(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 July 2020 (30.07.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/152389 A1

(51) International Patent Classification:
H04W 24/02 (2009.01) *G06N 20/00* (2019.01)
H04L 12/26 (2006.01) *G06N 3/02* (2006.01)
G05B 13/04 (2006.01)

(21) International Application Number:
PCT/FI2019/050049

(22) International Filing Date:
22 January 2019 (22.01.2019)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: **NOKIA SOLUTIONS AND NETWORKS OY** [FI/FI]; Karakaari 7, 02610 Espoo (FI).

(72) Inventors: **UUSITALO, Mikko**; c/o Nokia Solutions and Networks Oy, Karaportti 3, 02610 Espoo (FI). **HONKALA, Mikko**; c/o Nokia Solutions and Networks Oy, Karaportti 3, 02610 Espoo (FI). **KÄRKKÄINEN, Leo**; c/o Nokia Solutions and Networks Oy, Karaportti 3, 02610 Espoo (FI).

(74) Agent: **KOLSTER OY AB**; (Salmisaarenaukio 1), P.O.Box 204, 00181 Helsinki (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, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

WO 2020/152389 A1

(54) Title: MACHINE LEARNING FOR A COMMUNICATION NETWORK

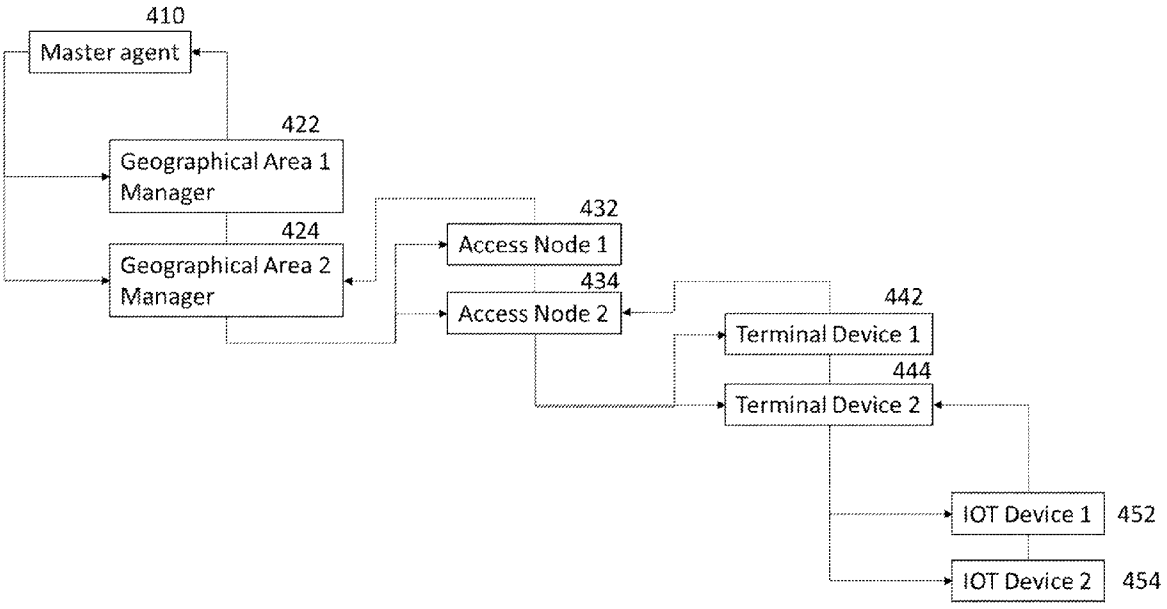

Figure 4

(57) Abstract: Disclosed is a method comprising executing a master learning algorithm that uses a master reward definition as an input; updating a sub-reward definition based on one or more variables that have been determined by executing the master learning algorithm; sharing the sub-reward definition to a plurality of sub-algorithms executed by their respective network elements; receiving, from the sub-algorithms, information regarding their respective sub-policies that were obtained based on, at least partly, the shared sub-reward definition; determining which sub-policies are associated with a first category; determining which of the sub-policies belonging to the first category has the best performance in comparison to a first criteria; and updating the sub-policies associated with the first category according to the sub-policy that has the best performance.

[Continued on next page]

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, 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))*
— *in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE*

Machine Learning for a Communication Network

Field

The following example embodiments relate to optimization of communication networks.

Background

As resources are limited, it is desirable to optimize the usage of networks, wired or wireless. To optimize the usage effectively, artificial intelligence and machine learning may be utilized to decide when and how to make trade-offs needed to achieve the optimal usage of resources. Different algorithms may be utilized at different parts of the networks.

Brief description of the invention

According to an aspect there is provided a system comprising means for executing a master learning algorithm that uses a master reward definition as an input, means for updating a sub-reward definition based on one or more variables that have been determined by executing the master learning algorithm, means for sharing the sub-reward definition to a plurality of sub-algorithms executed by their respective network elements, means for receiving, from the sub-algorithms, information regarding their respective sub-policies that were obtained based on, at least partly, the shared sub-reward definition, means for determining which sub-policies are associated with a first category, means for determining which of the sub-policies belonging to the first category has the best performance in comparison to a first criteria, and means for updating the sub-policies associated with the first category according to the sub-policy that has the best performance.

According to another aspect there is provided an apparatus comprising means for executing a master learning algorithm that uses a master reward definition as an input, means for updating a sub-reward definition based on one or more variables that have been determined by executing the master learning algorithm, means for sharing the sub-reward definition to a plurality of sub-algorithms executed by their respective network elements, means for receiving, from the sub-algorithms,

information regarding their respective sub-policies that were obtained based on, at least partly, the shared sub-reward definition, means for determining which sub-policies are associated with a first category, means for determining which of the sub-policies belonging to the first category has the best performance in comparison to a first criteria, and means for updating the sub-policies associated with the first category according to the sub-policy that has the best performance.

According to another aspect there is provided a method comprising executing a master learning algorithm that uses a master reward definition as an input, updating a sub-reward definition based on one or more variables that have been determined by executing the master learning algorithm, sharing the sub-reward definition to a plurality of sub-algorithms executed by their respective network elements, receiving, from the sub-algorithms, information regarding their respective sub-policies that were obtained based on, at least partly, the shared sub-reward definition, determining which sub-policies are associated with a first category, determining which of the sub-policies belonging to the first category has the best performance in comparison to a first criteria, and updating the sub-policies associated with the first category according to the sub-policy that has the best performance.

According to another aspect there is provided a computer program product readable by a computer and, when executed by the computer, configured to cause the computer to execute a computer process comprising executing a master learning algorithm that uses a master reward definition as an input, updating a sub-reward definition based on one or more variables that have been determined by executing the master learning algorithm, sharing the sub-reward definition to a plurality of sub-algorithms executed by their respective network elements, receiving, from the sub-algorithms, information regarding their respective sub-policies that were obtained based on, at least partly, the shared sub-reward definition, determining which sub-policies are associated with a first category, determining which of the sub-policies belonging to the first category has the best performance in comparison to a first criteria, and updating the sub-policies associated with the first category according to the sub-policy that has the best performance.

According to another aspect there is provided an apparatus comprising at least one processor, and at least one memory including a computer program code, wherein the at least one memory and the computer program code are configured, with the at least one processor, to cause the apparatus to execute a master learning algorithm that uses a master reward definition as an input update a sub-reward definition based on one or more variables that have been determined by executing the master learning algorithm share the sub-reward definition to a plurality of sub-algorithms executed by their respective network elements receive, from the sub-algorithms, information regarding their respective sub-policies that were obtained based on, at least partly, the shared sub-reward definition determine which sub-policies are associated with a first category determine which of the sub-policies belonging to the first category has the best performance in comparison to a first criteria and update the sub-policies associated with the first category according to the sub-policy that has the best performance.

According to another aspect there is provided a computer program product comprising computer-readable medium bearing computer program code embodied therein for use with a computer, the computer program code comprising code for executing a master learning algorithm that uses a master reward definition as an input, updating a sub-reward definition based on one or more variables that have been determined by executing the master learning algorithm, sharing the sub-reward definition to a plurality of sub-algorithms executed by their respective network elements, receiving, from the sub-algorithms, information regarding their respective sub-policies that were obtained based on, at least partly, the shared sub-reward definition, determining which sub-policies are associated with a first category, determining which of the sub-policies belonging to the first category has the best performance in comparison to a first criteria, and updating the sub-policies associated with the first category according to the sub-policy that has the best performance.

List of drawings

In the following, the invention will be described in greater detail with reference to the embodiments and the accompanying drawings, in which

Figure 1 illustrates an example embodiment of a communication system.

Figure 2 is a flow chart of an example embodiment.

Figures 3 and 4 are examples of hierarchical reinforcement learning.

Figure 5 is a block diagram of an example embodiment of an apparatus.

Description of embodiments

The following embodiments are exemplifying. Although the specification may refer to "an", "one", or "some" embodiment(s) in several locations of the text, this does not necessarily mean that each reference is made to the same embodiment(s), or that a particular feature only applies to a single embodiment. Single features of different embodiments may also be combined to provide other embodiments.

Embodiments described herein may be implemented in a communication system, such as in at least one of the following: Global System for Mobile Communications (GSM) or any other second generation cellular communication system, Universal Mobile Telecommunication System (UMTS, 3G) based on basic wideband-code division multiple access (W-CDMA), high-speed packet access (HSPA), Long Term Evolution (LTE), LTE-Advanced, a system based on IEEE 802.11 specifications, a system based on IEEE 802.15 specifications, and/or a fifth generation (5G) mobile or cellular communication system. The embodiments are not, however, restricted to the system given as an example but a person skilled in the art may apply the solution to other communication systems provided with necessary properties.

Figure 1 depicts examples of simplified system architectures only showing some elements and functional entities, all being logical units, whose implementation may differ from what is shown. The connections shown in Figure 1 are logical connections; the actual physical connections may be different. It is apparent to a person skilled in the art that the system typically comprises also other functions and structures than those shown in Figure 1. The example of Figure 1 shows a part of an exemplifying radio access network.

Figure 1 shows terminal devices 100 and 102 configured to be in a wireless connection on one or more communication channels in a cell with an access

node (such as (e/g)NodeB) 104 providing the cell. The physical link from a terminal device to a (e/g)NodeB is called uplink or reverse link and the physical link from the (e/g)NodeB to the terminal device is called downlink or forward link. It should be appreciated that (e/g)NodeBs or their functionalities may be implemented by using any node, host, server or access point etc. entity suitable for such a usage.

A communication system typically comprises more than one (e/g)NodeB in which case the (e/g)NodeBs may also be configured to communicate with one another over links, wired or wireless, designed for the purpose. These links may be used for signalling purposes. The (e/g)NodeB is a computing device configured to control the radio resources of communication system it is coupled to. The NodeB may also be referred to as a base station, an access point or any other type of interfacing device including a relay station capable of operating in a wireless environment. The (e/g)NodeB includes or is coupled to transceivers. From the transceivers of the (e/g)NodeB, a connection is provided to an antenna unit that establishes bi-directional radio links to user devices. The antenna unit may comprise a plurality of antennas or antenna elements. The (e/g)NodeB is further connected to core network 110 (CN or next generation core NGC). Depending on the system, the counterpart on the CN side can be a serving gateway (S-GW, routing and forwarding user data packets), packet data network gateway (P-GW), for providing connectivity of terminal devices (UEs) to external packet data networks, or mobile management entity (MME), etc.

The terminal device (also called UE, user equipment, user terminal, user device, etc.) illustrates one type of an apparatus to which resources on the air interface are allocated and assigned, and thus any feature described herein with a terminal device may be implemented with a corresponding apparatus, such as a relay node. An example of such a relay node is a layer 3 relay (self-backhauling relay) towards the base station. Another example of such a relay node is a layer 2 relay. Such a relay node may contain a terminal device part and a Distributed Unit (DU) part. A CU (centralized unit) may coordinate the DU operation via F1AP -interface for example.

The terminal device typically refers to a portable computing device that

includes wireless mobile communication devices operating with or without a subscriber identification module (SIM), or an embedded SIM, eSIM, including, but not limited to, the following types of devices: a mobile station (mobile phone), smartphone, personal digital assistant (PDA), handset, device using a wireless modem (alarm or measurement device, etc.), laptop and/or touch screen computer, tablet, game console, notebook, and multimedia device. It should be appreciated that a user device may also be a nearly exclusive uplink only device, of which an example is a camera or video camera loading images or video clips to a network. A terminal device may also be a device having capability to operate in Internet of Things (IoT) network which is a scenario in which objects are provided with the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction. The terminal device may also utilise cloud. In some applications, a terminal device may comprise a small portable device with radio parts (such as a watch, earphones or eyeglasses) and the computation is carried out in the cloud. The terminal device (or in some embodiments a layer 3 relay node) is configured to perform one or more of user equipment functionalities.

Various techniques described herein may also be applied to a cyber-physical system (CPS) (a system of collaborating computational elements controlling physical entities). CPS may enable the implementation and exploitation of massive amounts of interconnected ICT devices (sensors, actuators, processors microcontrollers, etc.) embedded in physical objects at different locations. Mobile cyber physical systems, in which the physical system in question has inherent mobility, are a subcategory of cyber-physical systems. Examples of mobile physical systems include mobile robotics and electronics transported by humans or animals.

Additionally, although the apparatuses have been depicted as single entities, different units, processors and/or memory units (not all shown in Figure 1) may be implemented.

5G enables using multiple input – multiple output (MIMO) antennas, many more base stations or nodes than the LTE (a so-called small cell concept), including macro sites operating in co-operation with smaller stations and employing a variety of radio technologies depending on service needs, use cases and/or

spectrum available. 5G mobile communications supports a wide range of use cases and related applications including video streaming, augmented reality, different ways of data sharing and various forms of machine type applications such as (massive) machine-type communications (mMTC), including vehicular safety, different sensors and real-time control. 5G is expected to have multiple radio interfaces, namely below 6GHz, cmWave and mmWave, and also being integratable with existing legacy radio access technologies, such as the LTE. Integration with the LTE may be implemented, at least in the early phase, as a system, where macro coverage is provided by the LTE and 5G radio interface access comes from small cells by aggregation to the LTE. In other words, 5G is planned to support both inter-RAT operability (such as LTE-5G) and inter-RI operability (inter-radio interface operability, such as below 6GHz – cmWave, below 6GHz – cmWave – mmWave). One of the concepts considered to be used in 5G networks is network slicing in which multiple independent and dedicated virtual sub-networks (network instances) may be created within the same infrastructure to run services that have different requirements on latency, reliability, throughput and mobility.

The current architecture in LTE networks is fully distributed in the radio and fully centralized in the core network. The low latency applications and services in 5G require to bring the content close to the radio which leads to local break out and multi-access edge computing (MEC). 5G enables analytics and knowledge generation to occur at the source of the data. This approach requires leveraging resources that may not be continuously connected to a network such as laptops, smartphones, tablets and sensors. MEC provides a distributed computing environment for application and service hosting. It also has the ability to store and process content in close proximity to cellular subscribers for faster response time. Edge computing covers a wide range of technologies such as wireless sensor networks, mobile data acquisition, mobile signature analysis, cooperative distributed peer-to-peer ad hoc networking and processing also classifiable as local cloud/fog computing and grid/mesh computing, dew computing, mobile edge computing, cloudlet, distributed data storage and retrieval, autonomic self-healing networks, remote cloud services, augmented and virtual reality, data caching, Internet of Things

(massive connectivity and/or latency critical), critical communications (autonomous vehicles, traffic safety, real-time analytics, time-critical control, healthcare applications).

The communication system is also able to communicate with other networks, such as a public switched telephone network or the Internet 112, or utilise services provided by them. The communication network may also be able to support the usage of cloud services, for example at least part of core network operations may be carried out as a cloud service (this is depicted in Figure 1 by "cloud" 114). The communication system may also comprise a central control entity, or a like, providing facilities for networks of different operators to cooperate for example in spectrum sharing.

Edge cloud may be brought into radio access network (RAN) by utilizing network function virtualization (NFV) and software defined networking (SDN). Using edge cloud may mean access node operations to be carried out, at least partly, in a server, host or node operationally coupled to a remote radio head or base station comprising radio parts. It is also possible that node operations will be distributed among a plurality of servers, nodes or hosts. Application of cloudRAN architecture enables RAN real time functions being carried out at the RAN side (in a distributed unit, DU 104) and non-real time functions being carried out in a centralized manner (in a centralized unit, CU 108).

It should also be understood that the distribution of labour between core network operations and base station operations may differ from that of the LTE or even be non-existent. Some other technology probably to be used are Big Data and all-IP, which may change the way networks are being constructed and managed. 5G (or new radio, NR) networks are being designed to support multiple hierarchies, where MEC servers can be placed between the core and the base station or nodeB (gNB). It should be appreciated that MEC can be applied in 4G networks as well.

5G may also utilize satellite communication to enhance or complement the coverage of 5G service, for example by providing backhauling. Possible use cases are providing service continuity for machine-to-machine (M2M) or Internet of Things (IoT) devices or for passengers on board of vehicles, or ensuring service

availability for critical communications, and future railway/maritime/aeronautical communications. Satellite communication may utilise geostationary earth orbit (GEO) satellite systems, but also low earth orbit (LEO) satellite systems, in particular mega-constellations (systems in which hundreds of (nano)satellites are deployed). Each satellite 106 in the mega-constellation may cover several satellite-enabled network entities that create on-ground cells. The on-ground cells may be created through an on-ground relay node 104 or by a gNB located on-ground or in a satellite.

It is to be noted that the depicted system is only an example of a part of a radio access system and in practice, the system may comprise a plurality of (e/g)NodeBs, the terminal device may have an access to a plurality of radio cells and the system may comprise also other apparatuses, such as physical layer relay nodes or other network elements, etc. At least one of the (e/g)NodeBs may be a Home(e/g)nodeB. Additionally, in a geographical area of a radio communication system a plurality of different kinds of radio cells as well as a plurality of radio cells may be provided. Radio cells may be macro cells (or umbrella cells) which are large cells, usually having a diameter of up to tens of kilometers, or smaller cells such as micro-, femto- or picocells. The (e/g)NodeBs of Figure 1 may provide any kind of these cells. A cellular radio system may be implemented as a multilayer network including several kinds of cells. Typically, in multilayer networks, one access node provides one kind of a cell or cells, and thus a plurality of (e/g)NodeBs are required to provide such a network structure.

For fulfilling the need for improving the deployment and performance of communication systems, the concept of "plug-and-play" (e/g)NodeBs has been introduced. Typically, a network which is able to use "plug-and-play" (e/g)Node Bs, includes, in addition to Home (e/g)NodeBs (H(e/g)nodeBs), a home node B gateway, or HNB-GW (not shown in Figure 1). A HNB Gateway (HNB-GW), which is typically installed within an operator's network may aggregate traffic from a large number of HNBs back to a core network.

In a communication network, which may be wireless or wired, artificial intelligence and/or machine learning algorithms may be utilized to optimize the functions of the network. For example, for a smartphone to function properly on a

cellular communication network, a trade-off between the amount of spectrum used for uplink control channel and the amount of spectrum used for data transmission is needed. The uplink control channel may be used to provide feedback regarding the channel quality for example to the access node. In some circumstances the need for this feedback may be larger thereby justifying a larger proportion of the available spectrum to be used for the control channel while in some other situations the amount of spectrum needed for control channel may be less thereby leaving more spectrum for the data transmission. It may be beneficial to be able to automatically adapt to the needed trade-off and further, to predict how the functions of the network should be adapted in each situation. Machine-learning algorithms may be used to facilitate such automatic adaptation and prediction.

An example of automation technology designed to make the planning, configuration, management, optimization and healing of a mobile communication network is a self-organizing network, SON. SON functionalities may be divided into three categories: self configuration, self optimization and self healing. Self configuration enables an access node to become essentially a plug and play unit by needing as little manual configuration as possible in a configuration process. Self optimization enables to optimize the operational characteristics such that they optimally meet the needs of the overall network. Self healing enables a network to change its characteristics such that it may temporarily mask a fault developed within the network thereby hiding effects of the fault from end-users. These functionalities may be implemented using software algorithms that include machine learning algorithms.

Machine learning algorithms may be categorized as supervised or unsupervised. While supervised algorithms require that each data sample has a corresponding label or a corresponding output of interest, unsupervised algorithms are able to develop the capability of producing the desired outcome without the labels. This can be achieved for example by using an iterative approach to review data to recognize one or more patterns and thereby providing outcomes. A further category is reinforcement learning, in which data are not provided beforehand, but an computing agent interacts with an environment to collect the data and labels.

Supervised and unsupervised machine learning may be divided into several process steps. The data first need to be gathered and after that it should be prepared such that it is useful. The quality and quantity of the data affect how well the machine-learning may function. Data preparation may involve storing the data in a certain way, modifying the data to, for example, remove unnecessary or extra aspects. Once the data is ready, a machine learning method is chosen. Various methods have been created by researches and data scientist that may be used for machine-learning purposes. The methods may have been developed with a certain purpose in mind, like image data, text or music, or the methods may be generic methods. Once the method has been chosen, the method is utilized in training a machine learning model. In training, the data gathered and prepared are used to incrementally improve the machine learning model in recognizing patterns and thereby providing correct outcomes. In the next step validation may be used to test the trained machine learning model to see how well the machine learning model generalizes, that is, can classify data which it has not been introduced to before. This provides an indication regarding how well the machine learning model may function in usage of classifying new data. After evaluation it is possible to further improve the training of the machine learning model. This may be done by tuning the parameters that are used for training the machine learning model. Prediction is the final phase in which the machine learning model that has been trained is used to provide predictions regarding new data it is given as an input.

Reinforcement learning algorithm, which is an example of a learning algorithm that may be used in machine learning, refers to a goal-oriented algorithm that learns how to attain a complex objective or maximize along a particular dimension over many steps. Such a reinforcement learning algorithm has access to a stateful environment in which input depends on previous actions taken, and it gets its training data from the stateful environment based on the actions it performs, and the state and rewards it observes. Reinforcement learning algorithms may also be able to improve over time, also when used in prediction. Additionally, or alternatively, reinforcement algorithms may also adapt to changes in the stateful environment. A reinforcement learning algorithm may start from a blank slate, and

under the right conditions achieve a desired performance. In a reinforcement learning algorithm, there may be an agent that takes actions. The agent may be the algorithm itself. The action taken by the agent may be one of many various alternative actions. The agent is in a state which may be understood as a concrete and immediate situation such as a specific place and moment. A reward may then be used as feedback by which the agent may measure the success of its action. The reward may be immediate or delayed. The reward is a measure against which the results from the actions taken by the agent are measured thereby making it a tool for evaluating the performance of the agent. A policy may be understood as the strategy the agent employs to determine its next action based on the current state it is in. The policy thereby associates states to actions that are predicted to have the best performance against the reward. An environment may be understood as a function that transforms an action taken in a current state into a next state and reward. Agents on the other hand are functions that transform a new state and reward into a next action. A function of an agent may be known but a function of the environment is not known in some examples. In addition, the environment may contain known or unknown stochastic elements.

One example of reinforcement learning algorithms is deep Q-learning. It is based on estimated Q-values using neural networks. A Q-value may be understood as a maximum achievable future reward that is determined by an action and a state. The Q-value thereby maps state-action pairs to rewards. A neural network comprises neurons, that are computational units that are connected to each other and may form one or more computational layers. A neural network thereby provides a framework that can be utilized by machine-learning algorithms. The structure of the neural network, the number of neurons and layers, may be determined by a developer for example. Each connection between two neurons may be assigned a weight factor that may describe the importance of the connection between those neurons.

A reinforcement learning algorithm thereby evaluates actions based on the estimated results they produce. It may be understood as a goal-oriented algorithm that is to learn actions that achieve its goal or maximizes its objective function.

In an example in which a reinforcement learning agent does not know the environment beforehand, it may comprise an exploration in its algorithm. The exploration can be guided by the learned model (e.g., Bolzmann exploration) or it can be random (E.g., epsilon-greedy). The ratio of exploration vs. exploitation is pre-determined, but it can be dynamically adjusted by the operator of the agent as well. The exploration enables the agent to learn new behaviors through (controlled) trial and error. It is to be noted that in live networks, there may be some predefined or dynamic boundaries on what kinds of explorations are allowed since there is a possibility of misconfiguring networks so that they no longer perform adequately.

In a communication network, there may be various network elements that run algorithms, which may be machine learning based algorithms, to optimize the functions of that network entity. As all network entities may be part of the larger communication network, the overall performance of the communication network is to be optimized as well. A master agent, which may be learning algorithm such as a reinforcement learning algorithm, may be utilized for optimizing the overall functions of the communication network. The master agent may be given, by an operator, a reward definition that defines an optimal outcome. The master agent then receives data from the network elements, which may be mobile network elements. The data represent functions of its respective network element. This step is illustrated in step S1 of the Figure 2. The master agent may, based on the data received and executing the reinforcement learning algorithm, determine which network element, among similar network elements, has performed best against the criteria of the optimal overall performance of the communication network, which may be indicated by the reward definition given to the master agent. This step, S2, is illustrated in Figure 2. Once the master agent has determined the best performing network element among similar network elements, it may then determine the algorithms to be used when operating the various network entities determined to be similar as illustrated in step S3, of Figure 2. Finally, the master agent communicates to the network entities that are determined to be similar the algorithms to be used as illustrated in step S4 of Figure 2. By exploring different policies and sub-policies the master agent may converge to good solutions that are suitable to various network

entities sharing similar network environments. The master agent thereby may select the best algorithm and/or reward definition for each network element controllable by it.

A reinforcement learning algorithm -based learning may be considered as a hierarchical learning in which the master agent runs the main algorithms and is considered to be at the top of the hierarchical structure. On lower levels of the hierarchical structure there may then be a sub-agent that consist of a reinforcement learning algorithm. The master agent provides a sub-reward definition to the sub-agent which may then use it as it defined its sub-policy. The sub-agent on the other hand provides information regarding its sub-policy and its key performance indicators to the master agent which can then evaluate the performance of the sub-agent and determine new sub-reward definition to be provided to the sub-agent.

In the hierarchical structure there may be a plurality of hierarchical levels and on each level there may be one or more sub-agents, each sub-agent representing a part of the communication network. The master agent and each sub-agent may then execute a respective reinforcement learning algorithm. The master agent determines the overall performance of the communication network. The master agent may evaluate the performance of each sub-agent and the share the sub-policy of the sub-agent that has the best performance with other sub-agents. In some examples, some sub-agents are determined to function in a similar state thereby being in a same category. The master agent may then determine which sub-agent has the best sub-policy in that category and share that sub-policy with other sub-agents in the same category. In some example embodiments the category of a sub-agent may be identified. A category that may be identified may be for example a category that has been pre-defined by a human or has been recognized by a machine. A machine may be able to recognize a category based on pattern recognition for example.

In one example embodiment, the agent may learn the notion of similarity and best performance. It may further be configured to learn to combine reward definitions. In such an example embodiment, the actions that the agent may perform are e.g., copy reward definition, merge reward definitions, or it may define new reward definitions from scratch. It may be beneficial to enable exploration (e.g.,

random) with the agent so that it may learn to manage reward definitions through trial and error.

In an example embodiment in which a neural network is utilized, the machine learning algorithms executed using the neural network may learn what the notion of similarity and subagent performance is at the learning phase. In such an example embodiment a master agent may receive as input state descriptions, which may be compressed, of sub-agents such that it may learn the notion of similarity.

Figure 3 illustrates an example of hierarchical reinforcement learning algorithm used to optimize the performance of a communication network. The master agent (310) determines master policy (312). The master policy is provided with a master reward definition (314) that is not optimized by the reinforcement learning algorithm. Instead it may be pre-determined, understandable to a human and it may be static or controllable by an operator. The master reward definition may comprise one or more high-level key performance indicators. The master policy, which can be e.g., an algorithm utilizing neural network framework with learnable weight factors, may be updated based on the master reward definition as well as based on the state the master agent is in (314). Based on the master policy (312) the master agent then takes one or more actions (316) and the environment (318) then turns those actions into a new state. Information regarding the new state may then be collected again (314) and based on that the master policy may again be updated and that affects the actions taken by the master agent and so on. In other words, there is a feedback loop to constantly monitor and update the master policy and thereby actions taken by the master agent.

The hierarchical structure on the other hand enables the master agent to determine sub-reward definitions for the sub-agents (320) and (330) as well. Although in this example embodiment there are only two sub-agents, there could be more sub-agents as well. The master agent (310) provides a sub-reward definition and, optionally, also a sub-policy (322) and (332) to the sub-agents (320) and (330). The sub-agents (320) and (330) then update their sub-policies (324) and (334) based on the received sub-reward definition and optionally also the sub-policy (322) and (332). The sub-agents (320) and (330) then take actions (326) and (336) based on

the updated sub-policies (324) and (334). The environments (328) and (338) then turn those actions into new states. Information regarding the new states may then be collected (329) and (339) and based on that the sub-policy may again be updated and that affects the actions taken by the sub-agents (320) and (330) and so on. In other words, there is a feedback loop to constantly monitor and update the sub-policies and thereby actions taken by the sub-agents (320) and (330). Information regarding the new states may also be shared with the master agent (310) thereby providing a feedback loop across the hierarchical levels as well.

Figure 4 illustrates another example with various hierarchical levels of the reinforcement learning algorithm. The amount of the hierarchical levels is not restricted nor the number of sub-agents in each hierarchical level. Figure 4 is thereby only an example for the purpose of illustrating the various hierarchical levels. In the example of Figure 4, the master agent (410) is considered to be at the top level of the hierarchy. On the next level there are sub-agents (422) and (424). The sub-agent (422) acts as a geographical area manager for area 1 of the communication network thereby optimizing the performance of the communication network in the area 1. The sub-agent (424) acts as a geographical area manager for area 2 of the communication network thereby optimizing the performance of the communication network in the area 2. It is to be noted that each level in the hierarchy is responsible for optimizing its own performance as well as the performance of the levels below in the hierarchy.

On a third level of the hierarchy in the example of Figure 4 there are sub-agents (432) and (434). The sub-agent (432) is an access node 1 and it optimizes the functioning of the communication network within the cell it provides. The sub-agent (434) is an access node 2 and it optimizes the performance of the communication network within the cell it provides. On the fourth level of the hierarchy there are sub-agents (442) and (444). The sub-agent (442) is a terminal device 1 and it optimizes the performance of itself and the devices connected to it. The sub-agent (444) is a terminal device 2 and it optimizes the functioning of itself and the devices connected to it. At the lowest level of hierarchy in the example of Figure 4 there are sub-agents (452) and (454). The sub-agent (452) is an IOT device 1 and it optimizes the functioning of itself. The sub-agent (454) is an IOT device 2 and it optimizes the

functioning of itself.

In the example of Figure 4 the sub-agents (432) and (434) are controlled by the sub-agent (424) due to close geographical location and a network connection that is fast enough for reacting to changes in the sub-states of the sub-agents (432) and (434). Similarly, the sub-agent (434) controls the sub-agents (442) and (444) as the terminal device 1 and terminal device 2 reside within the cell provided by the access node the sub-agent (434) represents. Further, the sub-agent (444) controls the sub-agents (452) and (454) as there is a connection between the terminal device 2 the sub-agent (444) represents and the IOT device 1 the sub-agent (452) represents and the IOT device 2 the sub-agent (454) represents.

The interaction between the master agent (410) at the top level of the hierarchy and the sub-agents (422) – (454) at the lower levels of the hierarchy enables optimizing the overall performance of the communication network. The master agent receives feedback from the lower levels of the hierarchy and recognizes the context of each sub-agents. The context may be understood as the local state representing the network conditions such as the amount of traffic, pathloss etc. The master agent may then determine which sub-agents are in a similar state such that they can be understood to be in a same category.

Within the category, the master agent may then evaluate the best-performing sub-agent and share the sub-policy of the best-performing sub-agent to other sub-agents in the same category. Yet, each sub-agent may continuously improve their performance through the reinforcement learning algorithm they execute against the sub-reward definition received from the controlling agent, which may be the master agent of another sub-agent. The sub-agents then send their policies and its performance to the controlling agent. This way there is a constant feedback loop allowing the master agent to have information regarding policies that provide best results all the time. The performance of a sub-agent may be understood in terms of key performance indicators such as reliability or correctness.

It is to be noted that the master agent may learn to evaluate the performance of sub-agents and may additionally learn to combine, merge or share sub-reward definitions (and even sub-policies). In this example embodiment, it is not

needed to predefine the similarity metric.

Rewards for the reinforcement learning algorithms may be understood as reward definitions. For the reward definitions to be suitable for use in the context of a communication network such as a cellular radio communication network, the reward definitions may be parametrized. The parametrization may depend on the level of the hierarchy. For example, a master reward definition may be static, or it may contain parameters that are understood by a human operator and may also be modified by the human operator. For example, the master reward definition may be defined as

$$R(Master) = g1*ThroughputUL[ti...ti+1] + g2* ThroughputDL[ti...ti+1]$$

In this definition g1 and g2 are static numbers or they may be otherwise determined by the human operator. However, g1 and g2 are not learned by executing the reinforcement algorithm. ThroughputUL defines uplink throughput and ThroughputDL defines downlink throughput. Also, the timescale [ti...ti+1] over which the throughput is computed is predefined in this example.

A sub-reward definition may be defined for example in the manner below:

$$RD(1) = a1* ThroughputULFreq1[ti...ti+1] + a2* ThroughputDLFreq1[ti...ti+1] + a3* ThroughputULFreq2[ti...ti+1] + a4* ThroughputDLFreq2[ti...ti+1]$$

In this example, a1, a2, a3 and a4 are variables that may be learned by the master agent. The sub-reward definition RD(1) may then be shared by the master agent to a sub-agent and the sub-agent utilizes it in its reinforcement learning algorithm to maximize the reward given by this sub-reward definition RD(1). In this example, ThroughputULFreq1 and ThroughputULFreq2 are key performance indicators indicating the performance in terms of uplink throughput at different frequencies. ThroughputDLFreq1 and ThroughputDLFreq2 are key performance indicators indicating the performance in terms of downlink throughput at different frequencies. In this example, the master agent may learn to instruct sub-agents to focus on different frequency bands. However, in some other examples there may be different

reward definition parameters such that the reward definition parameters are respective to multiple timescales, frequency ranges, antenna directions, beam forming key performance indicators and so on.

It is to be noted that any kind of reinforcement learning algorithm could be used by the agents. Examples of such algorithms include deep Q-Learning and A3C.

There may be a situation in which each sub-agent must learn the parameters of the Q network from scratch, which would make the process rather inefficient. To improve the efficiency, a summary of the state description (Sn), collected reward (Rn) and Policy (Pn) of each sub-agent may be periodically, or responsive to a trigger, sent to the master agents. The state description may be high-dimensional which would require a lot of resource to communicate over the network. To make the communication of the state description (Sn) more efficient, a dimensionality reduction method such as a bottleneck autoencoder or PCA may be applied to the state description (Sn).

Once the master agent receives a trigger to update or share policies, it may execute a clustering algorithm that outputs a distance between sub-agents that may be derived from the summaries sent by the sub-agents. Based on the distance the master agent may then replace sub-policies of sub-agent with better-performing sub-policies. Additionally, this enables a new sub-agent to perform better from the start as the so called cold-start is not needed but a known well-performing sub-policy may b taken into use directly without a need to learn it first.

In order to be able to share information between various network elements, data structure needs to be defined. The data structure elements could comprise for example application ID, current machine learning algorithms used, list of potential available algorithms, device ID, predicted location ID, predicted time, predicted load, operator ID etc. The data structure elements could be combined in different parts of the communication network, for example in a cloud. The data could then be provided as an input to the master agent that may then, based on the data, trigger changes as needed in the selection of right algorithms, or a combination of algorithms, to be used for controlling the functionality of the network better. The changes may be such that they are applied to those sub-agents that are determined

to be in a similar, or substantially similar, state thereby belonging to a same category. One criterion for the similarity may be location. It may be assumed that sub-agent that are geographically located close to each other have higher chances of similarity of the context in other ways as well. A distance threshold may be defined that determines if the sub-agents are considered to be close enough to each other so that they are assumed to have similar states. Therefore, location may have a weighted role in the deficient of a context. Similarity can also be determined by e.g., comparing the internal states or a compressed representation thereof, and e.g., using clustering or some other machine learning method.

In an example, secondary cells in LTE Carrier Aggregation, CA, and high band 5G/NR can provide localized enhanced capacity and may need help from low band LTE primary and/or macro cell for control traffic. 4G LTE Advanced CA or channel aggregation enables multiple LTE carriers to be used together to provide the high data rates required for 4G LTE Advanced. A need for more capacity is a motivation for enhancements in LTE and 5G. Local optimizations may enable more capacity to be available. An example of a localized solution is to have secondary cells on top of an LTE primary cell via CA or via higher band localized 5G on top of lower band wider coverage LTE or 5G cell. In such examples, control information may be provided via the primary LTE cell or the wider coverage lower band LTE or 5G cell. Yet, in all these examples a trade-off between downlink and uplink throughput and resources dedicated to control signalling is still needed. The reinforcement learning algorithm may be used to optimize the network performance both locally and also in the context of the overall communication network. For example, the number of physical resource blocks, PRBs, for a physical uplink control channel, PUCCH, is to be optimized locally by a sub-agent. Another example is the reporting period for channel quality indicator CQI, which is to be optimized locally to achieve an optimal trade-off between downlink and uplink throughput and resources dedicated to control signalling. For example, in a stable environment there may not be as much need for channel quality information compared to a local environment with a crowd moving in a difficult propagation environment like a dense city area with many participants of the crowd sending data via uplink.

The examples introduced above have various advantages. Some of the advantages include optimal performance of the network via optimal combination of the supporting and enabling algorithms and faster reaction to changes in a local environment to optimize the radio resource management, quality of experience, scheduling and beam forming.

The apparatus 500 of Figure 5 illustrates an example embodiment of an apparatus that may be an access node or be comprised in an access node. The apparatus 500 may be configured to execute a reinforcement learning algorithm. The apparatus may be, for example, a circuitry or a chipset applicable to an access node to realize the described embodiments. The apparatus (500) may be an electronic device comprising one or more electronic circuitries. The apparatus (500) may comprise a communication control circuitry (510) such as at least one processor, and at least one memory (520) including a computer program code (software) (522) wherein the at least one memory and the computer program code (software) (522) are configured, with the at least one processor, to cause the apparatus (500) to carry out any one of the example embodiments of the access node described above.

The memory (520) may be implemented using any suitable data storage technology, such as semiconductor-based memory devices, flash memory, magnetic memory devices and systems, optical memory devices and systems, fixed memory and removable memory. The memory may comprise a configuration database for storing configuration data. For example, the configuration database may store current neighbour cell list, and, in some example embodiments, structures of the frames used in the detected neighbour cells.

The apparatus (500) may further comprise a communication interface (530) comprising hardware and/or software for realizing communication connectivity according to one or more communication protocols. The communication interface (530) may provide the apparatus with radio communication capabilities to communicate in the cellular communication system. The communication interface may, for example, provide a radio interface to terminal devices. The apparatus (500) may further comprise another interface towards a core network such as the network coordinator apparatus and/or to the access nodes of the cellular communication

system. The apparatus (500) may further comprise a scheduler (540) that is configured to allocate resources.

As used in this application, the term 'circuitry' refers to all of the following: (a) hardware-only circuit implementations, such as implementations in only analog and/or digital circuitry, and (b) combinations of circuits and software (and/or firmware), such as (as applicable): (i) a combination of processor(s) or (ii) portions of processor(s)/software including digital signal processor(s), software, and memory(ies) that work together to cause an apparatus to perform various functions, and (c) circuits, such as a microprocessor(s) or a portion of a microprocessor(s), that require software or firmware for operation, even if the software or firmware is not physically present. This definition of 'circuitry' applies to all uses of this term in this application. As a further example, as used in this application, the term 'circuitry' would also cover an implementation of merely a processor (or multiple processors) or a portion of a processor and its (or their) accompanying software and/or firmware. The term 'circuitry' would also cover, for example and if applicable to the particular element, a baseband integrated circuit or applications processor integrated circuit for a mobile phone or a similar integrated circuit in a server, a cellular network device, or another network device. The above-described embodiments of the circuitry may also be considered as embodiments that provide means for carrying out the embodiments of the methods or processes described in this document.

The techniques and methods described herein may be implemented by various means. For example, these techniques may be implemented in hardware (one or more devices), firmware (one or more devices), software (one or more modules), or combinations thereof. For a hardware implementation, the apparatus(es) of embodiments may be implemented within one or more application-specific integrated circuits (ASICs), digital signal processors (DSPs), digital signal processing devices (DSPDs), programmable logic devices (PLDs), field programmable gate arrays (FPGAs), graphics processing units (GPUs), processors, controllers, micro-controllers, microprocessors, other electronic units designed to perform the functions described herein, or a combination thereof. For firmware or software, the

implementation can be carried out through modules of at least one chipset (e.g. procedures, functions, and so on) that perform the functions described herein. The software codes may be stored in a memory unit and executed by processors. The memory unit may be implemented within the processor or externally to the processor. In the latter case, it can be communicatively coupled to the processor via various means, as is known in the art. Additionally, the components of the systems described herein may be rearranged and/or complemented by additional components in order to facilitate the achievements of the various aspects, etc., described with regard thereto, and they are not limited to the precise configurations set forth in the given figures, as will be appreciated by one skilled in the art.

Embodiments as described may also be carried out in the form of a computer process defined by a computer program or portions thereof. Embodiments of the methods described in connection with Figures 2 to 4 may be carried out by executing at least one portion of a computer program comprising corresponding instructions. The computer program may be in source code form, object code form, or in some intermediate form, and it may be stored in some sort of carrier, which may be any entity or device capable of carrying the program. For example, the computer program may be stored on a computer program distribution medium readable by a computer or a processor. The computer program medium may be, for example but not limited to, a record medium, computer memory, read-only memory, electrical carrier signal, telecommunications signal, and software distribution package, for example. The computer program medium may be a non-transitory medium. Coding of software for carrying out the embodiments as shown and described is well within the scope of a person of ordinary skill in the art.

Even though the invention has been described above with reference to an example according to the accompanying drawings, it is clear that the invention is not restricted thereto but can be modified in several ways within the scope of the appended claims. Therefore, all words and expressions should be interpreted broadly and they are intended to illustrate, not to restrict, the embodiment. It will be obvious to a person skilled in the art that, as technology advances, the inventive concept can be implemented in various ways. Further, it is clear to a person skilled

in the art that the described embodiments may, but are not required to, be combined with other embodiments in various ways.

Claims

1. An apparatus comprising:

at least one processor, and

at least one memory including a computer program code, wherein the at least one memory and the computer program code are configured, with the at least one processor, to cause the apparatus to:

execute a master learning algorithm that uses a master reward definition as an input;

update a sub-reward definition based on one or more variables that have been determined by executing the master learning algorithm;

share the sub-reward definition to a plurality of sub-algorithms executed by their respective network elements;

receive, from the sub-algorithms, information regarding their respective sub-policies that were obtained based on, at least partly, the shared sub-reward definition;

determine which sub-policies are associated with a first category;

determine which of the sub-policies belonging to the first category has the best performance in comparison to a first criteria; and

update the sub-policies associated with the first category according to the sub-policy that has the best performance.

2. An apparatus according to claim 1, wherein the master reward definition is pre-determined.

3. An apparatus according to any of claims 1 or 2, wherein the master reward definition is based, at least, on one or more high-level key performance indicators.

4. An apparatus according to any previous claim, wherein the sub-policies are further configured to automatically respond to changes in their respective state.

5. An apparatus according to any previous claim, wherein the master reward definition and the sub-reward definition are parametrized and applicable to communication network control.

6. A system comprising:

means for executing a master learning algorithm that uses a master reward definition as an input;

means for updating a sub-reward definition based on one or more variables that have been determined by executing the master learning algorithm;

means for sharing the sub-reward definition to a plurality of sub-algorithms executed by their respective network elements;

means for receiving, from the sub-algorithms, information regarding their respective sub-policies that were obtained based on, at least partly, the shared sub-reward definition;

means for determining which sub-policies are associated with a first category;

means for determining which of the sub-policies belonging to the first category has the best performance in comparison to a first criteria; and

means for updating the sub-policies associated with the first category according to the sub-policy that has the best performance.

7. A system according to claim 6, wherein the master reward definition is pre-determined.

8. A system according to any of claims 6 or 7, wherein the master reward definition is based, at least, on one or more high-level key performance indicators.

9. A system according to any previous claim, wherein the sub-policies are further configured to automatically respond to changes in their respective state.

10. A system according to any previous claim, wherein the master reward definition and the sub-reward definition are parametrized and applicable to communication network control.

11. A method comprising:

executing a master learning algorithm that uses a master reward definition as an input;

updating a sub-reward definition based on one or more variables that have been determined by executing the master learning algorithm;

sharing the sub-reward definition to a plurality of sub-algorithms executed by their respective network elements;

receiving, from the sub-algorithms, information regarding their respective sub-policies that were obtained based on, at least partly, the shared sub-reward definition;

determining which sub-policies are associated with a first category;

determining which of the sub-policies belonging to the first category has the best performance in comparison to a first criteria; and

updating the sub-policies associated with the first category according to the sub-policy that has the best performance.

12. A method according to claim 11, wherein the master reward definition is pre-determined.

13. A method according to any of claims 11 or 12, wherein the master reward definition is based, at least, on one or more high-level key performance indicators.

14. A method according to any previous claim, wherein the sub-policies are further configured to automatically respond to changes in their respective state.

15. A method according to any previous claim, wherein the master reward definition and the sub-reward definition are parametrized and applicable to communication network control.

16. A computer program product readable by a computer and, when executed by the computer, configured to cause the computer to execute a computer process comprising:

executing a master learning algorithm that uses a master reward definition as an input;

updating a sub-reward definition based on one or more variables that have been determined by executing the master learning algorithm;

sharing the sub-reward definition to a plurality of sub-algorithms executed by their respective network elements;

receiving, from the sub-algorithms, information regarding their respective sub-policies that were obtained based on, at least partly, the shared sub-reward definition;

determining which sub-policies are associated with a first category;

determining which of the sub-policies belonging to the first category has the best performance in comparison to a first criteria; and

updating the sub-policies associated with the first category according to the sub-policy that has the best performance.

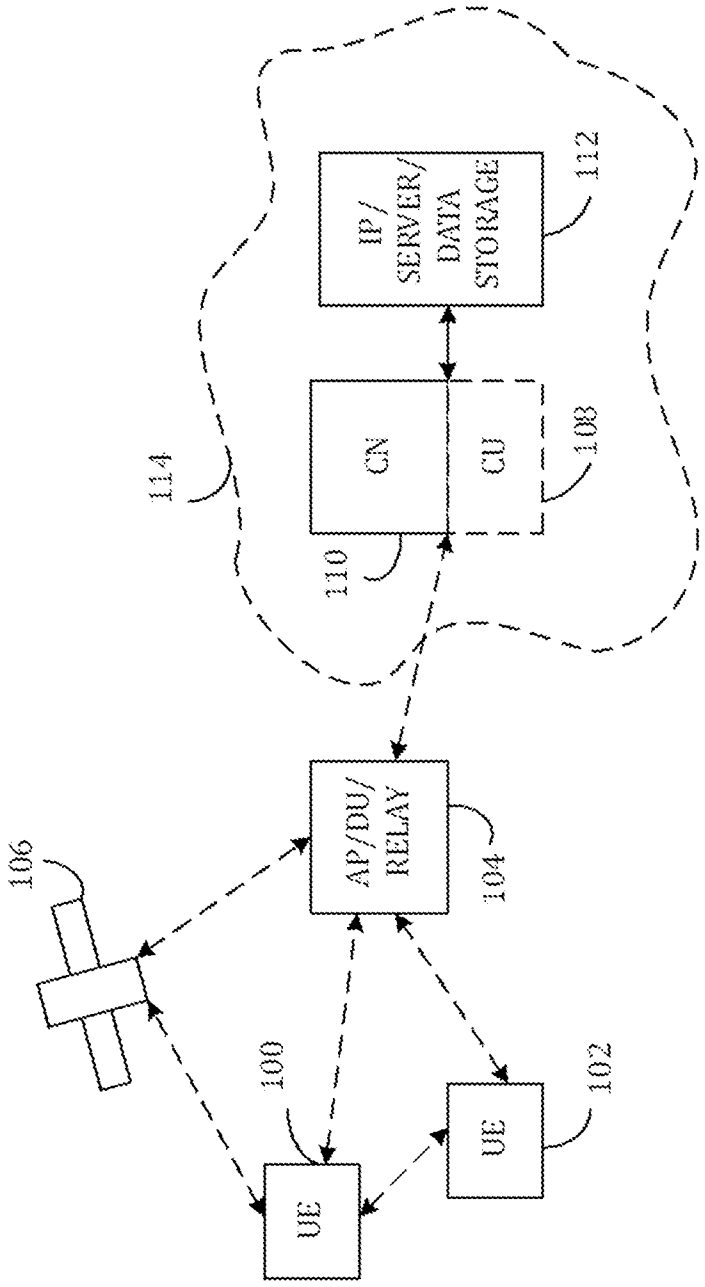
IP/
SERVER/
DATA
STORAGE
112
CN
CU
108
114
110
AP/DU/
RELAY
104
106
100
UE
UE
102

Figure 1

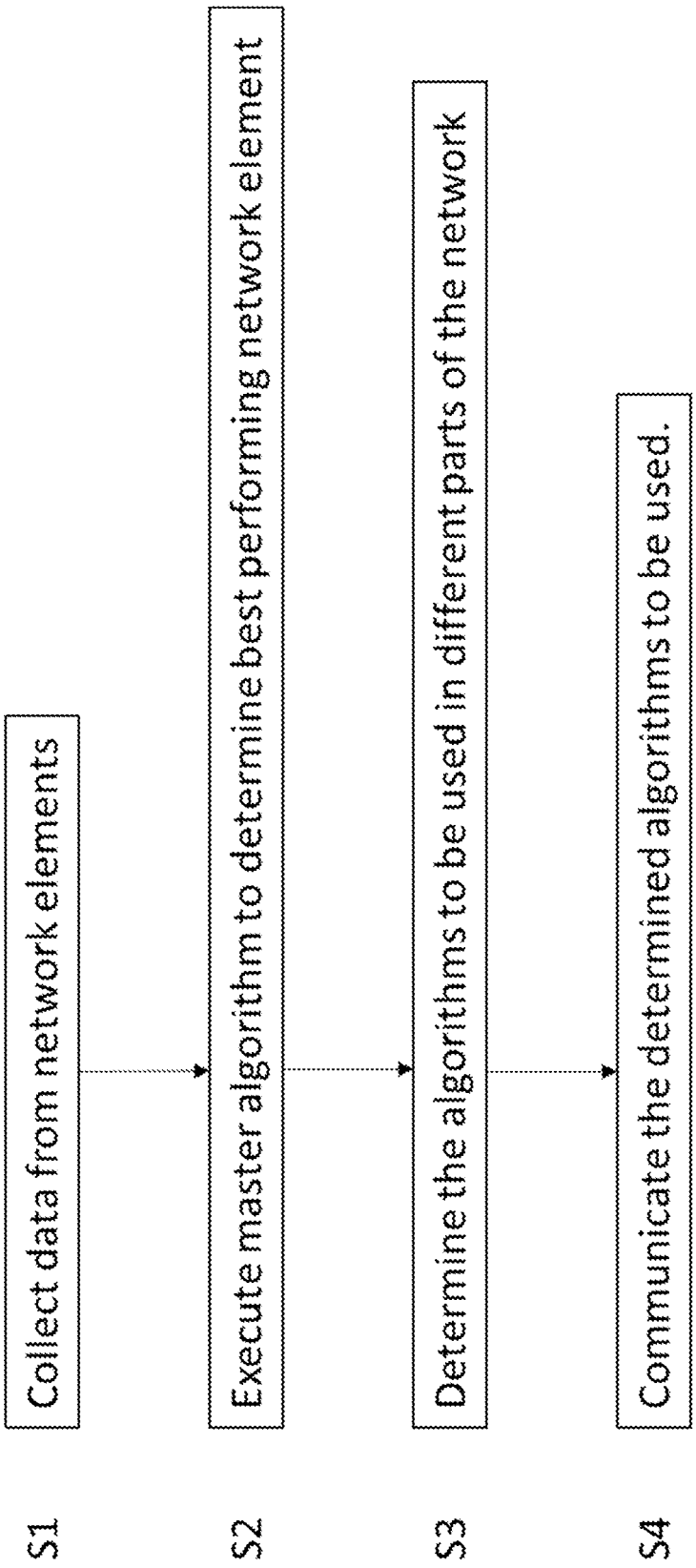
S1
Collect data from network elements
S2
Execute master algorithm to determine best performing network element
S3
Determine the algorithms to be used in different parts of the network
S4
Communicate the determined algorithms to be used.

Figure 2

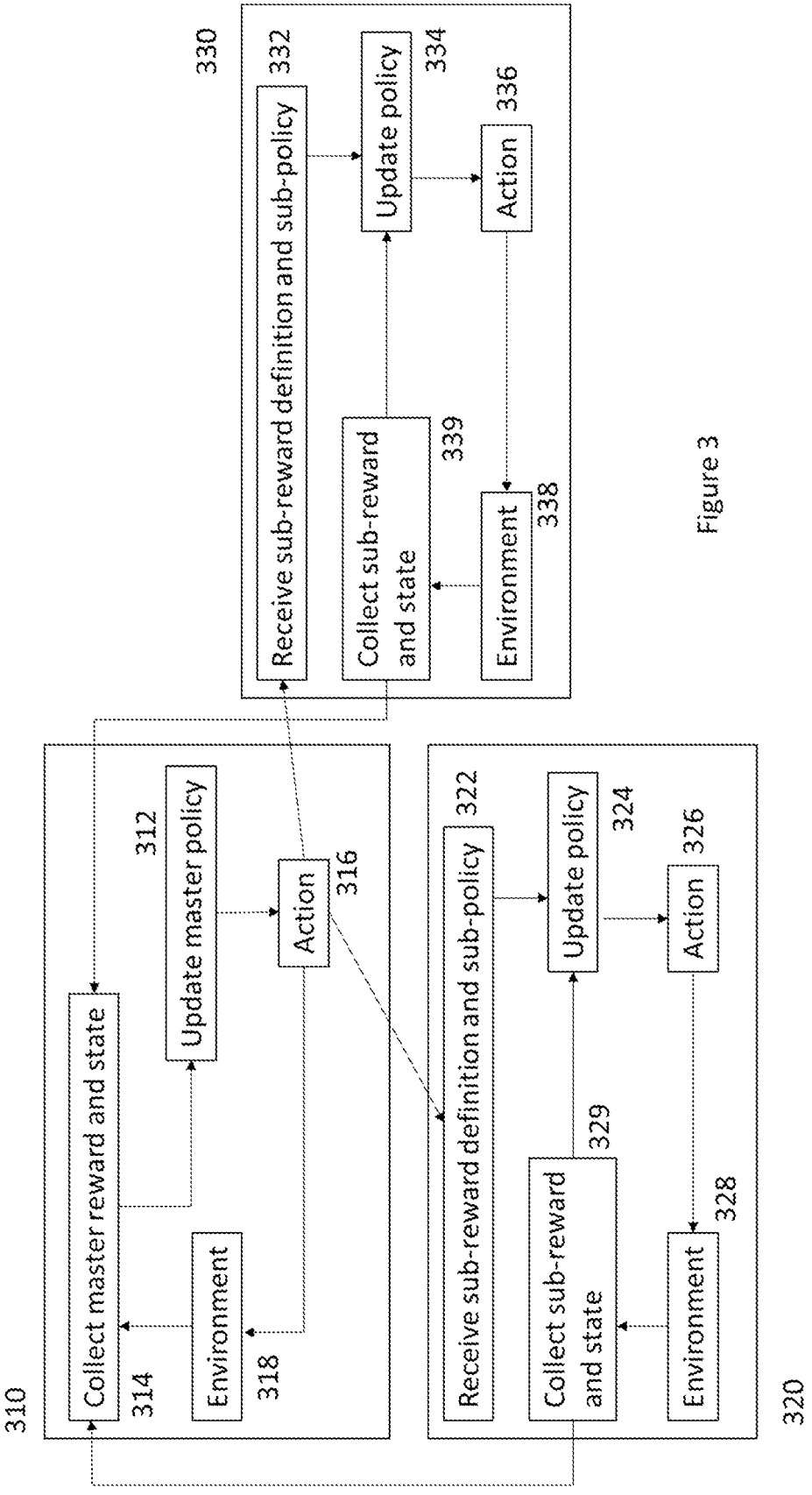
310
Collect master reward and state
314
Environment
318
Update master policy
312
Action
316
320
Receive sub-reward definition and sub-policy
322
Collect sub-reward and state
329
Environment
328
Update policy
324
Action
326
330
Receive sub-reward definition and sub-policy
332
Collect sub-reward and state
339
Environment
338
Update policy
334
Action
336

Figure 3

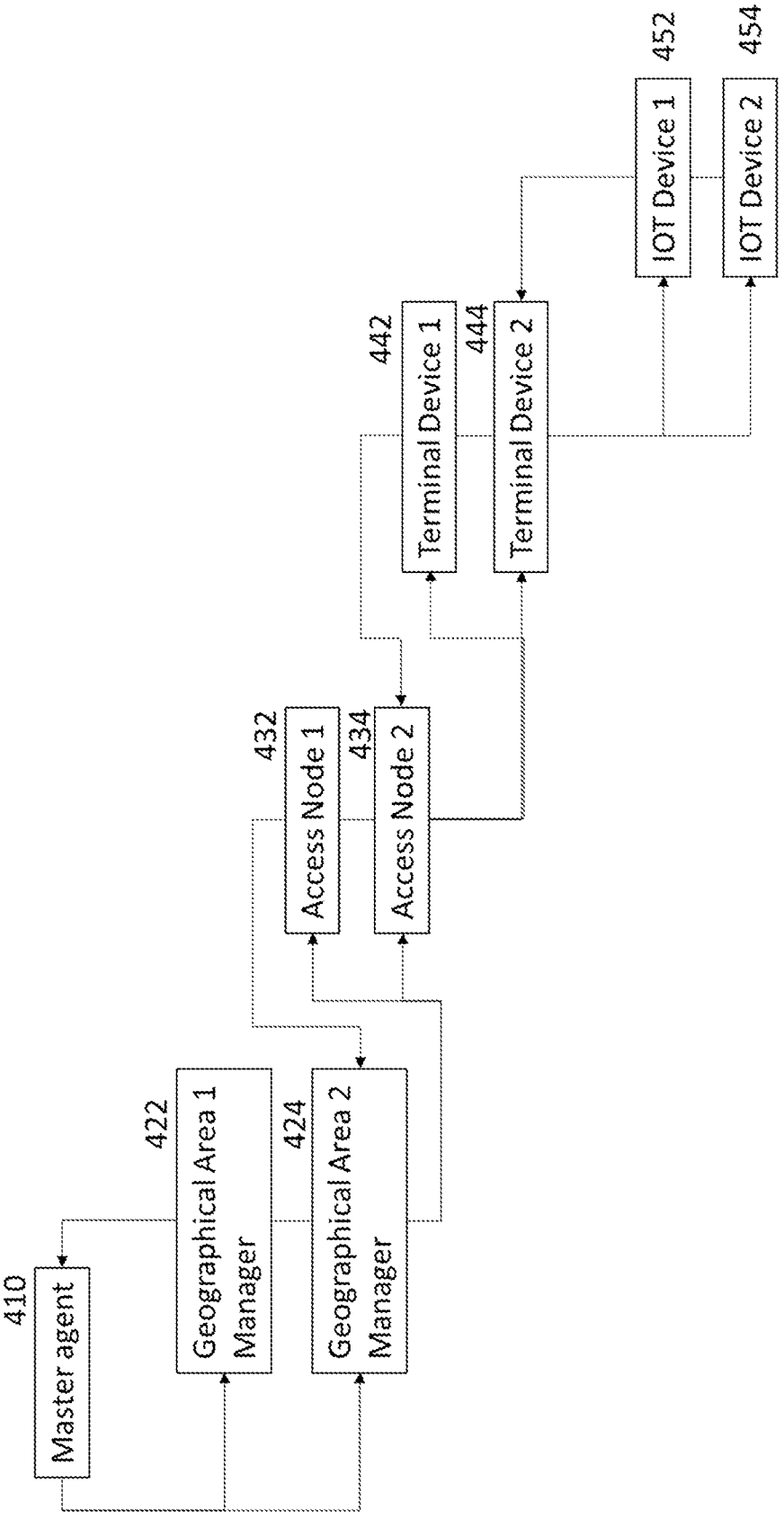
410
Master agent
422
Geographical Area 1 Manager
424
Geographical Area 2 Manager
432
Access Node 1
434
Access Node 2
442
Terminal Device 1
444
Terminal Device 2
452
IOT Device 1
454
IOT Device 2

Figure 4

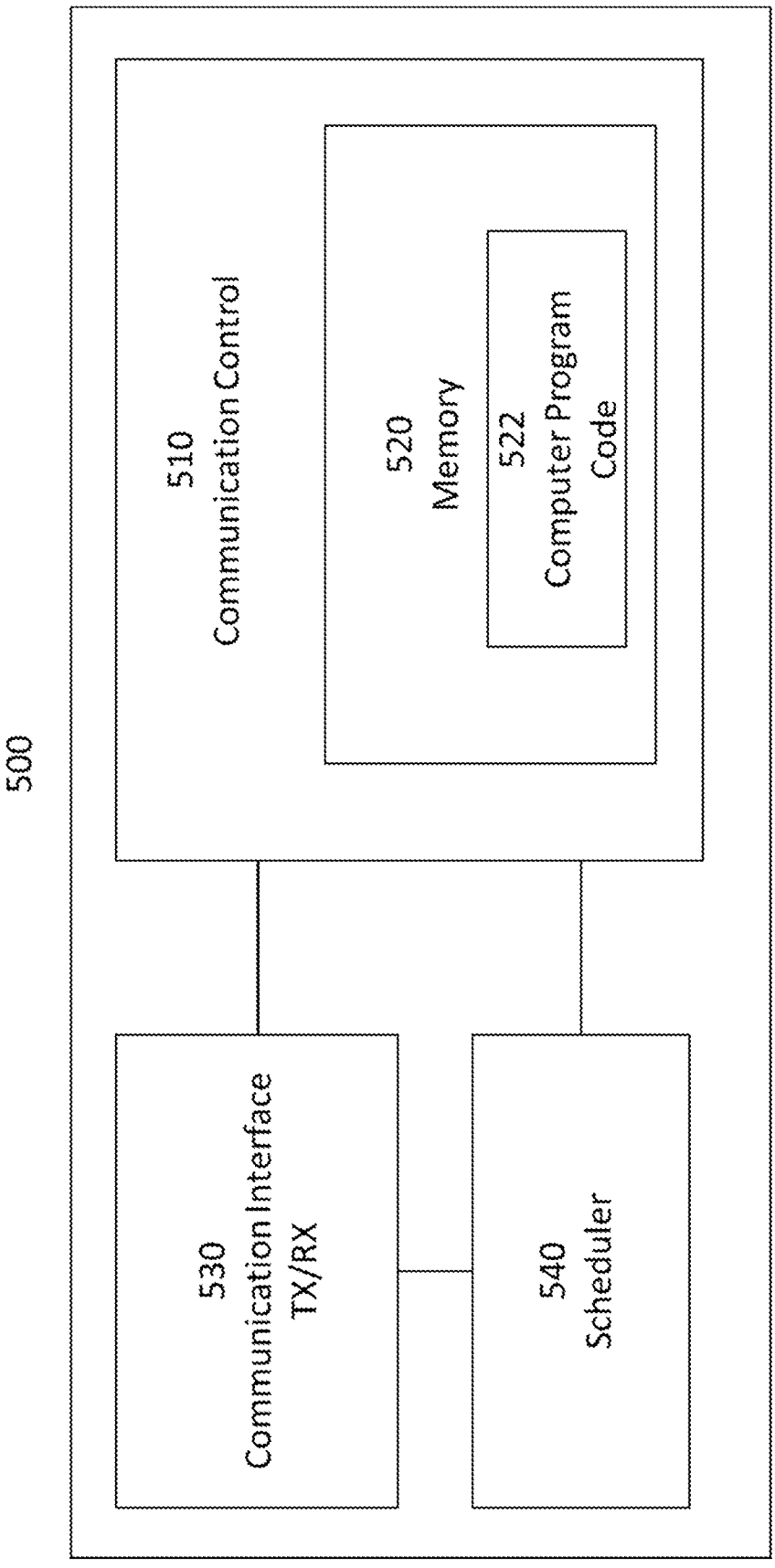
500
510
Communication Control
520
Memory
522
Computer Program
Code
530
Communication Interface
TX/RX
540
Scheduler

Figure 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/FI2019/050049

A. **CLASSIFICATION OF SUBJECT MATTER**

See extra sheet

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)

IPC: H04W, H04L, G05B, G06N

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

FI, SE, NO, DK

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

EPODOC, EPO-Internal full-text databases, Full-text translation databases from Asian languages, WPIAP, PRH-Internal, XP3GPP, XPAIP, XPESP, XPETSI, XPI3E, XPIEE, XPIETF, XPIOP, XPIPCOM, XPJPEG, XPMISC, XPOAC, XPRD, XPTK, COMPDX, INSPEC, TDB, NPL, Internet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2013122885 A1 (KOJIMA YUJI [JP]) 16 May 2013 (16.05.2013) the whole document, in particular abstract; paragraphs [0004], [0008], [0028]-[0030], [0039], [0044]-[0046], [0050]-[0051], [0055]-[0058], [0079]-[0082]; figs. 1-7	1-16
X	WO 2019007388 A1 (HUAWEI TECH CO LTD [CN]) 10 January 2019 (10.01.2019) the whole document, in particular abstract; paragraphs [6]-[8], [30]-[32], [41], [62]-[64], [69]-[71], [86], [89], [96]; figs. 3, 5, 10, 13, 18	1-16
X	WO 2012073059 A1 (NOKIA CORP [FI]) 07 June 2012 (07.06.2012) the whole document, in particular abstract; page 15 line 26-page 18 line 4, page 21 lines 1-34, page 22 lines 30-34, page 24 line 1-page 26 line 14; fig. 4	1-16

☐ 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 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

"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	Date of mailing of the international search report
22 May 2019 (22.05.2019)	22 May 2019 (22.05.2019)

Name and mailing address of the ISA/FI
Finnish Patent and Registration Office
FI-00091 PRH, FINLAND
Facsimile No. +358 29 509 5328

Authorized officer
Ari Hottinen
Telephone No. +358 29 509 5000

INTERNATIONAL SEARCH REPORT
Information on Patent Family Members

International application No.
PCT/FI2019/050049

Patent document cited in search report	Publication date	Patent family members(s)	Publication date
US 2013122885 A1	16/05/2013	US 8897767 B2	25/11/2014
		JP 2013106202 A	30/05/2013
		JP 5733166 B2	10/06/2015
WO 2019007388 A1	10/01/2019	US 2019014487 A1	10/01/2019
		US 2019014488 A1	10/01/2019
		WO 2019007386 A1	10/01/2019
WO 2012073059 A1	07/06/2012	None	

INTERNATIONAL SEARCH REPORT

International application No.
PCT/FI2019/050049

CLASSIFICATION OF SUBJECT MATTER

IPC
H04W 24/02 (2009.01)
H04L 12/26 (2006.01)
G05B 13/04 (2006.01)
G06N 20/00 (2019.01)
G06N 3/02 (2006.01)