



- (51) **International Patent Classification:**
G06F 9/50 (2006.01)
- (21) **International Application Number:**
PCT/US2016/068301
- (22) **International Filing Date:**
22 December 2016 (22.12.2016)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
62/387,556 23 December 2015 (23.12.2015) US
- (71) **Applicant:** IDAC HOLDINGS, INC. [US/US]; 200 Bellevue Parkway, Suite 300, Wilmington, DE 19809 (US).
- (72) **Inventor:** SAHIN, Onur; 33 Kleine Wharf, 14 Orsman Road, London N1 5QL (GB).
- (74) **Agents:** ROCCIA, Vincent J. et al.; Condo Roccia Kop-tiw LLP, 1800 JFK Blvd., Suite 1700, Philadelphia, PA 19103 (US).
- (81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM,

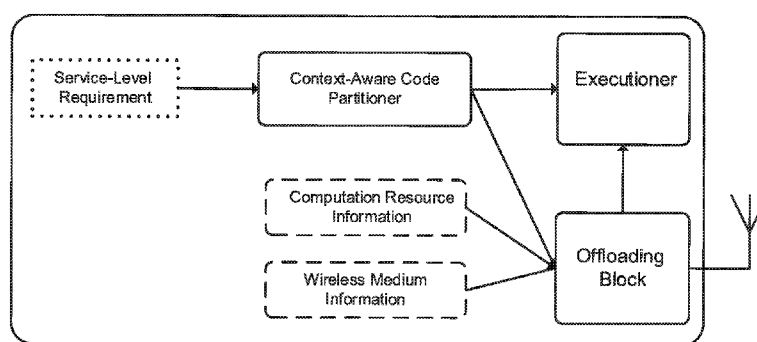
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, 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, 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).

Published:

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

- (54) **Title:** METHODS OF OFFLOADING COMPUTATION FROM MOBILE DEVICE TO CLOUD

**FIG. 3**

(57) **Abstract:** Systems, methods and instrumentalities are disclosed for offloading computation for an application on a local device in a wireless network, comprising determining service-level requirements for the application, wherein the service-level requirements comprise one or more of latency, reliability, and power consumption for processing; determining wireless network conditions; determining local device conditions; partitioning the application into one or more tasks; comparing a processing time of a task at the local device to a latency requirement of the task, and: if the processing time of the local device would exceed the latency requirement of the task, determining to offload the task; and if the processing time of the local device would not exceed the latency requirement of the task, comparing a power consumed at the local device to offload the task to a power consumed at the local device to execute the task, wherein if the power consumed at the local device to offload the task is less than the power consumed (e.g., while satisfying the latency and reliability requirements of the task) at the local device to execute the task, determining to offload the task.



METHODS OF OFFLOADING COMPUTATION FROM MOBILE DEVICE TO CLOUD

CROSS REFERENCE TO RELATED APPLICATION

[0001] This application claims the benefit of Provisional U.S. Patent Application No. 62/387,556 filed December 23, 2015, the contents of which are hereby incorporated by reference herein.

BACKGROUND

[0002] Recent development and popularity of smart phones and/or mobile applications (*e.g.*, virtual and augmented reality, speech recognition, and/or high definition video services) may enable mobile devices to use the resources of cloud computing (*e.g.*, computing, storage, operating systems, software, *etc.*) via wired and/or wireless networks. Mobile cloud computing (MCC) aims at including mobile devices in a cloud computing-like scenario, for example, such that the computation and/or power limitations of mobile applications and services may be overcome. The challenges in MCC may be different from those in traditional cloud computing technology. Frameworks and methods may be desirable to realize the potential of MCC.

SUMMARY

[0003] Systems, methods and instrumentalities are disclosed for offloading computation for an application on a local device in a wireless network. The local device (*e.g.*, a WTRU) may determine service-level requirements for the application, where the service-level requirements comprise one or more of latency, reliability, and power consumption for processing. The local device may then determine wireless network conditions and local device conditions. The local device may determine to partition the application into one or more tasks, and compare a processing time of a task at the local device to a latency requirement of the task. The latency requirement may include an amount of time incurred to upload the task to a remote entity, an amount of time incurred to process the task at the remote entity, and/or an amount of time incurred to download the executed task once processed by the remote entity. The local device may obtain the latency requirement from the application (*e.g.*, service-level). If the local device determines that the processing time of the local device would exceed the latency requirement of the task, then the local device may determine to offload the task. For example, the latency incurred once the local entity opts for offloading may include the total amount of time for the

task upload, processing at the remote entity, and the time it takes to download the result processed at the remote entity. If the local device determines that the processing time of the local device would not exceed the latency requirement of the task, then the local device may compare a power consumed at the local device to offload the task to a power consumed at the local device to execute the task, where if the power consumed at the local device to offload the task is less than the power consumed (*e.g.*, while satisfying the latency and reliability requirements of the task) at the local device to execute the task, then the local device may determine to offload the task. And if the local device determines that the power consumed at the local device to offload the task is greater than the power consumed (*e.g.*, while satisfying the latency and reliability requirements of the task) at the local device to execute the task, then the local device may determine to execute the task locally.

BRIEF DESCRIPTION OF THE DRAWINGS

- [0004] FIG. 1 depicts an example mobile cloud computing (MCC) architecture.
- [0005] FIG. 2 depicts an example of local and remote task partitioning.
- [0006] FIG. 3 depicts an example task offloading architecture at a local device.
- [0007] FIG. 4 depicts an example flow chart for task offloading.
- [0008] FIG. 5 depicts an example flow chart for task offloading at an offloading block.
- [0009] FIG. 6 depicts an example flow chart for task offloading at an offloading block with acknowledgment.
- [0010] FIG. 7 depicts an example flow chart for task offloading with system configuration parameter provisioning.
- [0011] FIG. 8A depicts a diagram of an example communications system in which one or more disclosed embodiments may be implemented.
- [0012] FIG. 8B depicts a system diagram of an example wireless transmit/receive unit (WTRU) that may be used within the communications system illustrated in FIG. 8A.
- [0013] FIG. 8C depicts a system diagram of an example radio access network and an example core network that may be used within the communications system illustrated in FIG. 8A.
- [0014] FIG. 8D depicts a system diagram of another example radio access network and an example core network that may be used within the communications system illustrated in FIG. 8A.
- [0015] FIG. 8E depicts a system diagram of another example radio access network and an example core network that may be used within the communications system illustrated in FIG. 8A.

DETAILED DESCRIPTION

[0016] A detailed description of illustrative embodiments will now be described with reference to the various Figures. Although this description provides a detailed example of possible implementations, it should be noted that the details are intended to be exemplary and in no way limit the scope of the application.

[0017] FIG. 1 shows an example mobile cloud computing (MCC) architecture. MCC may utilize cloud computing technologies. MCC may collaboratively utilize the underlying computation power of contributing system elements in a wireless network. The wireless network may include one or more wireless transmit/receive units (WTRUs). Examples of WTRUs depicted in FIG. 1 include a smartphone (WTRU 1) and a laptop computer (WTRU 2). The wireless network may include a network device. Examples of network devices are servers and other networking resources.

[0018] MCC may face challenges such as mobility and unreliability in the wireless links. The design of MCC may consider mobile device integration, quality of service, service level requirements of the mobile applications (*e.g.*, latency, reliability, and/or power consumption), *etc.* Cloud computing techniques (*e.g.*, power techniques such as application partitioning and code offloading) may be built upon distributed computing. For example, application partitioning may divide an application (*e.g.*, software blocks of such an application) into one or more blocks (*e.g.*, serializable blocks). The blocks may be denoted as tasks. The tasks may be serialized. Code offloading may transmit the blocks (or tasks) from the local device(s) (*e.g.*, WTRU 1 and/or WTRU 2) to the remote network entity/device over the wireless medium.

[0019] MCC design may include one or more of the cloud computing techniques (*e.g.*, the application partitioning and/or code offloading techniques described herein). The design may take the underlying system and/or wireless resources (*e.g.*, CPU power, link capacity, and/or reliability) into account.

[0020] MCC solutions may take a systematic approach. Mobile computing offloading (MCO) may be an enabler of MCC. MCO may allow distributed and/or parallel computation of applications. A code offloading framework may be developed between a local device (*e.g.*, a mobile device) and network elements. An example framework may include one or more modules (*e.g.*, combination of modules) configured to handle device virtualization, client-server communication, mobile agents, and/or other tasks. For example, device virtualization may enable parallel and/or independent execution of the offloaded tasks. The virtual machines may be used for task storage and/or task execution, for example. An example implementation may

configure a guest operating system in a remote entity to process offloaded tasks. Client-server communication may be responsible for code offloading between a local device and remote entities. Mobile agents may be developed to execute and collate the tasks distributed in the offloading framework.

[0021] FIG. 2 shows an example of local and remote task partitioning. Execution flow of an application may be partitioned into multiple tasks. One or more of these tasks (*e.g.*, Task 2 and Task 3) may be executed locally (*e.g.*, at the device). One or more of these tasks (*e.g.*, Task 1 and Task 4) may be executed remotely from the perspective of the local device, for example, by one or more remote entities.

[0022] FIG. 3 shows an example task offloading architecture at a local device. The example architecture may include one or more functions (which may be referred to herein as blocks) such as a Code Partitioner, an Executioner, and Offloading Block. The device may have one or more applications (not depicted). An application may require computational resources. For example, an application may be a mobile application residing on a WTRU (*e.g.*, cell-phone, tablet, *etc.*), a desktop application residing on a personal computer, a cloud application, and/or the like.

[0023] The Code Partitioner may be context aware. The Code Partitioner may receive service level requirements for an application. Examples of service level requirements include latency, reliability, and/or power consumption. The Code Partitioner may partition an application into one or more tasks. The Code Partitioner may forward tasks to the Executioner and/or the Offloading Block. For example, the Executioner block may be activated for local processing. The Code Partitioner may forward tasks to the Offloading Block. For example, the Offloading Block may be involved in remote execution of the task. The Offloading Block may identify the system configuration and/or wireless transmission parameters (*e.g.*, using the algorithms for identifying the parameters as described herein). The Offloading Block may send the task to a remote entity. The Offloading Block may receive an executed task from the remote entity. The Offloading Block may forward the executed task to the Executioner. The execute task and the application execution flow may be combined (*e.g.*, by the Executioner).

[0024] One or more of the following considerations may be taken into account in the design of the offloading framework. Framework modules (*e.g.*, partitioning, cloning/virtualization, task offloading, *etc.*) may collectively utilize computation and wireless resources (*e.g.*, CPU, total power such as battery power, transmission scheduling, transmission duration, modulation-coding schemes, *etc.*). The utilization of the system resources may aim to increase the efficiency of the utilization and/or to reduce conflicting resource assignments (*e.g.*, high data rate offloading in low signal-to-noise ratio (SNR), high CPU allocation in erroneous code reception, *etc.*). Server-

client communication procedures (*e.g.*, information exchange between the local device and remote entities) may incorporate the dynamics of wireless links among the network entities (*e.g.*, during the offloading decision-making process). Code partitioning and offloading procedures may be robust to errors in the wireless communication medium between the offloading (*e.g.*, a source or local device) and offloaded (*e.g.*, a remote device) entities.

[0025] Code partitioning at the local device and/or offloading of the corresponding tasks may be realized by configuring computation and/or transmission parameters. The configuration may be based on the system and/or communication medium characteristics. An acknowledgement procedure between the local device and the remote entity may adaptively update computation and/or transmission parameters.

[0026] The example architecture may include other functions, *e.g.*, computation resource profiling and/or wireless medium information provisioning. The example architecture may be used at a local device and/or at a remote entity. If used at a remote entity, the Code Partitioner block may be removed. One or more modules of the offloading framework (*e.g.*, a controller, a mobile agent, a partitioner, *etc.*) may partition an application (*e.g.*, a video compression application) into multiple tasks (*e.g.*, serializable tasks). The tasks may be distributed among various remote entities.

[0027] The local device (*e.g.*, a WTRU) may determine whether and how to offload computation for an application. For example, the local device may determine service-level requirements for the application, where the service-level requirements comprise one or more of latency, reliability, and power consumption for processing.

[0028] The local device may determine wireless network conditions and/or local device conditions. The wireless network conditions may include information relating to the wireless medium, such as one or more of bandwidth, signal-to-noise ratio (SNR), latency, channel state information (CSI), packet error rate (PER), bit error rate (BER), interference profile of the channel, *etc.*, for example. The local device conditions may include information relating to computation resource information at the local device, such as the total and/or available processing power, total and/or available battery life, wired/wireless connectivity, information relating to other tasks/processes being performed by the local device, *etc.*, for example.

[0029] The local device may determine to partition the application into one or more tasks, and compare a processing time of a task at the local device to a latency requirement of the task. For example, the latency requirement of the task may be determined by the local device (*e.g.*, one or more of the functions described herein) and/or determined by the application. For example, the local device may obtain the latency requirement from the application (*e.g.*, service-level). The

latency requirement of the task may include an amount of time incurred to upload the task to a remote entity, an amount of time incurred to process the task at the remote entity, and/or an amount of time incurred to download the executed task once processed by the remote entity.

[0030] If the local device determines that the processing time of the local device would exceed the latency requirement of the task, then the local device may determine to offload the task. For example, the latency incurred once the local entity opts for offloading may include the total amount of time for the task upload, processing at the remote entity, and the time it takes to download the result processed at the remote entity. If the local device determines that the processing time of the local device would not exceed the latency requirement of the task, then the local device may compare a power consumed at the local device to offload the task to a power consumed at the local device to execute the task, where if the power consumed at the local device to offload the task is less than the power consumed (*e.g.*, while satisfying the latency and reliability requirements of the task) at the local device to execute the task, then the local device may determine to offload the task. And if the local device determines that the power consumed at the local device to offload the task is greater than the power consumed (*e.g.*, while satisfying the latency and reliability requirements of the task) at the local device to execute the task, then the local device may determine to execute the task locally.

[0031] FIG. 4 shows an example flow chart for task offloading. Service, system, and wireless medium information may provide service-level requirements for an application, wireless network conditions (*e.g.*, wireless medium information), and/or local device conditions (*e.g.*, computation resource information). The Code Partitioner may partition the application into one or more tasks. The Code Partitioner may have knowledge of the application context (*e.g.*, be context-aware), which for example, may be used to determine how to partition the application into the one or more tasks.

[0032] The local device may use the wireless network conditions and/or local device conditions to determine whether to offload the task, for example, as described herein. The Offloading Block may determine to execute the task locally or offload the task. A task that is not offloaded may be sent to the Executioner for local processing (*e.g.*, sent by the Code Partitioner).

[0033] The Offloading Block may receive inputs from the computation resource and/or wireless medium information blocks. The information blocks may be utilized in an offloading decision and/or system configuration. If the local device determines that remote execution is selected for the execution of the task, the Offloading Block may identify the system configuration and/or wireless transmission parameters (*e.g.*, using the algorithms for identifying the parameters, for example, as described herein). The task may be sent/offloaded to a remote

entity. The Offloading Block may receive the executed task from the remote entity. The output may be conveyed to the Executioner. In one or more examples, the task and the application execution flow may be combined (*e.g.*, by the Executioner).

[0034] The local device may perform adaptive code offloading and/or wireless medium error handling. Errors in the wireless medium (*e.g.*, due to fading or interference) may result in corrupted bit streams at the receiving end. In an offloading framework, erroneous reception of the offloaded tasks (*e.g.*, code partitions) may result in execution latency and/or the termination of the application or tasks. The local device may employ techniques to identify errors in the received bits/packets corresponding to the offloaded task. For example, the local device may implement a form of robust transmission in the wireless medium, which for example, may utilize an acknowledgement procedure for error handling. The remote entity may perform adaptive computation resource optimization.

[0035] The offloading framework may configure and/or utilize system parameters. An example offloading, which may include retransmission, may be initiated by the local device, where one or more of the following may be performed. An offloading block may receive a Task i (T_i) from the output of the code partitioner. The Task i may be in raw format. The Offloading Block may convert the raw format into an offloading format. The conversion may involve converting the task into packets, which may be composition of bits transmittable over the wireless medium. The Offloading Block may configure the transmission parameters based on system and/or medium related values.

[0036] The local device may consider the bit size of the task. b_i may represent the number of bits corresponding to Task i (T_i) in the offloading block.

[0037] An offloading indicator may be provided: $I_i \in (0,1)$, which may be the indicator for local (*e.g.*, $I_i=0$) or remote/offloaded (*e.g.*, $I_i=1$) execution of the task.

[0038] The local device may consider one or more wireless medium parameters. Examples of wireless medium parameters include, but are not limited to, channel state information (h_{lr}), packet error rate (PER), bit error rate (BER), and/or interference profile of the channel (h_{INT}).

[0039] The local device may consider computing resource parameters, such as processing cycle at the local device ($f_{loc,i}$) and/or remote entity ($f_{rem,i}$).

[0040] The local device may consider reliability requirements of the task (*e.g.*, the reliability of transmission for T_i in the wireless medium ($p_{i,req}$)). The reliability requirements may be selected from a pre-determined set such as $p_{i,req} \in \{p_{i,min}, \dots, p_{i,max}\}$. $p_{i,min}$ and $p_{i,max}$ may correspond to minimum and maximum reliability requirements, respectively. The total reliability may be a function of uplink ($p_{i,ul}$) and downlink wireless link reliabilities ($p_{i,dl}$).

[0041] The local device may consider the power consumption at the local device ($P_{loc,i}$). Such power consumption may include power consumed for uplink transmission (e.g., from device to remote entity) ($P_{ul,i}$), for downlink transmission (e.g., from remote entity to device) ($P_{dl,i}$), for local processing ($P_{loc,cpu,i}$), and/or for remote processing ($P_{rem,cpu,i}$).

[0042] The local device may consider the latency requirements of the task T_i ($L_{i,req}$). These requirements may be associated with the latency incurred in one or more of the following: the uplink ($L_{ul,i}$) transmission, the downlink transmission ($L_{dl,i}$), the time spent during the CPU processing at the local device ($L_{cpu,loc,i}$), and/or the time spent during the CPU processing at the remote entity ($L_{cpu,rem,i}$). The CPU processing times $L_{cpu,loc,i}$ and $L_{cpu,rem,i}$ may be related (e.g., directly related) to the CPU.

[0043] The offloading block may determine (e.g., based on system and wireless medium parameters) whether to process the task locally or to offload the task to the remote entity.

[0044] FIG. 5 shows an example flow chart for task offloading at an offloading block. The offloading block may check the latency requirement $L_{i,req}$ of the task T_i , and compare it with local processing capabilities. Based on the comparison outcome, the actions performed by the offloading block may vary, for example, as illustrated below.

[0045] In example *case A*, if the time required to execute the task locally is more than the latency requirement of the task (e.g., $L_{cpu,loc} > L_{i,req} \rightarrow I_i = 1$), the offloading block may activate (e.g., determine to offload the task). The task may be offloaded to the remote entity over the wireless medium. The offloading block may select a reliability requirement (e.g., a highest reliability requirement $t(\rho_{i,max})$, which may be associated with a highest level of reliability in the wireless medium). The offloading block may consider the uplink and downlink transmissions when configuring the transmission parameters (e.g., to meet the reliability requirement in the wireless transmission). The reliability constraint may be identified by the following adaptive coding abstraction. Parameters such as signal-to-noise ratio (SNR) and/or uplink/downlink channel state information (CSI) may be taken into account.

$$\rho_{i,ul} = \Pr(W \log(1 + h_{ul} P_{ul}) \geq b_i \setminus L_{ul})$$

$$\rho_{i,dl} = \Pr(W \log(1 + h_{dl} P_{dl}) \geq b_i \setminus L_{dl}) \quad \text{Eq. (1)}$$

$$\text{Select}(L_{ul}, L_{dl}, L_{cpu,rem}, P_{ul}) \quad \text{Eq. (2)}$$

such that $P_{loc,i} \rightarrow \min P_{loc,i}; L_{ul} + L_{dl} + L_{cpu,rem} \leq L_{i,req}; \rho_{i,ul}, \rho_{i,dl} \geq \rho_{i,max}$. The offloading block may determine one or more of the following: the uplink and/or downlink transmission

duration of the task; the computation time at the remote entity for the execution of the task; minimum uplink transmission power (*e.g.*, so that the parameters jointly satisfy the latency requirement ($L_{i,req}$); and/or that the uplink and downlink transmissions are under a reliability class ($\rho_{i,max}$). The parameters in Eq. (2) may be determined by the offloading block (*e.g.*, by taking the constraints in Eq. (1) and Eq. (2) into account). The reliability of the offloading may be a combination of the wireless medium transmission reliability (*e.g.*, as in Eq. (1)), and/or the service level reliability of the task T_i , $\rho_{SL,i}$. The overall reliability requirement in Eq. (2) may be a function of one or both of the aforementioned reliability requirements (*e.g.*, $\rho_{i,ul} \cdot \rho_{i,ul} \geq \max\{\rho_{i,max}, \rho_{SL,i}\}$). The parameters in Eq. (2) may be identified and/or updated based on the set of reliability constraints.

[0046] With respect to transmit configuration, the configuration set Eq. (2) determined by the offloading block (*e.g.*, P_{ul}) may be mapped to a corresponding MCS scheme at the local device uplink transmission. The MCS scheme may be denoted as MCS_{ul} . The bit stream of T_i may be encoded with MCS_{ul} . The bit stream may be transmitted over the wireless medium within a duration L_{ul} . The processed task T_i at the remote entity may be transmitted from the remote entity to the local device in the downlink. The time allocated for such transmission may be denoted as L_{dl} . This information may be conveyed from the local device to the remote entity (*e.g.*, via an offload configuration message) along with other parameters.

[0047] With respect to processing power configuration (*e.g.*, CPU configuration), the optimization of the parameters given in Eq. (2) may provide resource provisioning for the processing power. For example, the resource provisioning may be performed as a function of CPU power allocated for task T_i . The CPU power allocated at the remote entity, $L_{cpu,rem}$, may be the CPU time requested to be configured at the remote entity for T_i . The local device may send this parameter to the remote entity within the offload_config message, for example.

[0048] The remote entity may send information about the power (*e.g.*, CPU power) it could allocate for the task T_i to the local device, for example, in advance. The information may be sent through a configuration message (*e.g.*, the CPU_{rem} remote_resource message). The local device may use the information in identifying the corresponding system configuration parameters (*e.g.*, as in Eq. (2)). There may be an upper bound on the power (*e.g.*, CPU power) that the remote entity may allocate for task T_i . The configuration parameters of Eq. (2) may be determined by putting additional constraint on the remote entity power provision (*e.g.*, CPU provision).

[0049] In example *case B*, the local processor may be able to execute the task within the latency requirement (*e.g.*, $L_{cpu,loc} < L_{i,req} \rightarrow I_i = 0$ or 1). The task may be executed locally or

offloaded. The determination may be made upon considering the system and/or wireless medium conditions. For example, offloading of the task may be selected if the total power consumed at the local device is lower compared to executing the task locally at the device.

[0050] The system and/or wireless parameter configuration in example *case B* may be similar to that in example *case A*. The optimization framework may include the option for local execution. The optimization performed by the offloading block may be illustrated by the following.

$$\text{Select } (I_i, L_{ul}, L_{dl}, L_{cpu,loc}, L_{cpu,rem}, P_{ul}) \quad \text{Eq. (3)}$$

such that

$$\min\{(1 - I_i)P_{loc,cpu} + I_i P_{ul}\}; I_i(L_{ul} + L_{dl} + L_{cpu,rem}) + (1 - I_i)L_{cpu,loc}; \rho_{i,ul}, \rho_{i,dl} \geq \rho_{i,req}.$$

The offloading block may make one or more determinations including, for example, whether to offload the task (T_i), the uplink and/or downlink transmission duration of the task, local and remote computation time for the execution of the task, uplink transmission power, and/or minimum total power at the local device corresponding to T_i . The constraints of the parameters may jointly satisfy the latency and reliability requirements. The reliability of the offloading procedure, $\rho_{i,req}$, may be set-up by the service level requirement and/or the parameters (e.g., system and/or wireless medium parameters). If bit-error-rate (BER) of the wireless link between the local and remote devices is above a threshold (e.g., BER_{th}), the offloading block may select a high reliability requirement. If the BER is low, the offloading block may select a lower reliability requirement. The lower reliability may result in lower transmission time in wireless transmissions. Corresponding CPU allocation times at the local and/or remote devices may also be lower.

[0051] The transmit configuration and/or processing power (e.g., CPU) configuration may depend on the offloading decision parameter I_i . For example, if the value of I_i is 1, it may indicate that the offloading mechanism is enabled for the task and that transmit and remote CPU provisioning configurations may follow similar steps as discussed in the example *case A*. If the value of I_i is 0, it may indicate that local processing is enabled and that the task may be executed locally with the CPU provisioning of the task determined by $P_{loc,cpu}$, for example.

[0052] The offloading framework may enable a flexible assignment of reliability requirements. The assignment may be on a per task basis. The reliability requirement may be determined based on the service level requirement of the task, the reliability requirement of the wireless

medium, and/or a combination of both. Lower reliability constraint on the wireless transmission may incur error in the uplink and/or downlink transmission. An acknowledgement and retransmission procedure may be developed for the offloading framework.

[0053] The offloading framework may call for remote execution of the task T_i (e.g., $I_i=1$). An error handling mechanism may be implemented. The following mechanism may be used to trigger retransmission or local execution of the task, e.g., when an event or error is detected in the wireless channel between the local device and remote entity. The remote entity may send feedback to the local entity. The feedback may include wireless transmission parameters (e.g., updated latency requirements) and/or system parameters (e.g., power and/or CPU resources). The local device may determine whether to retransmit the tasks for remote execution (e.g., with updated CPU resource and/or wireless transmission parameter configuration) or to execute the tasks locally. The determination may be based on the requirements of the tasks including, for example, latency, reliability, and/or system parameters (e.g., wireless channel conditions such as CSI, local and remote computation resources, etc.). The local device may inform the remote entity regarding updated resource requests such as CPU allocation and/or transmission time. The error scenario in the uplink channel (e.g., the channel of transmission from the local device to remote entity) may be considered. NACK feedback due to erroneous reception may be conveyed in the downlink (e.g., from the remote entity to the local device).

[0054] An example retransmission implementation may be provided. The local device offloading block may transmit the task to the remote entity with one or more parameters including (I_i , L_{ul} , L_{dl} , $L_{cpu,loc}$, $L_{cpu,rem}$, P_{ul}). The parameters may be determined as described herein. If the offloading block opts for remote execution and sends the task over the wireless medium, the parameters may be selected as $I_i=1$ and $L_{cpu,loc}=0$. A timer corresponding to the task (denoted as L_{Ti}) may be started at the output of the task partitioner or when the whole task is transferred into the offloading block. L_{Ti} may determine the time consumed since the beginning of the task execution (including the offloading procedure). The time consumed may be within the latency constraint of the task, $L_{req,i}$.

[0055] The remote entity may check the received stream to identify error(s) incurred during the transmission. Reed-Solomon codes may be used to determine such errors. If the errors cannot be corrected in the decoding phase, the remote entity may send a NACK message to the local device. The NACK message may indicate the errors. The NACK message may include additional information about the computing resources at the remote device. The computer resources may be utilized in the next retransmission for the task. The NACK message may include one or more of the following: an NACK header/identifier, the CPU resources (e.g., in

terms of the duration that the remote entity may allocate in the next retransmission $L_{rem,itr}$), updated wireless channel measurement results (*e.g.*, CSI, interference level, BER, *etc.*).

[0056] The local device may receive the NACK packet. The local device may check the current timer, L_{Ti} , of the task T_i to identify the time left for execution within the task latency constraint (*e.g.*, whether $L_{Ti} < L_{req,i}$ or $L_{Ti} > L_{req,i}$). If L_{Ti} is less than $L_{req,i}$, the local device may initiate re-computation of one or more of the parameters ($I_i, L_{ul}, L_{dl}, L_{cpu,loc}, L_{cpu,rem}, P_{ul}$) with an updated time constraint $L_{ul} + L_{dl} + L_{cpu,rem} \leq L_{i,req} - L_{Ti}$. The updated time constraint may demonstrate the time for execution of the task that may satisfy the latency constraint. The parameter configuration framework given in Eq. (3) may be updated and performed as follows.

$$Select(I_i, L_{ul}, L_{dl}, L_{cpu,loc}, L_{cpu,rem}, P_{ul}) \quad \text{Eq. (4)}$$

such that $\min\{(1 - I_i)P_{loc,cpu} + I_i P_{ul}\}; I_i(L_{ul} + L_{dl} + L_{cpu,rem}) + (1 - I_i)L_{cpu,loc} \leq L_{i,req} - L_{Ti}; \rho_{i,ul}, \rho_{i,dl} \geq \rho_{i,max}$.

[0057] At a retransmission step (*e.g.*, at each retransmission step), the CPU allocation provisioning at the remote entity for the new offloading step may change. The change may be due to the time consumed during the previous uplink and downlink transmission. If an offloading decision is made at the offloading block after the NACK message reception, the local device may inform the remote entity with the updated remote CPU allocation request $L_{cpu,rem}^*$. The information may be part of the offloaded code. The information may be transmitted in a separate transmission slot before the remote execution of the task. The offloading block may run the configuration framework and determine the parameters.

[0058] FIG. 6 shows an example flow chart for task offloading at an offloading block with acknowledgment. A processing power (*e.g.*, CPU) configuration/acknowledgment implementation may be implemented for a retransmission. One or more of the following may apply. An initial configuration may be set as in Eq. (4). The local device may receive a NACK message. The local device may check the task timer, L_{Ti} . If offloading is enabled, the offloading block may inform the remote entity with a CPU power provisioning request (*e.g.*, via the updated upload_config message). In an option, the remote entity may send a feedback message to confirm or reject the request (*e.g.*, via a CPU_prov_ACK message). The offloading block may proceed with reconfiguring the transmission parameters based on the latest configuration framework. The offloading block may adaptively select reliability requirement for a retransmission (*e.g.*, each retransmission), $\rho_{i,req}$. The offloading block may update system

configuration parameters. The reliability requirement of a transmission (*e.g.*, each transmission) may be determined adaptively based on the service level requirement and/or the reliability request at the wireless medium given a set of configuration parameters (*e.g.*, $(I_i, L_{ul}, L_{dl}, P_{ul})$).

[0059] An example retransmission implementation for the downlink erroneous reception may be provided. The local device may detect errors in the message that may not be correctable. The local device may send a NACK to the remote entity. The remote entity may retransmit the executed task in the downlink with higher reliability in the downlink channel. In an option, the required reliability level for the downlink transmission may be provided to the remote entity by the local device. The local device may identify this parameter by considering the remaining time for the task execution: $L_{i,req} - L_{Ti}$. The reliability level may be obtained by assuming adaptive modulation in the downlink, which may be abstracted as follows.

$$\rho_{i,dl} = \Pr(W \log(1 + h_{dl}P_{dl}) \geq b_i \setminus L_{dl}^*) \quad \text{Eq. (4)}$$

L_{dl}^* may denote the updated downlink transmission time. The transmission time may be re-calculated based on the remaining task execution time $L_{i,req} - L_{Ti}$.

[0060] The offloading framework may identify the system configuration and/or wireless transmission parameters for the initial transmit opportunity and/or the retransmission opportunities. Optimization and configuration may be carried out as in Eq. (3). Retransmission may be performed due to potential errors in the wireless link. The framework may identify the parameters before an error event occurs in the wireless link. The offloading block may inform the remote entity about offloading characteristics during the task execution (*e.g.*, whether the task will be offloaded or executed locally, the requested remote CPU resources, downlink and uplink transmit durations, *etc.*) in various transmit opportunities. The parameters may be related to the potential wireless link errors. An average statistics may be determined at the local device and/or remote entity.

[0061] With a resource provision request during a transmit opportunity (*e.g.*, each transmit opportunity), the remote entity may provision its resources (*e.g.*, CPU power) and/or transmit the configuration in advance for resource configuration. For example, the resource provision request may be $(I_i, L_{ul}, L_{dl}, L_{cpu,loc}, L_{cpu,rem}, P_{ul})$ for an original transmission,

$(I_i^{rtx1}, L_{ul}^{rtx1}, L_{dl}^{rtx1}, L_{cpu,loc}^{rtx1}, L_{cpu,rem}^{rtx1}, P_{ul}^{rtx1})$ for retransmission opportunity 1, and/or

$(I_i^{rtxi}, L_{ul}^{rtxi}, L_{dl}^{rtxi}, L_{cpu,loc}^{rtxi}, L_{cpu,rem}^{rtxi}, P_{ul}^{rtxi})$ for retransmission opportunity i . The parameters for the provisioning of resources in a retransmission opportunity (*e.g.*, each retransmission

opportunity) may be computed. FIG. 7 shows an example flow chart for task offloading with system configuration parameter provisioning.

[0062] With respect to transmit opportunity 1 (e.g., an original transmission), (I_i , L_{ul} , L_{dl} , $L_{cpu,loc}$, $L_{cpu,rem}$, P_{ul}) may be selected with the constraint in Eq. (4). Local or remote execution (e.g., $I_i = 0$ or 1) may be identified. Feedback of the parameters $L_{cpu,rem}$, L_{dl} may be included. With respect to transmission opportunity 2 (e.g., a first retransmission opportunity), the retransmission may occur when offloading is performed in transmit opportunity 1 (e.g., $I_i=1$) and an error occurs in the uplink of the transmission, p_{e1} (e.g., p_{e1} may denote the probability of an error event in the uplink transmission). The error event may refer to packet error rate, a function of bit error rate, or a combination of both. The system configuration and/or wireless parameters may be determined by the offloading entity (e.g., assuming $I_i = 1$ and an uplink error event occurred) as the following.

$$Select \left(I_i^{rtx1}, L_{ul}^{rtx1}, L_{dl}^{rtx1}, L_{cpu,loc}^{rtx1}, L_{cpu,rem}^{rtx1}, P_{ul}^{rtx1} \right) \quad \text{Eq. (5)}$$

such that $\min\{(1 - I_i^{rtx1})P_{loc,cpu}^{rtx1} + I_i^{rtx1}P_{ul}^{rtx1}\}, I_i^{rtx1}(L_{ul}^{rtx1} + L_{dl}^{rtx1} + L_{cpu,rem}^{rtx1}) + (1 - I_i^{rtx1})L_{cpu,loc}^{rtx1} + L_{NACK} \leq L_{i,req} - L_{ToT1}; \rho_{i,ul}^{rtx1} \cdot \rho_{i,ul}^{rtx1} \geq \rho_{i,req}$.

L_{NACK} may refer to the time spent in successful reception of the NACK message at the local device. The NACK message may be sent by the remote entity. L_{ToT1} may refer to the total time spent in the first offloading transmit opportunity. The output of Eq. (6) may demonstrate whether offloading will be performed (e.g., $I_i=0$ or 1). The output of Eq. (6) may demonstrate the resources that will be requested at the remote entity, (L_{dl}^{rtx1} , $L_{cpu,rem}^{rtx1}$), if an error event occurs, for example.

[0063] The retransmission parameters (I_i^{rtx1} , L_{dl}^{rtx1} , $L_{cpu,rem}^{rtx1}$) and/or the probability of an error event in the uplink transmission, p_{e1} , may be included in a configuration message (e.g., the offload_config message). The configuration message may contain the parameters included in the original transmission (e.g., (I_i , L_{ul} , L_{dl} , $L_{cpu,loc}$, $L_{cpu,rem}$, P_{ul})). The message may be conveyed from the local device to the remote entity before the task is transmitted in the uplink channel. The remote entity may receive these parameters. The remote entity may provision its resources accordingly pertaining to transmit opportunity 1 and transmit opportunity 2. System configuration and/or wireless parameters may be obtained in an iterative fashion similar to the procedure described in Eq. (6). The system configuration and/or wireless parameters may include

$(I_i^{rtx2}, L_{ul}^{rtx2}, L_{dl}^{rtx2}, L_{cpu,loc}^{rtx2}, L_{cpu,rem}^{rtx2}, P_{ul}^{rtx2}), \dots, (I_i^{rtxk}, L_{ul}^{rtxk}, L_{dl}^{rtxi}, L_{cpu,loc}^{rtxk}, L_{cpu,rem}^{rtxk}, P_{ul}^{rtxk})$ for the retransmissions $rtx2, \dots, rtxk$, where $rtx2, \dots, rtxk$ may correspond to the transmission opportunities $2, \dots, k$, respectively.

[0064] FIG. 8A is a diagram of an example communications system 100 in which one or more disclosed embodiments may be implemented. The communications system 100 may be a multiple access system that provides content, such as voice, data, video, messaging, broadcast, *etc.*, to multiple wireless users. The communications system 100 may enable multiple wireless users to access such content through the sharing of system resources, including wireless bandwidth. For example, the communications systems 100 may employ one or more channel access methods, such as code division multiple access (CDMA), time division multiple access (TDMA), frequency division multiple access (FDMA), orthogonal FDMA (OFDMA), single-carrier FDMA (SC-FDMA), and the like.

[0065] As shown in FIG. 8A, the communications system 100 may include wireless transmit/receive units (WTRUs) 102a, 102b, 102c and/or 102d (which generally or collectively may be referred to as WTRU 102), a radio access network (RAN) 103/104/105, a core network 106/107/109, a public switched telephone network (PSTN) 108, the Internet 110, and other networks 112, though it will be appreciated that the disclosed embodiments contemplate any number of WTRUs, base stations, networks, and/or network elements. Each of the WTRUs 102a, 102b, 102c, 102d may be any type of device configured to operate and/or communicate in a wireless environment. By way of example, the WTRUs 102a, 102b, 102c, 102d may be configured to transmit and/or receive wireless signals and may include user equipment (UE), a mobile station, a fixed or mobile subscriber unit, a pager, a cellular telephone, a personal digital assistant (PDA), a smartphone, a laptop, a netbook, a personal computer, a wireless sensor, consumer electronics, and the like. A DASH client may be run on a WTRU.

[0066] The communications systems 100 may also include a base station 114a and a base station 114b. Each of the base stations 114a, 114b may be any type of device configured to wirelessly interface with at least one of the WTRUs 102a, 102b, 102c, 102d to facilitate access to one or more communication networks, such as the core network 106/107/109, the Internet 110, and/or the networks 112. By way of example, the base stations 114a, 114b may be a base transceiver station (BTS), a Node-B, an eNode B, a Home Node B, a Home eNode B, a site controller, an access point (AP), a wireless router, and the like. While the base stations 114a, 114b are each depicted as a single element, it will be appreciated that the base stations 114a, 114b may include any number of interconnected base stations and/or network elements.

[0067] The base station 114a may be part of the RAN 103/104/105, which may also include other base stations and/or network elements (not shown), such as a base station controller (BSC), a radio network controller (RNC), relay nodes, *etc.* The base station 114a and/or the base station 114b may be configured to transmit and/or receive wireless signals within a particular geographic region, which may be referred to as a cell (not shown). The cell may further be divided into cell sectors. For example, the cell associated with the base station 114a may be divided into three sectors. Thus, in one embodiment, the base station 114a may include three transceivers, i.e., one for each sector of the cell. In an embodiment, the base station 114a may employ multiple-input multiple output (MIMO) technology and, therefore, may utilize multiple transceivers for each sector of the cell.

[0068] The base stations 114a, 114b may communicate with one or more of the WTRUs 102a, 102b, 102c, 102d over an air interface 115/116/117, which may be any suitable wireless communication link (*e.g.*, radio frequency (RF), microwave, infrared (IR), ultraviolet (UV), visible light, *etc.*). The air interface 115/116/117 may be established using any suitable radio access technology (RAT).

[0069] More specifically, as noted above, the communications system 100 may be a multiple access system and may employ one or more channel access schemes, such as CDMA, TDMA, FDMA, OFDMA, SC-FDMA, and the like. For example, the base station 114a in the RAN 103/104/105 and the WTRUs 102a, 102b, 102c may implement a radio technology such as Universal Mobile Telecommunications System (UMTS) Terrestrial Radio Access (UTRA), which may establish the air interface 115/116/117 using wideband CDMA (WCDMA). WCDMA may include communication protocols such as High-Speed Packet Access (HSPA) and/or Evolved HSPA (HSPA+). HSPA may include High-Speed Downlink Packet Access (HSDPA) and/or High-Speed Uplink Packet Access (HSUPA).

[0070] In an embodiment, the base station 114a and the WTRUs 102a, 102b, 102c may implement a radio technology such as Evolved UMTS Terrestrial Radio Access (E-UTRA), which may establish the air interface 115/116/117 using Long Term Evolution (LTE) and/or LTE-Advanced (LTE-A).

[0071] In other embodiments, the base station 114a and the WTRUs 102a, 102b, 102c may implement radio technologies such as IEEE 802.16 (*i.e.*, Worldwide Interoperability for Microwave Access (WiMAX)), CDMA2000, CDMA2000 1X, CDMA2000 EV-DO, Interim Standard 2000 (IS-2000), Interim Standard 95 (IS-95), Interim Standard 856 (IS-856), Global System for Mobile communications (GSM), Enhanced Data rates for GSM Evolution (EDGE), GSM EDGE (GERAN), and the like.

[0072] The base station 114b in FIG. 8A may be a wireless router, Home Node B, Home eNode B, or access point, for example, and may utilize any suitable RAT for facilitating wireless connectivity in a localized area, such as a place of business, a home, a vehicle, a campus, and the like. In one embodiment, the base station 114b and the WTRUs 102c, 102d may implement a radio technology such as IEEE 802.11 to establish a wireless local area network (WLAN). In an embodiment, the base station 114b and the WTRUs 102c, 102d may implement a radio technology such as IEEE 802.15 to establish a wireless personal area network (WPAN). In yet an embodiment, the base station 114b and the WTRUs 102c, 102d may utilize a cellular-based RAT (*e.g.*, WCDMA, CDMA2000, GSM, LTE, LTE-A) to establish a picocell or femtocell. As shown in FIG. 8A, the base station 114b may have a direct connection to the Internet 110. Thus, the base station 114b may not be required to access the Internet 110 via the core network 106/107/109.

[0073] The RAN 103/104/105 may be in communication with the core network 106/107/109, which may be any type of network configured to provide voice, data, applications, and/or voice over internet protocol (VoIP) services to one or more of the WTRUs 102a, 102b, 102c, 102d. For example, the core network 106/107/109 may provide call control, billing services, mobile location-based services, pre-paid calling, Internet connectivity, video distribution, *etc.*, and/or perform high-level security functions, such as user authentication. Although not shown in FIG. 8A, it will be appreciated that the RAN 103/104/105 and/or the core network 106/107/109 may be in direct or indirect communication with other RANs that employ the same RAT as the RAN 103/104/105 or a different RAT. For example, in addition to being connected to the RAN 103/104/105, which may be utilizing an E-UTRA radio technology, the core network 106/107/109 may also be in communication with another RAN (not shown) employing a GSM radio technology.

[0074] The core network 106/107/109 may also serve as a gateway for the WTRUs 102a, 102b, 102c, 102d to access the PSTN 108, the Internet 110, and/or other networks 112. The PSTN 108 may include circuit-switched telephone networks that provide plain old telephone service (POTS). The Internet 110 may include a global system of interconnected computer networks and devices that use common communication protocols, such as the transmission control protocol (TCP), user datagram protocol (UDP) and the internet protocol (IP) in the TCP/IP internet protocol suite. The networks 112 may include wired or wireless communications networks owned and/or operated by other service providers. For example, the networks 112 may include another core network connected to one or more RANs, which may employ the same RAT as the RAN 103/104/105 or a different RAT.

[0075] Some or all of the WTRUs 102a, 102b, 102c, 102d in the communications system 100 may include multi-mode capabilities, i.e., the WTRUs 102a, 102b, 102c, 102d may include multiple transceivers for communicating with different wireless networks over different wireless links. For example, the WTRU 102c shown in FIG. 8A may be configured to communicate with the base station 114a, which may employ a cellular-based radio technology, and with the base station 114b, which may employ an IEEE 802 radio technology.

[0076] FIG. 8B is a system diagram of an example WTRU 102. As shown in FIG. 8B, the WTRU 102 may include a processor 118, a transceiver 120, a transmit/receive element 122, a speaker/microphone 124, a keypad 126, a display/touchpad 128, non-removable memory 130, removable memory 132, a power source 134, a global positioning system (GPS) chipset 136, and other peripherals 138. It will be appreciated that the WTRU 102 may include any sub-combination of the foregoing elements while remaining consistent with an embodiment. Also, embodiments contemplate that the base stations 114a and 114b, and/or the nodes that base stations 114a and 114b may represent, such as but not limited to transceiver station (BTS), a Node-B, a site controller, an access point (AP), a home node-B, an evolved home node-B (eNodeB), a home evolved node-B (HeNB or HeNodeB), a home evolved node-B gateway, and proxy nodes, among others, may include some or all of the elements depicted in FIG. 8B and described herein.

[0077] The processor 118 may be a general purpose processor, a special purpose processor, a conventional processor, a digital signal processor (DSP), a plurality of microprocessors, one or more microprocessors in association with a DSP core, a controller, a microcontroller, Application Specific Integrated Circuits (ASICs), Field Programmable Gate Array (FPGAs) circuits, any other type of integrated circuit (IC), a state machine, and the like. The processor 118 may perform signal coding, data processing, power control, input/output processing, and/or any other functionality that enables the WTRU 102 to operate in a wireless environment. The processor 118 may be coupled to the transceiver 120, which may be coupled to the transmit/receive element 122. While FIG. 8B depicts the processor 118 and the transceiver 120 as separate components, it will be appreciated that the processor 118 and the transceiver 120 may be integrated together in an electronic package or chip.

[0078] The transmit/receive element 122 may be configured to transmit signals to, or receive signals from, a base station (e.g., base station 114a) over the air interface 115/116/117. For example, in one embodiment, the transmit/receive element 122 may be an antenna configured to transmit and/or receive RF signals. In another embodiment, the transmit/receive element 122 may be an emitter/detector configured to transmit and/or receive IR, UV, or visible light signals,

for example. In yet another embodiment, the transmit/receive element 122 may be configured to transmit and receive both RF and light signals. It will be appreciated that the transmit/receive element 122 may be configured to transmit and/or receive any combination of wireless signals.

[0079] In addition, although the transmit/receive element 122 is depicted in FIG. 8B as a single element, the WTRU 102 may include any number of transmit/receive elements 122. More specifically, the WTRU 102 may employ MIMO technology. Thus, in one embodiment, the WTRU 102 may include two or more transmit/receive elements 122 (*e.g.*, multiple antennas) for transmitting and receiving wireless signals over the air interface 115/116/117.

[0080] The transceiver 120 may be configured to modulate the signals that are to be transmitted by the transmit/receive element 122 and to demodulate the signals that are received by the transmit/receive element 122. As noted above, the WTRU 102 may have multi-mode capabilities. Thus, the transceiver 120 may include multiple transceivers for enabling the WTRU 102 to communicate via multiple RATs, such as UTRA and IEEE 802.11, for example.

[0081] The processor 118 of the WTRU 102 may be coupled to, and may receive user input data from, the speaker/microphone 124, the keypad 126, and/or the display/touchpad 128 (*e.g.*, a liquid crystal display (LCD) display unit or organic light-emitting diode (OLED) display unit). The processor 118 may also output user data to the speaker/microphone 124, the keypad 126, and/or the display/touchpad 128. In addition, the processor 118 may access information from, and store data in, any type of suitable memory, such as the non-removable memory 130 and/or the removable memory 132. The non-removable memory 130 may include random-access memory (RAM), read-only memory (ROM), a hard disk, or any other type of memory storage device. The removable memory 132 may include a subscriber identity module (SIM) card, a memory stick, a secure digital (SD) memory card, and the like. In other embodiments, the processor 118 may access information from, and store data in, memory that is not physically located on the WTRU 102, such as on a server or a home computer (not shown).

[0082] The processor 118 may receive power from the power source 134, and may be configured to distribute and/or control the power to the other components in the WTRU 102. The power source 134 may be any suitable device for powering the WTRU 102. For example, the power source 134 may include one or more dry cell batteries (*e.g.*, nickel-cadmium (NiCd), nickel-zinc (NiZn), nickel metal hydride (NiMH), lithium-ion (Li-ion), *etc.*), solar cells, fuel cells, and the like.

[0083] The processor 118 may also be coupled to the GPS chipset 136, which may be configured to provide location information (*e.g.*, longitude and latitude) regarding the current location of the WTRU 102. In addition to, or in lieu of, the information from the GPS chipset

136, the WTRU 102 may receive location information over the air interface 115/116/117 from a base station (*e.g.*, base stations 114a, 114b) and/or determine its location based on the timing of the signals being received from two or more nearby base stations. It will be appreciated that the WTRU 102 may acquire location information by way of any suitable location-determination implementation while remaining consistent with an embodiment.

[0084] The processor 118 may further be coupled to other peripherals 138, which may include one or more software and/or hardware modules that provide additional features, functionality and/or wired or wireless connectivity. For example, the peripherals 138 may include an accelerometer, an e-compass, a satellite transceiver, a digital camera (for photographs or video), a universal serial bus (USB) port, a vibration device, a television transceiver, a hands free headset, a Bluetooth® module, a frequency modulated (FM) radio unit, a digital music player, a media player, a video game player module, an Internet browser, and the like.

[0085] FIG. 8C is a system diagram of the RAN 103 and the core network 106 according to an embodiment. The RAN 103 may employ a UTRA radio technology to communicate with the WTRUs 102a, 102b, 102c over the air interface 115. The RAN 103 may also be in communication with the core network 106. As shown in FIG. 8C, the RAN 103 may include Node-Bs 140a, 140b, 140c, which may each include one or more transceivers for communicating with the WTRUs 102a, 102b, 102c over the air interface 115. The Node-Bs 140a, 140b, 140c may each be associated with a particular cell (not shown) within the RAN 103. The RAN 103 may also include RNCs 142a, 142b. It will be appreciated that the RAN 103 may include any number of Node-Bs and RNCs while remaining consistent with an embodiment.

[0086] As shown in FIG. 8C, the Node-Bs 140a, 140b may be in communication with the RNC 142a. Additionally, the Node-B 140c may be in communication with the RNC 142b. The Node-Bs 140a, 140b, 140c may communicate with the respective RNCs 142a, 142b via an Iub interface. The RNCs 142a, 142b may be in communication with one another via an Iur interface. Each of the RNCs 142a, 142b may be configured to control the respective Node-Bs 140a, 140b, 140c to which it is connected. In addition, each of the RNCs 142a, 142b may be configured to carry out or support other functionality, such as outer loop power control, load control, admission control, packet scheduling, handover control, macrodiversity, security functions, data encryption, and the like.

[0087] The core network 106 shown in FIG. 8C may include a media gateway (MGW) 144, a mobile switching center (MSC) 146, a serving GPRS support node (SGSN) 148, and/or a gateway GPRS support node (GGSN) 150. While each of the foregoing elements are depicted as part of the core network 106, it will be appreciated that any one of these elements may be owned

and/or operated by an entity other than the core network operator.

[0088] The RNC 142a in the RAN 103 may be connected to the MSC 146 in the core network 106 via an IuCS interface. The MSC 146 may be connected to the MGW 144. The MSC 146 and the MGW 144 may provide the WTRUs 102a, 102b, 102c with access to circuit-switched networks, such as the PSTN 108, to facilitate communications between the WTRUs 102a, 102b, 102c and traditional land-line communications devices.

[0089] The RNC 142a in the RAN 103 may also be connected to the SGSN 148 in the core network 106 via an IuPS interface. The SGSN 148 may be connected to the GGSN 150. The SGSN 148 and the GGSN 150 may provide the WTRUs 102a, 102b, 102c with access to packet-switched networks, such as the Internet 110, to facilitate communications between and the WTRUs 102a, 102b, 102c and IP-enabled devices.

[0090] As noted above, the core network 106 may also be connected to the networks 112, which may include other wired or wireless networks that are owned and/or operated by other service providers.

[0091] FIG. 8D is a system diagram of the RAN 104 and the core network 107 according to an embodiment. As noted above, the RAN 104 may employ an E-UTRA radio technology to communicate with the WTRUs 102a, 102b, 102c over the air interface 116. The RAN 104 may also be in communication with the core network 107.

[0092] The RAN 104 may include eNode-Bs 160a, 160b, 160c, though it will be appreciated that the RAN 104 may include any number of eNode-Bs while remaining consistent with an embodiment. The eNode-Bs 160a, 160b, 160c may each include one or more transceivers for communicating with the WTRUs 102a, 102b, 102c over the air interface 116. In one embodiment, the eNode-Bs 160a, 160b, 160c may implement MIMO technology. Thus, the eNode-B 160a, for example, may use multiple antennas to transmit wireless signals to, and receive wireless signals from, the WTRU 102a.

[0093] Each of the eNode-Bs 160a, 160b, 160c may be associated with a particular cell (not shown) and may be configured to handle radio resource management decisions, handover decisions, scheduling of users in the uplink and/or downlink, and the like. As shown in FIG. 8D, the eNode-Bs 160a, 160b, 160c may communicate with one another over an X2 interface.

[0094] The core network 107 shown in FIG. 8D may include a mobility management gateway (MME) 162, a serving gateway 164, and a packet data network (PDN) gateway 166. While each of the foregoing elements are depicted as part of the core network 107, it will be appreciated that any one of these elements may be owned and/or operated by an entity other than the core network operator.

[0095] The MME 162 may be connected to each of the eNode-Bs 160a, 160b, 160c in the RAN 104 via an S1 interface and may serve as a control node. For example, the MME 162 may be responsible for authenticating users of the WTRUs 102a, 102b, 102c, bearer activation/deactivation, selecting a particular serving gateway during an initial attach of the WTRUs 102a, 102b, 102c, and the like. The MME 162 may also provide a control plane function for switching between the RAN 104 and other RANs (not shown) that employ other radio technologies, such as GSM or WCDMA.

[0096] The serving gateway 164 may be connected to each of the eNode-Bs 160a, 160b, 160c in the RAN 104 via the S1 interface. The serving gateway 164 may generally route and forward user data packets to/from the WTRUs 102a, 102b, 102c. The serving gateway 164 may also perform other functions, such as anchoring user planes during inter-eNode B handovers, triggering paging when downlink data is available for the WTRUs 102a, 102b, 102c, managing and storing contexts of the WTRUs 102a, 102b, 102c, and the like.

[0097] The serving gateway 164 may also be connected to the PDN gateway 166, which may provide the WTRUs 102a, 102b, 102c with access to packet-switched networks, such as the Internet 110, to facilitate communications between the WTRUs 102a, 102b, 102c and IP-enabled devices.

[0098] The core network 107 may facilitate communications with other networks. For example, the core network 107 may provide the WTRUs 102a, 102b, 102c with access to circuit-switched networks, such as the PSTN 108, to facilitate communications between the WTRUs 102a, 102b, 102c and traditional land-line communications devices. For example, the core network 107 may include, or may communicate with, an IP gateway (*e.g.*, an IP multimedia subsystem (IMS) server) that serves as an interface between the core network 107 and the PSTN 108. In addition, the core network 107 may provide the WTRUs 102a, 102b, 102c with access to the networks 112, which may include other wired or wireless networks that are owned and/or operated by other service providers.

[0099] FIG. 8E is a system diagram of the RAN 105 and the core network 109 according to an embodiment. The RAN 105 may be an access service network (ASN) that employs IEEE 802.16 radio technology to communicate with the WTRUs 102a, 102b, 102c over the air interface 117. As will be further discussed below, the communication links between the different functional entities of the WTRUs 102a, 102b, 102c, the RAN 105, and the core network 109 may be defined as reference points.

[0100] As shown in FIG. 8E, the RAN 105 may include base stations 180a, 180b, 180c, and an ASN gateway 182, though it will be appreciated that the RAN 105 may include any number

of base stations and ASN gateways while remaining consistent with an embodiment. The base stations 180a, 180b, 180c may each be associated with a particular cell (not shown) in the RAN 105 and may each include one or more transceivers for communicating with the WTRUs 102a, 102b, 102c over the air interface 117. In one embodiment, the base stations 180a, 180b, 180c may implement MIMO technology. Thus, the base station 180a, for example, may use multiple antennas to transmit wireless signals to, and receive wireless signals from, the WTRU 102a. The base stations 180a, 180b, 180c may also provide mobility management functions, such as handoff triggering, tunnel establishment, radio resource management, traffic classification, quality of service (QoS) policy enforcement, and the like. The ASN gateway 182 may serve as a traffic aggregation point and may be responsible for paging, caching of subscriber profiles, routing to the core network 109, and the like.

[0101] The air interface 117 between the WTRUs 102a, 102b, 102c and the RAN 105 may be defined as an R1 reference point that implements the IEEE 802.16 specification. In addition, each of the WTRUs 102a, 102b, 102c may establish a logical interface (not shown) with the core network 109. The logical interface between the WTRUs 102a, 102b, 102c and the core network 109 may be defined as an R2 reference point, which may be used for authentication, authorization, IP host configuration management, and/or mobility management.

[0102] The communication link between each of the base stations 180a, 180b, 180c may be defined as an R8 reference point that includes protocols for facilitating WTRU handovers and the transfer of data between base stations. The communication link between the base stations 180a, 180b, 180c and the ASN gateway 182 may be defined as an R6 reference point. The R6 reference point may include protocols for facilitating mobility management based on mobility events associated with each of the WTRUs 102a, 102b, 102c.

[0103] As shown in FIG. 8E, the RAN 105 may be connected to the core network 109. The communication link between the RAN 105 and the core network 109 may be defined as an R3 reference point that includes protocols for facilitating data transfer and mobility management capabilities, for example. The core network 109 may include a mobile IP home agent (MIP-HA) 184, an authentication, authorization, accounting (AAA) server 186, and a gateway 188. While each of the foregoing elements are depicted as part of the core network 109, it will be appreciated that any one of these elements may be owned and/or operated by an entity other than the core network operator.

[0104] The MIP-HA may be responsible for IP address management, and may enable the WTRUs 102a, 102b, 102c to roam between different ASNs and/or different core networks. The MIP-HA 184 may provide the WTRUs 102a, 102b, 102c with access to packet-switched

networks, such as the Internet 110, to facilitate communications between the WTRUs 102a, 102b, 102c and IP-enabled devices. The AAA server 186 may be responsible for user authentication and for supporting user services. The gateway 188 may facilitate interworking with other networks. For example, the gateway 188 may provide the WTRUs 102a, 102b, 102c with access to circuit-switched networks, such as the PSTN 108, to facilitate communications between the WTRUs 102a, 102b, 102c and traditional land-line communications devices. In addition, the gateway 188 may provide the WTRUs 102a, 102b, 102c with access to the networks 112, which may include other wired or wireless networks that are owned and/or operated by other service providers.

[0105] Although not shown in FIG. 8E, it will be appreciated that the RAN 105 may be connected to other ASNs and the core network 109 may be connected to other core networks. The communication link between the RAN 105 and the other ASNs may be defined as an R4 reference point, which may include protocols for coordinating the mobility of the WTRUs 102a, 102b, 102c between the RAN 105 and the other ASNs. The communication link between the core network 109 and the other core networks may be defined as an R5 reference, which may include protocols for facilitating interworking between home core networks and visited core networks.

[0106] The processes described above may be implemented in a computer program, software, and/or firmware incorporated in a computer-readable medium for execution by a computer and/or processor. Examples of computer-readable media include, but are not limited to, electronic signals (transmitted over wired and/or wireless connections) and/or computer-readable storage media. Examples of computer-readable storage media include, but are not limited to, a read only memory (ROM), a random access memory (RAM), a register, cache memory, semiconductor memory devices, magnetic media such as, but not limited to, internal hard disks and removable disks, magneto-optical media, and/or optical media such as CD-ROM disks, and/or digital versatile disks (DVDs). A processor in association with software may be used to implement a radio frequency transceiver for use in a WTRU, UE, terminal, base station, RNC, and/or any host computer.

CLAIMS

1. A method of offloading computation for an application on a local device in a wireless network, comprising:
 - determining service-level requirements for the application, wherein the service-level requirements comprise one or more of latency, reliability, and power consumption for processing;
 - determining wireless network conditions;
 - determining local device conditions;
 - partitioning the application into one or more tasks;
 - comparing a processing time of a task at the local device to a latency requirement of the task, and:
 - if the processing time of the local device would exceed the latency requirement of the task, determining to offload the task; and
 - if the processing time of the local device would not exceed the latency requirement of the task, comparing a power consumed at the local device to offload the task to a power consumed at the local device to execute the task, wherein if the power consumed at the local device to offload the task is less than the power consumed at the local device to execute the task, determining to offload the task.
2. The method of claim 1, further comprising provisioning processing power to execute the task.
3. The method of claim 1, wherein a latency incurred with offloading the task comprises an amount of time incurred to upload the task to a remote entity, an amount of time incurred to process the task at the remote entity, and an amount of time incurred to download the executed task once processed by the remote entity.
4. The method of claim 2, wherein the task is executed locally, further comprising allocating a portion of the local device's central processing unit (CPU) power for the task.
5. The method of claim 2, wherein the task is offloaded, further comprising allocating a portion of a remote entity's CPU power for the task.

6. The method of claim 5, further comprising informing the remote entity regarding the CPU power allocation.
7. The method of claim 1, further comprising configuring wireless transmission parameters for offloading the task.
8. The method of claim 7, further comprising informing the remote entity regarding the wireless transmission parameters.
9. The method of claim 1, further comprising converting the task into packets to be transmitted over the wireless medium for offloading the task.
10. The method of claim 1, further comprising selecting a reliability requirement for offloading based upon wireless network conditions.
11. The method of claim 10, further comprising configuring the CPU power and wireless transmission parameters at the local device and the remote entity based on the selected reliability requirements.
12. The method of claim 1 or 10, further comprising selecting a reliability requirement for offloading based upon service-level requirements for the application.
13. The method of claim 12, further comprising configuring the CPU power and wireless transmission parameters at the local device and the remote entity based on the selected reliability requirements.
14. The method of claim 1, wherein the task is offloaded, further comprising transmitting the task to a remote entity, and receiving from the remote entity a NACK message in the event that potential errors are incurred during the transmission.
15. The method of claim 14, wherein the local device re-configures CPU power allocation, wireless transmission, and reliability requirements parameters at the local device and remote entity upon receipt of a NACK message for the task.

16. The method of claim 15, wherein the local device informs the remote entity regarding the updated re-configuration parameters.

17. The method of claim 1, wherein the task is offloaded, further comprising receiving an executed task from the remote entity and collating the executed task with one or more other executed tasks of the application.

18. A WTRU in a wireless network, comprising:

a memory for storing an application; and

a processor configured to:

determine service-level requirements for the application, wherein the service-level requirements comprise one or more of latency, reliability, and power consumption for processing;

determine wireless network conditions;

determine WTRU conditions;

partition the application into one or more tasks;

compare a processing time of a task at the WTRU to a latency requirement of the task, and:

if the processing time of the WTRU would exceed the latency requirement of the task, determining to offload the task; and

if the processing time of the WTRU would not exceed the latency requirement of the task, comparing a power consumed at the WTRU to offload the task to a power consumed at the WTRU to execute the task, wherein if the power consumed at the WTRU to offload the task is less than the power consumed at the WTRU to execute the task, determining to offload the task.

19. The WTRU of claim 18, wherein the processor is further configured to provision processing power to execute the task.

20. The WTRU of claim 18, wherein a latency incurred with offloading the task comprises an amount of time incurred to upload the task to a remote entity, an amount of time incurred to process the task at the remote entity, and an amount of time incurred to download the executed task once processed by the remote entity.

21. The WTRU of claim 19, wherein the task is executed locally, wherein the processor is further configured to allocate a portion of the WTRU's central processing unit (CPU) power for the task.
22. The WTRU of claim 19, wherein the task is offloaded, wherein the processor is further configured to allocate a portion of a remote entity's CPU power for the task.
23. The WTRU of claim 22, wherein the processor is further configured to inform the remote entity regarding the CPU power allocation.
24. The WTRU of claim 18, wherein the processor is further configured to configure wireless transmission parameters for offloading the task.
25. The WTRU of claim 24, wherein the processor is further configured to inform the remote entity regarding the wireless transmission parameters.
26. The WTRU of claim 18, wherein the processor is further configured to convert the task into packets to be transmitted over the wireless medium for offloading the task.
27. The WTRU of claim 18, wherein the processor is further configured to select a reliability requirement for offloading based upon wireless network conditions.
28. The WTRU of claim 27, wherein the processor is further configured to configure the CPU power and wireless transmission parameters at the WTRU and the remote entity based on the selected reliability requirements.
29. The WTRU of claim 18 or 28, wherein the processor is further configured to select a reliability requirement for offloading based upon service-level requirements for the application.
30. The WTRU of claim 29, wherein the processor is further configured to configure the CPU power and wireless transmission parameters at the WTRU and the remote entity based on the selected reliability requirements.

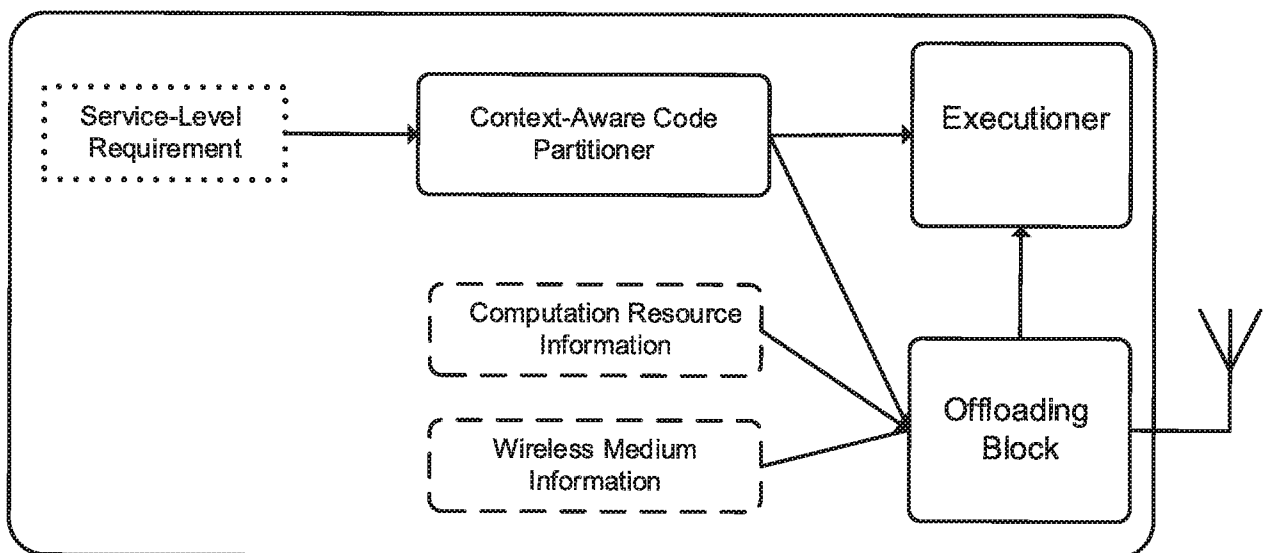
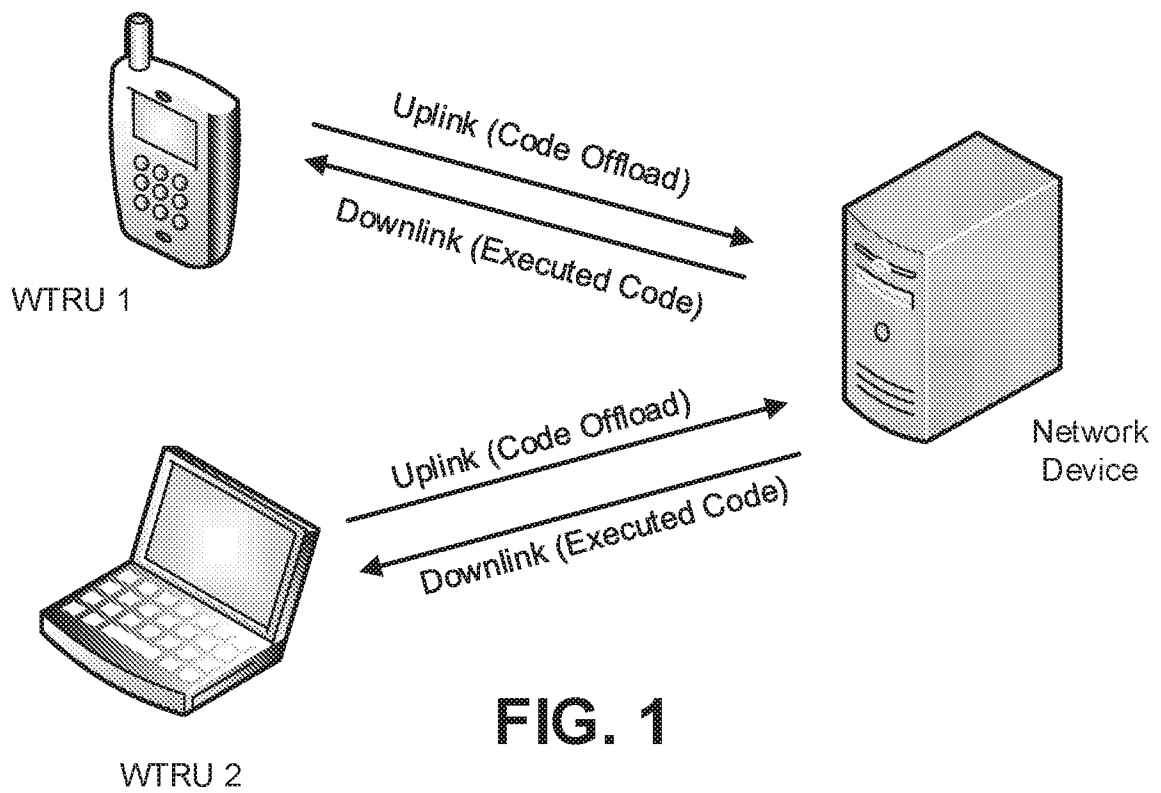
31. The WTRU of claim 18, wherein the task is offloaded, wherein the processor is further configured to transmit the task to a remote entity, and receive from the remote entity a NACK message in the event that potential errors are incurred during the transmission.

32. The WTRU of claim 31, wherein the processor is further configured to re-configure CPU power allocation, wireless transmission, and reliability requirements parameters at the WTRU and remote entity upon receipt of a NACK message for the task.

33. The WTRU of claim 32, wherein the processor is further configured to inform the remote entity regarding the updated re-configuration parameters.

34. The WTRU of claim 18, wherein the task is offloaded, wherein the processor is further configured to receive an executed task from the remote entity and collate the executed task with one or more other executed tasks of the application.

1/11



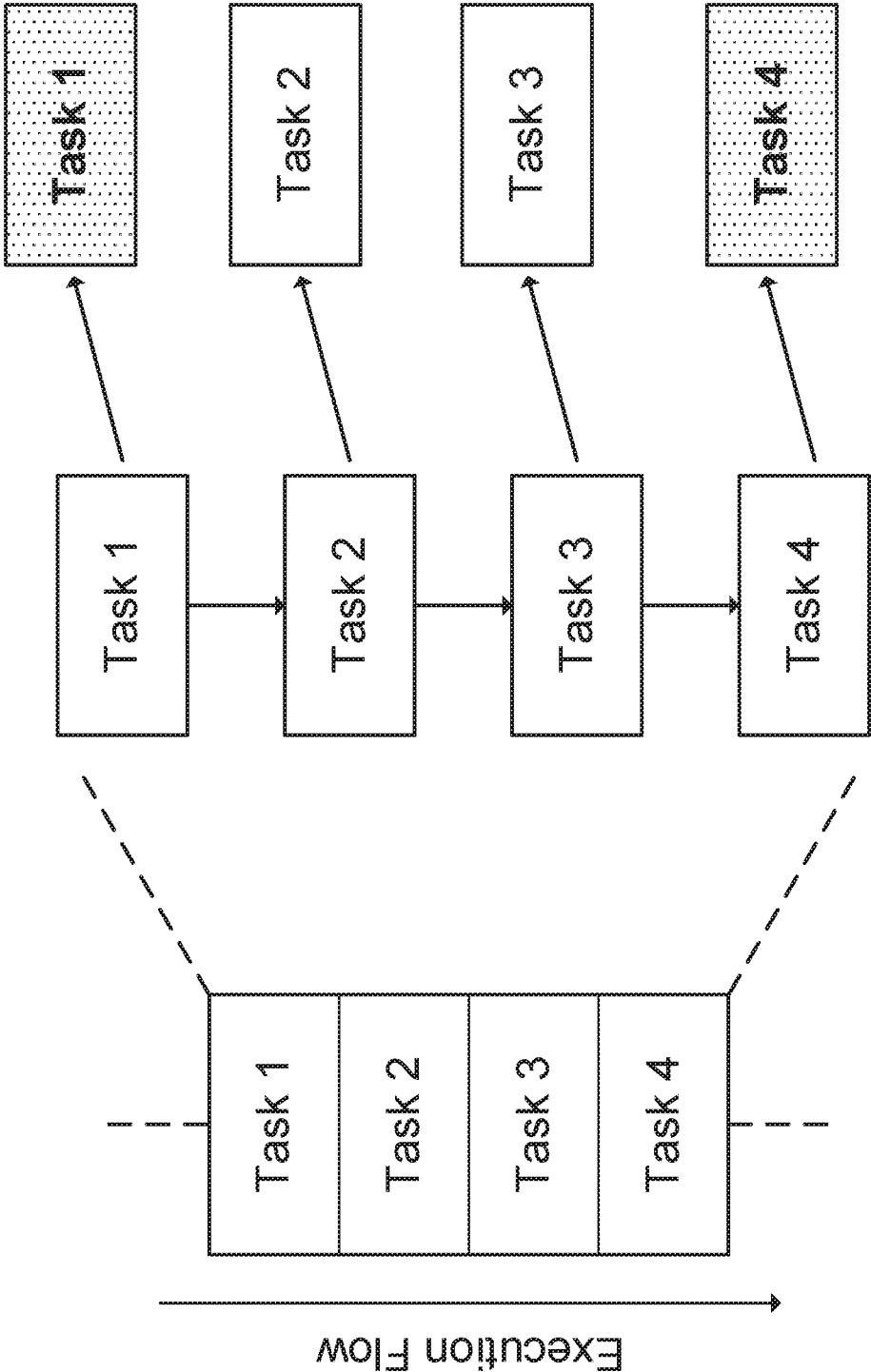
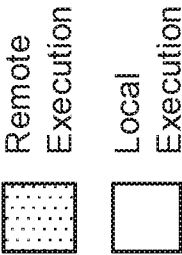


FIG. 2

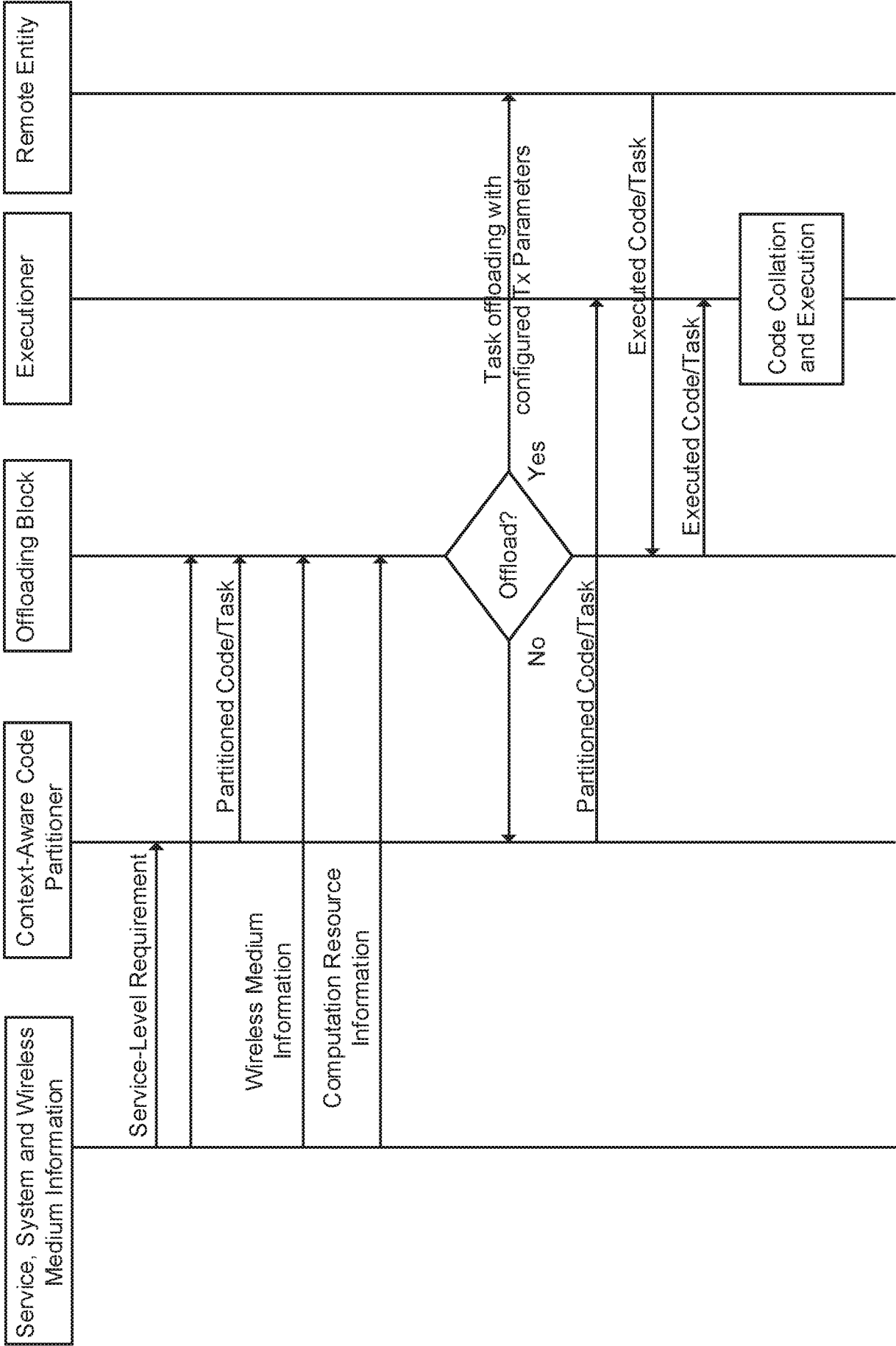


FIG. 4

4/11

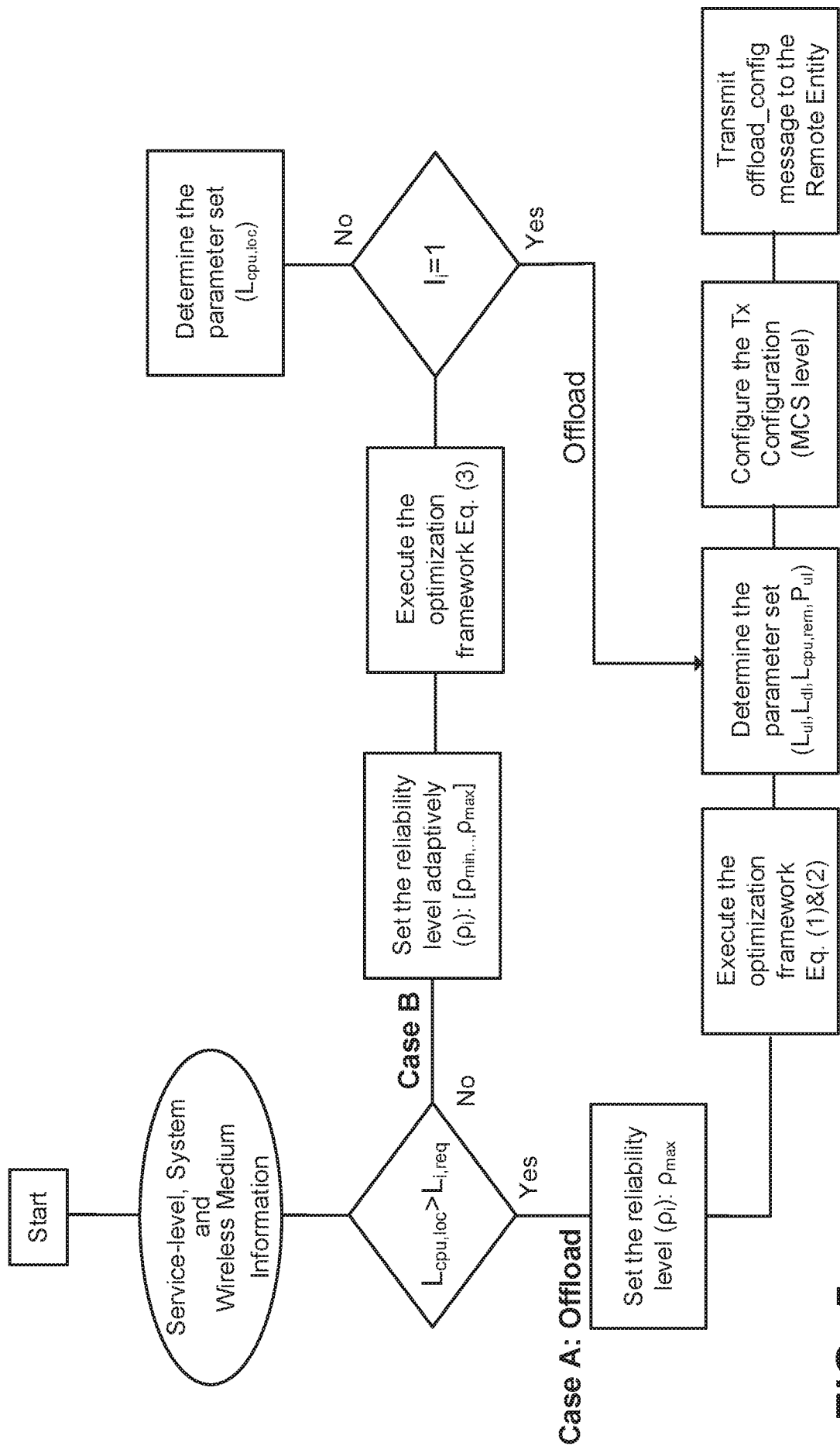
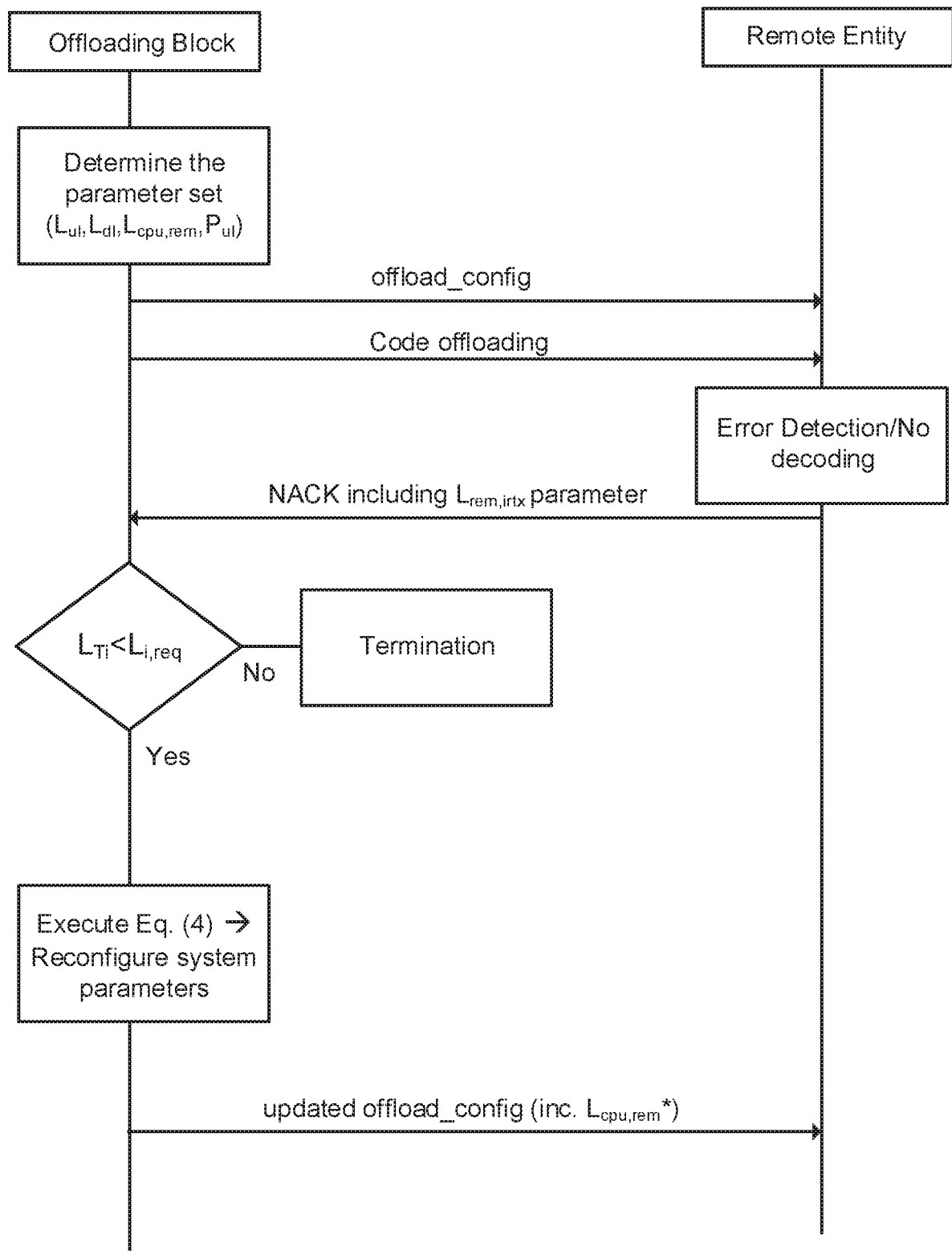


FIG. 5

5/11

**FIG. 6**

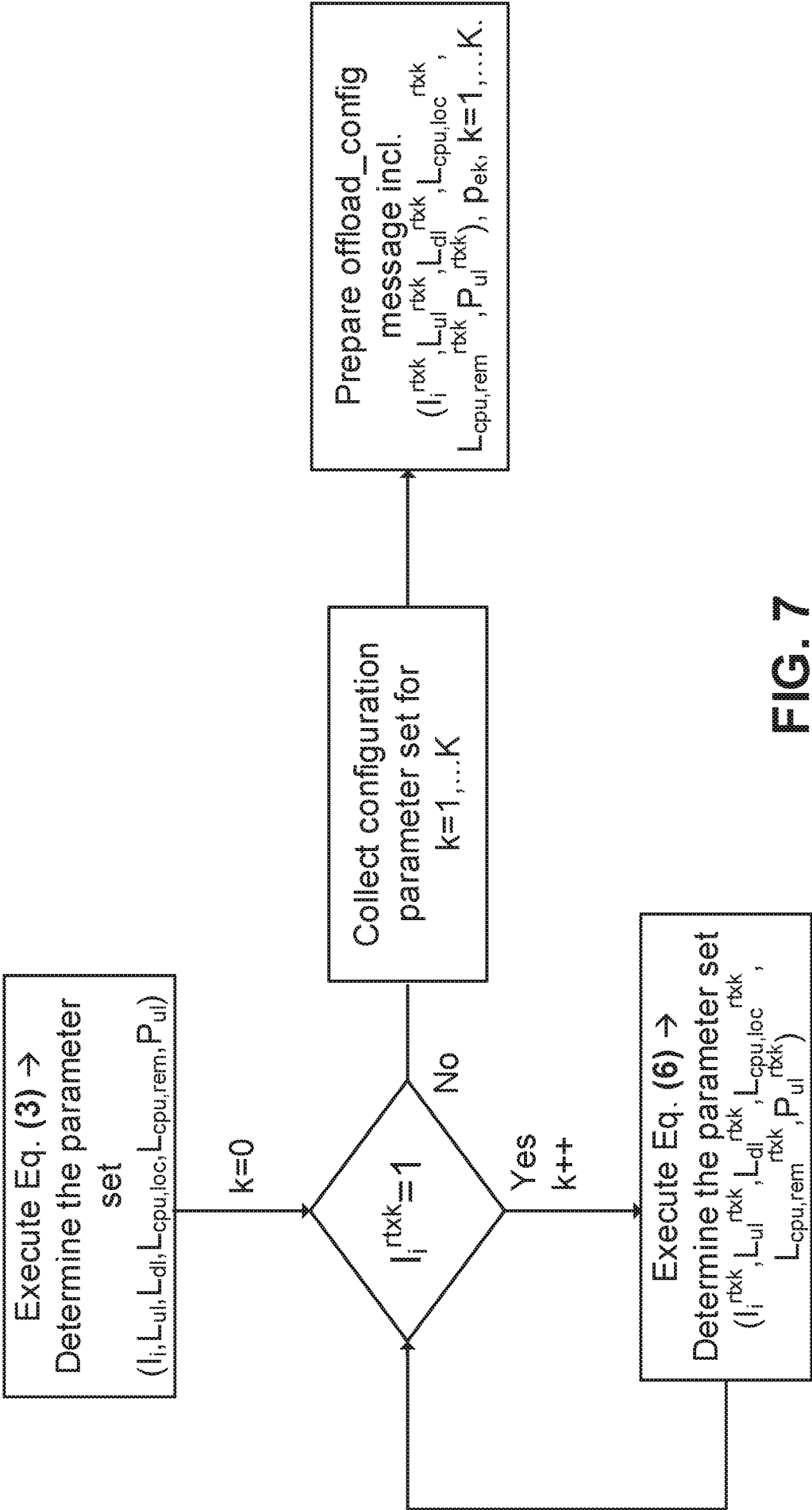


FIG. 7

100

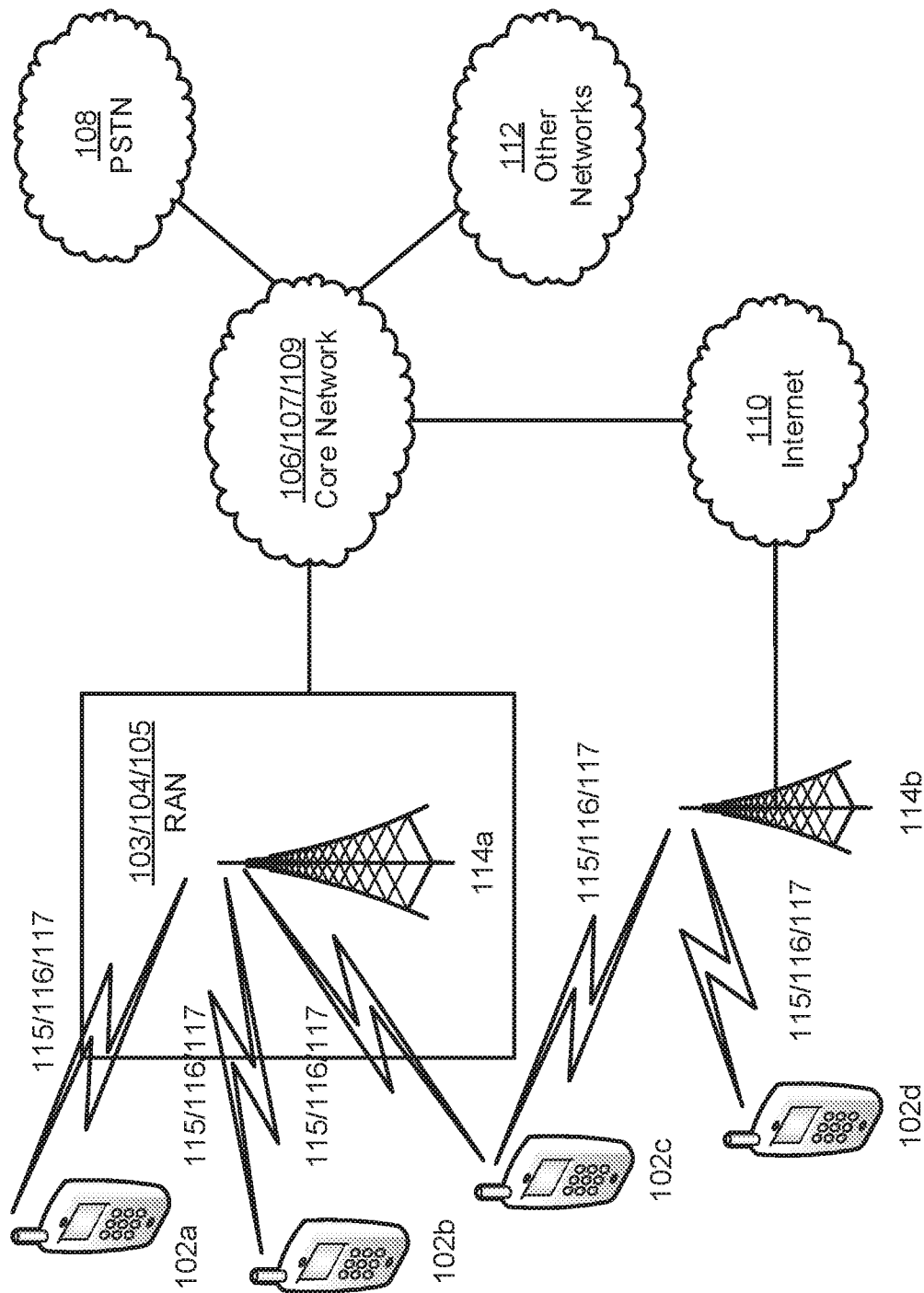
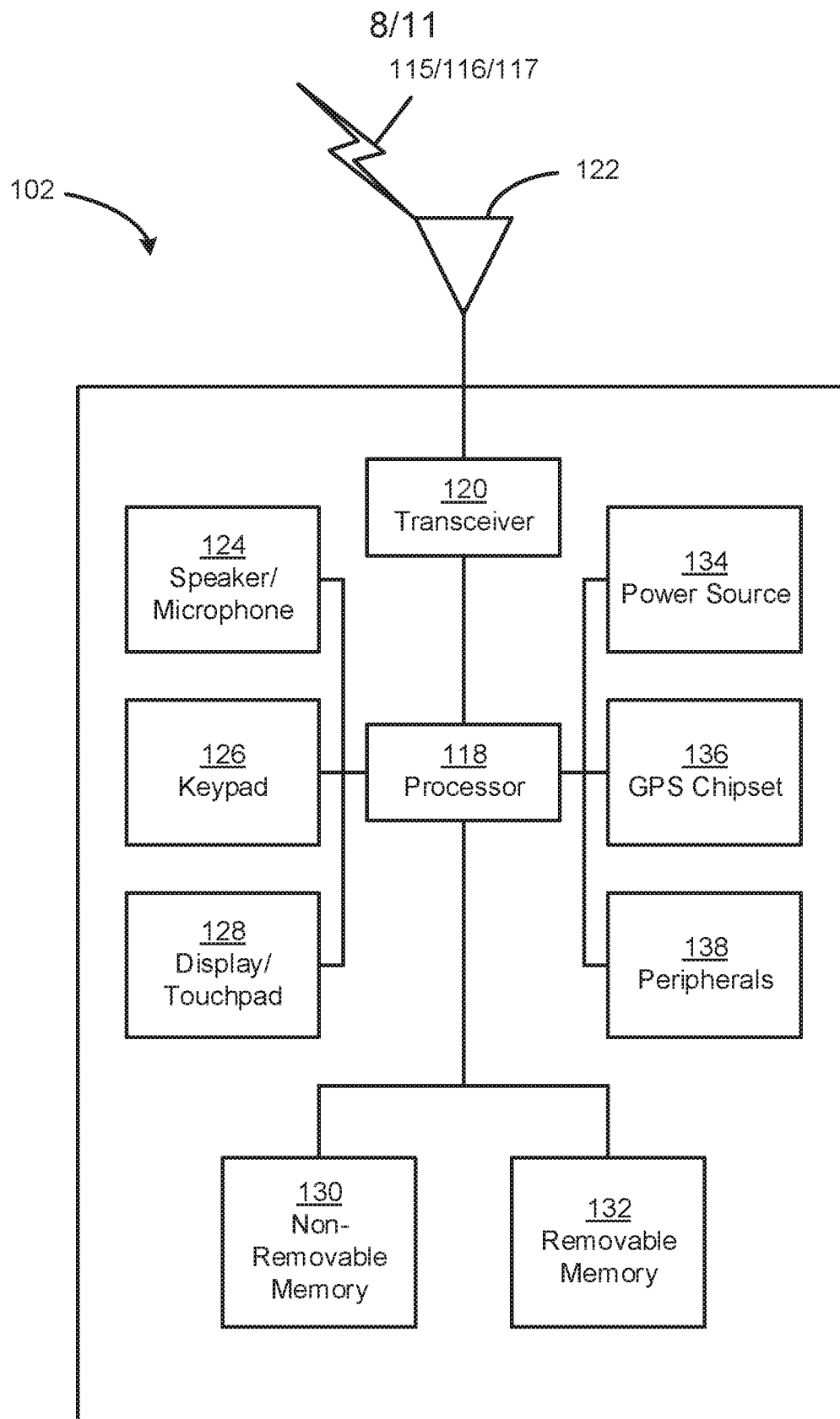


FIG. 8A

**FIG. 8B**

9/11

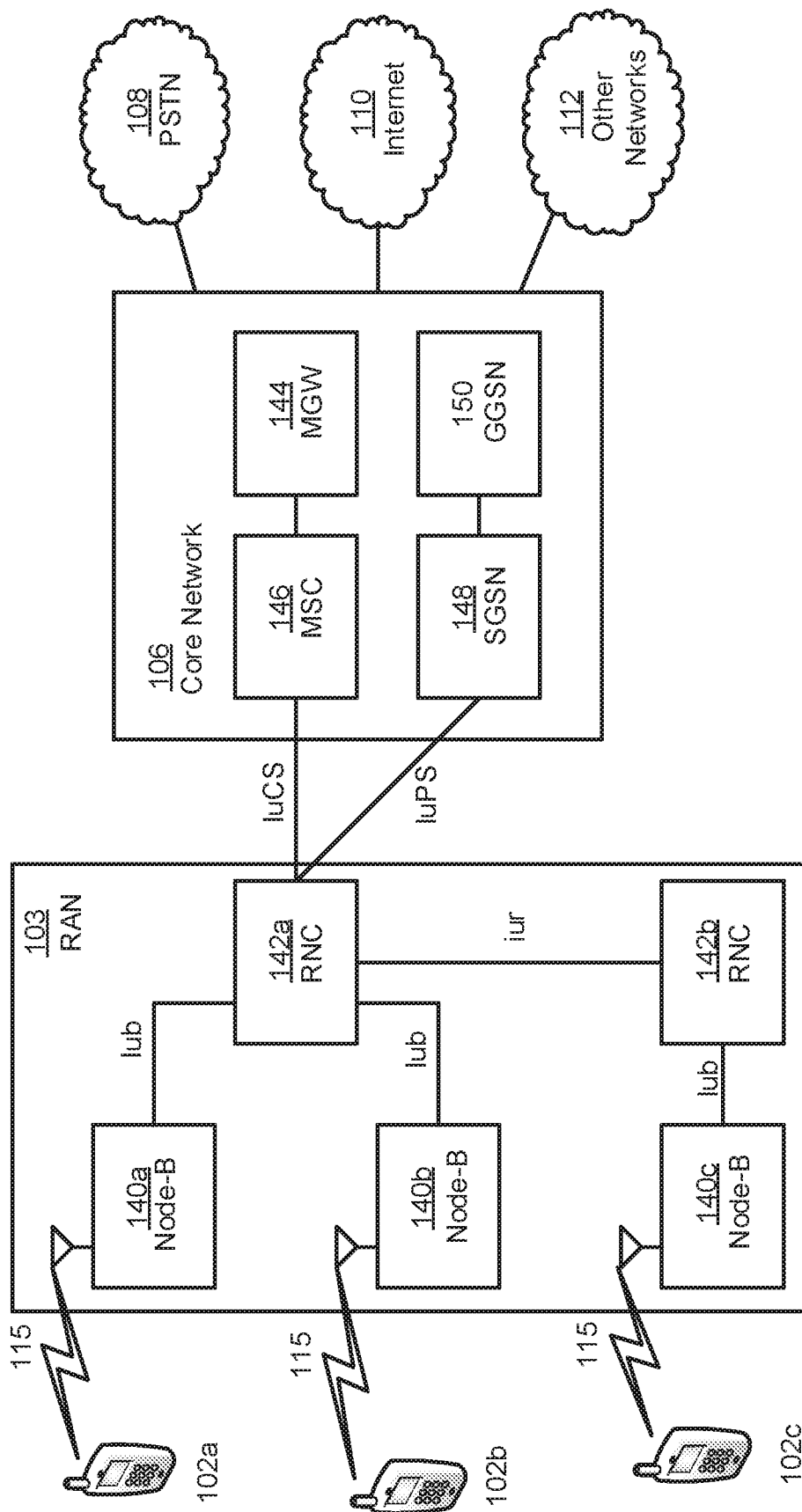


FIG. 8C

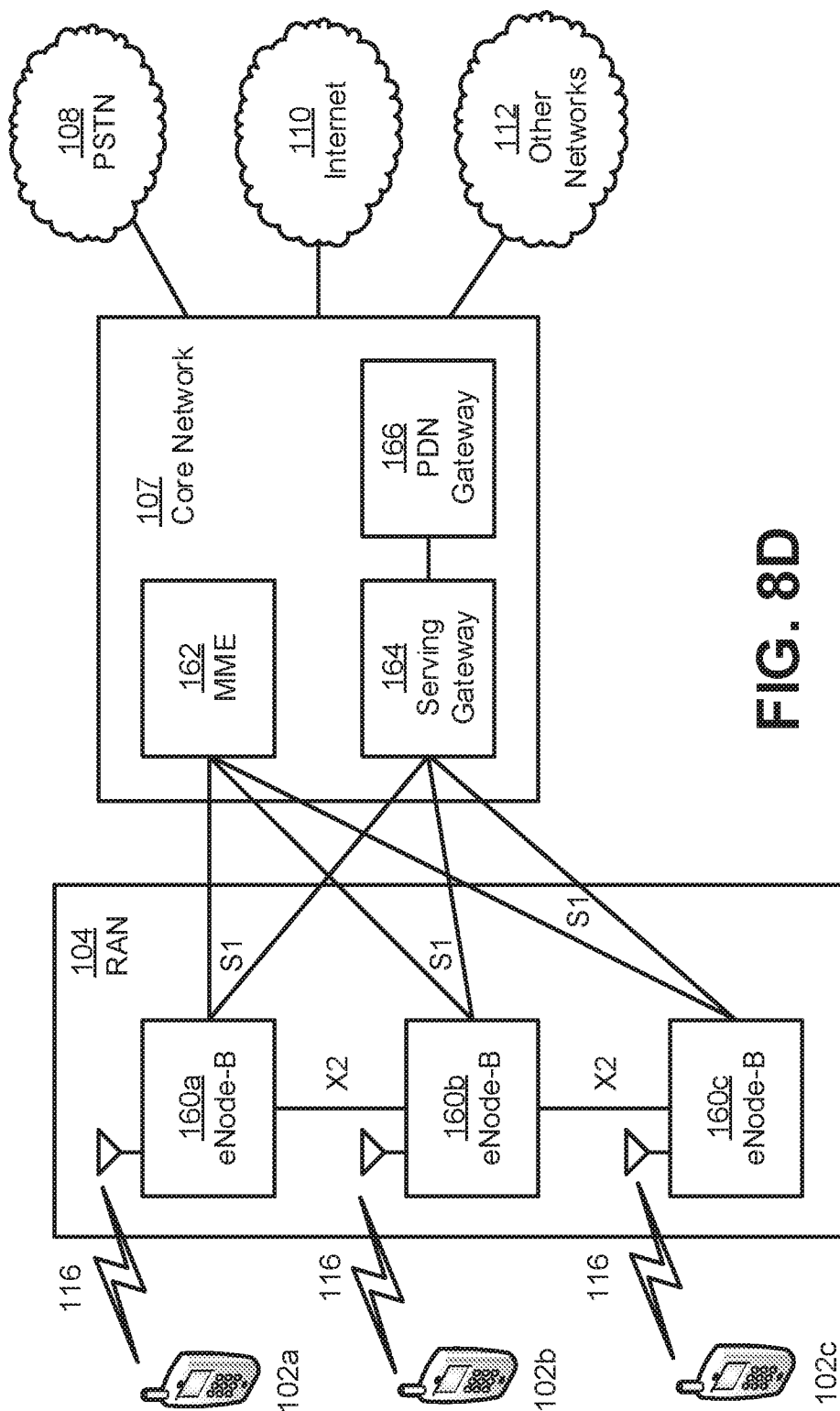


FIG. 8D

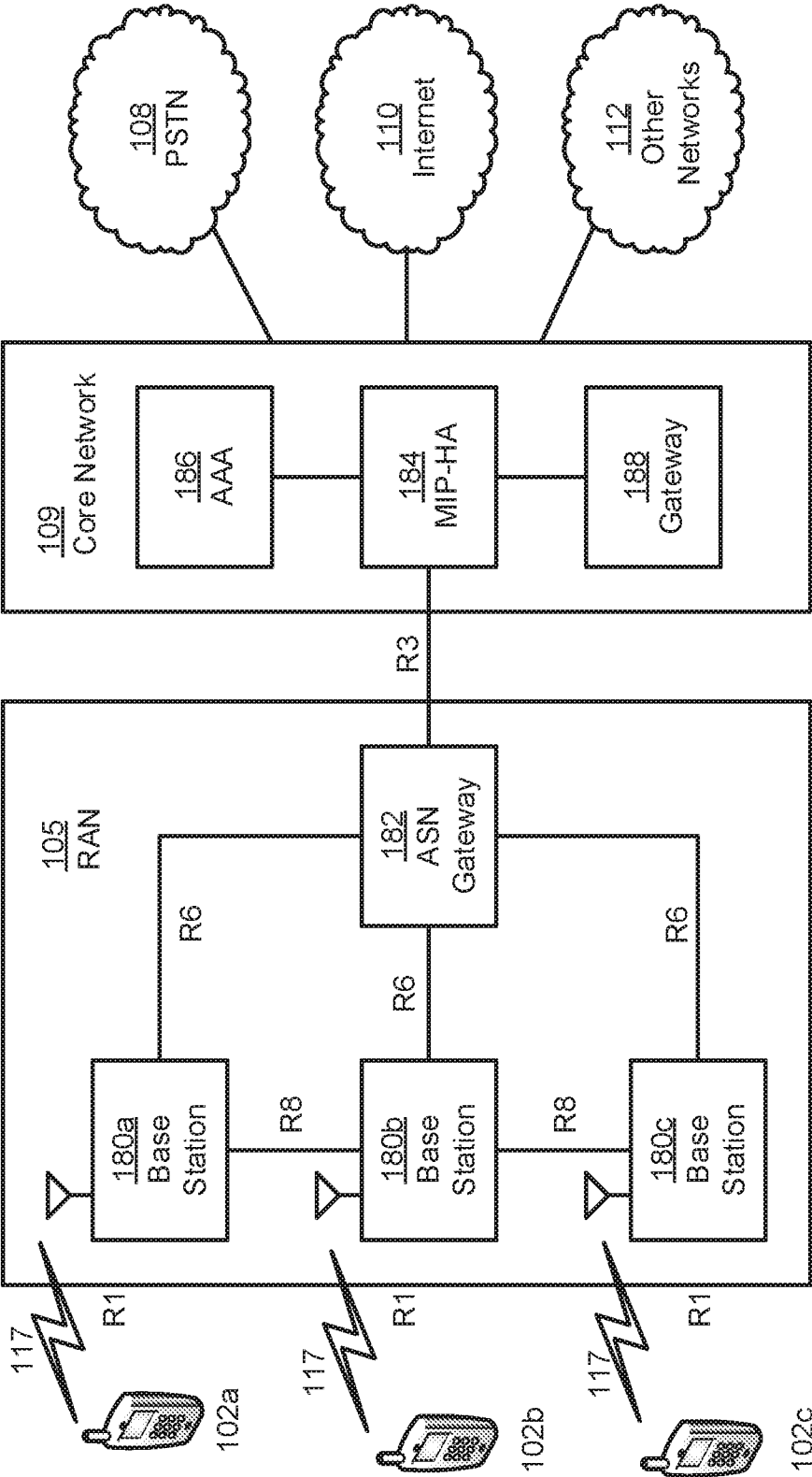


FIG. 8E

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2016/068301

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F9/50
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06F

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

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

EP0-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2014/045485 A1 (JAIN NITIN [IN] ET AL) 13 February 2014 (2014-02-13) paragraph [0006] - paragraph [0008] paragraph [0042] paragraph [0041] paragraph [0026] - paragraph [0027] -----	1-34
X	US 2011/320520 A1 (JAIN NAVENDU [US]) 29 December 2011 (2011-12-29) paragraph [0003] - paragraph [0004] paragraph [0020] paragraph [0027] paragraph [0030] paragraph [0032] - paragraph [0033] paragraph [0039] paragraph [0064] paragraph [0077] ----- -/--	1-34



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

10 March 2017

Date of mailing of the international search report

20/03/2017

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Alecu, Mihail

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2016/068301

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	GURUN AND C KRINTZ S: "Addressing the Energy Crisis in Mobile Computing with Developing Power Aware Software", INTERNET CITATION, 2003, pages 1-24, XP002660613, Retrieved from the Internet: URL:http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.58.9023 [retrieved on 2011-10-05] paragraph [Introduction] paragraph [4.Computationoffloading] - paragraphs [0006] - [RemoteExecution] -----	1-34
A	US 2011/119364 A1 (ZARKESH AMIR MASOUD [US]) 19 May 2011 (2011-05-19) paragraph [0040] paragraph [0024] paragraph [0038] -----	1-34

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2016/068301

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2014045485 A1	13-02-2014	EP 2696632 A2 US 2014045485 A1	12-02-2014 13-02-2014
US 2011320520 A1	29-12-2011	NONE	
US 2011119364 A1	19-05-2011	EP 2502154 A2 JP 5775523 B2 JP 2013511775 A US 2011119364 A1 WO 2011062994 A2	26-09-2012 09-09-2015 04-04-2013 19-05-2011 26-05-2011