



## (51) International Patent Classification:

H04W 24/02 (2009.01) H04L 45/121 (2022.01)

H04L 43/0852 (2022.01) H04L 47/283 (2022.01)

## (21) International Application Number:

PCT/IN2024/050819

## (22) International Filing Date:

14 June 2024 (14.06.2024)

## (25) Filing Language:

English

## (26) Publication Language:

English

## (30) Priority Data:

202321044640 04 July 2023 (04.07.2023) IN

(71) Applicant: **JIO PLATFORMS LIMITED** [IN/IN]; Office - 101, Saffron, Nr. Centre Point, Panchwati 5 Rasta, Ambawadi, Ahmedabad - 380006, Gujarat (IN).

(72) Inventors: **BISHT, Sandeep**; Office - 101, Saffron, Nr. Centre Point, Panchwati 5 Rasta, Ambawadi, Ahmedabad - 380006, Gujarat (IN). **BHATNAGAR, Aayush**; Office - 101, Saffron, Nr. Centre Point, Panchwati 5 Rasta, Ambawadi, Ahmedabad - 380006, Gujarat (IN). **SINHA, Anurag**; Office - 101, Saffron, Nr. Centre Point, Panchwati 5 Rasta, Ambawadi, Ahmedabad - 380006, Gujarat (IN). **ANSARI, Ezaj**; Office - 101, Saffron, Nr. Centre Point,

Panchwati 5 Rasta, Ambawadi, Ahmedabad - 380006, Gujarat (IN). **YADAV, Ravindra**; Office - 101, Saffron, Nr. Centre Point, Panchwati 5 Rasta, Ambawadi, Ahmedabad - 380006, Gujarat (IN). **PANDEY, Prashant**; Office - 101, Saffron, Nr. Centre Point, Panchwati 5 Rasta, Ambawadi, Ahmedabad - 380006, Gujarat (IN).

(74) Agent: **SAHNEY, Garima**; SAIKRISHNA & ASSOCIATES ADVOCATES, B-140, Sector 51, Noida-201301, Uttar Pradesh (IN).

(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, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST,

## (54) Title: METHOD AND SYSTEM FOR OPTIMISING LATENCY ASSOCIATED WITH A NETWORK

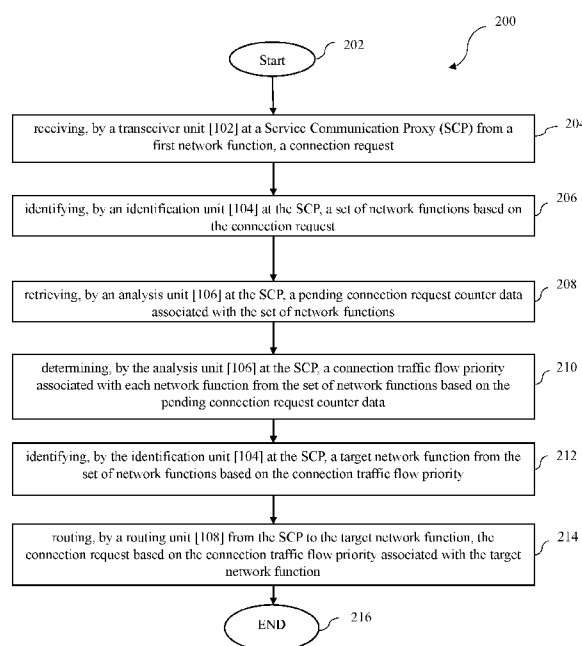


FIG. 2

(57) Abstract: The present disclosure relates to a method [200] and a system [100] for optimising latency associated with a network. The disclosure encompasses, receiving, by a transceiver unit [102] from a first network function, a connection request, identifying, by an identification unit [104], a set of network functions based on the connection request, retrieving, by an analysis unit [106], a pending connection request counter data associated with the set of network functions, determining, by the analysis unit [106], a connection traffic flow priority associated with each network function from the set of network functions based on the pending connection request counter data, identifying, by the identification unit [104], a target network function from the set of network functions based on the connection traffic flow priority; and routing, by a routing unit [108] the connection request based on the connection traffic flow priority associated with the target network function.

SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, 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, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

**Declarations under Rule 4.17:**

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *of inventorship (Rule 4.17(iv))*

**Published:**

- *with international search report (Art. 21(3))*

## METHOD AND SYSTEM FOR OPTIMISING LATENCY ASSOCIATED WITH A NETWORK

### FIELD OF THE DISCLOSURE

5

[001] The present disclosure relates generally to the field of wireless communication systems. More particularly, the present disclosure relates to methods and systems for optimising latency associated with a network.

### 10 BACKGROUND

[002] The following description of related art is intended to provide background information pertaining to the field of the disclosure. This section may include certain aspects of the art that may be related to various features of the present disclosure. However, it should be appreciated that  
15 this section be used only to enhance the understanding of the reader with respect to the present disclosure, and not as admissions of prior art.

[003] Wireless communication technology has undergone rapid evolution over the past few decades, with each generation bringing significant improvements and advancements. The first  
20 generation of wireless communication technology was based on analog technology and primarily offered voice services. However, the advent of second-generation (2G) technology introduced digital communication and data services, along with the introduction of text messaging.

[004] The introduction of third generation (3G) technology marked a significant milestone,  
25 enabling high-speed internet access, mobile video calling, and location-based services. Subsequently, the fourth generation (4G) technology revolutionized wireless communication with faster data speeds, broader network coverage, and enhanced security features.

[005] In the field of telecommunication networks, one persistent challenge is the issue of average  
30 latency. Latency refers to the delay experienced in data transmission between network endpoints, directly impacting the quality and responsiveness of real-time applications. High latency can result in various issues, including delays in voice and video communication, lag in online gaming, and buffering in streaming services. These problems can lead to a frustrating user experience, hinder productivity, and limit the potential of telecommunication technologies.

35

[006] Over time, various solutions have been developed to improve the performance of communication devices and optimize the average latency associated with networks. However, existing solutions often face challenges such as ineffective prioritization and routing of data packets based on latency requirements. Additionally, prior solutions lack the ability to dynamically  
5 adapt to changing network conditions, resulting in suboptimal latency performance during peak traffic periods or in geographically dispersed networks. Furthermore, existing solutions do not adequately address the impact of network congestion, inefficient data compression techniques, or subpar signal processing algorithms, all of which contribute to latency issues. Moreover, prior art does not sufficiently consider the impact of latency on different types of applications, such as real-  
10 time voice and video communication, where even minimal delays can significantly degrade user experience.

[007] Thus, there exists an imperative need in the art to optimise latency associated with a network, which the present disclosure aims to address.

15

## **OBJECTS OF THE DISCLOSURE**

[008] Some of the objects of the present disclosure, which at least one embodiment disclosed herein satisfies are listed herein below.

20

[009] It is an object of the present disclosure to provide a system and a method for optimising latency associated with the network.

[010] It is another object of the present disclosure to provide a solution that identifies a target  
25 network function from one or more network functions based on a number of pending requests associated with each network function from the one or more network functions.

[011] It is yet another object of the present disclosure to provide a solution to determine a number of pending requests associated with each network function from the one or more network functions  
30 based on an active stream data associated with each connection point from the one or more connection points present in a network.

[012] It is yet another object of the disclosure to reduce overall latency in 5G core network.

[013] It is yet another object of the disclosure to further reduce pending request queue at network functions (servers) which saves from any type of anomaly at network function (server) end which increases overall request success rate in the system.

- 5 [014] It is yet another object of the disclosure to provide a method for optimizing latency that reduces the overall request processing time.

## SUMMARY OF THE DISCLOSURE

- 10 [015] This section is provided to introduce certain aspects of the present disclosure in a simplified form that are further described below in the detailed description. This summary is not intended to identify the key features or the scope of the claimed subject matter.

- [016] An aspect of the present disclosure may relate to a method for optimising latency associated with a network. The method comprises receiving, by a transceiver unit at a Service Communication Proxy (SCP) from a first network function, a connection request. The method further comprises identifying, by an identification unit at the SCP, a set of network functions based on the connection request. Further the method comprises retrieving, by an analysis unit at the SCP, a pending connection request counter data associated with the set of network functions. The method further comprises determining, by the analysis unit at the SCP, a connection traffic flow priority associated with each network function from the set of network functions based on the pending connection request counter data. The method further comprises identifying, by the identification unit at the SCP, a target network function from the set of network functions based on the connection traffic flow priority. The method further comprises routing, by a routing unit from the SCP to the target network function, the connection request based on the connection traffic flow priority associated with the target network function.

- [017] In an aspect of the present disclosure the pending connection request counter data associated with the set of network functions comprises at least a total number of pending connection requests associated with each network function from the set of network functions.

[018] In an aspect of the present disclosure, the connection traffic flow priority is one of a highest connection traffic flow priority and a lowest connection traffic flow priority.

[019] In an aspect of the present disclosure, the highest connection traffic flow priority associated with at least one network function from the set of network functions is determined by the analysis unit in an event the total number of pending connection request associated with said network function is determined as lowest value associated with the total number of pending connection request.

[020] In an aspect of the present disclosure, the lowest connection traffic flow priority associated with at least one network function from the set of network functions is determined by the analysis unit in an event the total number of pending connection request associated with said network function is determined as highest value associated with the total number of pending connection request.

[021] In an aspect of the present disclosure, the target network function from the set of network functions is identified by the identification unit based on at least the highest connection traffic flow priority.

[022] Another aspect of the present disclosure may relate to a system for optimising a latency associated with a network. The system is configured at a Service Communication Proxy (SCP). The system further comprising a transceiver unit, wherein the transceiver unit is configured to receive from a first network function, a connection request. The system further comprising an identification unit, wherein the identification unit is configured to identify a set of network functions based on the connection request. The system further comprising an analysis unit, wherein the analysis unit is configured to retrieve a pending connection request counter data associated with the set of network functions. The analysis unit is further configured to determine a connection traffic flow priority associated with each network function from the set of network functions based on the pending connection request counter data. Furthermore, the identification unit is further configured to identify a target network function from the set of network functions based on the connection traffic flow priority. The system further comprising a routing unit, wherein the routing unit configured to route to the target network function, the connection request based on the connection traffic flow priority associated with the target network function.

[023] Yet another aspect of the present disclosure may relate to a non-transitory computer readable storage medium storing instruction for optimising latency associated with a network. The instructions include an executable code which, when executed by one or more units of a system, causes a transceiver unit [102] of the system to receive a connection request from a first network

function; an identification unit [104] of the system to a set of network functions based on the connection request; an analysis unit [106] of the system to retrieve a pending connection request counter data associated with the set of network functions and to determine a connection traffic flow priority associated with each network function from the set of network functions based on the pending connection request counter data; the identification unit [104] further identify a target network function from the set of network functions based on the connection traffic flow priority and a routing unit [108] of the system to route to the target network function, the connection request based on the connection traffic flow priority associated with the target network function.

## BRIEF DESCRIPTION OF DRAWINGS

[024] The accompanying drawings, which are incorporated herein, and constitute a part of this disclosure, illustrate exemplary embodiments of the disclosed methods and systems in which like reference numerals refer to the same parts throughout the different drawings. Components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present disclosure. Some drawings may indicate the components using block diagrams and may not represent the internal circuitry of each component. It will be appreciated by those skilled in the art that disclosure of such drawings includes disclosure of electrical components, electronic components or circuitry commonly used to implement such components.

[025] FIG. 1A illustrates an exemplary block diagram representation of 5th generation core (5GC) network architecture [101].

[026] FIG.1B illustrates an exemplary block diagram of a system [100] for optimising latency associated with a network, in accordance with exemplary embodiments of the present disclosure.

[027] FIG.2 illustrates an exemplary method flow diagram indicating the process [200] for optimising latency associated with a network, in accordance with exemplary embodiments of the present disclosure.

[028] FIG.3 illustrates an exemplary scenario block diagram [300] of a system for maintaining a number of connection request for optimising latency associated with a network, in accordance with exemplary embodiments of the present disclosure.

[029] FIG.4 illustrates an exemplary scenario method flow diagram indicating the process [400] for optimising latency associated with a network, in accordance with exemplary embodiments of the present disclosure.

5 [030] FIG. 5 illustrates an exemplary block diagram of a computing device [1000] upon which an embodiment of the present disclosure may be implemented.

[031] The foregoing shall be more apparent from the following more detailed description of the disclosure.

10

## **DETAILED DESCRIPTION**

[032] In the following description, for the purposes of explanation, various specific details are set forth in order to provide a thorough understanding of embodiments of the present disclosure.

15 It will be apparent, however, that embodiments of the present disclosure may be practiced without these specific details. Several features described hereafter can each be used independently of one another or with any combination of other features. An individual feature may not address any of the problems discussed above or might address only some of the problems discussed above. Some of the problems discussed above might not be fully addressed by any of the features described  
20 herein. Example embodiments of the present disclosure are described below, as illustrated in various drawings in which like reference numerals refer to the same parts throughout the different drawings.

[033] The ensuing description provides exemplary embodiments only, and is not intended to limit  
25 the scope, applicability, or configuration of the disclosure. Rather, the ensuing description of the exemplary embodiments will provide those skilled in the art with an enabling description for implementing an exemplary embodiment. It should be understood that various changes may be made in the function and arrangement of elements without departing from the spirit and scope of the disclosure as set forth.

30

[034] It should be noted that the terms "mobile device", "user equipment", "user device", "communication device", "device" and similar terms are used interchangeably for the purpose of describing the disclosure. These terms are not intended to limit the scope of the disclosure or imply any specific functionality or limitations on the described embodiments. The use of these terms is  
35 solely for convenience and clarity of description. The disclosure is not limited to any particular

type of device or equipment, and it should be understood that other equivalent terms or variations thereof may be used interchangeably without departing from the scope of the disclosure as defined herein.

5 [035] Specific details are given in the following description to provide a thorough understanding of the embodiments. However, it will be understood by one of ordinary skill in the art that the embodiments may be practiced without these specific details. For example, circuits, systems, networks, processes, and other components may be shown as components in block diagram form in order not to obscure the embodiments in unnecessary detail. In other instances, well-known  
10 circuits, processes, algorithms, structures, and techniques may be shown without unnecessary detail in order to avoid obscuring the embodiments.

[036] Also, it is noted that individual embodiments may be described as a process which is depicted as a flowchart, a flow diagram, a data flow diagram, a structure diagram, or a block  
15 diagram. Although a flowchart may describe the operations as a sequential process, many of the operations can be performed in parallel or concurrently. In addition, the order of the operations may be re-arranged. A process is terminated when its operations are completed but could have additional steps not included in a figure.

20 [037] In addition, each block may indicate some of modules, segments, or codes including one or more executable instructions for executing a specific logical function(s). Further, functions mentioned in the blocks occur regardless of a sequence in some alternative embodiments. For example, two blocks that are contiguously illustrated may be simultaneously performed in fact or be performed in a reverse sequence depending on corresponding functions.

25 [038] One or more modules, units, components (including but not limited to analysis unit, identification unit, alert unit, determination unit and fetching unit) used herein may be software modules configured via hardware modules/processor, or hardware processors, the processors being a general-purpose processor, a special purpose processor, a conventional processor, a digital signal  
30 processor, a plurality of microprocessors, one or more microprocessors in association with a DSP core, a controller, a microcontroller, Application Specific Integrated Circuits, Field Programmable Gate Array circuits, any other type of integrated circuits, etc.

[039] The word “exemplary” and/or “demonstrative” is used herein to mean serving as an  
35 example, instance, or illustration. For the avoidance of doubt, the subject matter disclosed herein

is not limited by such examples. In addition, any aspect or design described herein as “exemplary” and/or “demonstrative” is not necessarily to be construed as preferred or advantageous over other aspects or designs, nor is it meant to preclude equivalent exemplary structures and techniques known to those of ordinary skill in the art. Furthermore, to the extent that the terms “includes,” “has,” “contains,” and other similar words are used in either the detailed description or the claims, such terms are intended to be inclusive—in a manner similar to the term “comprising” as an open transition word—without precluding any additional or other elements.

[040] As used herein, an “electronic device”, or “portable electronic device”, or “user device” or “communication device” or “user equipment” or “device” refers to any electrical, electronic, electromechanical and computing device. The user device is capable of receiving and/or transmitting one or parameters, performing function/s, communicating with other user devices and/or systems, and transmitting data to the other user devices and/or systems. The user equipment may have a processor, a display, a memory, a battery and an input-means such as a hard keypad and/or a soft keypad. The user equipment may be capable of operating on any radio access technology including but not limited to IP-enabled communication, Zig Bee, Bluetooth, Bluetooth Low Energy, Near Field Communication, Z-Wave, Wi-Fi, Wi-Fi direct, etc. For instance, the user equipment may include, but not limited to, a mobile phone, smartphone, virtual reality (VR) devices, augmented reality (AR) devices, laptop, a general-purpose computer, desktop, personal digital assistant, tablet computer, mainframe computer, or any other device as may be obvious to a person skilled in the art for implementation of the features of the present disclosure.

[041] Further, the user device may also comprise a “processor” or “processing unit” includes processing unit, wherein processor refers to any logic circuitry for processing instructions. The processor may be a general-purpose processor, a special purpose processor, a conventional processor, a digital signal processor, a plurality of microprocessors, one or more microprocessors in association with a DSP core, a controller, a microcontroller, Application Specific Integrated Circuits, Field Programmable Gate Array circuits, any other type of integrated circuits, etc. The processor may perform signal coding data processing, input/output processing, and/or any other functionality that enables the working of the system according to the present disclosure. More specifically, the processor is a hardware processor.

[042] As portable electronic devices and wireless technologies continue to improve and grow in popularity, the advancing wireless technologies for data transfer are also expected to evolve and replace the older generations of technologies. In the field of wireless data communications, the

dynamic advancement of various generations of cellular technology are also seen. The development, in this respect, has been incremental in the order of second generation (2G), third generation (3G), fourth generation (4G), and now fifth generation (5G), and more such generations are expected to continue in the forthcoming time.

5

[043] The 5G core is responsible for managing a wide variety of network functions within the mobile network that make it possible for users to communicate. The network functions include but not limited to mobility management, authentication, authorization, data management, policy management, and quality of service (QOS) for end users.

10

[044] Radio Access Technology (RAT) refers to the technology used by mobile devices/ user equipment (UE) to connect to a cellular network. It refers to the specific protocol and standards that govern the way devices communicate with base stations, which are responsible for providing the wireless connection. Further, each RAT has its own set of protocols and standards for communication, which define the frequency bands, modulation techniques, and other parameters used for transmitting and receiving data. Examples of RATs include GSM (Global System for Mobile Communications), CDMA (Code Division Multiple Access), UMTS (Universal Mobile Telecommunications System), LTE (Long-Term Evolution), and 5G. The choice of RAT depends on a variety of factors, including the network infrastructure, the available spectrum, and the mobile device's/device's capabilities. Mobile devices often support multiple RATs, allowing them to connect to different types of networks and provide optimal performance based on the available network resources.

15

20

[045] It is to be noted that the term “latency” as used herein associated with a telecommunication network refers to the time delay experienced when data is transmitted from one point to another within the network. Further, it is a crucial measure of network performance and directly affects the user experience in real-time applications such as voice and video communication, online gaming, and streaming services.

25

[046] As used herein, “first network function” and “target network function” refers to a specific network function that is provided within a 5G network architecture such as Access and Mobility Management Function (AMF), Session Management Function (SMF) and User plane function (UPF) etc.

30

[047] As discussed in the background section, the current known solutions for optimising latency associated with the network have several shortcomings such as the lack of adaptability to varying network conditions and traffic patterns. Existing solutions often employ static configurations and routing algorithms, which fail to dynamically respond to changing network loads and congestion levels. Consequently, this leads to suboptimal latency performance during periods of high demand. Moreover, some prior solutions focus primarily on optimizing latency for specific types of applications, such as data transfer, while neglecting the latency requirements of other critical applications like real-time communication or video streaming. This selective approach limits their overall effectiveness in providing a comprehensive solution. Additionally, the prior known solutions do not adequately address the issue of latency in geographically dispersed networks, where data packets may traverse long distances and encounter additional delays. As a result, latency issues persist in these scenarios.

[048] The present disclosure aims to overcome the above-mentioned and other existing problems in this field of technology by providing a novel solution for efficiently allocating network function service instances in a network environment to optimise latency.

[049] Further, the present solution provides a significant technical advancement by optimizing latency within the network via intelligently distributing incoming connection requests based on the number of pending requests, the solution significantly minimizes delays and enhances the overall responsiveness of the network. This optimization strategy enhances the user experience, reduces the processing time, and further improves the throughput within the network infrastructure. Moreover, the ability of present solution to intelligently allocate resources based on real-time data contributes to the efficient utilization of network function service instances, resulting in a technically superior and high-performing network environment.

[050] In 5G core Network, Service Communication Proxy (SCP) is used for selection and Routing of request from consumer NF (client) to producer NF (server). The SCP decides the appropriate producer Network Function for the requests so as to reduce overall latency. For 5G core Network, the system and method of the present disclosure may be implemented via the SCP, however the implementation may also be used for any Hypertext Transfer Protocol 2 (http2) based client-server-based communication in direct mode or an indirect communication between client & server involving proxies. Further, in the present disclosure, the one or more network functions whose response is received in less time gets share of load in comparison to the one or more network functions whose response is received late to send response to the SCP. Also, the SCP maintains a

counter for number of pending requests that are derived from one or more current active streams for each connection.

[051] The present disclosure provides a solution for optimising latency associated with a network that comprises receiving, by a transceiver unit at a Service Communication Proxy (SCP) from a first network function, a connection request. Thereafter, a set of network functions are identified by an identification unit based on the connection request. Further, an analysis unit retrieves a pending connection request counter data associated with the set of network functions. Thereafter, a connection traffic flow priority associated with each network function is determined from the set of network functions based on the pending connection request counter data via a determination unit. Further, upon determination of the connection traffic flow priority, the identification unit identifies a target network function from the set of network functions based on the connection traffic flow priority. Additionally, the connection request based on the connection traffic flow priority associated with the target network function are routed via a routing unit from the SCP to the target network function, thereby optimizing latency associated with the network. Overall, the present solution provides an efficient method for allocating network function service instances in a network, considering pending requests and optimizing latency, thereby improving the performance and responsiveness of the network.

[052] Hereinafter, exemplary embodiments of the present disclosure will be described with reference to the accompanying drawings.

[053] Referring to FIG. 1A, an exemplary block diagram representation of 5th generation core (5GC) network architecture is shown. As shown in FIG. 1A, the 5GC network architecture [101] includes a user equipment (UE) [101a], a radio access network (RAN) [101b], and a 5G Core Network and a Data Network (DN) [101p]. The 5G Core Network includes an access and mobility management function (AMF) [101c], a Session Management Function (SMF) [101d], a Service Communication Proxy (SCP) [101e], an Authentication Server Function (AUSF) [101f], a Network Slice Specific Authentication and Authorization Function (NSSAAF) [101g], a Network Slice Selection Function (NSSF) [101h], a Network Exposure Function (NEF) [101i], a Network Repository Function (NRF) [101j], a Policy Control Function (PCF) [101k], a Unified Data Management (UDM) [101l], an application function (AF) [101m], and a User Plane Function (UPF) [101n].

[054] The User Equipment (UE) [101a] interfaces with the network via the Radio Access Network (RAN) [101b]. The RAN [101b] in the 5G architecture is also called as New Radio or nG-RAN, and these terms may be interchangeably used herein. Radio Access Network (RAN) [101b] is the part of a mobile telecommunications system that connects user equipment (UE) [101a] to the core network (CN) and provides access to different types of networks (e.g., 5G, LTE). It consists of radio base stations and the radio access technologies that enable wireless communication.

[055] The Access and Mobility Management Function (AMF) [101c] manages connectivity and mobility. When a UE [101a] is active, i.e. it is interacting with the 5G network, e.g., by using data/call functionalities, the AMF [101c] knows and maintains the location of the UE [101a] within the network. The AMF [101c] is configured to maintain the tracking area or registration area of the UE [101a], in case the UE is inactive. The AMF [101c] is configured to communicate with other network functions/ elements such as the Session Management Function (SMF) [101d], etc. to ensure that the UE [101a] is allowed and is able to avail the services by the network.

[056] Particularly, the Access and Mobility Management Function (AMF) [101c] is a 5G core network function responsible for managing access and mobility aspects, such as UE registration, connection, and reachability etc. It also handles mobility management procedures like handovers and paging.

[057] The Session Management Function (SMF) [101d] is a 5G core network function responsible for managing session-related aspects, such as establishing, modifying, and releasing sessions. It coordinates with the User Plane Function (UPF) for data forwarding and handles IP address allocation and QoS enforcement.

[058] The Service Communication Proxy (SCP) [101e] is a network function in the 5G core that facilitates communication between other network functions by providing a secure and efficient messaging service. It acts as a mediator for service-based interfaces.

[059] The Authentication Server Function (AUSF) [101f] is a network function in the 5G core responsible for authenticating UEs during registration and providing security services. It generates and verifies authentication vectors and tokens.

[060] The Network Slice Specific Authentication and Authorization Function (NSSAAF) [101g] is a network function that provides authentication and authorization services specific to network slices. It ensures that UEs can access only the slices for which they are authorized.

5 [061] The Network Slice Selection Function (NSSF) [101h] is a network function responsible for selecting the appropriate network slice for a UE based on factors such as subscription, requested services, and network policies.

[062] The Network Exposure Function (NEF) [101i] is a network function that exposes  
10 capabilities and services of the 5G network to external applications, enabling integration with third-party services and applications.

[063] The Network Repository Function (NRF) [101j] is a network function that acts as a central repository for information about available network functions and services. It facilitates the  
15 discovery and dynamic registration of network functions.

[064] The Policy Control Function (PCF) [101k] is a network function responsible for policy control decisions, such as QoS, charging, and access control, based on subscriber information and network policies.

20

[065] The Unified Data Management (UDM) [101l] is a network function that centralizes the management of subscriber data, including authentication, authorization, and subscription information.

25 [066] The Application Function (AF) [101m] is a network function that represents external applications interfacing with the 5G core network to access network capabilities and services.

[067] The User Plane Function (UPF) [101n] is a network function responsible for handling user data traffic, including packet routing, forwarding, and QoS enforcement.

30

[068] The Data Network (DN) [101p] represents external networks or services that users connect to through the mobile network, such as the internet or enterprise networks.

[069] Referring to FIG.1B, an exemplary block diagram of a system [100] for optimising a  
35 latency associated with a network is shown, in accordance with the exemplary embodiments of the

present disclosure. The system [100] comprises a transceiver unit [102], an identification unit [104], an analysis unit [106] and a routing unit [108]. Also, all of the components/ units of the system [100] are assumed to be connected to each other unless otherwise indicated below. Also, in FIG. 1B only a few units are shown, however, the system [100] may comprise multiple such units or the system [100] may comprise any such numbers of said units, as required to implement the features of the present disclosure.

[070] Additionally, the identification unit [104], the analysis unit [106] and the routing unit [108] are processors. The processor may be a general-purpose processor, a special purpose processor, a conventional processor, a digital signal processor, a plurality of microprocessors, one or more microprocessors in association with a DSP (digital signal processor) core, a controller, a microcontroller, Application Specific Integrated Circuits, Field Programmable Gate Array circuits, any other type of integrated circuits, etc.

[071] Also, the transceiver unit [102] includes a transmitter having capabilities to transmit data/signals and optionally also a receiver unit having capabilities to receive data/signals.

[072] The system [100] is configured at network node (such as at a Service Communication Proxy (SCP)) of a network for optimising latency associated with the network, with the help of the interconnection between the components/units of the system [100].

[073] Further, in accordance with the present disclosure, it is to be acknowledged that the functionality described for the various the components/units can be implemented interchangeably. While specific embodiments may disclose a particular functionality of these units for clarity, it is recognized that various configurations and combinations thereof are within the scope of the disclosure. The functionality of specific units as disclosed in the disclosure should not be construed as limiting the scope of the present disclosure. Consequently, alternative arrangements and substitutions of units, provided they achieve the intended functionality described herein, are considered to be encompassed within the scope of the present disclosure.

[074] Particularly, according to the present disclosure for optimising latency associated with a network, the system [100] is configured at Service Communication Proxy (SCP) node of the network. Initially, the transceiver unit [102] of the system [100] is configured to receive at the Service Communication Proxy (SCP) from a first network function, a connection request. The first network function may refer to a specific network function that is provided within a 5G network

architecture such as Access and Mobility Management Function (AMF), Session Management Function (SMF) or User plane function (UPF), etc. The connection request is received to initiate a connection between a consumer network function (i.e. client or user) to a producer network function (i.e. server).

5

[075] Upon receiving the connection request, the identification unit [104] that is connected at least with the transceiver unit [102], identifies a set of network functions based on the connection request. For example, to identify the set of network functions, the identification unit [104] is configured to analyse the connection request by using a set of parameters like protocol type, source and destination addresses, and/or payload content to extract the set of network functions etc.

10

[076] For example, a Hypertext Transfer Protocol (HTTP) i.e., the protocol type, include one or more path headers (contexts), wherein each path represents a service belonging to a NF. For instance, in HTTP2 header include one or more Public Land Mobile Network (PLMN) that identifies circle names. Also, the NF registrations & NF profile are stored at the SCP controller shared with the SCP instances, based on received NF profile, thereafter the SCP identify where to route the received request.

15

[077] The set of network functions refers to one or more individual network functions that are working independently or in connection with other network functions within a network architecture. For example, the set of network functions may include at least one of a User Plane Function (UPF), Session Management Function (SMF), Access and Mobility Management Function (AMF), Policy Control Function (PCF) and Network Exposure Function (NEF).

20

[078] Thereafter, upon identification of the set of network functions, the analysis unit [106] that is connected at least with the identification unit [104] is configured to retrieve a pending connection request counter data associated with the set of network functions. For example, the analysis unit [106] process one or more details associated with the set of network functions to extract the pending connection request counter data.

25

30

[079] Additionally, the present disclosure encompasses that the pending connection request counter data associated with the set of network functions comprises at least a total number of pending connection requests associated with each network function from the set of network functions.

35

[080] Furthermore, the analysis unit [106] is also configured to determine a connection traffic flow priority associated with each network function from the set of network functions based on the pending connection request counter data. Additionally, the present disclosure encompasses that the connection traffic flow priority is one of a highest connection traffic flow priority and a lowest connection traffic flow priority.

[081] Further, the connection traffic flow priority refers to an urgency assigned to different network functions based on their respective pending connection request loads. This priority determines the order in which connection requests are processed or allocated network resources, with higher-priority functions receiving preferential treatment over lower-priority ones.

[082] Furthermore, the present disclosure encompasses that the analysis unit [106] is configured to determine the highest connection traffic flow priority associated with at least one network function from the set of network functions, in an event the total number of pending connection request associated with said network function is determined as lowest value associated with the total number of pending connection request.

[083] Also, the present disclosure encompasses that the analysis unit [106] is configured to determine the lowest connection traffic flow priority associated with at least one network function from the set of network functions, in an event the total number of pending connection request associated with said network function is determined as highest value associated with the total number of pending connection request.

[084] In addition to this, the identification unit [104] is further configured to identify a target network function from the set of network functions based on the connection traffic flow priority. Further, the present disclosure encompasses that the identification unit [104] is configured to identify the target network function from the set of network functions, based on at least the highest connection traffic flow priority.

[085] Further, upon identification of the target network function from the set of network functions based on the connection traffic flow priority, the routing unit [108] routes to target network function, the connection request based on the connection traffic flow priority associated with the target network function. Additionally, the routing unit [108] is connected at least to the analysis unit [106] and the identification unit [104].

[086] For example, if the system [100] of the present disclosure is implemented in a 5G network with two network instances i.e. network instance A and network instance B which have same function type. The analysis unit [106] determines the traffic flow priority based on a pending connection request counter request data. If the network instance A has large number of pending request and the network instance B has a lesser number of pending requests. The analysis unit [106] shall assign the highest priority to the network instance B, as the network instance B has fewest pending requests.

[087] Referring to FIG.2, an exemplary method flow diagram [200] for optimising latency associated with a network, in accordance with exemplary embodiments of the present disclosure is shown. The disclosure encompasses that the method [200] is performed by the system [100] at Service Communication Proxy (SCP) node. disclosure Also, as shown in FIG.2, the method [200] starts at step [202].

[088] At step [204], the method [200] as disclosed by the present disclosure comprises receiving, by a transceiver unit [102] at the Service Communication Proxy (SCP) from a first network function, a connection request. The first network function may refer to a specific network function that is provided within a 5G network architecture such as Access and Mobility Management Function (AMF), Session Management Function (SMF) or User plane function (UPF) etc. The connection request is received to initiate a connection between a consumer network function (i.e. client or user) to a producer network function (i.e. server).

[089] At step [206], the method [200] as disclosed by the present disclosure comprises identifying, by the identification unit [104] at the SCP, a set of network functions associated with the network based on the connection request. The set of network functions refers to a collection or group of individual network functions within a network architecture. For instance, in a 5G network architecture, the set of network functions may be selected from a User Plane Function (UPF), Session Management Function (SMF), Access and Mobility Management Function (AMF), Policy Control Function (PCF) and Network Exposure Function (NEF).

[090] For example, the identification unit [104] may analyse the connection request by using a set of parameters like protocol type, source and destination addresses, and payload content to extract the set of network functions.

[091] At step [208], the method [200] as disclosed by the present disclosure comprises retrieving, by analysis unit [106] at the SCP a pending connection request counter data associated with the set of network functions.

5 [092] Further, the present disclosure encompasses that the pending connection request counter data associated with the set of network functions comprises at least a total number of pending connection request associated with each network function from the set of network functions.

[093] For example, the analysis unit [106] processes the set of network functions to extract the  
10 pending connection request counter data.

[094] At step [210], the method [200] as disclosed by the present disclosure comprises determining, by the analysis unit [106] at the SCP in the network, a connection traffic flow priority associated with each network function from the set of network functions based on the pending  
15 connection request counter data.

[095] Further, the connection traffic flow priority refers to an urgency assigned to different network functions based on their respective pending connection request loads. This priority determines the order in which connection requests are processed or allocated network resources,  
20 with higher-priority functions receiving preferential treatment over lower-priority ones.

[096] The present disclosure encompasses that, the connection traffic flow priority is one of a highest connection traffic flow priority and a lowest connection traffic flow priority.

25 [097] Further the present disclosure encompasses that the highest connection traffic flow priority associated with at least one network function from the set of network functions is determined by the analysis unit [106], in an event the total number of pending connection request associated with said network function is determined as lowest value associated with the total number of pending connection request.

30

[098] Also, the lowest connection traffic flow priority associated with at least one network function from the set of network functions is determined by the analysis unit [106], in an event the total number of pending connection request associated with said network function is determined as highest value associated with the total number of pending connection request.

35

[099] At step [212], the method [200] as disclosed by the present disclosure comprises identifying, by the identification unit [104] at the SCP in the network, a target network function from the set of network functions based on the connection traffic flow priority. The present disclosure encompasses that the target network function from the set of network functions is identified by the identification unit [104] based on at least the highest connection traffic flow priority.

[100] Further, upon identification of the target network function from the set of network functions based on the connection traffic flow priority, the method proceeds to step [214], in which the method [200] as disclosed by the present disclosure comprises routing, by a routing unit [108] from the SCP in the network to the target network function, the connection request based on the connection traffic flow priority associated with the target network function.

[101] Thereafter, the method [200] terminates at step [216].

[102] Referring to FIG.3, an exemplary scenario block diagram of a system for maintaining a number of connection requests for optimising latency associated with a network is shown, in accordance with exemplary embodiments of the present disclosure. Also, in FIG. 3 only a few units are shown, however, the system [300] may comprise multiple such units or the system [300] may comprise any such numbers of said units, as required to implement the features of the present disclosure in 5G architecture. The disclosure encompasses that, the system [300] as depicted in FIG.3 works in conjunction with system [100] as depicted in FIG.1B to perform the method [200] as depicted in FIG.2.

[103] FIG. 3 depicts an exemplary scenario where a 5G NF consumer [302] (i.e. client or consumer) sends a connection request at a 5G Service Communication Proxy (SCP) [304] for requesting a service from at least one 5G NF producer [306, 308]). Further, the 5G SCP [304] maintains a counter for number of pending requests derived from current active streams for each connection. For every endpoint, the pending request is calculated by summing number of active streams of all the connection established to a particular endpoint. Further, basis the pending request of each endpoint, client decides to send the request to the endpoint having lowest pending request. When the pending request queue size is low, the traffic flow has the highest priority (as shown between 5G SCP-Proxy [304] and 5G NF Producer Instance A [306]) and when the pending request queue size is high, the traffic flow has the lowest priority (as shown between 5G SCP-Proxy [304] and 5G NF Producer Instance B [308]).

[104] Further, there are multiple endpoints at each producer instance and each endpoint is related to specific service provided by a producer. For instance, a NF producer such as 5G NF producer [306, 308] has a plurality of endpoints for each service. When a consumer requests the service of producer using endpoint, the SCP decides endpoint based on queue at the endpoint of each Producer instances (A and B).

[105] Referring to FIG.4 an exemplary scenario method flow diagram [400], for optimising latency associated with a network, in accordance with exemplary embodiments of the present disclosure is shown. The present disclosure encompasses that the method [400] is performed by the system [100]. As shown in FIG.4, the method [400] starts at step [402].

[106] At step [404], the method [400] as disclosed by the present disclosure comprises receiving, by a transceiver unit [102], a target network function service instance request (similar to the connection request) associated with a network. For example, the transceiver unit [102] receives a request from a 5G Network Function (NF) consumer for accessing a service a 5G NF producer (i.e. service provider).

[107] Next, at step [406], the method [400] as disclosed by the present disclosure comprises identifying, via an identification unit [104], a network function data associated with one or more network functions, wherein the network function data comprises at least a network function ID associated with each network function from the one or more network functions. Further, network function ID refers to a unique identifier that helps in uniquely identifying and distinguishing different network functions from each other.

[108] The present disclosure encompasses that, each network function from the one or more network functions is associated with at least a network function service instance. Further, the network service instance provides services in a network.

[109] For instance, a service named “Nudm\_UECM” provides the NF consumer of the information related to the User Equipment (UE)’s transaction information. For example, UE’s serving NF identifier, UE status and alike. Further the Nudm\_UECM (network function service instance) allows the NF consumer to register and deregister the information for serving UE in the UDM.

[110] Next, at step [408], the method [400] as disclosed by the present disclosure comprises receiving, by an analysis unit [106], a data associated with a set of connection point, wherein each connection point from the set of connection point is associated with at least one network function from the one or more network functions.

5

[111] Next, at step [410], the method [400] as disclosed by the present disclosure comprises retrieving via an analysis unit [106] a connection point data associated with each connection point from the one or more connection points. The connection point data comprises at least an active stream data associated with each connection point from the one or more connection points. Further,  
10 active stream data comprises number of pending network function service instances associated with each connection point from the one or more connection points.

[112] Next, at step [412], the method [400] as disclosed by the present disclosure comprises determining, by the analysis unit [106], a number of pending requests associated with each network  
15 function from the one or more network functions based on the active stream data associated with each connection point from the one or more connection points.

[113] The present disclosure encompasses that the number of pending requests is associated with a number of pending network function service instances associated with each network function  
20 from the one or more network functions.

[114] Next, at step [414], the method [400] as disclosed by the present disclosure comprises identifying, by the identification unit [104], a target network function from the one or more network functions based on the number of pending requests associated with each network function  
25 from the one or more network functions. For example, the identification unit [104] may utilize the network function data that was identified at step [406] for identification of the target network function.

[115] Next, at step [416], the method [400] as disclosed by the present disclosure comprises  
30 allocating, by a routing unit [108], the target network function service instance to the target network function from the one or more network functions.

[116] Thereafter, the method [400] terminates at step [418].

[117] FIG. 5 illustrates an exemplary block diagram of a computing device [1000] (also referred herein as computing system [1000]) upon which an embodiment of the present disclosure may be implemented. In an implementation, the computing device [1000] implements the method [200] for optimising latency associated with a network using the system [100]. In another  
5 implementation, the computing device [1000] itself implements the method [200] for optimising latency associated with a network using one or more units configured within the computing device [1000], wherein said one or more units are capable of implementing the features as disclosed in the present disclosure.

10 [118] The computing device [1000] may include a bus [1002] or other communication mechanism for communicating information, and a hardware processor [1004] coupled with the bus [1002] for processing information. The hardware processor [1004] may be, for example, a general-purpose microprocessor. The computing device [1000] may also include a main memory [1006], such as a random-access memory (RAM), or other dynamic storage device, coupled to the  
15 bus [1002] for storing information and instructions to be executed by the processor [1004]. The main memory [1006] also may be used for storing temporary variables or other intermediate information during execution of the instructions to be executed by the processor [1004]. Such instructions, when stored in non-transitory storage media accessible to the processor [1004], render the computing device [1000] into a special-purpose machine that is customized to perform the  
20 operations specified in the instructions. The computing device [1000] further includes a read only memory (ROM) [1008] or other static storage device coupled to the bus [1002] for storing static information and instructions for the processor [1004].

[119] A storage device [1010], such as a magnetic disk, optical disk, or solid-state drive is  
25 provided and coupled to the bus [1002] for storing information and instructions. The computing device [1000] may be coupled via the bus [1002] to a display [1012], such as a cathode ray tube (CRT), for displaying information to a computer user. An input device [1014], including alphanumeric and other keys, may be coupled to the bus [1002] for communicating information and command selections to the processor [1004]. Another type of user input device may be a cursor  
30 controller [1016], such as a mouse, a trackball, or cursor direction keys, for communicating direction information and command selections to the processor [1004], and for controlling cursor movement on the display [1012]. This input device typically has two degrees of freedom in two axes, a first axis (e.g., x) and a second axis (e.g., y), that allow the device to specify positions in a plane.

35

[120] The computing device [1000] may implement the techniques described herein using customized hard-wired logic, one or more ASICs or FPGAs, firmware and/or program logic which in combination with the computing device [1000] causes or programs the computing device [1000] to be a special-purpose machine. According to one embodiment, the techniques herein are performed by the computing device [1000] in response to the processor [1004] executing one or more sequences of one or more instructions contained in the main memory [1006]. Such instructions may be read into the main memory [1006] from another storage medium, such as the storage device [1010]. Execution of the sequences of instructions contained in the main memory [1006] causes the processor [1004] to perform the process steps described herein. In alternative embodiments, hard-wired circuitry may be used in place of or in combination with software instructions.

[121] The computing device [1000] also may include a communication interface [1018] coupled to the bus [1002]. The communication interface [1018] provides a two-way data communication coupling to a network link [1020] that is connected to a local network [1022]. For example, the communication interface [1018] may be an integrated services digital network (ISDN) card, cable modem, satellite modem, or a modem to provide a data communication connection to a corresponding type of telephone line. As another example, the communication interface [1018] may be a local area network (LAN) card to provide a data communication connection to a compatible LAN. Wireless links may also be implemented. In any such implementation, the communication interface [1018] sends and receives electrical, electromagnetic or optical signals that carry digital data streams representing various types of information.

[122] The computing device [1000] can send messages and receive data, including program code, through the network(s), the network link [1020] and the communication interface 1018. In the Internet example, a server [1030] might transmit a requested code for an application program through the Internet [1028], the ISP [1026], the host [1024], the local network [1022] and the communication interface [1018]. The received code may be executed by the processor [1004] as it is received, and/or stored in the storage device 1010, or other non-volatile storage for later execution.

[123] Also, the present disclosure may relate to a non-transitory computer readable storage medium storing instruction for optimising latency associated with a network. The instructions include an executable code which, when executed by one or more units of a system, causes a transceiver unit of the system to receive a connection request from a first network function; an

identification unit of the system to a set of network functions based on the connection request; an analysis unit of the system to retrieve a pending connection request counter data associated with the set of network functions and to determine a connection traffic flow priority associated with each network function from the set of network functions based on the pending connection request counter data; the identification unit further identify a target network function from the set of network functions based on the connection traffic flow priority and a routing unit of the system to route to the target network function, the connection request based on the connection traffic flow priority associated with the target network function.

[124] Moreover, it is pertinent to note that the method and system as encompassed by the present disclosure may be applicable in a telecommunication organization that is responsible for handling a large volume of voice calls, video calls, and data transfers. The present disclosure may assist in optimizing latency and ensuring a smooth communication experience for users. For instance, in an exemplary scenario, initially, the SCP in the 5G network receives one or more connection requests via the transceiver unit [102] of the system [100]. As users initiate one or more calls or data transfers, the user-devices send the connection requests to the network that is managed by the telecommunication organization. Further, the identification unit [104] of the system [100] analyses these connection requests and identifies various network functions required to establish and maintain communication sessions. These functions may include routing calls, managing data transfers, and handling signalling protocols. Further the analysis unit [106] of the system [100] retrieves data of pending connection requests associated with each network function. The analysis unit [106] may also monitor the number of active calls, ongoing data transfers, and queued requests waiting, to be processed. Based on the processing, the analysis unit [106] determines the traffic flow priority for each network function. For example, if there's a surge in voice call requests during peak hours, prioritizing voice call routing becomes essential to minimize call setup time and ensure clear communication. Using this prioritization, the identification unit [104] selects the target network function with the highest priority. For instance, if there's congestion in the voice call routing function, prioritizing this function ensures that incoming voice calls are promptly routed without delay. Finally, the routing unit [108] of the system [100] directs incoming connection requests to the target network function based on its priority. If voice call routing has the highest priority, the routing unit [108] ensures that new voice call requests are expedited to the routing function, optimizing latency and maintaining call quality. Hence, by implementing the method and system of the present disclosure, the telecommunication networks may assist in efficiently managing network traffic, prioritizing critical communication functions, and optimizing latency to deliver a seamless communication experience for users.

[125] As is evident from the above, the present disclosure provides a technically advanced solution in the field of network function service instance allocation. By introducing a novel approach that considers the number of pending requests associated with each network function, the solution enables more efficient resource utilization and improved network performance. The utilization of active stream data and network function IDs enhances the accuracy and effectiveness of the allocation process. This technical advancement addresses the challenges of latency optimization in network environments, providing a more streamlined and responsive system. Furthermore, by considering the handshake limit for total pending requests in each client-server communication for Http2 connections, the present disclosure aims to mitigate scenarios where the server becomes over-loaded by pending requests. The present disclosure helps prevent instances where the server reaches its maximum capacity for pending requests, thereby avoiding the need for the client to discard additional request messages.

[126] Further, the present solution provides a technical effect by providing a solution for optimization of latency within a network by dynamically allocating the target network function service instance to the appropriate network function based on the number of pending requests, therefore the solution effectively reduces delays and enhances the overall responsiveness of the network. This optimization leads to improved user experience, reduced processing times, and increased throughput within the network infrastructure. The solution's ability to intelligently allocate resources based on real-time data contributes to the efficient utilization of network function service instances, resulting in a technically superior and high-performing network environment.

[127] While considerable emphasis has been placed herein on the disclosed embodiments, it will be appreciated that many embodiments can be made and that many changes can be made to the embodiments without departing from the principles of the present disclosure. These and other changes in the embodiments of the present disclosure will be apparent to those skilled in the art, whereby it is to be understood that the foregoing descriptive matter to be implemented is illustrative and non-limiting.

**I/We claim:**

1. A method [200] for optimising latency associated with a network, the method comprising:
  - receiving, by a transceiver unit [102] at a Service Communication Proxy (SCP) from a first network function, a connection request;
  - 5 - identifying, by an identification unit [104] at the SCP, a set of network functions based on the connection request;
  - retrieving, by an analysis unit [106] at the SCP, a pending connection request counter data associated with the set of network functions;
  - determining, by the analysis unit [106] at the SCP, a connection traffic flow priority associated with each network function from the set of network functions based on the pending connection request counter data;
  - 10 - identifying, by the identification unit [104] at the SCP, a target network function from the set of network functions based on the connection traffic flow priority; and
  - routing, by a routing unit [108] from the SCP to the target network function, the connection request based on the connection traffic flow priority associated with the target network function.
  - 15
2. The method [200] as claimed in claim 1, wherein the pending connection request counter data associated with the set of network functions comprises at least a total number of pending connection requests associated with each network function from the set of network functions.
- 20
3. The method [200] as claimed in claim 1, wherein the connection traffic flow priority is one of a highest connection traffic flow priority and a lowest connection traffic flow priority.
- 25
4. The method [200] as claimed in claim 3, wherein the highest connection traffic flow priority associated with at least one network function from the set of network functions is determined by the analysis unit [106], in an event the total number of pending connection requests associated with said network function is determined as lowest value associated with the total number of pending connection requests.
- 30
5. The method [200] as claimed in claim 3, wherein the lowest connection traffic flow priority associated with at least one network function from the set of network functions is determined by the analysis unit [106], in an event the total number of pending connection requests associated with said network function is determined as highest value associated with the total number of pending connection requests.
- 35

6. The method [200] as claimed in claim 3, wherein the target network function from the set of network functions is identified by the identification unit [104] based on at least the highest connection traffic flow priority.
- 5
7. A system [100] for optimising a latency associated with a network, the system [100] is configured at a Service Communication Proxy (SCP), the system further comprising:
- a transceiver unit [102], wherein the transceiver unit [102] is configured to receive from a first network function, a connection request;
  - 10 - an identification unit [104] connected at least to the transceiver unit [102], wherein the identification unit [104] is configured to identify a set of network functions based on the connection request;
  - an analysis unit [106], connected at least to the identification unit [104], wherein the analysis unit [106] is configured to:
    - 15       o retrieve a pending connection request counter data associated with the set of network functions, and
    - o determine a connection traffic flow priority associated with each network function from the set of network functions based on the pending connection request counter data,
  - 20       wherein the identification unit [104] is further configured to identify a target network function from the set of network functions based on the connection traffic flow priority; and
  - a routing unit [108], connected at least to the analysis unit [106] and the identification unit [104], the routing unit [108] is configured to route to the target network function, the
  - 25       connection request based on the connection traffic flow priority associated with the target network function.
8. The system [100] as claimed in claim 7, wherein the pending connection request counter data associated with the set of network functions comprises at least a total number of pending
- 30       connection requests associated with each network function from the set of network functions.
9. The system [100] as claimed in claim 7, wherein the connection traffic flow priority is one of a highest connection traffic flow priority and a lowest connection traffic flow priority.

10. The system [100] as claimed in claim 9, wherein the analysis unit [106] is configured to determine the highest connection traffic flow priority associated with at least one network function from the set of network functions, in an event the total number of pending connection requests associated with said network function is determined as lowest value associated with the total number of pending connection requests.
11. The system [100] as claimed in claim 9, wherein the analysis unit [106] is configured to determine the lowest connection traffic flow priority associated with at least one network function from the set of network functions, in an event the total number of pending connection requests associated with said network function is determined as highest value associated with the total number of pending connection requests.
12. The system [100] as claimed in claim 9, wherein the identification unit [104] is configured to identify the target network function from the set of network functions, based on at least the highest connection traffic flow priority.
13. A non-transitory computer readable storage medium storing instructions for optimising latency associated with a network, the instructions including executable code, the executable code when executed, may cause:
- a transceiver unit [102] to receive from a first network function, a connection request,
  - an identification unit [104] to identify, a set of network functions based on the connection request,
  - an analysis unit [106] to retrieve, a pending connection request counter data associated with the set of network functions,
  - the analysis unit [106] to determine, a connection traffic flow priority associated with each network function from the set of network functions based on the pending connection request counter data,
  - the identification unit [104] to identify, a target network function from the set of network functions based on the connection traffic flow priority, and
  - a routing unit [108] to route, the connection request based on the connection traffic flow priority associated with the target network function.

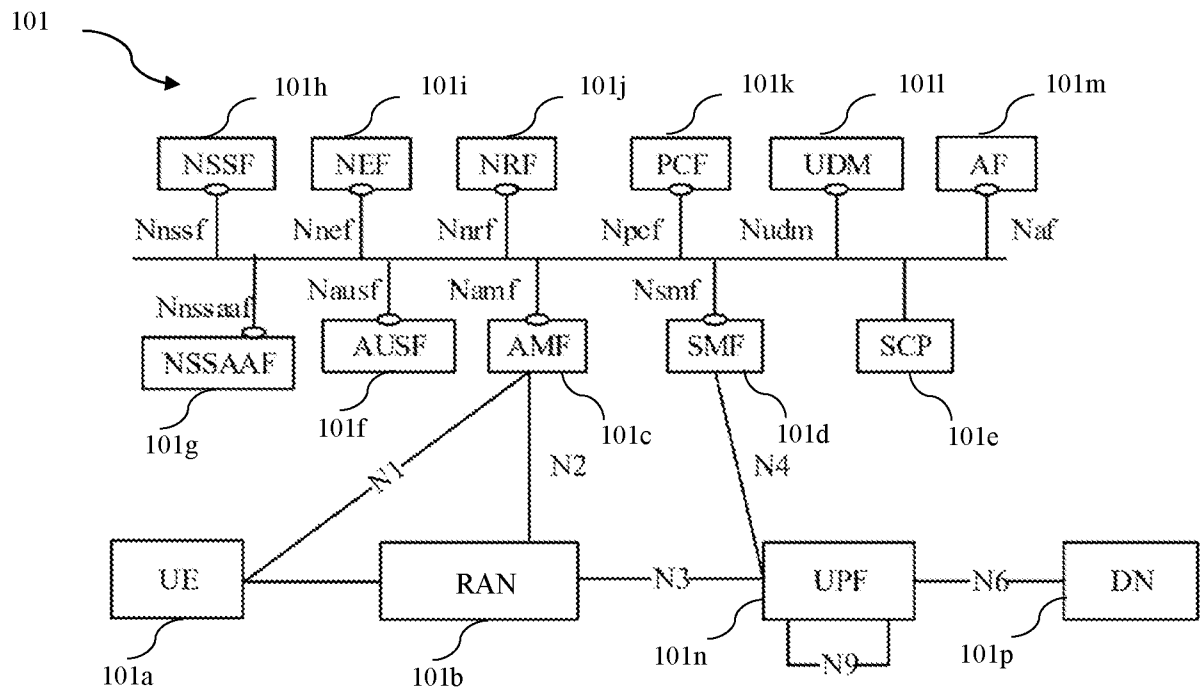


FIG. 1A

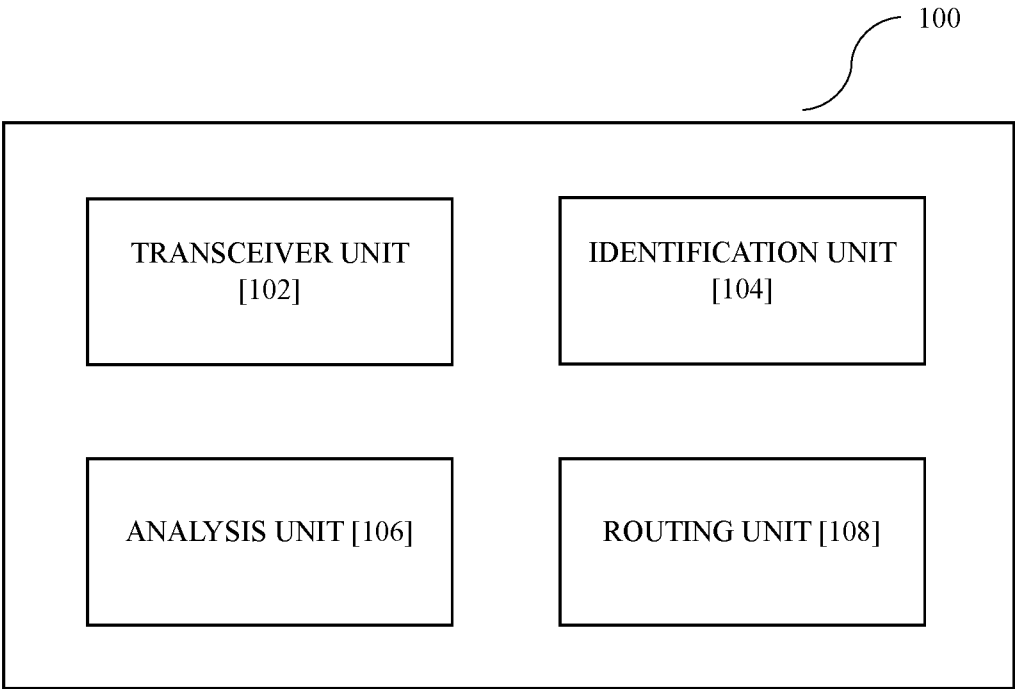


FIG. 1B

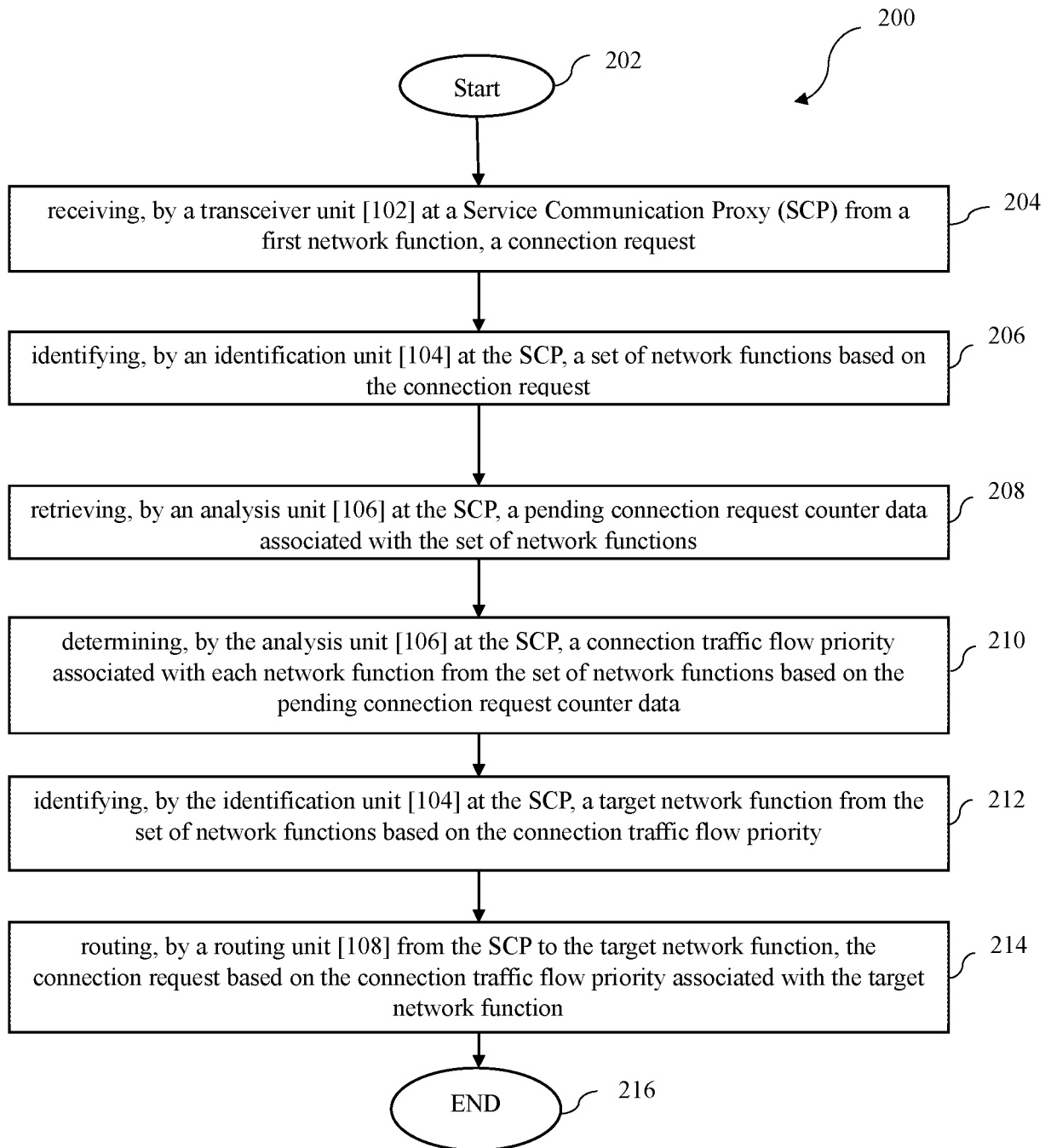


FIG. 2

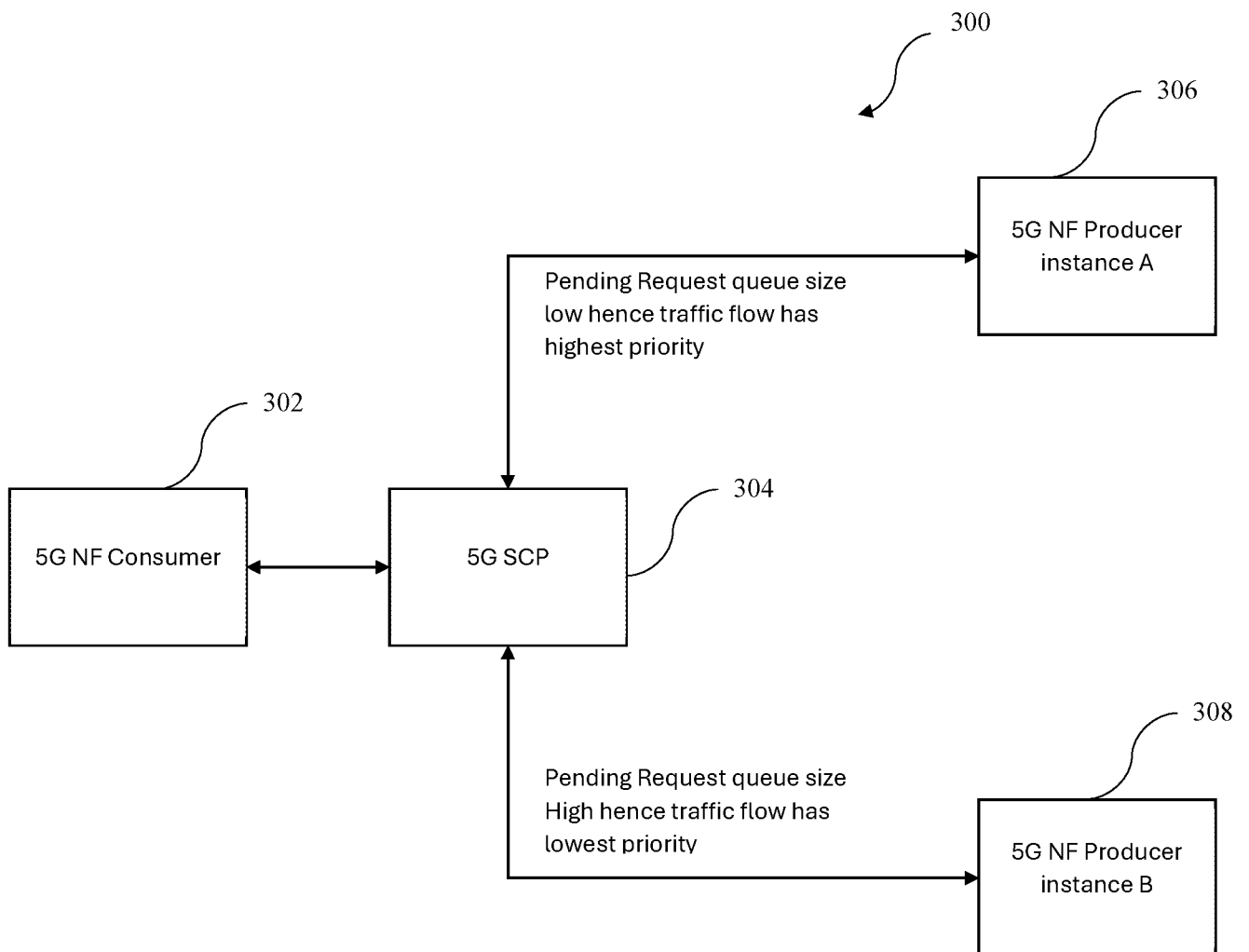


FIG. 3

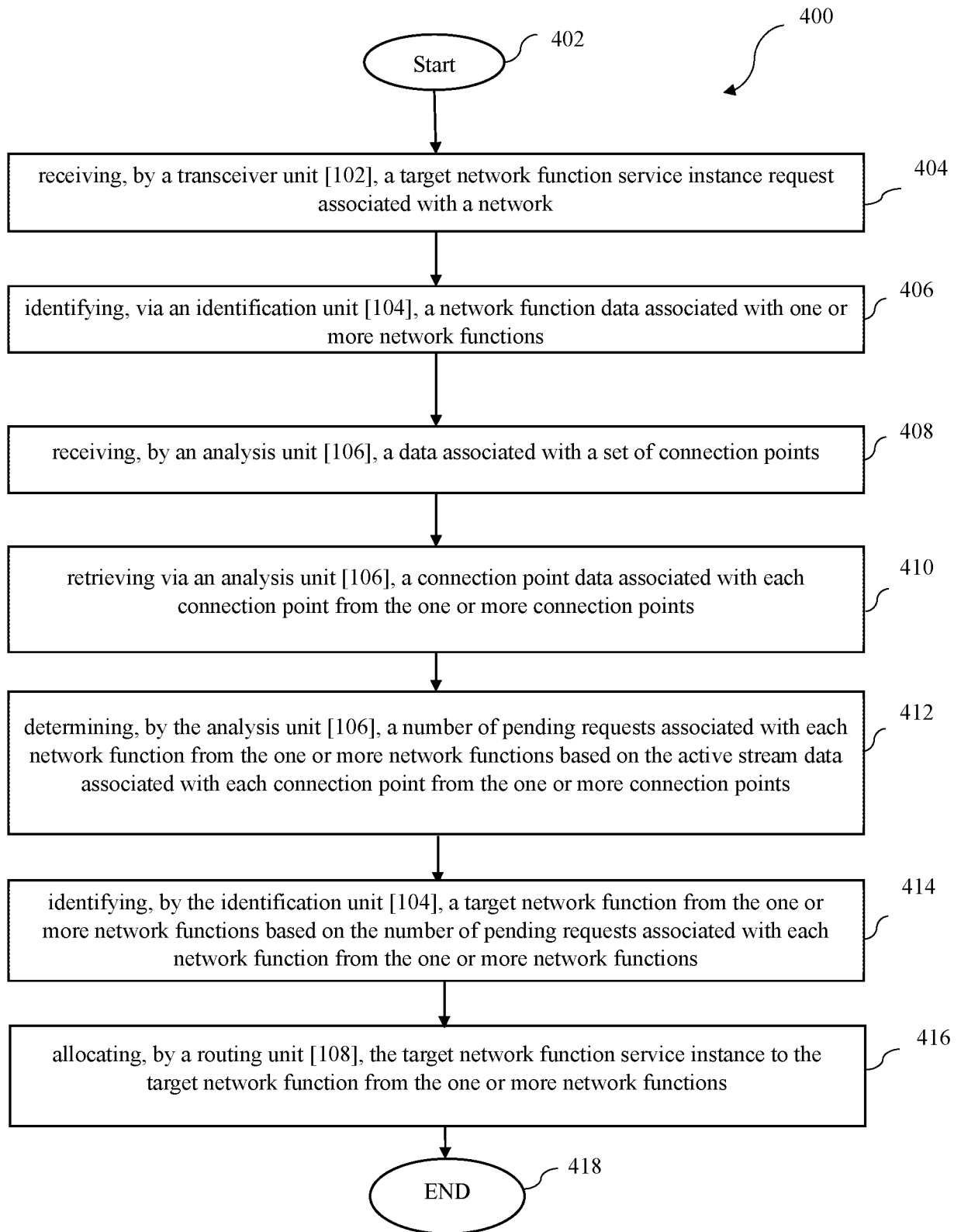


FIG. 4

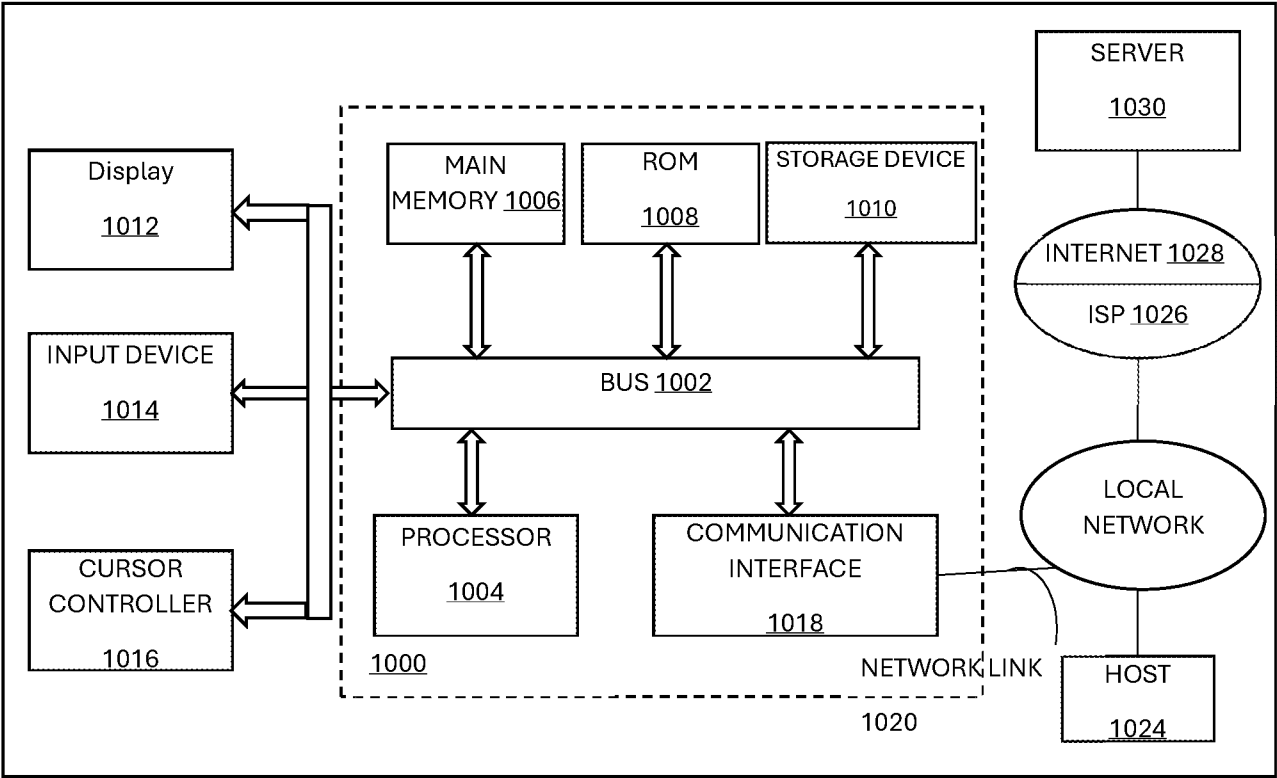


FIG. 5

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/IN2024/050819

## A. CLASSIFICATION OF SUBJECT MATTER

H04W24/02, H04L43/0852, H04L45/121, H04L47/283 Version=2024.01

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 &amp; H04L

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

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

PatSeer, IPO Internal Database

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                            | Relevant to claim No. |
|-----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| Y         | CN112968936 A; (UNICLOUD TECH. CO. LTD.); 15 JUNE 2021 (15/06/2021) (English Translation Copy from Google Patents) Page 2, Lines 008-013 of disclosure of invention and Fig. 1 & related text | 1-13                  |
| Y         | CN111770545 A; (HUAWEI TECHNOLOGIES CO. LTD.); 13 OCTOBER 2020 (13/10/2020) (English Translation Copy from Google Patents) Page 15 lines 010-015 & 048-056                                    | 1, 3-7 & 9-13         |
| Y         | US2022200899 A1; (ARRIS ENTERPRISES LLC); 23 JUNE 2022 (23/06/2022) Paragraphs [0003]-[0017], [0051]-[0056], Figs. 3A-4                                                                       | 1, 3-7 & 9-13         |
| A         | EP2638683 B1; (USABLENET INC.); 15 MARCH 2017 (15/03/2017) Paragraphs [0006]-[0008] & [0062]-[0064]                                                                                           | 1-13                  |
| A         | WO2016144414 A1; (APPLE INC.); 15 SEPTEMBER 2016 (15/09/2016) Paragraphs [0006]-[0007], [0057]-[0059] & Fig. 3 and related text                                                               | 1, 7 & 13             |



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

"D" document cited by the applicant in the international application

"E" earlier application or patent 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

"I" 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

"&amp;" document member of the same patent family

Date of the actual completion of the international search

08-10-2024

Date of mailing of the international search report

08-10-2024

Name and mailing address of the ISA/

Indian Patent Office

Plot No.32, Sector 14, Dwarka, New Delhi-110075

Facsimile No.

Authorized officer

Jugal Kishor

Telephone No. +91-1125300200

INTERNATIONAL SEARCH REPORT  
Information on patent family members

International application No.  
PCT/IN2024/050819

| Citation         | Pub.Date   | Family             | Pub.Date   |
|------------------|------------|--------------------|------------|
| CN 111770545 A   | 13-10-2020 | CN 111770545 B     | 26-08-2022 |
|                  |            | WO 2020199896 A1   | 08-10-2020 |
| US 2022200899 A1 | 23-06-2022 | US 11689445 B2     | 27-06-2023 |
|                  |            | WO 2022132374 A1   | 23-06-2022 |
| EP 2638683 B1    | 15-03-2017 | US 8868638 B2      | 21-10-2014 |
|                  |            | CA 2816338 A1      | 18-05-2012 |
|                  |            | WO 2012064857 A2   | 18-05-2012 |
| WO 2016144414 A1 | 15-09-2016 | US 10375620 B2     | 06-08-2019 |
|                  |            | DE 112015006261 T5 | 14-12-2017 |
|                  |            | CN 107409269 B     | 01-06-2021 |