



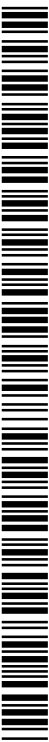
- (51) International Patent Classification:
H04W 48/00 (2009.01)
- (21) International Application Number:
PCT/EP2011/060598
- (22) International Filing Date:
24 June 2011 (24.06.2011)
- (25) Filing Language: English
- (26) Publication Language: English
- (71) Applicant (for all designated States except US): **NOKIA SIEMENS NETWORKS OY** [FI/FI]; Karaportti 3, FI-02610 Espoo (FI).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **HALME, Jarno Antero** [FI/FI]; Keijunpolku 5, FI-03100 Nummela (FI). **KAUPPINEN, Risto Antero** [FI/FI]; Keihäsrimne 22 C, FI-02610 Espoo (FI).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ,

CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))



WO 2012/175140 A1

(54) Title: GATEWAY SELECTION FOR LOAD BALANCING

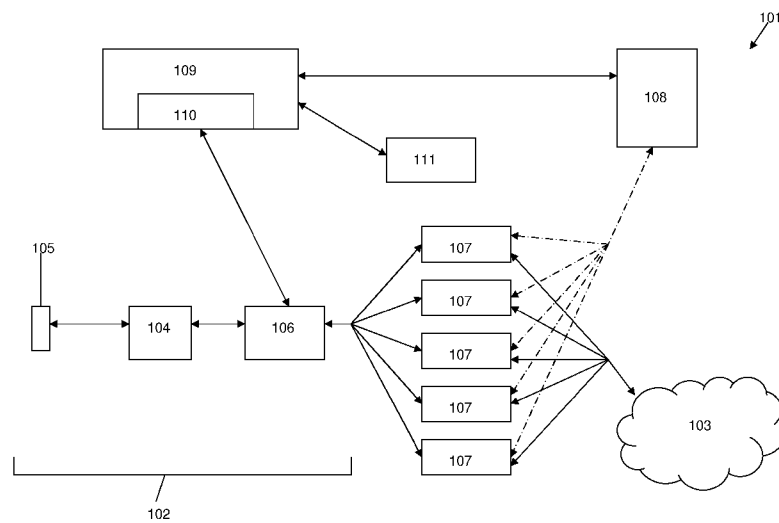


Figure 1

(57) Abstract: The present invention relates to load balancing and in particular to load balancing in gateways. One or more performance measurements relating one or more gateways (107) may be received and a load value for each of the gateways (107) may be determined based on the one or more performance measurements relating to each gateway (107). A request for a gateway (107) may be received from a network entity (106) and an available gateway (107) may be identified based on the determined load value for each of the gateways (107). An identification of the identified available gateway (107) may then be transmitted to the network entity (106).

GATEWAY SELECTION FOR LOAD BALANCING

5

The present invention relates to load balancing and, in particular, to load balancing in Gateways in a telecommunication network.

10 In conventional telecommunication networks a Gateway (GW) refers to a network entity that interfaces from one network to another or one part of a network to another part of the network. In other words, it acts as a gateway between adjoining or operatively adjoining networks or parts of a network.

15 Considering, for example, a network in a Mobile telecommunication domain, the mobile network may typically include, in a simplified view, an access network and a core network.

20 The access network enables User Equipment (UE), for example in the mobile domain, a mobile telephone, mobile device, a tablet, and so on, to receive and transmit traffic, e.g. voice traffic, data traffic, multimedia traffic and so on. The access network in the mobile domain may include various entities including, for example depending on the type of mobile network, Base Transceiver Stations (BTS), Node Bs (NB), Evolved Node Bs (eNB) and so on, to enable the UE to access wirelessly the mobile network.

25

The core network may be one or more of a packet switched network, a circuit switched network, Public Switched Telephone Network (PSTN), Wide Area Network (WAN), Local Area Network (LAN), Internet Protocol (IP) network, or any other type of network.

A GW may be located (either virtually or physically) between the access network and the core network where the GW may handle all traffic traversing between UE, via the access network, and the core network. Accordingly, if a GW becomes
5 overloaded with then there will be a significant degradation in quality of service to the UE and therefore the end-users of the mobile network, e.g. the subscribers to a Mobile Network Operator (MNO).

In Next Generation Mobile Networks (NGMN), which includes the so-called Fourth
10 Generation (4G) networks, for example, Long Term Evolution (LTE) networks, it is expected that there will be a substantial increase in traffic, e.g. data traffic, multimedia traffic, and so on, along with a substantial increase in the number of user equipment sessions. A user equipment session starts when the user equipment connects to the network, for example, via a transceiver, and continues
15 until the user equipment is disconnected from the network. During the session the user equipment may request and/or receive traffic from the core network, e.g. data traffic, multimedia traffic, TV on demand, video streaming, and so on. The increase of user equipment sessions and the increase in traffic from and/or to the user equipment mean that the gateways between the access network and the core
20 network will be substantially more heavily utilised. It would not be feasible for a Network Operator to provision the network with a substantial number of gateways due to the cost and that for periods of time the gateways may not be fully utilised and thus have significant spare capacity.

25 In the NGMN there is a desire to enable self-organizing functionality, which may be referred to as Self-organizing Networks (SON) due to the size and complexity of NGMNs as manual configuration may not be possible. However, standardised specific use cases for NGMNs will take a substantial amount of time to propose and be accepted and thus, as NGMNs are becoming available for deployment and
30 implementation, there is a need to provide certain functionality prior to

standardised solutions are implemented and to efficiently manage gateways in a telecommunication network.

5 The present invention seeks to address, at least in part, some or all of the problems and drawbacks discussed hereinabove.

10 According to a first aspect of the invention there is provided a method comprising the steps of: receiving one or more performance measurements relating one or more gateways; determining a load value for each of the gateways based on the one or more performance measurements relating to each gateway; receiving a request for a gateway from a network entity; identifying an available gateway based on the determined load value for each of the gateways; and transmitting an identification of the identified available gateway to the network entity.

15 The method may be implemented by any entity such as, for example, a network node, server, system, computing device, module, and so on. The entity implementing the method may a separate entity, in other words a stand alone entity, or the entity may form part of integrated with another entity, for example, the method may be implemented by a module (where the module may be hardware, software or a combination thereof) that is integrated with a management system of
20 the network.

25 One or more performance measurements are received relating to one or more gateways, wherein a network may include one or more gateways. The gateways may act as a gateway between an access network and a core network. The network may be a mobile network.

The performance measurements for one or more of the gateways may be received periodically, for example, every 10 seconds, every 30 seconds, every minute, every 5 minutes, every 10 minutes, and so on. As will be appreciated, the period between receiving consecutive performance measurements may be any period
5 suitable for the size, type and implementation of the network.

The method may further comprise requesting the performance measurements from a management system. The request for performance measurements may be performed periodically, for example, every 10 seconds, every 30 seconds, every
10 minute, every 5 minutes and so on. As will be appreciated, the period for requesting the performance measurements may be any period suitable for the size, type and implementation of the network.

The performance measurements may include counter, load, throughput, processor
15 load, capacity and any other Key Performance Indicator (KPI) data for one or more of the gateways. Based on the received performance measurements a load value for each gateway may be determined.

The load value determined for each gateway may be a percentage value, a
20 weighting value or any other rating or type of value that may enable an available gateway to be selected or identified based on the determined load values.

A request may be received from a network entity for an available gateway. The network entity from which the request may be received may be a Mobility
25 Management Entity, or any other network entity in the network. The network entity may request an available gateway to enable a session for User Equipment to be established whilst efficiently balancing the load on each of the gateways to substantially ensure that gateways are not overloaded and that the service provided for the User Equipment session is not degraded.

The available gateway may be identified or selected based on the determined load values.

5 For example, if the load value indicates the actual load or percentage load on each gateway then the available gateway may be identified or selected by comparing the load values to identify the gateway with the lowest load value or highest load value depending on the scale, type or meaning of the load value determined. Similarly, the load value may indicate the actual spare capacity or percentage
10 spare capacity for each gateway and then the available gateway may be identified or selected by comparing the load values to identify the gateway with the highest spare capacity. Thus, the step of identifying the best available gateway may further comprise the step of comparing the load value for each of the gateways to identify the best available gateway, for example, as a gateway having a lowest
15 load value.

In a further example, the load value may be a probability value that may be used as a probability factor in identifying an available gateway.

20 For example, if a first gateway (GW1) has a determined load value of 44% probability (which may be based on the performance measurements for GW1, for example, current or predicted spare capacity and/or load on GW1), a second gateway (GW2) has a determined load value of 31% probability (which may be based on the performance measurements for GW2, for example, current or
25 predicted spare capacity and load on GW2) and a third gateway (GW3) has a determined load value of 25% probability (which may be based on the performance measurements for GW3, for example, current or predicted spare capacity and load on GW3) then, on average, using the determined load values, e.g. probabilities or percentages, GW1 will be identified 44 times out of 100

requests for a gateway as the available gateway, GW2 will be identified 31 times out of 100 requests for a gateway as the available gateway and GW3 will be identified 25 times out of 100 requests for a gateway as the available gateway. Thus, the determined load value may define a probability of identifying a gateway and the step of identifying an available gateway may be based on the probability.

Accordingly, over a period of time the load for the network of gateways will be shared and balanced amongst the gateways in the network depending on the determined load values for the gateways.

10

However, any type and/or scale, (e.g. percentage, absolute value, probability, etc), for the load value determined for each gateway (based on one or more of load, capacity, spare capacity, number of sessions, processor usage, etc), may be implemented to enable the available gateway to be identified or selected.

15

The identification or selection of the available gateway may be further based on actual parameters or predicted/estimated parameters relating to the session to be established, for example, the type of traffic (e.g. multimedia data, video streaming, and so on), the length of time of the session, bandwidth or capacity required, and so on. In other words, the identification of an available gateway may be further based on the estimated/predicted or actual parameters indicating the actual or estimated load that may be generated by the new session for a gateway.

20

Thus, the method may enable one or more gateways in a network to be efficiently and effectively managed in terms of load balancing across the gateways. At the time a gateway is required for a session establishment for User Equipment an available gateway can be identified and/or selected for that session which effectively and efficiently balances the load between the gateways in the network.

25

The step of determining the load value for each of the gateways may further comprise the step of determining a current load value for each of the gateways. Thus, the method may determine the current load of each gateway where the
5 current load is effectively related to the point in time the performance measurements were taken or collected.

The step of determining the load value for each of the gateways may further comprise the step of determining a predicted load value for each of the gateways.
10 As each set of performance measurements for each gateway are taken at a point in time and the performance measurements are collected and/or received periodically then the method may determine a predicted load value for each of the gateways for the period until the next consecutive set of performance measurements are collected and/or received. The predicted load value may be
15 determined based on, for example, one or more of the current performance measurements, a series of performance measurements for each gateway (e.g. using historical performance measurements for each gateway), estimated/predicted or actual trends/patterns (e.g. time of day), the type of sessions each gateway is handling, and so on.

20

The step of determining the load value for each of the gateways maybe performed periodically. For example, each time a new set of performance parameters for each of the one or more gateways is received, the load value may be determined. Thus, the load values for each gateway may be dynamic or variable and may be
25 determined periodically to ensure that the load balancing between gateways is maintained.

The method may further comprise the steps of receiving a list of the one or more gateways; and storing the list of one or more gateways. Thus, a list of the one or

more gateways in the network may be obtained so that it is known which gateways are in the network. This list of gateways may change depending on new gateways being provisioned, gateways that are faulty, undergoing maintenance, and so on. The list of gateways may be requested from a network entity (e.g. management system, Domain Name Server (DNS), and so on) and the list of gateways received
5 in response to the request, e.g. pulled. The list of gateways may be received, e.g. pushed, from a network entity when the list of gateways is generated, updated or changed.

10 According to a second aspect of the present invention there is provided an apparatus comprising: a first input adapted to receive one or more performance measurements relating one or more gateways; a first processor adapted to determine a load value for each of the gateways based on the one or more performance measurements relating to each gateway; a second input adapted to
15 receive a request for a gateway from a network entity; a second processor adapted to identify an available gateway based on the determined load value for each of the gateways; and an output adapted to transmit an identification of the identified available gateway to the network entity.

20 According to a third aspect of the present invention there is provided an apparatus adapted to: receive one or more performance measurements relating one or more gateways; determine a load value for each of the gateways based on the one or more performance measurements relating to each gateway; receive a request for a gateway from a network entity; identify an available gateway based on the
25 determined load value for each of the gateways; and transmit an identification of the identified available gateway to the network entity.

The apparatus may be a computing device, e.g. a server, personal computer, and so on. The apparatus may be a module or emulator that is separate to, attached

to or integrated with any other computing device, system (e.g. management system), or network.

5 The first processor may be further adapted to determine a current load value for each of the gateways. The first processor may be further adapted to determine a predicted load value for each of the gateways.

10 The first processor may be further adapted to determine the load value for each of the gateways periodically.

The second processor may be further adapted to compare the load value for each of the gateways to identify the available gateway.

15 The determined load value may define a probability of identifying a gateway and the second processor may be further adapted to identify the available gateway based on the probability.

20 The apparatus may further comprise a third input adapted to receive a list of the one or more gateways; and a memory adapted to store the list of one or more gateways.

The apparatus may be adapted to perform any or all of the functionality of the aspects of the present invention.

25 The apparatus may be adapted or configured using hardware, software or any combination thereof. The first input, second input and third input may be the same

input, different inputs or any combination thereof. The first processor and the second processor may be the same processor or different processors.

5 According to a fourth aspect of the present invention there is provided a computer program product comprising computer readable executable code for: receiving one or more performance measurements relating one or more gateways; determining a load value for each of the gateways based on the one or more performance measurements relating to each gateway; receiving a request for a gateway from a network entity; identifying an available gateway based on the determined load
10 value for each of the gateways; and transmitting an identification of the identified available gateway to the network entity.

The computer program product may further comprise computer readable executable code for performing any or all of the functions in accordance with the
15 aspects of the invention.

Embodiments of the present invention will now be described, by way of example only, and with reference to the accompanying drawings, in which:

20 Figure 1 shows a simplified block diagram of a mobile network in accordance with many of the embodiments of the present invention.

The mobile network 101 shown in Figure 1 relates to a fourth generation (4G), Long Term Evolution (LTE) mobile network which is a Next Generation Mobile
25 Network (NGMN). Thus, many of the embodiments will be described in relation to LTE networks and several nodes or entities may be referred to by the name associated with LTE networks and may include nodes or entities that are relevant to LTE networks. However, as will be appreciated, many of the embodiments may

be applicable to other mobile networks, e.g. second generation (2G) networks, third generation networks (3G), and so on.

5 In the other network types, different network nodes, entities, systems or servers may implement similar functionality to those described hereinbelow in relation to 4G LTE networks. The physical or virtual location and/or the names of the network nodes, entities, systems or servers may be different in the other mobile networks but the load balancing of many of the embodiments described hereinbelow may be implemented in the other types of networks.

10

Figure 1 shows a simplified block diagram of a 4G LTE mobile network 101, which comprises an access network 102 and a core network 103. The access network 102 may include, amongst other entities, one or more Evolved Node Bs (eNBs) 104 to enable mobile devices 105, e.g. mobile telephones, tablet devices, 15 notebooks, laptops, and so on, to connect to the LTE mobile network 101 in order to transmit and receive traffic. As will be appreciated, in other mobile networks the eNBs may instead be Base Transceiver Stations (BTS), Node Bs (NB), and so on.

20 Only one eNB is shown in Figure 1, but as will be appreciated, it is expected that a 4G LTE network may comprise tens, hundreds or even thousands of eNBs 104 in the access network 102 of the mobile network 101.

eNB 104 may be operatively connected to one or more Mobility Management Entity (MME) 106. The MME 106 is an LTE network node that is responsible for 25 various functions, services and processes in the LTE access network 102. For example, the MME 106 may be responsible, amongst others, for idle mode User Equipment (UE) 105 tracking, bearer activation/deactivation processes, authenticating the UE 105, and so on.

The MME 106 may also serve the purpose of establishing a session, or enabling session establishment, for a UE 105 that is requesting to transmit or receive traffic, e.g. multimedia data traffic, over the LTE mobile network 101. In this respect, the

5 MME 106 may receive a request from eNB 104 to establish a session for UE 105 that is operatively communicating with eNB 104. In order to enable the session establishment for UE 105 the MME 106 may request a selected Serving Gateway (S-GW) 107 through which the session for UE 105 may be channelled to the core network 103.

10

The S-GW 107 acts as a gateway between the access network 102 and the core network 103 and provide various functions and services. For example, the S-GW 107 may route and forward data packets from and to UE105, and so on. Thus, given the predicted or expected significant increase in UE 105 sessions and traffic

15 over the mobile network 101 that will be routed via the S-GWs 107 then it will be an important factor to be able to manage or select the appropriate S-GW 107 to handle each UE 105 session in order to efficiently manage and control the load balancing so as to not degrade a user's experience and quality of service.

20 The mobile network 101 may be managed by one or more management systems. For example, the mobile network 101 may include a Network Management System (NMS) 108, along with any number of other management systems, such as Element Management Systems (EMS), Charging and Billing Systems, and so on. For ease of reference only the NMS 108 is shown in Figure 1, however as will be

25 appreciated by a person skilled in the art, network systems typically comprise several management systems or management functions in order to ensure the operation of the network.

The NMS 108 may include various functions, services and processes to manage the mobile network 101 as a whole. For example, the NMS 108 may include, amongst others, Performance Management functions/modules, Fault Management functions/modules, Configuration Management functions/modules, and so on.

5

The NMS 108 may be operatively connected to other management systems, e.g. the EMS (not shown) in order to request and/or receive information and or data relating to the operation, maintenance and implementation of the mobile network 101. In this regard, the NMS 108 may be operatively connected to the S-GWs 107, directly or indirectly via another management system, network entity, network node, or module.

10

The NMS 108 may receive various performance measurements relating to each of the S-GWs 107 in the mobile network 101. The NMS 108 may receive the performance measurements automatically and periodically, for example, every 10 seconds, every 30 seconds, every minute, every 5 minutes and so on. As will be appreciated, the periodic time period may be set at any time period appropriate for the size and type of network to ensure that the NMS 108 has up-to-date and adequate performance measurements or data for the S-GWs 107 in the mobile network 101. The NMS 108 may request the performance measurements from the S-GWs 107, or from any other management system, entity, node or module.

15

20

The performance measurements may include any performance measurements relating to the performance and/or operation of the S-GW 107. For example, the performance measurements may include, for example, counter, load, number of active sessions, processor loads, and/or any other Key Performance Indicators (KPI) relating to each of the S-GWs 107 in the mobile network 101.

25

In many of the embodiments, the mobile network 101 may further include a Process Automation Engine (PAE) 109. In Figure 1, PAE 109 is shown as a separate server or system in the mobile network 101 that is operatively connected to the NMS 108 and the MME 106. However, the PAE 109 may be part of a management system, e.g. the NMS 108, part of another network entity or node, a module, or any other device or entity that may perform the function of the PAE 109 of many of the embodiments described herein.

The PAE 109 may also be operatively connected to a Domain Name Server (DNS) 111, where the PAE 109 may request and/or receive a list of the S-GWs 107 deployed or provisioned in the mobile network 101 from the DNS 111.

The PAE 109 may request and/or receive the performance measurements for each of the S-GWs 107 in the mobile network 101. The PAE 109 may receive the performance measurements periodically from the NMS 108, for example, every minute. However, as will be appreciated the period between which the PAE 109 receives the may be set at any period that is suitable or appropriate for the purpose of many of the embodiments. The PAE 109 may receive the performance measurements automatically from the NMS 108 or the PAE 109 may periodically request the performance measurements from the NMS 108.

Based on the received performance measurements relating to each S-GW 107 in the mobile network 101, the PAE 109 may determine or calculate a load value, for example, a load percentage, a weighting value or any other rating or differentiator value, for each of the S-GWs 107.

As a person skilled in the art will appreciate, there are many mechanisms, algorithms, etc, that may be utilised in order to determine a load value for each

gateway in the network based on performance measurements. For example, the load value may be a percentage, a weighting value, an absolute value, and so on.

In one example the load value for each S-GW 107 may be determined as the
5 overall load/spare capacity of each S-GWs 107 based on the current performance measurements for each S-GW 107.

In another example, the load value for each S-GW 107 may be determined as a
10 weighted value or percentage of the load/spare capacity of each S-GW 107 based on the current performance measurements for each S-GW 107 in relation to, or in respect of, the total load/spare capacity in the network of S-GWs 107.

In these examples, the available gateway may be identified or selected by
15 comparing the load values to identify the gateway with the lowest load value or highest load value depending on the scale, type or meaning of the load value determined.

In a further example, the load value may be determined as a probability value that
20 may be used as a probability factor in identifying an available gateway.

For example, if a first gateway (GW1) has a determined load value of 44%
probability (which may be based on the performance measurements for GW1, for
example, current or predicted spare capacity and/or load on GW1), a second
gateway (GW2) has a determined load value of 31% probability (which may be
25 based on the performance measurements for GW2, for example, current or predicted spare capacity and load on GW2) and a third gateway (GW3) has a determined load value of 25% probability (which may be based on the performance measurements for GW3, for example, current or predicted spare

capacity and load on GW3) then, on average, using the determined load values, e.g. probabilities or percentages, GW1 will be identified 44 times out of 100 requests for a gateway as the available gateway, GW2 will be identified 31 times out of 100 requests for a gateway as the available gateway and GW3 will be
5 identified 25 times out of 100 requests for a gateway as the available gateway.

However, as will be appreciated, any type and/or scale, (e.g. percentage, absolute value, probability, etc), for the load value determined for each gateway (based on one or more of load, capacity, spare capacity, number of sessions, processor
10 usage, etc), may be implemented to enable an available gateway to be identified or selected.

Accordingly, in all the examples then over a period of time the load for the network of gateways will be shared and balanced amongst the gateways in the network
15 depending on the determined load values for the gateways.

In many of the embodiments, depending on the period between receiving performance measurements for each of the S-GWs 107 in the mobile network 101, the PAE 109 may further determine or calculate a predicted load value, e.g. load
20 percentage, for each S-GW 107. The predicted load value relates to the expected load of each S-GW 107 for the period between receiving performance measurements for each S-GW 107 from the NMS 108. The predicted load value may be based on the performance measurements along with historical performance measurements, trends or patterns associated with the S-GWs 107 as
25 a group or each individual S-GW 107 in the mobile network 101.

Thus, in many of the embodiments the PAE 109 is continually or periodically determining or calculating a load value for each of the S-GWs 107 in the mobile network 101. In many of the embodiments, the PAE 109 may further or

alternatively determine or calculate a predicted load value for each of the S-GWs 107 in the mobile network 101 to predict the load of each S-GW 107 for the time period between the periodic determination or calculation of the load value for each S-GW 107.

5

When a UE 105 operatively connects wirelessly to the eNB 104, the eNB 104 may interact with the MME 106 in order to enable the session establishment for the UE 105. The MME 106 conventionally or typically may request a selected S-GW 107 from the DNS 111 and the DNS 111 would respond to the request with a selected
10 S-GW 107 for the UE 105 session establishment. Typically, the DNS 111 may select a S-GW 107 based on a simple round robin mechanism, in other words, the next S-GW in a list, which is an inefficient mechanism and may cause one or more S-GWs to become overloaded thereby degrading the service provided to the UE 105.

15

However, as current MMEs 106 are configured to request a selection of a S-GW 107 from a DNS then, in many of the embodiments, the PAE 109 may include a DNS emulator 110 to emulate a DNS. Therefore, in many of the embodiments, the MME 106 may now request a selected S-GW 107 from the DNS emulator 110.

20

By implementing a DNS emulator 110 then little or no changes are required in the present process for the MME 106 to request a gateway which enhances many of the embodiments. However, as will be appreciated, many of the embodiments may be implemented without the DNS emulator 110 where the MME 106 is
25 instructed to communicate directly with the PAE 109, the entity performing the function of the PAE 109 and/or the DNS emulator 110.

On receipt of the request for a S-GW 107 the DNS emulator 110 based on either or both of the determined load value and predicted load value for each S-GW 107

in the mobile network 101 may select the best available S-GW 107 to handle the session for the UE 105. The PAE 109 via the DNS emulator 110 may then inform or transmit an identification of the selected or identified S-GW 107 to the MME 106 so that the session for the UE 105 can be established using the selected S-GW
5 107 to handle, amongst other functions, the routing of the data packets between the UE 105 (via eNB 104) and the core network 103.

Accordingly, many of the embodiments enable an efficient and effective mechanism for load balancing of gateways in a network. By determining a load
10 value periodically for each gateway then the load from User Equipment sessions may be shared and balanced between the gateways. Thus, the gateways may be utilised more efficiently, may prevent gateways becoming overloaded and thereby maintaining a high level of service for the user's of the User Equipment.

15 Many of the embodiments have several advantages, for example, a fast solution to provide network resource efficient usage in networks, e.g. LTE networks, may be implemented now as part of a network management solution that also enables multi-vendor support without waiting for a standardised approach to be agreed, and determination of the load values for each gateway may be fine tuned
20 depending on usage of the gateways and gateway vendor specific implementation and characteristics.

Many of the embodiments were described in relation to Long Term Evolution networks and to Serving Gateways. However, a will be appreciated by a person
25 skilled in the art that the principles of the present invention may be applicable to other types of networks, e.g. 2G, 3G and future network technologies as well as to different types or functions of gateways.

While preferred embodiments of the invention have been shown and described, it will be understood that such embodiments are described by way of example only. Numerous variations, changes and substitutions will occur to those skilled in the art without departing from the scope of the present invention as defined by the
5 appended claims. Accordingly, it is intended that the following claims cover all such variations or equivalents as fall within the spirit and the scope of the invention.

Claims

1. A method comprising the steps of:
 - receiving one or more performance measurements relating one or more
 - 5 gateways;
 - determining a load value for each of said gateways based on said one or more performance measurements relating to each gateway;
 - receiving a request for a gateway from a network entity;
 - identifying an available gateway based on said determined load value for
 - 10 each of said gateways; and
 - transmitting an identification of said identified available gateway to said network entity.

2. The method as claimed in claim 1 in which said step of determining said
- 15 load value for each of said gateways further comprises the step of:
 - determining a current load value for each of said gateways.

3. The method as claimed in claim 1 or 2 in which said step of determining said load value for each of said gateways further comprises the step of:
- 20 determining a predicted load value for each of said gateways.

4. The method as claimed in any one of the preceding claims in which said step of determining said load value for each of said gateways is performed periodically.

5. The method as claimed in any one of the preceding claims in which said step of identifying said available gateway further comprises the step of:
- comparing said load value for each of said gateways to identify said available gateway.
- 5
6. The method as claimed in any one of the preceding claims in which said determined load value defines a probability of identifying a gateway and identifying said available gateway based on said probability.
- 10 7. The method as claimed in any one of the preceding claims further comprising the steps of:
- receiving a list of said one or more gateways; and
- storing said list of one or more gateways.
- 15 8. An apparatus comprising:
- a first input adapted to receive one or more performance measurements relating one or more gateways;
- a first processor adapted to determine a load value for each of said gateways based on said one or more performance measurements relating to each
- 20 gateway;
- a second input adapted to receive a request for a gateway from a network entity;
- a second processor adapted to identify an available gateway based on said determined load value for each of said gateways; and
- 25 an output adapted to transmit an identification of said identified available gateway to said network entity.

9. The apparatus as claimed in claim 8 in which said first processor is further adapted to:

determine a current load value for each of said gateways.

5

10. The apparatus as claimed in claim 8 or 9 in which said first processor is further adapted to:

determine a predicted load value for each of said gateways.

10 11. The apparatus as claimed in any one of claims 8 to 10 in which said first processor is further adapted to:

determine said load value for each of said gateways periodically.

12. The apparatus as claimed in any one of claims 8 10 11 in which said
15 second processor is further adapted to:

compare said load value for each of said gateways to identify said available gateway.

13. The apparatus as claimed in any one of claims 8 to 12 in which said
20 determined load value defines a probability of identifying a gateway and said second processor is further adapted to identify said available gateway based on said probability.

14. The apparatus as claimed in any one of claims 8 to 13 further comprising:
25 a third input adapted to receive a list of said one or more gateways; and

a memory adapted to store said list of one or more gateways.

15. A computer program product comprising computer readable executable code for:

5 receiving one or more performance measurements relating one or more gateways;

determining a load value for each of said gateways based on said one or more performance measurements relating to each gateway;

receiving a request for a gateway from a network entity;

10 identifying an available gateway based on said determined load value for each of said gateways; and

transmitting an identification of said identified available gateway to said network entity.

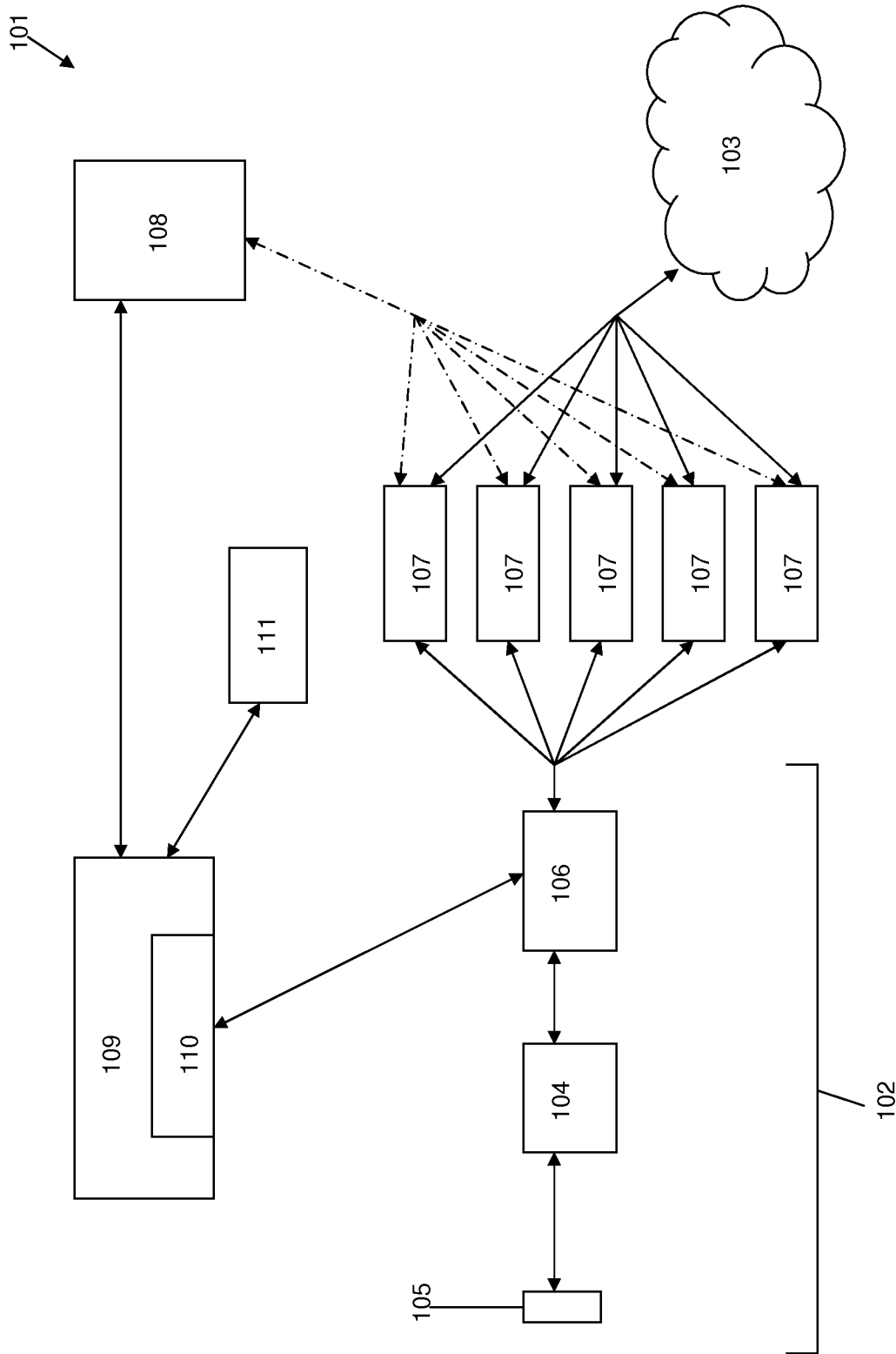


Figure 1

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2011/060598

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04W48/00
ADD.
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED
Minimum documentation searched (classification system followed by classification symbols)
H04W
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, COMPENDEX, INSPEC, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2009/092440 A1 (ERICSSON TELEFON AB L M [SE]; MIHALY ATTILA [HU]; TOTH GABOR [HU]; WES) 30 July 2009 (2009-07-30) page 1, line 5 - line 7 page 3, line 31 - page 6, line 19 page 15, line 8 - line 29 page 16, line 12 - line 30 page 18, lines 5,25 - page 19, line 23 page 21, line 26 - line 27 figures 3, 4, 5, 6	1-15
X	US 2009/285179 A1 (JONES D MARK [CA] ET AL) 19 November 2009 (2009-11-19) paragraph [0009] - paragraph [0011] paragraphs [0023], [0026] paragraph [0028] - paragraph [0036] figures 2, 3	1,2,4,5, 7-9,11, 12,14,15

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search 2 March 2012	Date of mailing of the international search report 09/03/2012
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer Rosenauer, Hubert

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2011/060598

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2010/102127 A1 (CISCO TECH INC [US]; RAMANKUTTY RAJESH [US]; CHOWDHURY KUNTAL [US]; GI) 10 September 2010 (2010-09-10) paragraphs [0002], [0007], [0009] paragraph [0016] - paragraph [0027] claim 1 figures 3, 4	1,2,4,5, 7-9,11, 12,14,15
A	----- US 2003/156537 A1 (SOULHI SAID [CA] ET AL) 21 August 2003 (2003-08-21) paragraphs [0002], [0005] paragraph [0020] - paragraph [0030] figure 2	1,8,15
A	----- EP 1 587 272 A1 (CIT ALCATEL [FR]) 19 October 2005 (2005-10-19) paragraph [0008] - paragraph [0009] paragraph [0011] - paragraph [0017] figure 1	1-15
A	----- WO 2010/132884 A1 (CISO TECHNOLOGY INC [US]; RAMANKUTTY RAJESH [US]; KOODLI RAJEEV [US];) 18 November 2010 (2010-11-18) paragraph [0023] paragraph [0028] - paragraph [0030] paragraph [0035] paragraph [0039] - paragraph [0040] paragraph [0053] - paragraph [0058] figures 2, 3	1-15
A	----- WO 2007/092687 A2 (MOTOROLA INC [US]; ALMENDROS FERNANDO [ES]; RIVAS IGNACIO [ES]) 16 August 2007 (2007-08-16) page 20, line 21 - page 21, line 16 page 26, line 24 - line 29 figure 3	1-15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2011/060598

Patent document cited in search report	Publication date	Patent family member(s)	Publication date	
WO 2009092440	A1	30-07-2009	CA 2711467 A1	30-07-2009
			CN 101926153 A	22-12-2010
			EP 2241087 A1	20-10-2010
			JP 2011512715 A	21-04-2011
			US 2010291943 A1	18-11-2010
			WO 2009092440 A1	30-07-2009

US 2009285179	A1	19-11-2009	NONE	

WO 2010102127	A1	10-09-2010	CN 102342144 A	01-02-2012
			US 2010281157 A1	04-11-2010
			WO 2010102127 A1	10-09-2010

US 2003156537	A1	21-08-2003	NONE	

EP 1587272	A1	19-10-2005	CN 1684448 A	19-10-2005
			EP 1587272 A1	19-10-2005
			US 2005226258 A1	13-10-2005

WO 2010132884	A1	18-11-2010	CN 102077636 A	25-05-2011
			EP 2430851 A1	21-03-2012
			US 2010299419 A1	25-11-2010
			WO 2010132884 A1	18-11-2010

WO 2007092687	A2	16-08-2007	GB 2435150 A	15-08-2007
			WO 2007092687 A2	16-08-2007
