

(19) **United States**

(12) **Patent Application Publication**
LI

(10) **Pub. No.: US 2016/0113018 A1**

(43) **Pub. Date: Apr. 21, 2016**

(54) **RESOURCE MANAGEMENT IN CLOUD-BASED RADIO ACCESS NETWORK**

(71) Applicant: **FUJITSU LIMITED**, Kawasaki-shi (JP)

(72) Inventor: **Zhaojun LI**, Guildford (GB)

(21) Appl. No.: **14/675,332**

(22) Filed: **Mar. 31, 2015**

(30) **Foreign Application Priority Data**

Oct. 20, 2014 (EP) 14189490.7

Publication Classification

(51) **Int. Cl.**

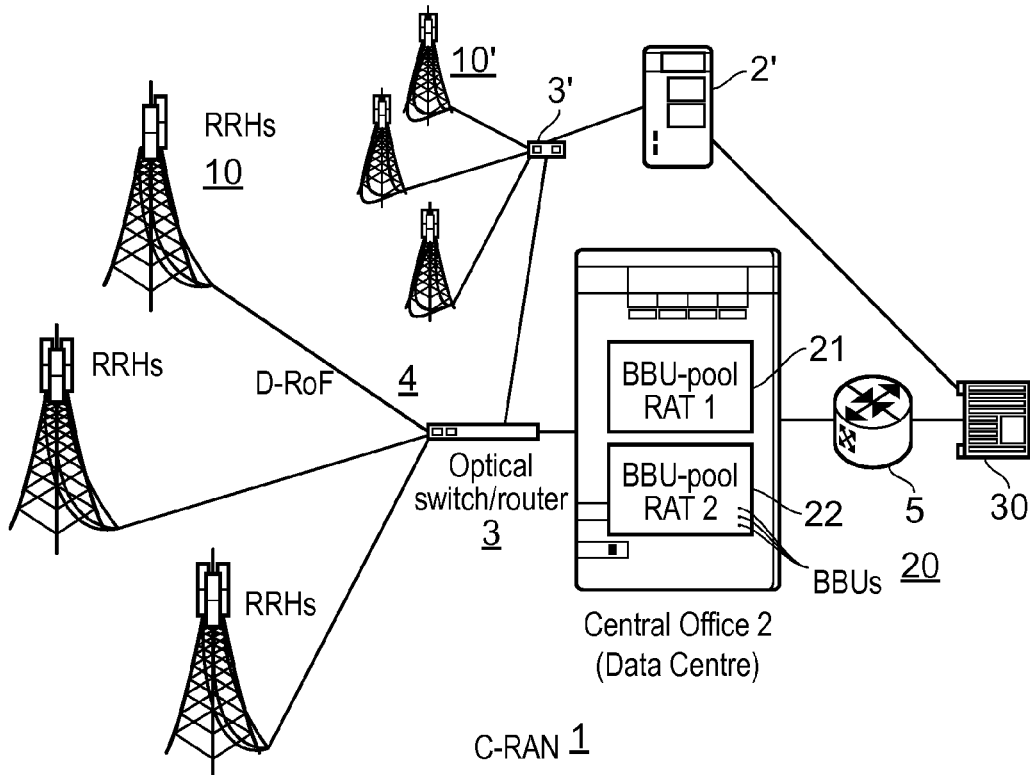
H04W 72/04 (2006.01)
H04L 29/08 (2006.01)

(52) **U.S. Cl.**

CPC **H04W 72/0493** (2013.01); **H04L 67/10** (2013.01)

(57) **ABSTRACT**

A cloud-based wireless communication system (1) comprising user devices (12) arranged to receive services through any of a plurality of cloud-based virtual networks. The system has a plurality of base stations or access points (14), a user device (12) being arranged to wirelessly connect to any of the virtual networks via at least one base station (14) in order to receive from a virtual network a specific service for which a user of the user device has subscribed. There are virtual service anchors or vSAs (26) in the cloud, each virtual service anchor provided for a respective service provided by its virtual network. Thus, one vSA (26) handles resource allocation for all users of one service in one virtual network. Each vSA (26) receives content access requests (S10) from user devices (12) indicating content to be accessed in that vSA's service, the content stored by at least one content delivery node in the cloud. The vSA (26), in response to the content access requests, identifies (S12, S14) at least one suitable Content Delivery Node based on the indicated content and base stations proximate to the user device, and transmits service requests (S16, S22). In response to service requests received from multiple vSAs (26), each base station (14) performs second resource allocation, to provide services to each user device (12) wirelessly connected to that base station (14).



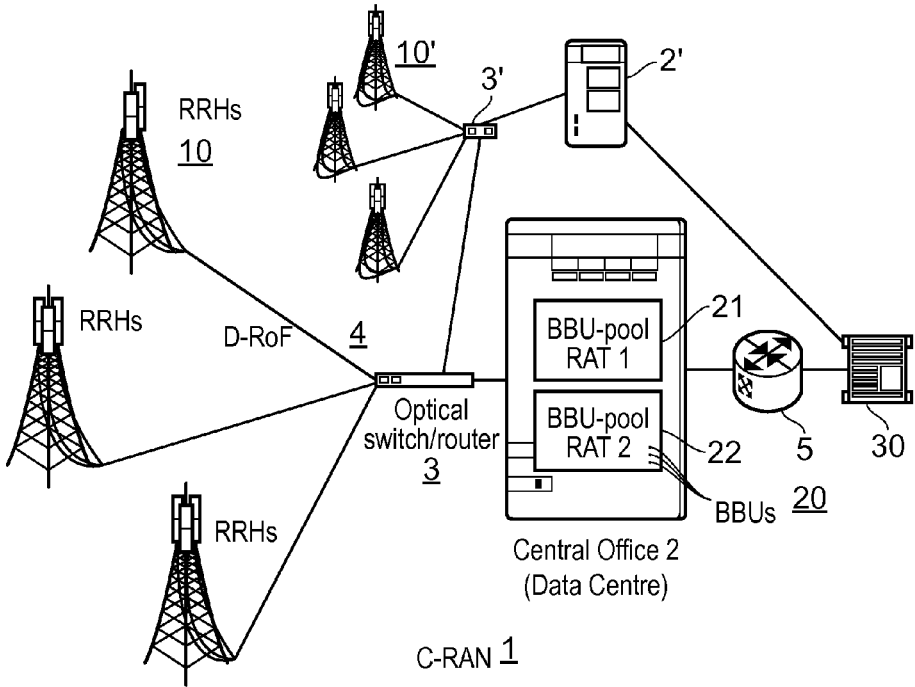


FIG. 1

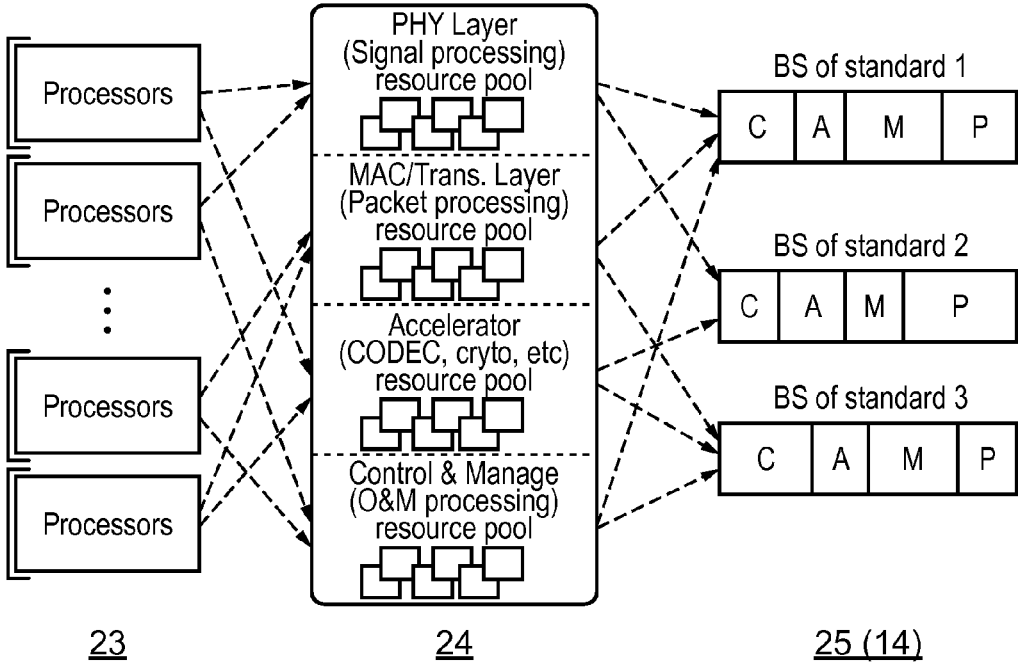


FIG. 2

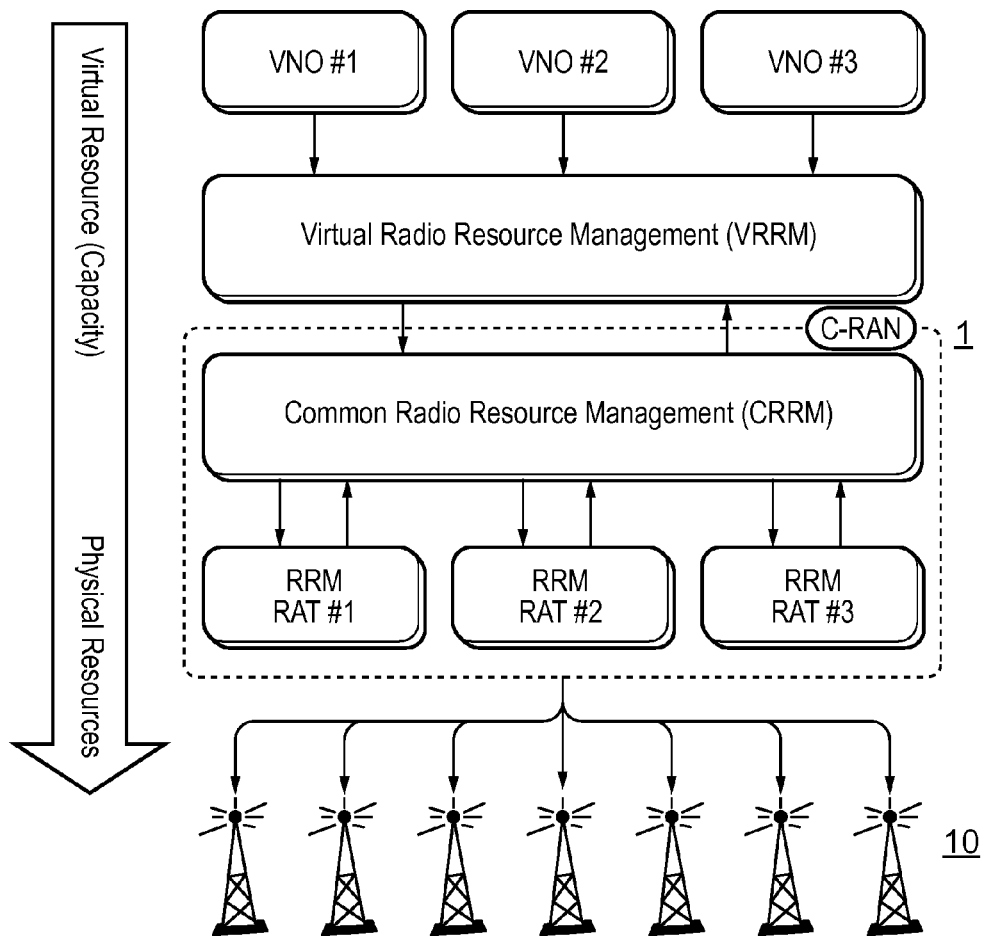


FIG. 3

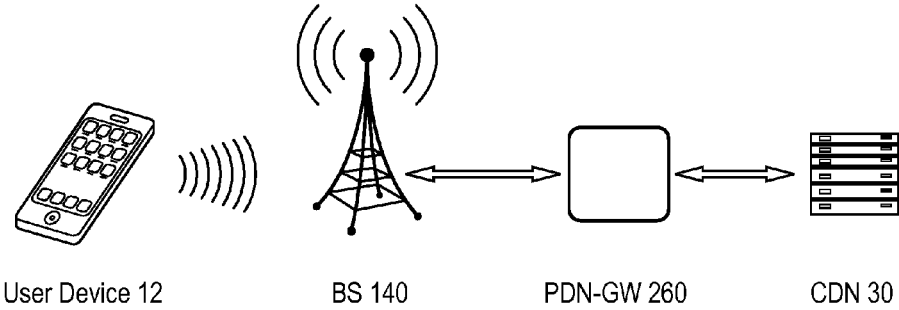


FIG. 4A

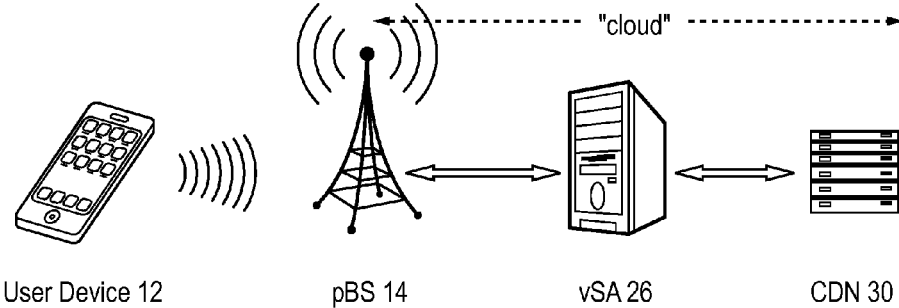


FIG. 4B

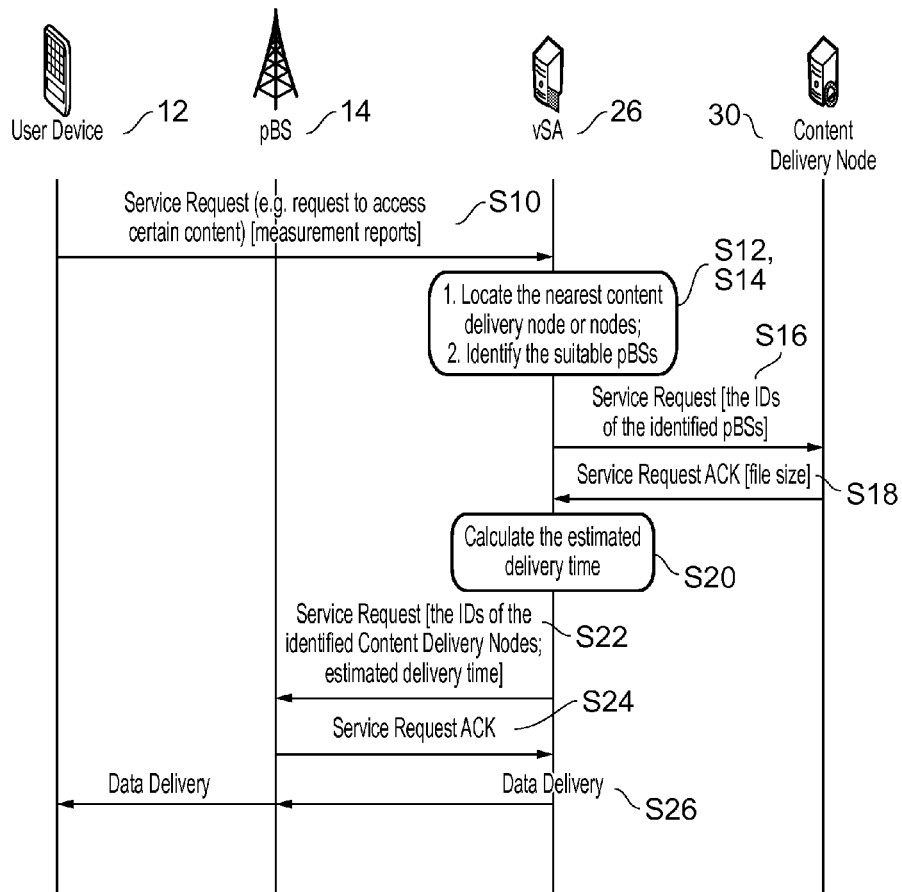


FIG. 5

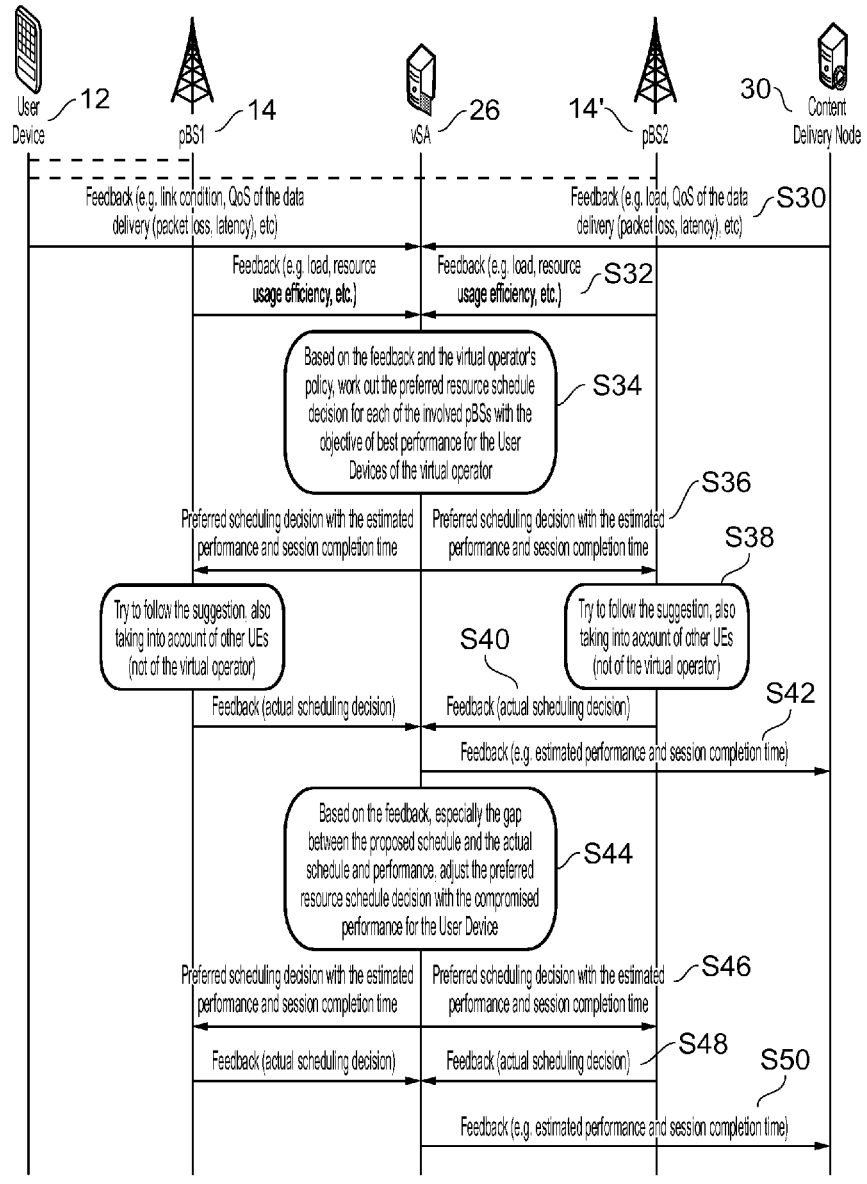


FIG. 6

RESOURCE MANAGEMENT IN CLOUD-BASED RADIO ACCESS NETWORK

FIELD OF THE INVENTION

[0001] The present invention relates to resource management in wireless networks and more particularly to wireless networks in which at least part of the radio access network (RAN) is implemented via cloud computing, or in other words “in the cloud”.

BACKGROUND OF THE INVENTION

[0002] The exponential increase of data rate demands today is tightly coupled with the exponential increase in available storage capacity and processing power, where more processing power requires more storage in order to store the processed data. To meet this demand, cloud computing is a disruptive technology which has changed the development of IT platforms significantly.

[0003] NIST (National Institute of Standards and Technology) defines cloud computing as “a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.” The essential characteristics of the cloud computing model are:

[0004] (i) on-demand self-service: Any cloud user can self-provision the needed computational resources, through a direct service entry point into the cloud platform, at the time and in the amount best fitting its specific needs, where there is no human assistance and the provisioning is automated;

[0005] (ii) broad network access: the computational resources are accessed and used uniquely through a standard network connection (no other types of connection are viable);

[0006] (iii) resource pooling: the computational resource pool is unique and shared among all the users, according to a multi-tenancy model. The user has no way to influence the mechanisms according to which the assigned resource location is selected by the cloud platform;

[0007] (iv) rapid elasticity: the resources are supposed to be seamlessly scalable to the user, and anyhow must be able to dynamically adapt to the real time demand; and

[0008] (v) measured service: the cloud platform must have a metering capability, using abstraction levels peculiar to each provisioned service, supporting monitoring and accounting of resource requests and usage from all the involved consumers.

[0009] A conventional mobile network employs dedicated network infrastructure to provide a RAN based on one specific radio access technology (RAT). Within the same geographical area there are typically multiple co-existing RANs based on radio access technologies such as GSM, UMTS and LTE, often together with wireless LANs based on the Wi-Fi group of standards. Base stations or access points of one network are usually of a proprietary nature and incompatible with the other systems, whilst the utilization rate of any one base station may be quite low. Overall capacity is likely to be limited by interference among the RANs, frustrating attempts to meet the ever-increasing demand for wireless data. Although attempts have been made to share hardware among

co-existing RANs, the overall scheme is basically non-collaborative, inflexible and wasteful of hardware resources.

[0010] Therefore, the possibility is being investigated of leveraging cloud-technology to improve telecommunication networks and address user needs raised by changing traffic demands.

[0011] Meanwhile, virtualisation techniques are being proposed to eliminate the dependency between a network function and its hardware, which results in the sharing of the physical hardware by multiple virtual network functions in the form of virtual machines.

[0012] Further pooling of the hardware might facilitate a massive and agile resource sharing: a phenomenon which is already seen in cloud computing infrastructures.

[0013] As part of the Next Generation Mobile Networks (NGMN) project, the so-called Cloud-RAN (C-RAN) has been proposed as a centralised processing, collaborative radio, real-time cloud computing and “clean” (in the sense of a simple architecture) RAN system. Centralised processing allows operators to use the cloud computing paradigm to handle incoming radio traffic in a collaborative manner, enabling higher throughput for the end users.

[0014] FIG. 1 shows a system architecture applicable to C-RAN (incidentally, in this specification the terms “system” and “network” are interchangeable unless demanded otherwise by the context). In this C-RAN 1, the conventional Base Station (BS) of a wireless communication network is split into a Remote Radio Head (RRH) 10 and a Base Band Unit (BBU) 20.

[0015] The RRH 10 is a remotely located radio unit and antenna, in charge of the radio functions from the RF transmission and reception with user devices (not shown), to digital baseband and adaptation to the transport network. There may be cluster of RRHs 10 as indicated in FIG. 1. Each RRH may provide at least one cell in the wireless communication network and the cluster, therefore, may form a network of overlapping cells in a certain geographical area.

[0016] The BBU 20 is the baseband digital processing equipment, composed of high-performance programmable processors and real-time virtualization technology. The BBU 20 carries out any action required at physical layer level (L1) function and congregates Layer 2 and 3 functions (control and management) of a BS. The BBU 20 and the RRH 10 are linked via a high bandwidth and low-latency optical “fronthaul” network which includes one or more optical switch/router 3.

[0017] Each BBU is not a standalone unit but rather, as shown in FIG. 1 it is provided as a virtual machine, part of a “BBU pool” implemented in a Central Office or Data Centre 2. The Central Office 2 may contain more than one BBU pool, one for each of a plurality of radio access technologies (radio access technologies). Within each Central Office, the respective BBUs (virtual machines) may be tightly interconnected, permitting high-speed of exchange of data. Linkage of BBUs within one RAT pool will generally be tighter than between different pools.

[0018] Incidentally, although the BBUs are shown in different pools on a per-RAT basis in FIG. 1, this is only one possible arrangement.

[0019] Referring to FIG. 2, an alternative concept of virtual machines in a Central Office is shown.

[0020] In this example, resource pools 24 are defined at various hierarchical levels of a PHY (physical) layer, a MAC/Trans. (Medium Access Control/Transport) layer, Acceleration

tor layer, and a Control and Management (Operation & Maintenance or O&M) layer. As indicated in FIG. 2, the resource pools 24 are formed from the hardware resources of multiple processors 23 of the Central Office. The resource pools at each protocol layer co-operate to define a plurality of (virtual) base stations (BS) 25 as indicated at the right-hand part of the Figure, corresponding to the BBUs in FIG. 1. As indicated the base stations may correspond to different standards 1, 2 and 3—in other words to different radio access technologies. Incidentally, although the base stations are depicted here as each belonging to one RAT, multi-RAT base stations are also possible. Base stations may also combine wireless cellular functions with WLAN capability (Wi-Fi), or unlicensed spectrum used for LTE-based wireless communication (so-called LTE-U), acting as an access point as well as a wireless cellular base station. A further possibility is to employ so-called device-to-device communication (D2D) using LTE spectrum for wireless communication between user devices, supported by a base station.

[0021] Referring again to FIG. 1, the optical switch/router 3, as well as providing routing to the BBU, permits connection to another region of the network having RRHs 10 coupled to a Central Office 2' via switch/router 3'. The Central Office 2' may serve a different geographical area from that covered by Central Office 2.

[0022] Connections of BBUs 20 to other nodes in the cloud are carried over a so-called “backhaul” network. This is—or at least may be thought of as—a broadband Internet connection, using conventional Internet protocols. Each Central Office 2 is connected via the backhaul network to Content Delivery Nodes 30, one of which is shown in the Figure, and in practice other network infrastructure, exemplified by a router 5, will be present.

[0023] Radio resource management in such architecture is known to be a non-trivial task, which becomes even more challenging in the context of the Virtual Radio Access Network (V-RAN) where multiple Virtual Network Operators' (VNOs) co-exist on the same physical infrastructure, defining multiple virtual networks.

[0024] FIG. 3 illustrates the management of radio resources in a V-RAN, which as shown is hierarchical, consisting of Virtual Radio Resource Management (VRRM) over Common Radio Resource Management (CRRM) and local RRM. The hierarchy is shown from top to bottom with the highest-level (most abstracted) functions at the top; the antenna symbols at the bottom in FIG. 3 correspond to the RRHs in FIG. 1. The RRM are roughly equivalent to the per-RAT BBU pools in FIG. 1.

[0025] VNOs, at the highest level of this hierarchy, are network operators who do not own the radio access infrastructure, but share the network resource by arrangement with the infrastructure provider, to provide wireless connectivity to subscribers of the VNOs. In this scheme, VNOs request a certain level of service quality from a virtual resource manager handling the physical infrastructure; this is in addition to quality of service (QoS) requirements of each service session. Such a request may be covered by a Service Level Agreement (SLA) between the virtual operator and the infrastructure provider, and/or may be a request issued dynamically based on current demand.

[0026] The VRRM, the highest-level manager, is in charge of translating VNOs' requirements and SLAs into sets of policies for lower levels. The VRRM optimises the usage of virtual radio resources and it does not deal with physical

resource. Nevertheless, reports and monitoring information (e.g. estimated remaining capacity) received from CRRM enable it to improve the policies.

[0027] CRRM, also called Joint RRM (JRRM), is the intermediate management level between VRRM and local RRM. CRRM maps the policies of VRRM from virtual resources to physical resources, in addition to issuing policies to manage radio resources in a heterogeneous access environment (heterogeneous in the sense of multiple-RAT). It also optimises the policies based on information from local RRM, and sends reports to the VRRM.

[0028] Local RRM, in the lowest resource management level, are liable for optimising radio resource usage in a single Radio Access Technology (RAT). They are in charge of assigning physical radio resource parameters (e.g. power, frequency bandwidth, time slots, etc.) to the end-users upon receiving request. The policies issued by the CRRM for each local RRM are used as decision guidelines for that RRM. In addition to policies set by the CRRM, the resource allocation in each local RRM has to meet the QoS requirement of each service, where “service” refers to a specific application being provided to an end user.

[0029] To date, virtual networks and the V-RAN concept have shared conventional network infrastructure without employing the principles of cloud computing. On the other hand, the virtualisation provided by C-RAN is a natural fit with the V-RAN concept, allowing resources to be flexibly allocated among the virtual operators in accordance with demand. As the dotted line in FIG. 3 shows, at least the RRM function for each RAT, and the CRRM, may be performed by C-RAN 1.

[0030] Consider now a cloud based network, where the basic infrastructure is provided by one or more operators, shared by several virtual operators. Basic characteristics of such a network may include:

[0031] Traffic dense area

[0032] Diversity of radio access technologies

[0033] Multi-mode Small Cells (LTE tier and LTE-U/WiFi tier) on the Macro Cell coverage layer

[0034] In such a cloud-based network, resources are shared by multiple operators (including virtual operators), which may have different business strategies/focuses and therefore different policies with different customers/subscriber bases. Resource management is essential in order to, for individual virtual operators, guarantee service policies to be enforced; and for the customers/subscribers of individual virtual operators, guarantee the QoS taking into account of the characteristics of the offered services. Meanwhile, from the viewpoint of physical network performance, it is important to maintain high resource utilisation with minimum cost.

[0035] In short, the problem requiring a solution is how to efficiently support resource management in a cloud based network.

SUMMARY OF THE INVENTION

[0036] According to a first aspect of the present invention, there is provided a method of managing resources in a cloud-based wireless communication system comprising:

[0037] defining, from common resources available in the cloud, a plurality of virtual networks, each virtual network offering one or more services and employing any of a plurality of serving stations, to which user devices are wirelessly connected to access a said virtual network;

- [0038] allowing the user devices to be able receive a specific service by users thereof subscribing to a said virtual network;
- [0039] in each virtual network, performing first resource allocation for each specific service offered by the virtual network, using the common resources for provisioning the service for all users of the virtual network, and transmitting service requests on the basis of the first resource allocation; and
- [0040] in each serving station, receiving service requests from each virtual network and in response, performing second resource allocation for providing services to each user device wirelessly connected to the serving station.
- [0041] Here, “the cloud” refers to hardware resources accessible via the Internet and the physical location of which is unimportant (and indeed may not be apparent to the serving stations or user devices). The serving stations may be common to the virtual networks but are not necessarily (or not necessarily wholly) in “the cloud”. At least part of the functions of the serving stations may however be implemented in the cloud; for example a serving station in the form of a base station having separate RRH and BBU as depicted in FIG. 1 may have the BBU function located in the cloud.
- [0042] Each “serving station” may be a base station or access point. Alternatively, or in addition, one or more serving stations may comprise other user devices configured for device-to-device (D2D) communication. As already mentioned such user devices are supported by base stations; therefore in this case it may be appropriate to regard the term “serving station” as covering a user device in combination with its supporting base station.
- [0043] Allowing user devices to be able to receive a service refers to setting a precondition to making services available, in particular requiring the user to subscribe to a virtual network (and possibly within that virtual network, to one or more service types). Actual delivery of services does not take place until after the first and second resource allocation and these in turn normally depend on a specific request from the user device (see below).
- [0044] The method preferably provides a respective virtual service anchor for each specific service offered by the virtual network, the virtual service anchor performing the resource allocation for its specific service and transmitting the service requests for its specific service.
- [0045] Normally, the method is triggered by specific requests for services (content access requests) from user devices. Thus, in a Service Request procedure which is one embodiment of the present invention, the method preferably further comprises each virtual service anchor receiving content access requests from user devices indicating content to be accessed in its specific service, each virtual service anchor performing first resource allocation for its specific service on the basis of the content access requests and preferably also at least one of a predetermined virtual operator policy and a stored user profile.
- [0046] In this case, preferably, each content access request further comprises measurement reports of the user device with respect to serving stations, and/or location information of the user device. Meanwhile, the service requests transmitted from each virtual service anchor may include a service request to at least one content delivery node holding content indicated in a content access request.
- [0047] Then, preferably, at least one said content delivery node responds to the service request by notifying the virtual service anchor of a data size of content indicated in the service request. This data size may be of only part of the content requested, if that is all that is available in the content delivery node concerned.
- [0048] Thereafter the virtual service anchor may calculate an estimated delivery time of the content to the user device based on:
- [0049] the data size notified by the or each content delivery node;
 - [0050] properties of one or more connections from a content delivery node to a serving station; and
 - [0051] the measurement reports and/or location information of the user device.
- [0052] In addition, the virtual service anchor may update the estimated delivery time during content delivery.
- [0053] Here, the “properties of one or more connections” refers to for example, the capacity or latency on a communications link between the content delivery node to the serving station concerned.
- [0054] In any method as defined above, preferably, the service requests transmitted from each virtual network include service requests to each of one or more serving stations identified as proximate to a user device requiring the specific service. Thus, providing the service to the user device may involve wireless communication of the user device with two or more serving stations simultaneously.
- [0055] Preferably the service requests to each of one or more serving stations identified as proximate to a user device identify at least one Content Delivery Node and the estimated delivery time. In this way the serving station knows to which node in the network to route requests for data, and an indication can be given to the user device of the expected delay and/or throughput for the content.
- [0056] Preferably also, each service request to at least one content delivery node includes information on the one or more serving stations identified as proximate to a user device requiring the specific service. The content delivery node may likewise be informed of an expected time (or duration) of content delivery, facilitating load management in the content delivery node.
- [0057] Then, each of the one or more serving stations, identified as proximate to a user device requiring the specific service, may establish a connection with the or each Content Delivery Node for receiving the content and transmitting the content to the user device. In this way the service is actually provided to the user device.
- [0058] Another embodiment of the present invention provides a Co-ordinated Data Delivery procedure for a plurality of serving stations. In this case, preferably, each virtual service anchor receives feedback from the user devices, the serving stations and the or each Content Delivery Node, the feedback including at least one of:
- [0059] from a said user device, status of wireless connections of the user device with proximate serving stations;
 - [0060] from a said serving station, load status of the serving station and/or interference experienced by the serving station; and
 - [0061] from a said Content Delivery Node, load status and/or status of connection with the virtual network.
- [0062] Then, in the first resource allocation, the virtual service anchor preferably determines a preferred resource

scheduling decision for each of a plurality of serving stations proximate to the user device, and transmitting the decision to each serving station concerned. In the second resource allocation, each serving station takes the decision into account, determining an achievable resource scheduling decision (or recommendation) which the serving station then notifies to the virtual service anchor.

[0063] A service session is started with the user device by delivering content from one or more content delivery node via at least one serving station to the user device. Preferably, the first and second resource allocation are repeated at intervals during the service session. This can include varying the serving stations employed for data delivery.

[0064] Here, preferably, the virtual service anchor notifies the Content Delivery Node of each resource scheduling decision relevant to that Content Delivery Node.

[0065] This embodiment preferably further comprises the virtual service anchor revising its preferred resource scheduling decision on the basis of achievable resource scheduling decisions from the serving stations and transmitting the revised decision to each serving station concerned, each serving station repeating the second resource allocation taking the revised decision into account and notifying results to the virtual service anchor.

[0066] According to a second aspect of the present invention, there is provided a cloud-based wireless communication system comprising:

[0067] user devices arranged to receive services through any of a plurality of virtual networks defined using common resources available in the cloud;

[0068] a plurality of serving stations, a said user device being arranged to wirelessly connect to any of the virtual networks via at least one said serving station in order to receive from a said virtual network a specific service for which a user of the user device has subscribed to said virtual network;

[0069] virtual service anchors in the virtual networks, each virtual service anchor provided for a respective said service provided by said virtual network, arranged to perform first resource allocation using the common resources, and arranged to transmit service requests on the basis of the first resource allocation; wherein

[0070] each serving station is arranged to perform second resource allocation in response to the service requests received from the virtual service anchors, to provide services to each user device wirelessly connected to the serving station.

[0071] Here, the virtual service anchors are in the cloud. Whether, and to what extent, serving stations can be said to be “in the cloud” will depend on the implementation; for example a serving station in the form of a user device configured for D2D is not itself in the cloud, but as already mentioned a serving station which is, or which comprises, a base station having a BBU and RRH may have the BBU part in the cloud. Elsewhere in this specification, the term “physical base station” is used to denote that there is at least some part of the base station (such as a RRH) which communicates with user devices over the physical air interface.

[0072] Preferably each virtual service anchor is further arranged to receive content access requests from user devices indicating content to be accessed in the respective service, the system further comprising at least one content delivery node, the service requests transmitted from each virtual service

anchor including service requests to at least one content delivery node based on the indicated content.

[0073] According to a third aspect of the present invention, there is provided a server in a computer network, configured to provide at least one of the virtual service anchors in the system defined above.

[0074] According to a fourth aspect of the present invention, there is provided software comprising computer-readable code which, when executed by processors of at least one networked computer and serving station, perform any of the methods defined above.

[0075] Thus, embodiments of the present invention can provide a cloud-based wireless communication system, in which user devices are arranged to receive services through any of a plurality of virtual networks defined using common resources in “the cloud”. The system has a plurality of base stations (physical base stations or access points), a user device being arranged to wirelessly connect to any of the virtual networks via at least one of the base stations in order to receive, from a virtual network, a specific service for which a user of the user device has subscribed. There are virtual service anchors or vSAs in the cloud, each virtual service anchor provided for a respective service offered by its virtual network. In other words, in each virtual network there are multiple virtual service anchors, one per service (or service type) and covering all users of the service in that virtual network.

[0076] Similar to the VRRM mentioned in the introduction, the vSA provides an extra layer of radio resource management on top of conventional RRM. The difference is that vSA is defined on a per-service and per-VNO basis. In addition, the vSA provides a direct logical interface to Content Delivery Nodes which allows the vSA to take into account feedback from Content Delivery Nodes when making scheduling requests.

[0077] Embodiments of the present invention are mainly aimed at services involving transmission of content (audio, video, files) to a user device. Examples of such services include video and audio streaming, downloading etc. Each vSA receives, via physical base stations or access points (henceforth denoted pBSs), content access or service requests from user devices indicating content to be accessed in that vSA’s service, the content concerned being stored by at least one Content Delivery Node in the cloud. The vSA, in response to the content access requests, identifies at least one relevant Content Delivery Node as well as base stations proximate to each requesting user device. Then the vSA transmits service requests, including service requests to at least one Content Delivery Node based on the indicated content, as well as service requests to the proximate base stations. In response to service requests received from multiple vSAs, each base station performs resource allocation at a local (cell) level, to provide services to each user device wirelessly connected to that base station.

[0078] In this way, embodiments of the present invention may provide a multi-level service-centric resource management mechanism in a cloud based virtualised network that is shared by multiple operators including virtual operators (VNOs). There are two levels of resource management in the proposed scheme: a Level 1 resource management at the Service Anchor (vSA), and a Level 2 resource management at the physical base stations or access points (pBSs).

[0079] A service request procedure is proposed to allow vSA to identify the suitable pBSs and Content Delivery

Nodes for a request raised by a user device of the virtual operator. A coordination scheme is also proposed that enables the vSA to coordinate the data delivery session of a user device among related nodes.

[0080] Embodiments of the present invention can optimise the resource management by coordinating the operation of nodes from the application layer (i.e. Content Delivery Node) to the access network layer (i.e. physical base station or access point)

[0081] The proposed resource management allows automated coordination which benefits not only the concerned user devices of virtual operators but overall network performance.

BRIEF DESCRIPTION OF THE DRAWINGS

[0082] FIG. 1 shows a network architecture proposed for a Cloud-based Radio Access Network (C-RAN);

[0083] FIG. 2 shows levels of resource management in a virtualized RAN, which can include a C-RAN like that of FIG. 1;

[0084] FIG. 3 shows processors of a Central Office providing resource pools for implementing functions of a C-RAN;

[0085] FIGS. 4A and 4B compare a simplified conventional network architecture (FIG. 4A) with an architecture proposed in the present invention (FIG. 4B);

[0086] FIG. 5 is a signalling diagram for a service request procedure in an embodiment of the present invention; and

[0087] FIG. 6 is a signalling diagram for a data delivery session in an embodiment of the present invention.

DETAILED DESCRIPTION

[0088] Embodiments of the present invention provide a multi-level service-centric resource management mechanism in a cloud based virtualised network that is shared by multiple operators including virtual operators (VNOs). It is assumed that a virtual network is created for a specific virtual operator. One virtual network comprises one or more virtual service anchors (vSAs), and the coverage footprint and the capacity of the virtual network vary corresponding to the demand of the subscribers of this virtual operator. In principle the coverage footprint of one virtual network could extend to an entire country, in which case the corresponding vSA(s) would be distributed over different geographical areas of the country.

[0089] The vSA, a novel concept in the present invention, is service-specific and operator-specific. In other words, one virtual operator can have multiple vSAs, and one vSA belongs to one and only one virtual operator (VNO). Normally there is only one vSA for one service of a given VNO. Examples of “service” here would include Video-on-Demand, music streaming, multimedia broadcast (MBMS), WWW access, and so forth. Thus, in this context, “service” equates with “service type”. Users may subscribe for one or more different kinds of service when they enter a contract with (“subscribe to”) the VNO. How many vSAs are provided is partly a question of definition; however, it may be said that using cloud computing, one logical vSA is sufficient for one specific service of a VNO regardless of the geographical area covered.

[0090] FIGS. 4A and 4B compare the architecture in the present invention with that of a conventional wireless network such as an LTE-based network.

[0091] The conventional wireless network in FIG. 4A has a user device 12 wirelessly connecting with a base station 140 which is a standalone, dedicated base station of one specific wireless network. The base station 140 communicates over a dedicated, network-specific backhaul with a PDN-GW 260 which acts as a gateway to the Internet, allowing content to be retrieved from a Content Delivery Node 30. The connection between the PDN-GW and Content Delivery Node 30 may thus a broadband Internet connection; alternatively dedicated backhaul may be used if CDN 30 is managed by the network operator. The PDN-GW 260 may also act as a service anchor for user devices; on the other hand it should be noted that the PDN-GW does not have any resource management function.

[0092] As shown in FIG. 4B, the architecture proposed for use in the present invention is superficially similar, and the user device and Content Delivery Node are unchanged. However, in this case the physical base station or pBS 14 is (for example) a base station of a virtual network partly implemented in the cloud, for example by the BBU pool shown in FIG. 1, and the service anchor is the above mentioned vSA 26. The connection between the virtual network base station 14 and the vSA 26 of that network, may either be via broadband Internet or via dedicated backhaul. Meanwhile the connection from vSA 26 to Content Delivery Node 30 may be made either over broadband Internet (where the Content Delivery Node 30 is located on the Internet) or via a dedicated link (if the Content Delivery Node 30 is proprietary to the operator). The requirement for expensive dedicated backhaul can therefore be reduced.

[0093] The vSA differs from prior proposals such as the VRRM shown in FIG. 3 in that the VRRM serves all VNOs in a (physical) network, which provides a generic level of radio resource management for all types of services provided by all VNOs. By contrast the vSA proposed here is service specific and VNO specific. Thus, where there are first and second VNOs in a network, both offering similar kinds of service, each would have their own (one or more) vSAs. Due to the differences between these two VNOs’ policies, the vSA for the first VNO would take different resource management decisions compared with the vSA for the second VNO.

[0094] It can be assumed that each user subscribes for a service in some way, for example by downloading an application program (“app”) provided by the VNO. The user (subscriber) initiates a service session by sending a service request from his or her user device. The virtual service anchor is provisioned only when there is at least one service request from the subscriber(s) of this virtual operator. One virtual service anchor may be associated (or linked) with multiple physical base stations or access points of different radio access technologies. One physical base station or access point may be shared by multiple virtual service anchors. Fixed mapping between a virtual service anchor and a physical base station is not necessary.

[0095] Relating the above to the BBU pool 21 or 22 discussed in the introduction, the BBU pool is mainly in charge of physical radio resource management, whereas the vSA 26 handles the virtual resource management for a specific service of a specific VNO. In this sense, vSA can be considered as a “higher” layer entity, which interacts with the RRM function provided in a BBU pool or provided in another way. The vSA is a logical entity independent of the BBU pool (if any), even though it may be physically co-located in the same Central Office 2 as the BBU pool 21, 22. For example both may be provided by software running on processors in the

cloud, i.e. on a server or in a server farm. Since the vSA 26 is service specific and VNO specific, there is no fixed mapping between vSAs and BBU pools.

[0096] It should be noted, however, that although the pBSs may be implemented partly in the cloud as already mentioned, the present invention can also be employed with conventional, “standalone” base stations and/or with user devices configured for D2D.

[0097] A virtual service anchor may “follow” the subscribers of the virtual operator. For example, when the subscribers move (e.g. move out of the coverage of a current physical base station associated with the virtual service anchor), the virtual service anchor may associate with different physical base stations or access points to provide sufficient coverage and capacity for the subscribers.

[0098] There are two levels of resource management in the proposed scheme:

[0099] (a) Level 1—at the Service Anchor (vSA 26)

[0100] This resource management level is for the subscribers of the virtual operator only, where these subscribers run applications on user devices to obtain services, the resource management taking into account of the application layer characteristics, e.g. the estimated completion time of a session and service class (based on the virtual operator’s policy). A key function of this level is to coordinate the involved physical base stations.

[0101] (b) Level 2—at the physical base station or access point (pBS 14)

[0102] This level applies to all the users in the coverage area (cell or range) of the pBS/access point, taking into account the physical link condition of wireless links between pBSs and user devices. Here, resource scheduling decisions are made to meet the requirements from Level 1 resource management component.

[0103] More particularly these levels of resource management provide a service request procedure as shown in FIG. 5, and a procedure for conducting a data deliver session as shown in FIG. 6.

[0104] Service Request Procedure

[0105] In this procedure, a user device 12 of a virtual operator initiates a service request to the vSA 26 of the virtual operator. The user device 12 can be a user equipment of a human user, e.g. mobile phone or tablet or a computer; it can also be a machine type device. By operating the user device, the user causes a Service Request to be generated. The requested service may be a certain content delivery which is part of the services offered by the virtual operator, for example for a particular video stream available from a store of the virtual operator.

[0106] All communications from and to the user device are made wirelessly via a base station (physical base station or pBS 14, which may also, or alternatively, be an access point). Usually, service requests from one user device will be directed to one particular base station (normally the closest), but it is possible for the user device to communicate with multiple pBSs.

[0107] FIG. 5 illustrates an example of the service request procedure, including the following steps.

[0108] In an initial step S10, the user device 12, the owner of which is a subscriber of a virtual operator’s services, sends a Service Request to the vSA 26 of the virtual operator via the pBS 14, in order to access certain content as part of the service offered by the virtual operator. This Service Request (also referred to as content access request) identifies, implicitly or

explicitly, the virtual network providing the service which is the subject of the request. An example of an implicit indication would be a subscriber ID contained in the request and identifying the sender as a subscriber of a specific virtual network. An example of an explicit indication would be if the Service Request is generated using an application program (“app”) provided by and specific to the virtual operator. Consequently the pBS knows where to forward the request. The vSA being a higher-level node in the network than a pBS, the pBS forwards the Service Request to the appropriate vSA via the backhaul network.

[0109] Measurement reports of the user device 12 may be included in the request, which indicate the quality of the signals received in the neighbouring base stations and access points. The location information of the user device may also be included, if known from a GPS function of the user device for example.

[0110] In step S12, upon receipt of the request, the vSA 26 checks the requested content and locates the most suitable Content Delivery Node or Nodes (e.g. nearest nodes available, in the sense of transmission time for example). After that, in S14 the vSA 26 identifies the suitable pBSs for the content delivery session. Determination of suitable pBSs may be based on the measurements and user device’s location information, and subscriber’s profile and or virtual operator’s policy, etc. Normally, one or more pBS proximate to the user device will be chosen. The pBSs can be regarded as common resources available for use of any of the virtual networks.

[0111] In S16, the vSA 26 then forwards the request (or a corresponding message generated from the request in the vSA) to the identified Content Delivery Node or Nodes 30 over the backhaul network. The IDs and address information of the identified pBS(s) are included in the message.

[0112] In the next step S18, the Content Delivery Node 30 responds with an acknowledgement message, with the file size of the requested content.

[0113] In one alternative case, multiple Content Delivery Nodes may be identified for this content delivery session, each of the selected nodes storing part of the requested content. Each of them may respond to the request with the file size of part of the requested content that is stored in this node.

[0114] The subsequent steps are described with respect to a single pBS for simplicity. However, it should be noted that it is possible for the vSA to employ a plurality of base stations to deliver the requested content to the user device. Thus “identified pBS” below should be understood as referring possibly to a plurality of base stations.

[0115] At S20, upon the receipt of the response from the Content Delivery Node, the vSA 26 calculates the estimated delivery time based on the file size and the link quality of the identified pBS. In case of multiple Content Delivery Nodes, the vSA 26 may calculate the estimated delivery time for the content available in each Content Delivery Node.

[0116] The vSA 26 then forwards the request to the identified pBS in step S22. The ID and address information of the identified Content Delivery Node are included in the message, as well as the estimated delivery time for the content available in the Content Delivery Node.

[0117] The pBS 14 acknowledges the request in S24. The data delivery path can then be set up from the Content Delivery Node to the pBS, and then to the user device 12, in step S26.

[0118] In the case of multiple Content Delivery Nodes, multiple paths may be set up between the pBS and the Content

Delivery Nodes. It may be up to the pBS how to deliver the data to the user device 12. Content delivery is direct from the (or each) Content Delivery Node to the pBS; the data does not pass through the virtual service anchor.

[0119] The vSA 26 may update the pBS 14 with the estimated delivery time based on the remaining file size and the link quality.

[0120] Coordinated Data Delivery

[0121] This procedure is a coordination scheme between vSA 26 and other related nodes for the data delivery session of a user device 12. One example procedure is shown in FIG. 6 where, as indicated by the dot-dash lines at the top of the Figure, a user device 12 is engaged in a data delivery session for the content it requests as part of the services offered by the virtual operator, receiving data from two pBSs 14 and 14'.

[0122] A vSA 26 is coordinating the data delivery involving the pBSs 14, 14' (also labelled as pBS1 and pBS2) and multiple Content Delivery Nodes 30 (only one shown in this example). During the course of the data delivery session, the network nodes provide feedback to the vSA as indicated by steps S30 and S32. It is assumed that vSA 26 will be updated frequently by the user device 12, pBSs 14, 14' and Content Delivery Nodes 30.

[0123] Some examples of feedback include:

[0124] From the user device 12: link conditions with respect to one or more pBS (including the detected interference), QoS of the data delivery, such as packet loss and latency

[0125] From pBSs 14 and 14': load status (or available capacity), resource usage efficiency, detected interference

[0126] From Content Delivery Nodes 30: load status, QoS of the data delivery, such as packet loss and latency

[0127] It should be noted that the feedback from the user device 12 is transmitted via pBS 14 in FIG. 6. On the other hand, the feedback from Content Delivery Node 30 is transmitted over the backhaul network direct to vSA 26.

[0128] In performing resource scheduling (resource allocation) the virtual service anchor follows a policy set by the virtual operator. An example of such a policy can be "to deliver in the most cost-efficient way (i.e. cheapest)". Based on this policy, the vSA may choose, as the pBS to involve in the content delivery, a Wi-Fi access point instead of an LTE base station as long as it can meet the basic QoS requirements, or vSA may instruct the pBS to use most cost efficient technique for the service (e.g. D2D communication, or unlicensed spectrum such as LTE-U).

[0129] Based on the feedback information in S30 and S32, and the virtual operator's policy, in step S34 the vSA 26 works out the preferred resource scheduling decision for each involved pBS in order to achieve best performance (e.g. throughput, data rate) for the user device. Then, in step S36 the vSA sends the suggestion to the involved pBSs 14 and 14', together with the estimated performance of the concerned user device and the estimated session completion time for the session.

[0130] In S38, upon the receipt of such suggestion, each pBS 14, 14' will assess whether it is possible to follow the preferred scheduling, taking into account the resources available (including choice of RAT with which to communicate with the user device, if applicable). Here the pBS will also take into account other user equipments apart from those subscribed to the virtual operator; and make the scheduling decision (e.g. for the purpose of best overall resource usage

efficiency) accordingly. In S40, as part of the feedback procedure each pBS then informs the vSA 26 of an "achievable scheduling decision" which is a resource allocation which the pBS thinks it can deliver on the basis of information available to it, including requests from other vSAs. Thus, it is up to the pBSs to make scheduling decisions (or recommendations) on the basis of possibly conflicting requests from multiple vSAs. In S42, the vSA then updates the Content Delivery Node 30 with the estimated performance and session completion time. Such feedback helps the Content Delivery Nodes to take appropriate action/decisions on resource management and load balancing if necessary.

[0131] Based on the information fed back from the pBSs, in S44 the vSA 26 checks the gap/difference, if any, between its own proposed scheduling decision and the achievable decision of each pBS, and identifies the performance difference resulting from the different scheduling decisions. Then if necessary the vSA adjusts the preferred scheduling, either on a per-user basis or collectively for multiple users of the same operator, with the compromised goal (e.g. the second best performance the user device would achieve); this suggestion may have better chance to be accepted by the pBS. One example for an individual user is that the vSA will try to get the best service for a user's service if the VNO's policy is set as the best user experience possible. However, due to the limitation of the available resource in a pBS (e.g. a pBS may have to serve several UEs with high QoS demand), this user might be scheduled with limited resource in the pBS. Based on this feedback, the vSA may lower its scheduling request as a compromise, for example setting a lower data rate which may result in streaming a video at reduced resolution. Thus, the vSA and pBS coordinate to arrive at an acceptable scheduling in the pBS for all users currently served by that pBS.

[0132] In step S46 the vSA 26 sends the suggestion (revised scheduling decision) to the involved pBSs 14, 14'. In S50, the vSA 26 further informs the Content Delivery Node 30 of the estimated performance and session completion time.

[0133] Each pBS accepts, or at least considers, the revised resource allocation (scheduling decision) from the vSA. Once accepted, content delivery from the Content Delivery Node(s) to the pBS(s) can commence. Although not shown in FIG. 6, the above coordination procedure can be repeated to refine the scheduling decision, which results in optimised resource management not only benefitting the concerned user devices of virtual operators but overall network performance.

[0134] This can include repeating the scheduling at intervals during content delivery. The timescale for such repetition generally should be greater than the physical layer scheduling interval (e.g. several milliseconds), which may be several seconds and may vary depending on the feedback (e.g. less frequent when the gap/difference between the vSA's proposed scheduling decision and the scheduling achievable in the pBS is sufficiently small). The result of revisiting the scheduling in this way may include involving a further (or different) pBS in the data delivery as loads on individual pBSs or the wireless link conditions change.

[0135] Especially in the cases where multiple virtual operators share the same resources, such coordination/negotiation is particularly important to automatically optimise the resource allocation and achieve the optimal and harmonised operating environment.

[0136] To summarise, embodiments of the present invention can provide a cloud-based wireless communication system 1 comprising user devices 12 arranged to receive services

through any of a plurality of virtual networks defined using common resources which exist at least partly in “the cloud” and which are available to the system. The common resources include a plurality of base stations (physical base stations or access points) **14**, a user device **12** being arranged to wirelessly connect to any of the virtual networks via at least one base station **14** in order to receive from a virtual network a specific service for which a user of the user device has subscribed. There are virtual service anchors or vSAs **26** in the virtual networks, each virtual service anchor provided for a respective service provided by its virtual network. Each vSA **26** receives content access requests from user devices **12** indicating content to be accessed in that vSA’s service, the content stored by at least one content delivery node in the cloud. The vSA **26**, in response to the content access requests, performs first resource allocation using the common resources, this including transmitting service requests to at least one content delivery node based on the indicated content, as well as service requests to base stations **14**. In response to service requests received from multiple vSAs **26**, each base station **14** performs second resource allocation, to provide services to each user device **12** wirelessly connected to that base station **14**.

[0137] In this way, embodiments of the present invention can provide a multi-level service-centric management scheme, especially at the service anchor level which represents the subscribers of the corresponding virtual operator only and coordinates the involved physical base stations, taking into account of the application layer characteristics. More particularly embodiments of the present invention provide a multi-level service-centric resource management mechanism in a cloud based virtualised network that is shared by multiple operators including virtual operators.

[0138] A service request procedure is proposed to allow a vSA to identify the suitable pBSs and Content Delivery Nodes for a request raised by a user device of the virtual operator. A coordination scheme is also proposed that enables the vSA to coordinate the data delivery session of a user device among related nodes.

[0139] Various modifications are possible within the scope of the invention.

[0140] Although the architecture described above involves only virtual networks, the present invention can co-exist with dedicated (non-virtual) networks, allowing efficient network sharing by multiple operators including both VNOs and non-VNOs. Users of the dedicated networks need not be provided with a vSA, but providing vSAs for users of at least the virtual networks will improve network usage efficiency whilst guaranteeing QoS for individual users based on VNO’s policies.

[0141] The above mentioned service request procedure and coordination scheme can be combined, or they may be employed separately.

INDUSTRIAL APPLICABILITY

[0142] Advantages of the invention include, firstly, optimising the resource management by coordinating the operation of nodes from application layer (i.e. Content Delivery Node) to the access network layer (i.e. physical base station or access point). Secondly, the proposed resource management allows automated coordination which benefits not only the concerned user devices of virtual operators but overall network performance.

1. A method of managing resources in a cloud-based wireless communication system comprising:

defining, from common resources available in the cloud, a plurality of virtual networks, each virtual network offering one or more services and employing any of a plurality of serving stations, to which user devices are wirelessly connected to access a said virtual network;

allowing the user devices to be able to receive a specific service by users thereof subscribing to a said virtual network;

in each virtual network, performing first resource allocation for each specific service offered by the virtual network, using said common resources for provisioning said service for all users of the virtual network, and transmitting service requests on the basis of said first resource allocation; and

in each serving station, receiving service requests from each virtual network and in response, performing second resource allocation for providing services to each user device wirelessly connected to the serving station.

2. The method according to claim 1 further comprising providing a respective virtual service anchor for each specific service offered by the virtual network, the virtual service anchor performing said resource allocation for its specific service and transmitting said service requests for its specific service.

3. The method according to claim 2 further comprising each virtual service anchor receiving content access requests from user devices indicating content to be accessed in its specific service, each virtual service anchor performing said first resource allocation for its specific service on the basis of: the content access requests and at least one of a predetermined virtual operator policy and a stored user profile.

4. The method according to claim 3 wherein each content access request further comprises measurement reports of the user device with respect to serving stations, and/or location information of the user device.

5. The method according to claim 3 wherein the service requests transmitted from each virtual service anchor include a service request to at least one content delivery node holding content indicated in a content access request.

6. The method according to claim 5 further comprising at least one said content delivery node responding to the service request by notifying the virtual service anchor of a data size of content indicated in the service request.

7. The method according to claim 6 wherein each content access request further comprises measurement reports of the user device with respect to serving stations, and/or location information of the user device, further comprising the virtual service anchor calculating an estimated delivery time of the content to the user device based on:

the data size notified by the or each content delivery node; properties of one or more connections from a content delivery node to a serving station; and

the measurement reports and/or location information of the user device;

wherein the virtual service anchor updates the estimated delivery time during content delivery.

8. The method according to claim 1 wherein the service requests transmitted from each virtual network include service requests to each of one or more serving stations identified as proximate to a user device requiring said specific service.

9. The method according to claim 7 wherein the service requests transmitted from each virtual network include service requests to each of one or more serving stations identified as proximate to a user device requiring said specific service,

and wherein the service requests to each of one or more serving stations identified as proximate to a user device identify at least one Content Delivery Node and the estimated delivery time.

10. The method according to claim **8** wherein the service requests transmitted from each virtual service anchor include a service request to at least one content delivery node holding content indicated in a content access request, and wherein each service request to at least one content delivery node includes information on the one or more serving stations identified as proximate to a user device requiring said specific service.

11. The method according to claim **9** further comprising each of the one or more serving stations, identified as proximate to a user device requiring said specific service, establishing a connection with the or each Content Delivery Node for receiving the content and transmitting the content to the user device.

12. The method according to claim **5** further comprising each virtual service anchor receiving feedback from the user devices, the serving stations and the or each Content Delivery Node, the feedback including at least one of:

from a said user device, status of wireless connections of the user device with proximate serving stations;

from a said serving station, load status of the serving station and/or interference experienced by the serving station; and

from a said Content Delivery Node, load status and/or status of connection with the virtual network.

13. The method according to claim **12** further comprising: in said first resource allocation, the virtual service anchor determining a preferred resource scheduling decision for each of a plurality of serving stations proximate to the user device, and transmitting the decision to each serving station concerned;

in said second resource allocation, each serving station taking said decision into account, determining an achievable resource scheduling decision which the serving station then notifies to the virtual service anchor;

starting a service session of the user device by delivering content from one or more content delivery node via at least one serving station to the user device; and

repeating the first and second resource allocation at intervals during the service session.

14. The method according to claim **13** further comprising the virtual service anchor notifying the Content Delivery Node of each resource scheduling decision relevant to that Content Delivery Node.

15. The method according to claim **13** further comprising the virtual service anchor revising its preferred resource scheduling decision on the basis of achievable resource scheduling decisions from the serving stations and transmitting the revised decision to each serving station concerned, each serving station repeating the second resource allocation taking said revised decision into account and notifying results to the virtual service anchor.

16. A cloud-based wireless communication system comprising:

user devices arranged to receive services through any of a plurality of virtual networks defined using common resources available in the cloud;

a plurality of serving stations, a said user device being arranged to wirelessly connect to any of the virtual networks via at least one said serving station in order to receive from a said virtual network a specific service for which a user of the user device has subscribed to said virtual network; and

virtual service anchors in the virtual networks, each virtual service anchor provided for a respective said service provided by said virtual network, arranged to perform first resource allocation using said common resources, and arranged to transmit service requests on the basis of said first resource allocation; wherein each serving station is arranged to perform second resource allocation in response to the service requests received from the virtual service anchors, and to provide services to each user device wirelessly connected to said serving station.

17. The wireless communication system according to claim **16** wherein each virtual service anchor is further arranged to receive content access requests from user devices indicating content to be accessed in the respective service, the system further comprising at least one content delivery node, said service requests transmitted from each virtual service anchor including service requests to at least one content delivery node based on the indicated content.

18. One or more non-transitive computer-readable recording media storing computer-readable code which, when executed by processors of at least one networked computer and serving station, perform the method according to claim **1**.

* * * * *