3GPP TR 29.807 V0.1.0 (2013-08)

Technical Report

3rd Generation Partnership Project; Technical Specification Group Core Network and Terminals; Study on GTP-C Overload Control Mechanisms (Release 12)

Keywords
<keyword[, keyword, …]>

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 Specification 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 investigates various aspects related to the support of the GTP-C signalling based load / overload control solution as specified in 3GPP TS 23.401 and TS 23.060, as also concluded in TR 23.843, with the main focus on:

- Definition of "Load Control" and "Overload Control" related information with enough precision to guarantee a common multi-vendor interpretation of this information allowing inter-operability between various GTP-C nodes;
- Mechanisms to address various "Notes", which are targeted to the stage 3, specified in clause 8.2.5, clause 8.2.6 & clause 10.2 of the 3GPP TR 23.843 v1.0.0.

This technical report will address the following aspects related to the "GTP-C signaling based load and overload control" feature in detail:

- Investigation and study of the following aspects related to the "Load Control Information" to fulfill the normative requirements or to produce recommendations.
  - Definition of the "Load Control Information" by evaluating various parameters which can be exchanged under this information.
  - Inclusion of "Load Control Information" in GTP-C messages.
  - Potential enhancements to the existing node selection algorithm to take information received from "Load Control Information" into account.
- Investigation and study of the following aspects related to the "Overload Control Information" to fulfill the normative requirements or to produce recommendations.
  - Definition of the "Overload Control Information" by evaluating various parameters which can be exchanged under this information.
  - Inclusion of the "Overload Control Information" in GTP-C messages.
  - Message throttling algorithms and message prioritization when congestion mitigation is applied.
  - Propagation of the MME/SGSN identity to the PGW to ensure that the overload mitigation is always applied to the currently serving MME/SGSN.
  - Potential interactions with the existing overload control mechanisms when the overload factor is received within the "Overload Control Information".
- Investigation and study of the following the other deployment related aspects to fulfill the normative requirements or to produce recommendations.
  - Applicability of this feature to 3GPP and non-3GPP based GTP-C interfaces.
  - Methods to discover the support of this feature by the peer node in the network.
  - Supporting the feature across the PLMN boundary.
  - Issues within the network with partial support of this feature.
  - Overload mitigation policies when this feature support is not enabled in the network.

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] 3GPP TS 23.060: "General Packet Radio Service (GPRS); Service description; Stage 2".

[4] 3GPP TR 23.843: "Study on Core Network Overload Solutions".

3 Definitions, symbols 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].

3.2 Symbols

For the purposes of the present document, the following symbols apply:

3.3 Abbreviations

For the purposes of the present document, the abbreviations given in TR 21.905 [1] and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in TR 21.905 [1].

4 Introduction to GTP-C overload control

4.1 Scenarios leading to GTP-C overload

Editor's Note: Description of various scenarios resulting into the overload of the network due to messaging over GTP-C based interfaces. Most of these scenarios are captured in TR 23.843 clause 4.

4.2 GTP-C signalling based Load and Overload Control solution

4.2.1 Description

Editor's Note: Introduction to "Load Control" and "Overload Control" concepts and the "GTP-C signalling based Load and Overload Control" solution as defined in 3GPP TS 23.401 clause 4.3.7.1a.

4.2.2 Principles of Load Control

Editor's Note: High level principles of "Load Control" as defined in 3GPP TS 23.401 clause 4.3.7.1a.1.

4.2.3 Principles of Overload Control

Editor's Note: High level principles of "Overload Control" as defined in 3GPP TS 23.401 clause 4.3.7.1a.2.

4.2.4 Applicability to 3GPP and non-3GPP interfaces

4.2.4.1 Description

While recommending the list of interfaces for which GTP-C load/overload mechanism should be applied, the decision on the exact list of applicable interfaces has been left by the stage 2 on the stage 3. Specifically, for 3GPP access based interfaces it needs to be assessed whether it is beneficial to support the overload control mechanisms on the interfaces such as Sm/Sn, S10, S3 or S16. For non-3GPP access based interfaces, the applicability of the overload control mechanisms over any of the interfaces, i.e. S2a or S2b, needs to be assessed.

5 Load Control Information

5.1 General

In order to guarantee a common interpretation in a multi-vendor network deployment, it is necessary to define the "Load Control Information" with enough precision such that coherent and homogeneous node selection algorithms are applied by different nodes of the same network such that an evenly load balanced network is realized. This clause investigates possible parameters and their definitions which can be exchanged under the "Load Control Information". Thus in turn, this clause aims at defining the exact format the "Load Control Information".

5.2 Definition

Editor's Note: This clause will have various alternatives each defining the structure of "Load Control Information". Each alternative will be evaluated to conclude the normative definition of the "Load Control Information".

5.2.1 Requirements

Editor's Note: This clause will capture the requirements guiding the definition of the "Load Control Information".

5.2.2 Alternative 1

5.2.2.1 Description

5.2.2.2 Evaluation

5.3 Frequency of inclusion

5.3.1 Requirements

This sub clause aims at defining how often/frequently the "Load Control Information" should be transferred, while ensuring the following requirements:

- The transfer of the load Information shall not add significant additional load to each peer node.

- The calculation of load Information should not severely impact the resource utilization of the node.

5.3.2 Alternative 1

5.4 Interaction with existing mechanisms

5.4.1 General

The parameters such as weight factor of the node are an essential input to the node selection algorithm. These parameters are either returned by the DNS or configured locally and hence they are mostly static or pseudo-dynamic type of information. On the other hand, the "Load Control Information", transferred within GTP-C messages, provides the current value of the load level representing the dynamic load condition of the sending node more accurately. Hence, the node selection algorithm should take both into account, the preference related information provided by the DNS or other mechanisms and the dynamic load level provided by the "Load Control Information", to calculate the effective load of the target node. This clause investigates the enhancements to the node selection algorithms which take "Load Control Information" and other existing parameters into account.

5.4.2 Information received from DNS

5.4.2.1 Description

6 Overload Control Information

6.1 General

In order to guarantee a common interpretation in a multi-vendor network deployment, it is necessary to define the "Overload Control Information" with enough precision such that coherent and homogeneous mitigation policies are enforced by different nodes of the same network alleviate the congestion, effectively. This clause investigates possible parameters and their definitions which can be exchanged under the "Overload Control Information". Thus in turn, this clause aims at defining the exact format the "Overload Control Information".

6.2 Definition

Editor's Note: This clause will have various alternatives each defining the structure of "Overload Control Information". Each alternative will be evaluated to conclude the normative definition of the "Overload Control Information".

6.2.1 Requirements

Editor's Note: This clause will capture the requirements guiding the definition of the "Overload Control Information".

6.2.2 Alternative 1

6.2.2.1 Description

6.2.2.2 Evaluation

6.3 Frequency of inclusion

6.3.1 Requirements

This sub clause aims at defining how often/frequently the "Overload Control Information" should be transferred, while ensuring the following requirements:

- The transfer of the overload Information shall not add significant additional load to each peer node.
- The calculation of overload Information should not severely impact the resource utilization of the node.

6.3.2 Alternative 1

6.4 Message throttling

6.4.1 Throttling algorithms

As a part of the overload mitigation, the GTP-C node is required to reduce the total number of messages, which would have been sent otherwise, towards the overloaded peer based on the information received within "Overload Control Information". This is achieved by discarding a fraction of the messages in proportion to the overload level of the target peer. This is called message throttling and there could be multiple ways (i.e. algorithms) to achieve the same. Correspondingly, this sub clause examines various message throttling algorithms, and for each algorithm evaluates various aspects such as effectiveness, efficiency, ease of implementation, etc. Finally, one of the algorithms will be recommended as the default algorithm which should be minimally implemented as the part of the support of the "GTP-C overload control" feature.

6.4.2 Message prioritization

6.4.2.1 Description

As part of overload mitigation mechanisms, based on the "Overload Control Information" received, the GTP-C node may start throttling the messages towards the overloaded peer. In the absence of any guidelines related to identifying the priority of the messages, the GTP-C node may perform random throttling and hence start discarding the message without any special consideration. This type of sub-optimal message throttling would result in the overall poor congestion mitigation mechanism. Correspondingly, this clause investigates various criteria for the prioritization of the messages so that the messages which are considered as low priority are considered for throttling before the other messages, when the message throttling is to be applied. These prioritization mechanisms can also be used by the overloaded node to discard the messages during overloaded condition, as a part of the self-protection mechanism.

6.4.2.2 Based on procedures

6.5 Propagation of MME/SGSN identity to PGW

6.5.1 Description

The PGW may not be aware about the identity of the currently serving MME/SGSN since there is no signalling over S5/S8 interface during the procedures involving inter-MME/SGSN and intra-SGW scenarios, when the reporting of RAT, ULI, UCI, Serving Network is not needed. In that case, if the PGW has received the "Overload Control

Information" from the old MME/SGSN, it may enforce the mitigation actions assuming that the old MME/SGSN is the currently serving the UE. This would result in incorrect enforcement of the overload control and hence it should be avoided while also ensuring that we do not overload S5/S8 interface with unnecessary signalling. This sub clause investigates possible methods for propagating the currently serving MME/SGSN identity to PGW when the overload control for the old MME/SGSN is to be applied by the PGW.

6.6 Interaction with existing mechanisms

It shall be possible to run the existing congestion control mechanisms in parallel and concurrently with the new congestion control mechanisms defined as part of "GTP-C Overload Control Mechanisms". However, there could be potential impact to these existing mechanisms due to the support of GTP-C overload control mechanism, e.g. potential interaction with the DDN throttling mechanism when the MME/SGSN sends an "Overload Control Information" to SGW/PGW. Correspondingly, the analysis of these impacts and possible interaction between the existing and new mechanisms are investigated in this clause.

6.6.1 DDN throttling

6.6.2 Congestion control using APN back-off timer

7 Deployment related considerations

7.1 General

In this clause, various deployment related considerations, for the support of load/overload control mechanism, are investigated.

7.2 Discovery of the support of the feature by the peer node

7.2.1 Description

In order to apply the overload control mechanisms and exchange the load/overload control related information, the node may need to be made aware about the support of the "GTP-C signalling based Load & Overload Control" feature of the peer node. Methods to realize the support of this feature by the peer node are investigated within this sub clause, keeping in mind the inter-PLMN and the intra-PLMN related considerations.

7.3 Supporting the feature across the PLMN boundary

7.3.1 Description

Editor's Note: Supporting the GTP-C Overload Control across the PLMN boundary is studied here. Currently, 3GPP TS 23.401 has following related text "*Based on local policies/configuration, a GTP-C node may support Overload Control feature and act upon or ignore Overload control Information in the VPLMN when received from HPLMN and in the HPLMN when received from VPLMN. When this feature is supported, a GTP-C node may decide to send different values of Overload control Information on inter-network (roaming) and on intra-network (non-roaming) interfaces based on local policies/configuration.*"

8 Heterogeneous network related considerations

8.1 General

The network with non-homogenous support of the "GTP-C overload control mechanisms", such that for a particular interface, some nodes are upgraded with the support for the feature while the others are not, are termed as heterogeneous network for this study, irrespective of the PLMN boundary. If the HPLMN or VPLMN operators do not support or activate the support of this feature in the whole network then potentially, other mechanisms which do not rely on the explicit exchange of load/overload control information could be considered to identify a possible overload condition of the GTP-C peer node. These are termed as "Implicit overload control mechanisms". On the other hand, if the operator enables the support of this feature in the heterogeneous network then there could be potential issues related to the handling of the overload mitigation, e.g. uncontrolled rate of signalling from the nodes which do not support this feature would require higher throttling from the nodes which support this feature in order control the overall rate of signalling towards the target node to avoid its meltdown. And this result into unfair advantage to the nodes not supporting the feature in the heterogeneous network compared to the nodes not supporting this feature in the homogeneous network. These and other related aspects are investigated in this clause.

8.2 Implicit overload control mechanisms

Editor's Note: Potential mechanisms which to do not rely on the explicit exchange of the load/overload control information could be considered to identify a possible overload condition of the GTP-C peer node. Additionally, for the self-protection, the overloaded node may take certain mitigation actions such as message prioritization to discard the lower priority messages. These "Implicit overload control mechanisms" are investigated here.

8.3 Issues in the network with partial support of the feature

9 Conclusion and recommendations

Annex A (informative): Impacts to Specifications

Annex B (informative): Change history

Change history							
Date	TSG #	TSG Doc.	CR	Rev	Subject/Comment	Old	New
08-2013					*Includes the following TDocs agreed during CT4#62: C4-131451, C4-131314, C4-131452, C4-131453.* *Some editorial corrections are done.*	*0.0.0*	*0.1.0*