3GPP TR 23.866 V1.0.0 (2012-06)

Technical Report

3rd Generation Partnership Project;
Technical Specification Group Services and System Aspects;
Study on System Improvements for Energy Efficiency
(Release 12)

Keywords

3GPP, Architecture, Energy efficiency

3GPP

Postal address

3GPP support office address

650 Route des Lucioles - Sophia Antipolis
Valbonne - FRANCE
Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Internet

http://www.3gpp.org

Contents

Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

1 Scope

The present document is for investigating deployment aspects that relate to energy efficiency, including potential system enhancements that support energy efficient deployments. System enhancements may be anticipated in the area of functions that have major influence on deployment like functions that support pools of CN nodes or functions that enable multiple CN nodes to serve the same or overlapping areas. The initial focus is on PS domain. This study should avoid any overlap with the work by RAN/CT/SA5 by taking into account the work that has been and is going on in these WGs. Proposals have to be well justified to be considered in this document.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[2] 3GPP TS 23.401: "GPRS enhancements for E-UTRAN access".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the terms and definitions given in TR 21.905 [1] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in TR 21.905 [1].

Energy conservation mode (for load re-distribution): an operational state, subject to implementation, of an EPC node with reduced capacity (e.g. in terms of number of users to be handled) or performance (e.g. in terms of data rate for user plane handling), requiring less power than in normal operational state. Subject to implementation, a node in energy conservation mode may reduce its capacity/performance to zero.

Suspension mode (for load re-distribution): an operational state of an EPC node where load needs to be reduced (by re-distribution to alternative nodes). While in suspension mode, this node will not be selected to handle further users. In the context of this technical report, suspension mode is used to enable load redistribution for energy saving reasons.

3.2 Abbreviations

For the purposes of the present document, the abbreviations given in TR 21.905 [1] apply.

4 Approaches for Energy Efficient Deployments

This clause describes high level approaches that enable or support energy efficient deployments. The energy efficient deployment scenarios that are described and evaluated later in this document are assumed to base on one or more of the mechanisms provided by the system architecture, which are:

- Separation of control and user plane allowing for independent scaling and deployment of related network entities and deployment at different locations, e.g. allowing for operating control functionality at locations with more energy efficient conditions, which could be locations with lower cooling requirements or where functionality can be shared with other applications or can be hosted in a more energy efficient way.

- Aggregation or pooling of network entities, allowing for higher multiplexing gains or economies of scale not only for PLMN specific functionality, but also for example for power supply or air conditioning. It can also reduce the amount of deployed spare capacity that is used, for example, to cope with load changes or for reliability reasons.

- Load balancing/(re)distribution approaches; allowing network nodes to operate at an energy efficient load status, or for offloading nodes followed by a change to some energy saving mode.

Further, it is assumed that energy efficient deployments shall have minimum impact on user experience.

5 Energy Efficient Deployment Scenarios

5.1 Scenario: Pooled deployment of MMEs

5.1.1 Scenario description

The separation into user and control plane network entities enables separate deployments of those network entities. In the scenario described here the user plane network entities are deployed more distributed, i.e. topologically closer to the RAN to optimise the routing and reduce the required transmission resources. The control plane network entities MME are deployed as a pool of MMEs to take advantage from the pooling effects, like better scaling or possibilities to distribute traffic between different nodes. On the other hand, pooled MMEs will cover a larger geographical area compared to a single MME deployment and therefore resulting in less MME changes compared to a single MME deployment. As these single MMEs most probably will be deployed at different locations, the single MME deployment will result in increased inter-MME transmission effort compared to a pooled deployment.

A drawback of that pooled deployment of MMEs is that some considerable signalling traffic may occur between eNBs, S-GWs and MMEs. In contrast to the transmission resource gains in the user plane this increases transmission efforts for the control plane as the user plane nodes and the MMEs are deployed at topologically different locations.

Traffic model for Service Request procedures

An interesting procedure of that deployment scenario is the Service Request procedure due to its high rate. The following estimation characterises the required signalling capacity and performance. For this estimation an MME with 1 Mio subscriber capacity is assumed. As a rough estimation it is assumed that only 50 % of the MME's subscriber capacity is used by attached and active UE. Assuming just 50 % compensates for some percentage of active UEs with lower activity or that some of the Service Requests are initiated for other purposes than just changing to active state. With this compensation it is assumed that each active UE generates a rate of 20 Service Requests per busy hour.

With 20 Service Requests per busy hour and user and 0.5 Mio active users that MME would have 10 Mio Service Requests per busy hour. These are around 2800 Service Request procedures per second. The UE Initiated Service Request procedure has 5 messages belonging to 3 messaging events, i.e. there are about 8400 Service Request related messaging events per second for that MME during a busy hour.

With about 50 octets per signalling message it is about 4.2 Mbit/s each for S1-MME and S11, together 8.4 Mbit/s of signalling traffic for idle to active transition of UEs between eNB or SGW and the pooled MME.

It is for further study whether the assumed 20 Service Requests per UE and busy hour imply usage of URA_PCH.

5.1.2 System enhancements

Enhancements may need to be considered to diminish the signalling traffic drawbacks of the pooled deployment of MMEs.

A potential system enhancement is to enable the SGW for handling Service Request procedures without always involving the MME, which may be done as typically all parameters from the earlier S1 connection apply also for the subsequent Service Request. The SGW may store that parameters for handling subsequent Service Requests without involving the MME. Thereby the signalling traffic estimated in the traffic model above may be reduced. For this the SGW needs additional functionality to store parameters required for establishing the S1 connection. More evaluations or also study by SA3 is required when the SGW needs to be enabled to manage security parameters. This because any idle to active transaction includes some security procedures. Depending on the approach this may impact the eNBs.

Another possible solution is to enhance the handling of the connected mode by the RAN, basically to extend the connection time and reduce signalling traffic thereby. Such approaches are under discussion in RAN groups.

5.1.3 Evaluation

The pooled deployment of MMEs is supported by already defined standards. Any potential optimisations for reducing the signalling traffic, that may run long distance with such deployments, need further study.

Concerns were raised for the Service Request change proposal excluding the MME and having the SGW to store the information to be used for the subsequent Service Requests. The SGW needs to support new protocols such as SCTP, S1-AP and NAS, if trying to avoid impacts for eNB.

5.2 Scenario: Energy efficient node utilization through load re-distribution during off-peak times

5.2.1 Scenario description

In this deployment scenario it is assumed that the operator has configured pools of MMEs and/or S-GWs to serve the EUTRAN. These resource pools are typically dimensioned for peak loads. This means that the number of user plane gateways (e.g. S-GW and P-GW) and control plane entities (e.g. MME) deployed and operated are dimensioned according to the maximum number of UEs and their traffic demand.

However, given that most of the time, the load in the network is far from the dimensioned peak rate, plus the fact that during certain times (e.g. night hours), the network load is rather low, operators can expect significant energy saving opportunities through switching unnecessary core network resources/nodes to energy conservation mode when they are not needed.

5.2.1.1 Achievable Energy Savings through load concentration during off-peak times

According to the typical daily traffic load variation taken from TR 32.826 (illustrated in Figure 5.2-1, left side), an estimation of the average utilization of the network capacity can be derived (illustrated in Figure 5.2-1, right side). The result shows approx. 60% under-utilization of capacity on average. Assuming that a typical day has a busy hour traffic curve well below the expectable "worst case" busy hour traffic overall (i.e. over the year, including special days and events, for which the NW would generally have to be dimensioned), it can be conclude that at a minimum 2/3 of the core NW capacity is not utilized over in average.

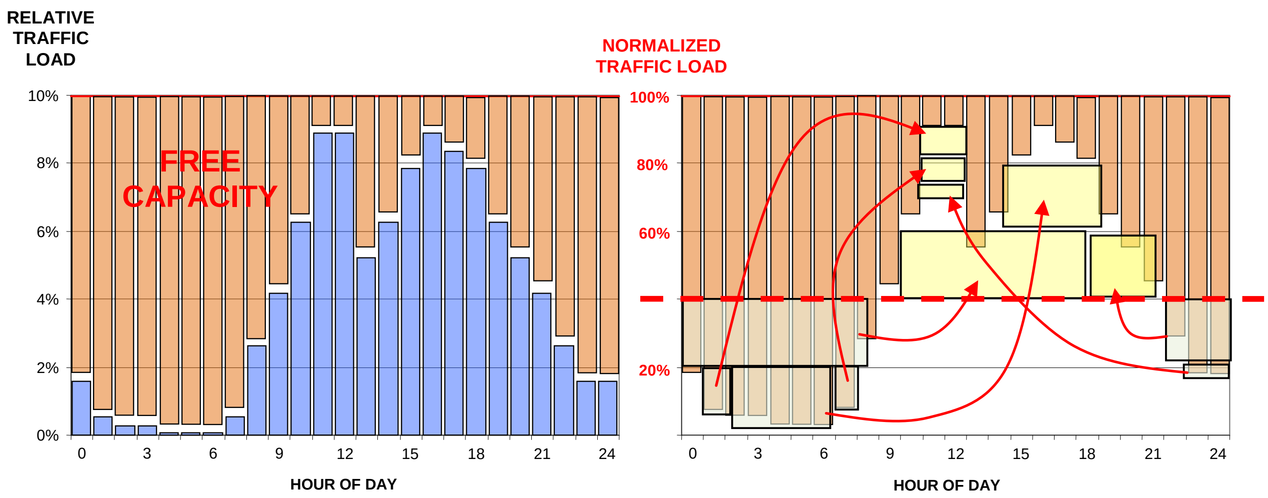

Figure 5.2-1: Typical daily traffic load variation (left) and estimation of average network load (right)

Figure 5.2-2 (top part) illustrates a simplified energy consumption curve for typical network nodes implementations. It consists of the following parts:

1) zero-load energy consumption (i.e. the minimum power drain whenever the node is turned on);

2) linear increase of the energy consumption as the load increases (up to 100% load).

Based on that model, Figure 5.2-2 shows the estimated energy saving factor that can be achieved through load concentration during off-peak times, depending on the overall system load (given in percentage on the abscissa). The set of curves refer to different zero-load energy consumptions of network elements. E.g. for network elements with a 25% zero-load energy consumption of a network node (5th curve from the bottom) at 20% overall system load leads to a potential for energy saving through load concentration of 100%.

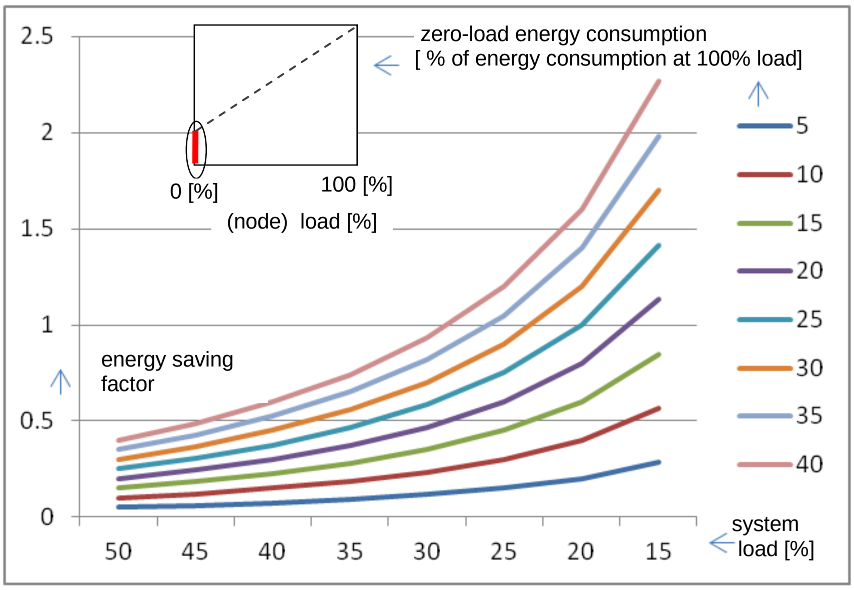

Figure 5.2-2: Estimation of energy saving factor through load concentration during off-peak times

The graph shows that the potential for energy saving depends on the design of network elements. For network elements that have a small zero-load energy consumption and support energy efficiency by design, the energy saving gains that can be achieved through load concentration during off-peak times are low, while for network elements with a medium-to-high zero-load energy consumption, the achievable energy saving gains are significant.

5.2.2 System enhancements

In order to allow energy savings by switching unnecessary core network resources/nodes to energy conservation mode during off-peak times, load re-distribution mechanisms are needed that allow the network to move users/sessions from underutilized core network nodes (e.g. P-GW, S-GW, MME) to other nodes.

5.2.2.1 General

Energy efficient node utilization can be achieved by means of node internal process optimizations or through load re-distribution according to the following basic steps:

1. Decision to switch unnecessary core network nodes to energy conservation mode: This can be controlled/triggered via the operator's O&M system.

2. Adjustment of the node selection: This ensures that a node which has been selected for suspension mode will not be considered by the node selection function until further notice.

3. Re-distribution of load: This involves execution of load re-distribution procedures, which allows a node in suspension mode to shift its load to one or more alternative nodes that run in normal mode.

4. Switch to energy conservation mode: Once the load of a node in suspension mode has been re-distributed, the node can be switched to energy conservation mode.

5.2.2.2 Node Selection Adjustment Procedures

5.2.2.2.1 Serving GW and PDN GW Selection

The PDN GW and Serving GW selection functions are used, according to clauses 4.3.8.1 and 4.3.8.2 of TS 23.401 [2] respectively, with the extension that gateway nodes operating in suspension mode are not considered.

The MME may be informed about the operational status of the nodes (i.e. normal mode or suspension mode) in one of the following ways:

- through O&M, or

- based on DNS.

5.2.2.3 Load Redistribution Procedures

5.2.2.3.1 Serving GW

Once a Serving GW has been selected for suspension mode, the following behaviour is expected:

- For idle mode UEs connected to that S-GW, the MME will upon receipt of a Tracking Area Update execute the Tracking Area Update with Serving GW change procedure (according to clause 5.3.3.1 of TS 23.401 [2]), i.e. UEs will be moved to another S-GW that operates in normal mode.

- For active mode UEs, the MME load re-balancing procedures as defined in clause 4.3.7.3 of TS 23.401 [2] can be used with the assumption that the MME will select another S-GW for the UE during the TAU procedures. Alternatively, a new S-GW relocation procedure can be defined.

5.2.2.3.2 PDN GW

Once a PDN GW has been selected for suspension mode, the following behaviour is expected:

- For idle mode UEs connected to that P-GW, the MME will deactivate in a scheduled manner the impacted PDN connections indicating "reactivation requested" as specified in clause 5.10.3 of TS 23.401 [2]. If all of the PDN connections of a UE need to be relocated, the MME may initiate the "explicit detach with reattach required" procedure as specified in clause 5.3.8.3 of TS 23.401 [2], i.e. UEs will be moved to another P-GW that operators in normal mode.

Editor's note: The handling of active mode UEs is for further study.

5.2.2.3.3 MME

Once a MME has been selected for suspension mode (e.g. triggered via the operator's O&M system), the following behaviour is expected:

- The MME notifies all eNBs in the service area about its mode change.
- For idle mode UEs served by that MME, any subsequent Tracking Area Update will be redirected to an MME running in normal mode based on existing eNB procedures. The target MME will then execute the default Tracking Area Update with MME change procedure (according to clause 5.3.3 of TS 23.401 [2]).
- For active mode UEs, either the MME load re-balancing procedure as defined in clause 4.3.7.3 of TS 23.401 [2] can be used, or a new MME relocation procedure can be defined. Additional signalling would be required to the UE if the procedure in clause 4.3.7.3 of TS 23.401 [2] is executed.

5.2.2.3.4 eNB

After the eNB has received from the MME the notification that the mode of the MME has changed to "suspended", it sets the target percentage for load on this MME to zero.

5.2.3 Evaluation

The energy saving potential through load re-distribution depends crucially on the zero load power consumption of a particular network node implementation. When the zero load power consumption of a network element is low, the gains that could be achieved through this mechanism may not be justified. However, for network elements that are not implemented in an energy efficient manner, the possible energy saving gains are significant.

The necessary enhancements for MME and SGW load re-distribution have not been fully analysed for active mode UEs. Also, the potential impacts on RAN have not been investigated for active mode UEs.

Potential drawbacks of the proposed solution described in clause 5.2.2 are:

- There is a possibility of increased signalling load, consuming additional energy, in the event of failure of a concentrated node which is acting as a single point of failure. An example is a failure of an MME causing increased signalling from a new MME in an energy conserved mode.
- Another limitation of the proposal to move UEs when switching core network nodes to/from energy conservation mode is that such load re-distribution in itself results in increased transmission efforts.

5.3 Scenario: Energy Efficiency by Network Sharing

5.3.1 Scenario description

The scenarios described here use the network sharing features for gaining energy efficiency.

In rural areas with lower dense of population PLMNs may need to provide coverage. However already the minimum configuration may provide more resources than actually needed. For such scenarios it can be advantageous when different PLMNs share network resources for covering such areas with low traffic.

In urban areas each PLMN may need its own capacity to serve the urban area. However when the capacity per site is scalable, e.g. in number of sectors or frequencies, then sharing may also help to improve energy efficiency as the combined traffic of multiple PLMNs should have a lower variance than the traffic of each individual PLMN. Therefore the maximum peak capacity per shared site could be lower.

Another variant is to share dynamically. During the time of the day with high traffic the PLMNs may want to use their own capacity and coverage. During the time of the day with low traffic every single PLMN may need only low capacity. The sharing may be done only during the low traffic time.

5.3.2 System enhancements

The scenarios with static sharing don't provide any new system or signalling specifics and may be accomplished by configuring already specified network sharing features.

The more dynamic sharing may need enhancements of the procedures that enable energy saving by inter-RAN-node mechanisms for use with multiple PLMNs.

5.3.3 Evaluation

Network sharing features as already specified may be deployed for gaining energy efficiency without any need for modifying or enhancing that sharing features. Enhanced sharing features like described above may be considered during the definition of inter- or intra-RAT energy saving. Basically it would need to be enabled for inter-PLMN scenarios.

The use of such sharing approaches may depend on regional, legal or licence conditions that may determine permission or obligations for network sharing in general.

The variant of dynamic sharing may cause peaks in signalling traffic during the change between sharing and non-sharing configurations.

5.4 Scenario: Energy Efficiency by Scheduled Communications

5.4.1 Scenario description

RAN developed approaches with inter-RAT energy saving. There may be however devices that are single RAT, e.g. some devices for machine type communications. Those devices cannot move to another RAT, but may need to communicate only during regular intervals, like considered under time controlled communications for MTC.

So far there is no feature "time controlled" specified for MTC or otherwise. It may be however assumed that the network could determine whether in some area there are only such single RAT devices that do not need to communicate. Or that have only low priority. The scheduling of deactivating a RAT may be supported by the knowledge about business hours for an industrial or office area/building. So devices that are less important, e.g. that serve some appliances, may lose communications. More mission critical devices may be assumed to have multi-RAT capabilities.

5.4.2 System enhancements

A scheduled RAT switch-off/on may happen controlled by O&M without any system enhancements needed. This approach may be improved by the network determining whether there are only low priority or other single-RAT devices that tolerate a RAT switch-off. For time controlled devices the time-control information may needs to be aligned with the switch-off period. Preferably the indication is by the RAT, e.g. a broadcast indication of the switch-off time so that e.g. time-controlled MTC devices can schedule their transmissions accordingly.

5.4.3 Evaluation

As it is not considered in the scope of this study to define such communications feature or change UE behaviour the usage of this deployment scenario depends on the introduction of scheduled communications like studied under M2M scenarios. When such a feature will be defined it should be considered to also enabling its applicability for energy efficiency purposes.

Legal obligations might prevent switching off RATs, e.g. due to potential emergency services.

6 Conclusions

The work on this study for identifying energy efficient deployment scenarios was launched in the beginning of 2011. Despite the limited time available to this study a few energy efficient deployment scenarios have been identified and evaluated. The following general conclusions can be derived from the work and the related discussions:

Approaches with no or only small system enhancements or modifications typically yield only small efficiency gains. Potentially more promising larger gains typically depend on enhancements or modifications with deeper system impacts. Larger system modifications require also bigger changes to deployed systems.

It was not possible during the study period to derive quantitative metrics for measuring the gains of approaches for improved energy efficiency as the base figures of energy consumptions are largely determined by the specific product implementations. The study concludes that features that are designed for load or network sharing, like pooling of core network nodes, can be deployed also for energy efficiency purposes. The efficiency is gained mainly from better scaling granularity, e.g. less needs for over-provisioning, and from sharing base functionality, like power supply or other site equipment, which reduces also base power consumption.

As a conclusion we may determine that it seems more promising to consider energy efficiency or also resource efficiency during the design of any new feature or functionality, e.g. with regard to required storage, signalling or processing resources. Larger effects may be obtained when designing a new system. Hence energy efficiency should be one of the main system design guidelines.

Annex A:
Change history

Change history							
Date	**TSG #**	**TSG Doc.**	**CR**	**Rev**	**Subject/Comment**	**Old**	**New**
2011-02	SA2#83	-	-	-	Initial skeleton draft by the editor and approved in S2-111139	-	0.0.0
2011-03	SA2#83	-	-	-	S2-110457 - Scope S2-111239 - approaches for energy efficient deployments S2-111141 - Pooled deployment of MMEs S2-111240 - Scenario for energy efficient node utilization during off-peak times	0.0.0	0.1.0
2011-03	SA2#84	-	-	-	Implementation of S2-112088, S2-112089, Some editorial updates	0.1.0	0.2.0
2012-04	SA2#90	-	-	-	Implementation of S2-121762, S2-121763, S2-121913	0.2.0	0.3.0
2012-06	SA2#91	-	-	-	Implementation of S2-122605, S2-122471, S2-122584, S2-122607, S2-122473, S2-122175 Version for information and approval by TSG SA	0.3.0	0.4.0
2012-06	SA2#91	SP-120282	-	-	MCC editorial update for presentation to TSG SA for information and approval	0.4.0	1.0.0