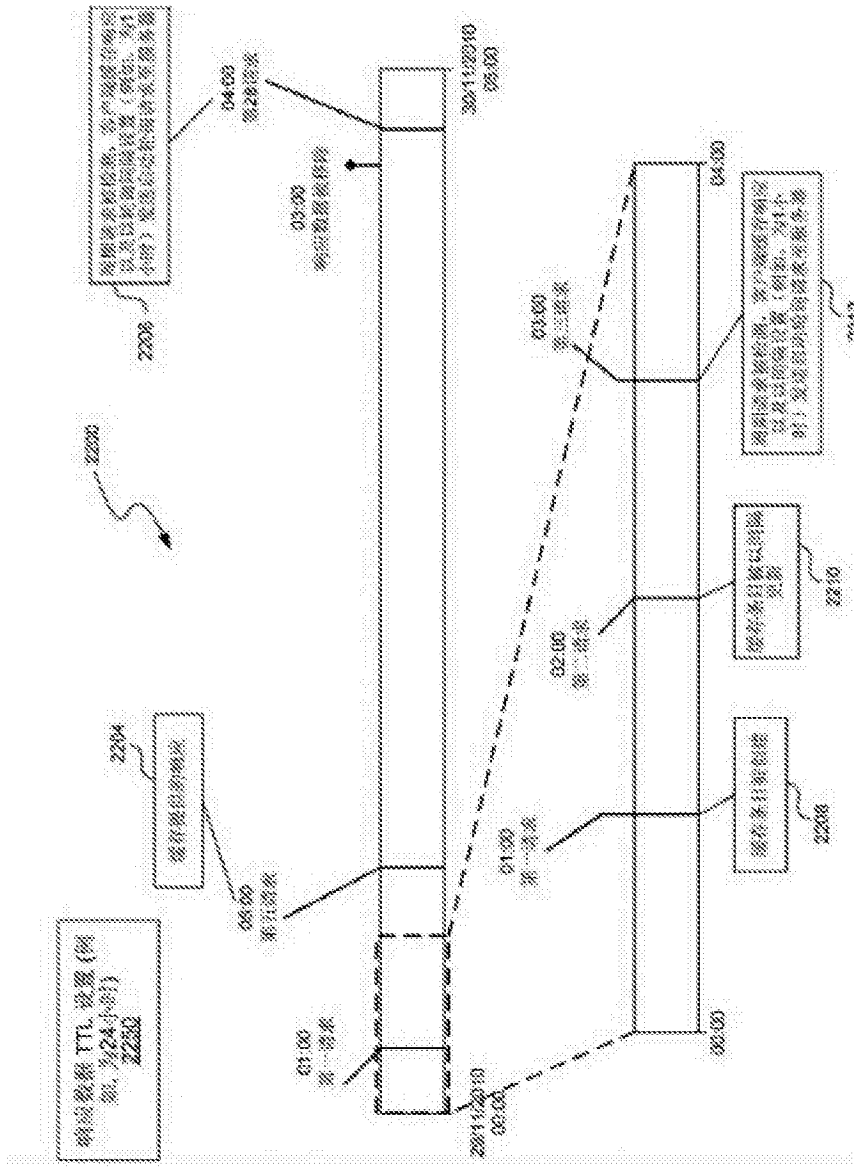


图 22

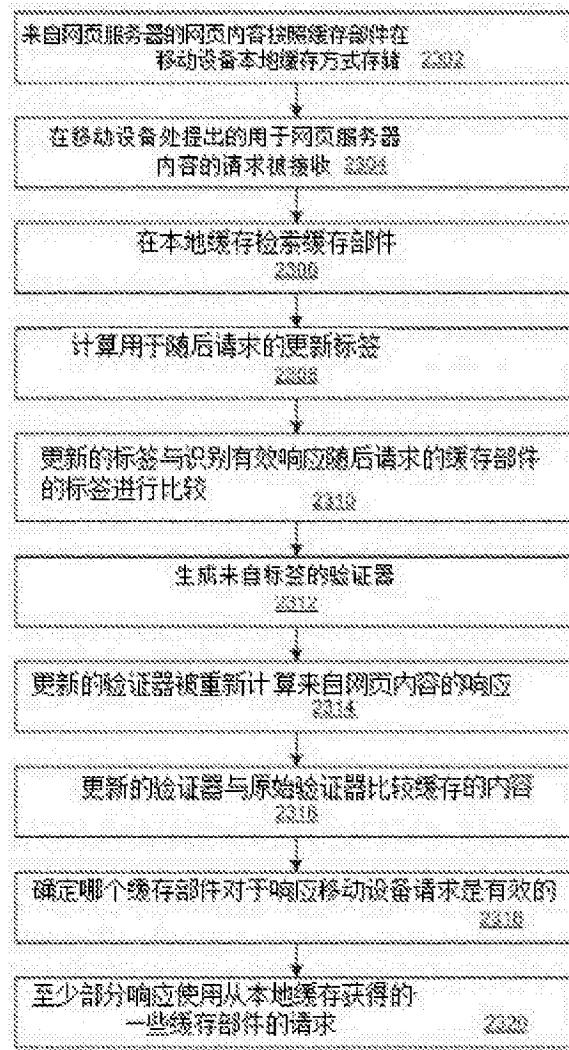


图23

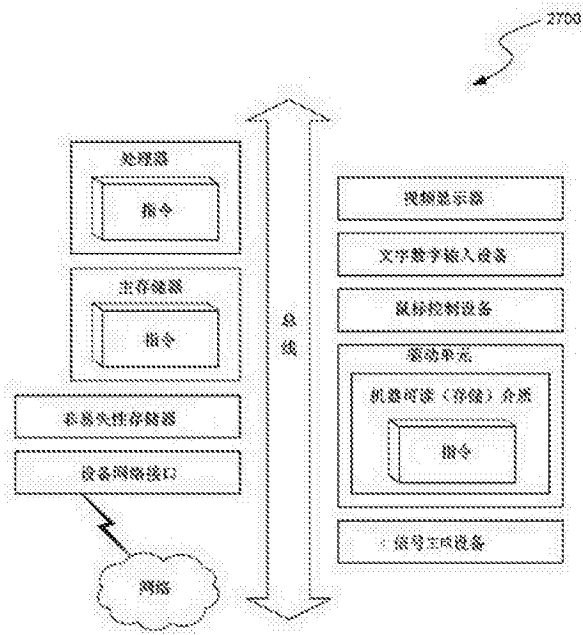


图24