Final draft ETSI ES 283 039-2 V3.1.0 (2009-11)

ETSI Standard

Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); NGN Congestion and Overload Control; Part 2: Core GOCAP and NOCA Entity Behaviours

Reference
DES/TISPAN-03034-2-NGN-R3

Keywords
control, quality, protocol

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

Individual copies of the present document can be downloaded from:
http://www.etsi.org

The present document may be made available in more than one electronic version or in print. In any case of existing or perceived difference in contents between such versions, the reference version is the Portable Document Format (PDF). In case of dispute, the reference shall be the printing on ETSI printers of the PDF version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at
http://portal.etsi.org/tb/status/status.asp

If you find errors in the present document, please send your comment to one of the following services:
http://portal.etsi.org/chaircor/ETSI_support.asp

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://webapp.etsi.org/IPR/home.asp).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This ETSI Standard (ES) has been produced by ETSI Technical Committee Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN), and is now submitted for the ETSI standards Membership Approval Procedure.

The present document is part 2 of a multi-part deliverable covering NGN Overload and Congestion Control as identified below:

Part 1: "Overview";

Part 2: "Core GOCAP and NOCA Entity Behaviours";

Part 3: "Overload and Congestion Control for H.248 MG/MGC";

Part 4: "Adaptative Control for the MGC";

Part 5: "ISDN overload control at the Access Gateway".

1 Scope

The present document describes the core features of the NGN Overload Control Architecture (NOCA) and the Generic Overload Control Application Protocol (GOCAP). While it is usual for the architectural components to be specified separately from the protocols that are used to communicate between them, the performance requirements of overload controls are such that the coupling between architecture, protocol and implementation is very strong. This means that the present document specifies the architecture, entity behaviours and protocol for the core NOCA/GOCAP together. The way GOCAP and the NOCA entities are deployed to control traffic that uses a specific application protocol is profiled via additional small shim specifications.

2 References

References are either specific (identified by date of publication and/or edition number or version number) or non-specific.

- For a specific reference, subsequent revisions do not apply.
- Non-specific reference may be made only to a complete document or a part thereof and only in the following cases:
  - if it is accepted that it will be possible to use all future changes of the referenced document for the purposes of the referring document;
  - for informative references.

Referenced documents which are not found to be publicly available in the expected location might be found at http://docbox.etsi.org/Reference.

NOTE: While any hyperlinks included in this clause were valid at the time of publication ETSI cannot guarantee their long term validity.

2.1 Normative references

The following referenced documents are indispensable for the application of the present document. For dated references, only the edition cited applies. For non-specific references, the latest edition of the referenced document (including any amendments) applies.

[1] ETSI TS 182 018: "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); Control of Processing Overload; Stage 2 Requirements".

[2] IETF RFC 3588: "Diameter Base Protocol".

[3] IETF RFC 4005: "Diameter Network Access Server Application".

[4] ETSI TS 133 210: "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); LTE; 3G security; Network Domain Security (NDS); IP network layer security (3GPP TS 33.210)".

[5] ETSI TS 129 329: "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); LTE; Sh interface based on the Diameter protocol; Protocol details (3GPP TS 29.329 version 8.4.0 Release 8)".

[6] IETF RFC 3265: "Session Initiation Protocol (SIP)-Specific Event Notification".

2.2 Informative references

The following referenced documents are not essential to the use of the present document but they assist the user with regard to a particular subject area. For non-specific references, the latest version of the referenced document (including any amendments) applies.

[i.1] ETSI ES 283 039-4: "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); NGN Overload Control Architecture; Part 4: Adaptative Control for the MGC".

[i.2] IETF RFC 4662: "A Session Initiation Protocol (SIP) Event Notification Extension for Resource Lists".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the terms and definitions given in TS 182 018 [1] and the following apply:

application: software component(s) running on a system to provide service to end users or support the management of the system

NOTE: In the present document the term application excludes those software components that implement the NOCA.

application protocol: protocol used to enable application instances to communicate

control variable: time-varying parameter used to control actuators in a feedback loop, calculated on the basis of the target and measured values of some system quantity

feedback loop: control mechanism where the result of changing an actuator is used ("fed back") into the algorithm used to calculate future changes

load control: mechanism for controlling the workload of a system

overload: system workload exceeds a defined threshold of the processing capacity of that system

source: system that generates workload for another system

target: system that receives workload from another system

workload: amount of processing work a system has to perform

3.2 Abbreviations

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

AAA	AA-Answer
API	Application Programming Interface
ASA	Abort-Session-Answer
ASR	Abort-Session-Request
AVP	Attribute-Value Pair
CA	Control Adaptor
CD	Control Distribution
CDR	CDRestriction
CEA	Capabilities-Exchange-Answer
CER	Capabilities-Exchange-Request
CM	Channel Manager
FQDN	Fully Qualified Domain Name
GOCAP	Generic Overload Control Application Protocol

IP	Internet Protocol
ISUP	Integrated Service Digital Network User Part
NGN	Next Generation Network
NOCA	NGN Overload Control Architecture
PUA	Profile-Update-Answer
PUR	Profile-Update-Request
RM	Restrictor Manager
SCTP	Stream Control Transmission Protocol
SDL	Specification and Description Language
SIP	Session Initiation Protocol
SLA	Service Level Agreement
STA	Session-Termination-Answer
STR	Session-Termination-Request
TCP	Transmission Control Protocol

4 Control Architecture

4.1 Description of NOCA Components

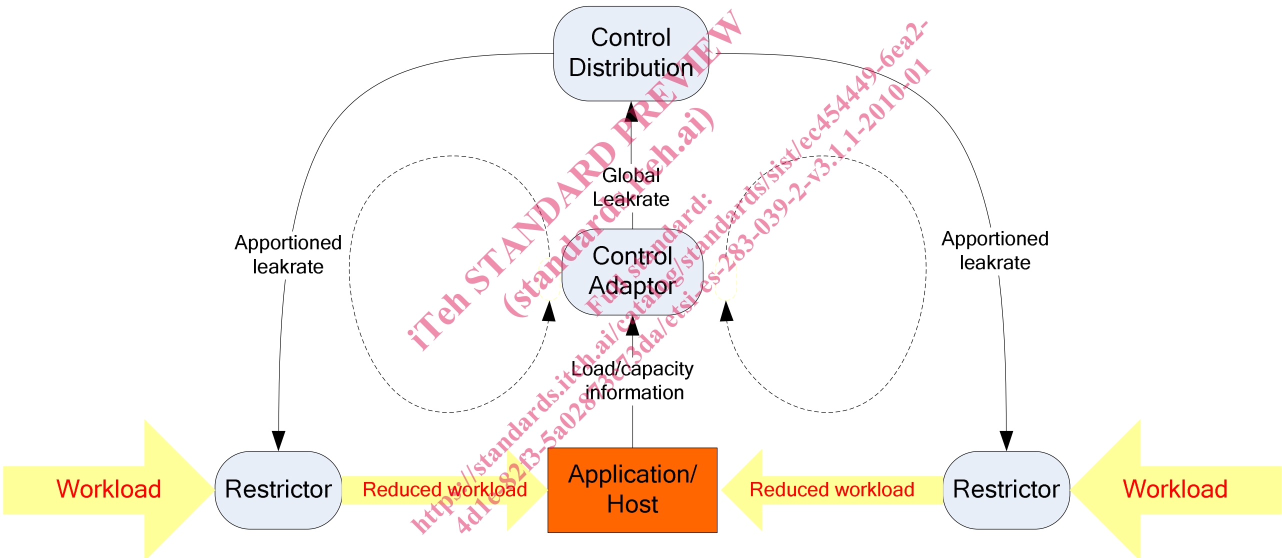

Figure 1: Control components implementing a feedback control path

The NGN Overload Control Architecture (NOCA) aims to provide feedback based processing load control for hosts that implement the functionality of NGN (and other) networks. Each feedback control loop indicated by the ovals in Figure 1, comprises three key NOCA components, the Control Adaptor, Control Distribution and Restrictor. The objective of the feedback loops is to enable the protected host to operate at the optimum load by restricting excess workload originating from its nearest neighbours, so the restrictors are usually (though not always) located at those nearest neighbours.

Control Adaptor:

The control adaptor receives data from the host system describing the current workload and system capacity and uses this to derive a global leakrate which is passed to the control distribution. The control adaptor then adjusts the global leak rate, in order that the current workload converges to a goal value equal to the system capacity.

Control Distribution:

> The control distribution component shares the global leak rate between the restrictors on the basis of simple local policies. These policies enable defined service levels or fairness requirements to be realised. The control distribution uses the Generic Overload Control Application Protocol (GOCAP) to transmit leak rate information to restrictors on remote hosts.

Restrictor:

> The restrictor is a leaky bucket rate limiter that is request priority aware. It is controlled by the leak rate received from the control distribution component and restricts the workload presented to the host which the control is protecting.

The components and functions shown in Figure 1 should be thought of as being located above the top level of the protocol stacks running at the overloaded target (the GOCAP Master) and sources (GOCAP Slaves). As an example, two communications application instances running on the source and target may use ISUP to communicate with each other in establishing and clearing-down voice calls. Beneath ISUP, the protocol stack (in descending order) could be SIP-I/TCP/IP.

In general, a host may be the source of excess processing load for other hosts as well as the target of excess processing load from those other hosts. This means that a single host typically implements all the control components, together with additional components to manage them. The components at a fully functional (from a GOCAP perspective) host are shown in Figure 2.

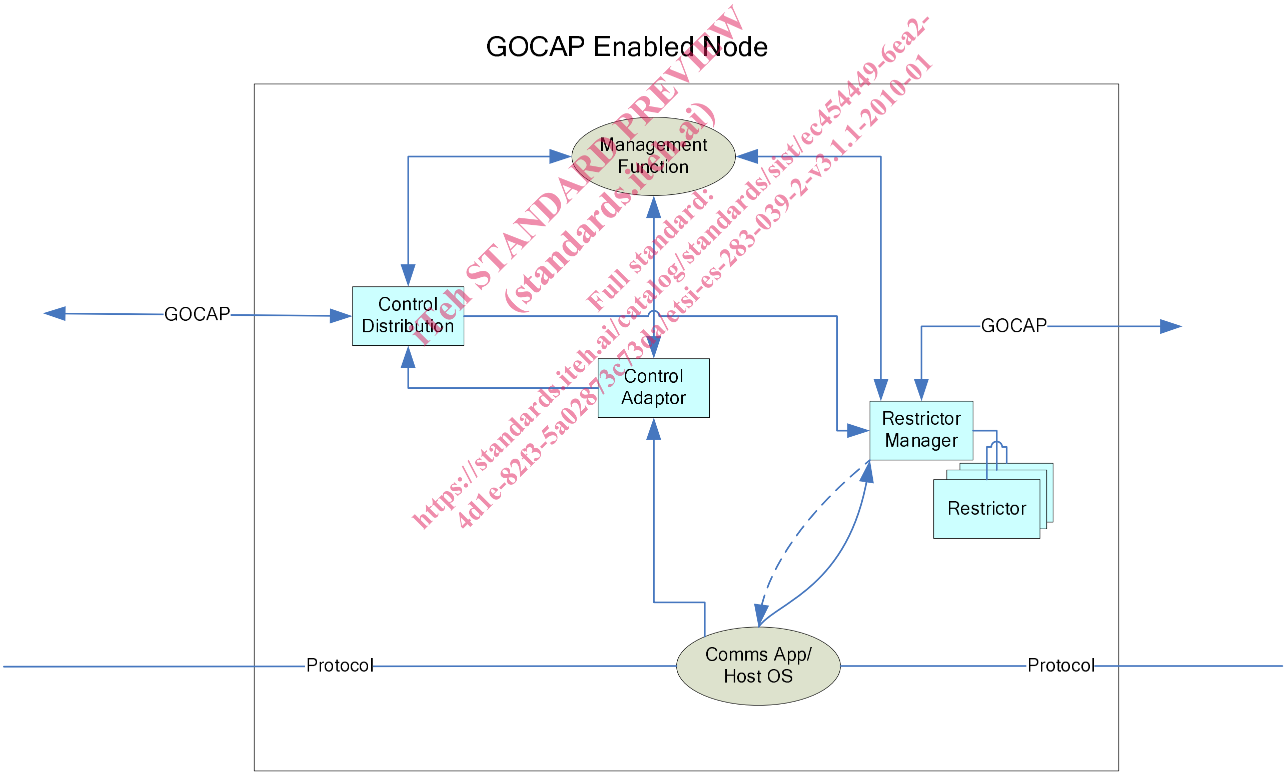

Figure 2: Structure of a GOCAP enabled host, arrows denote information flow - dashed arrows represent replies to requests

Figure 2 introduces an additional NOCA component, the restrictor manager. This component is responsible for co-ordinating multiple restrictors while providing a single interface to the host/application. All the NOCA components are described in greater detail in the next clause.

Figure 2 also shows that internal interfaces are required between NOCA components and the network management system for configuration, status enquiries, control statistics etc. Also shown are the interfaces between the host OS and/or communications applications which are used for admission queries (used to reduce workload) and for the protected system to send the data required to drive the control adaptor.

4.2 Detailed Description of NOCA Components and Behaviour

4.2.1 Overview

In the following clauses, the behaviours and attributes of the NOCA will be described in detail. To assist that description, SDL diagrams are included. The signals and data types used in the SDL diagrams are defined in Annex A.

NOTE: The present document uses SDL to describe the required behaviour, but does not enforce a particular method of realisation. SDL diagrams included in the present document are not meant to imply a particular implementation in terms of processes/threads, message passing or function/procedure calls. In many cases the signals given in the SDL diagrams may just be interpreted as function/procedure calls. The implementer is free to realise this behaviour in the most appropriate and efficient way.

Figure 3 shows a representation in SDL of a GOCAP system with the explicit addition of a GOCAP signalling channel between the GOCAP master and a remote slave (remote_gs). Figure 4 shows the key components of a GOCAP master - a host that is protected from overload by restrictions on GOCAP slaves. Note that the Control Distribution function has been split into two elements in Figure 4, the CDProcess and the CDRestriction. This is to help separate the description of the leakrate calculation elements of the Control Distribution (in the CDProcess) from those aspects that relate to the management of a particular restriction (the CDRestriction). Figure 5 gives a similar view of a GOCAP slave - a host that implements restrictors to limit the flow of requests.

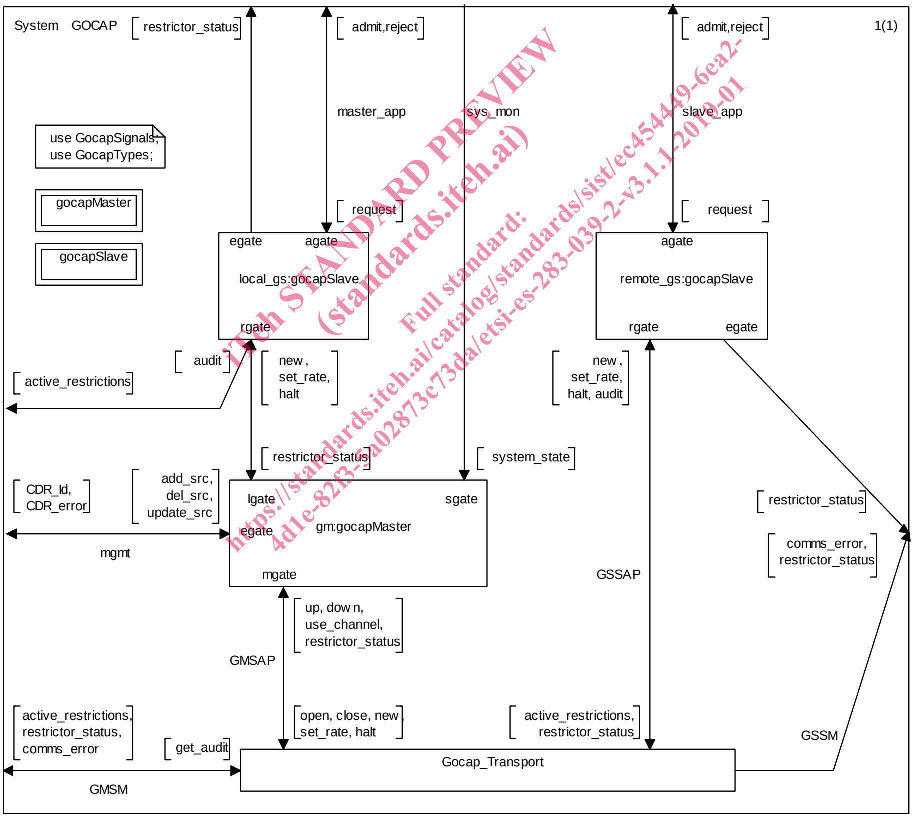

Figure 3: A representation of the GOCAP system using SDL

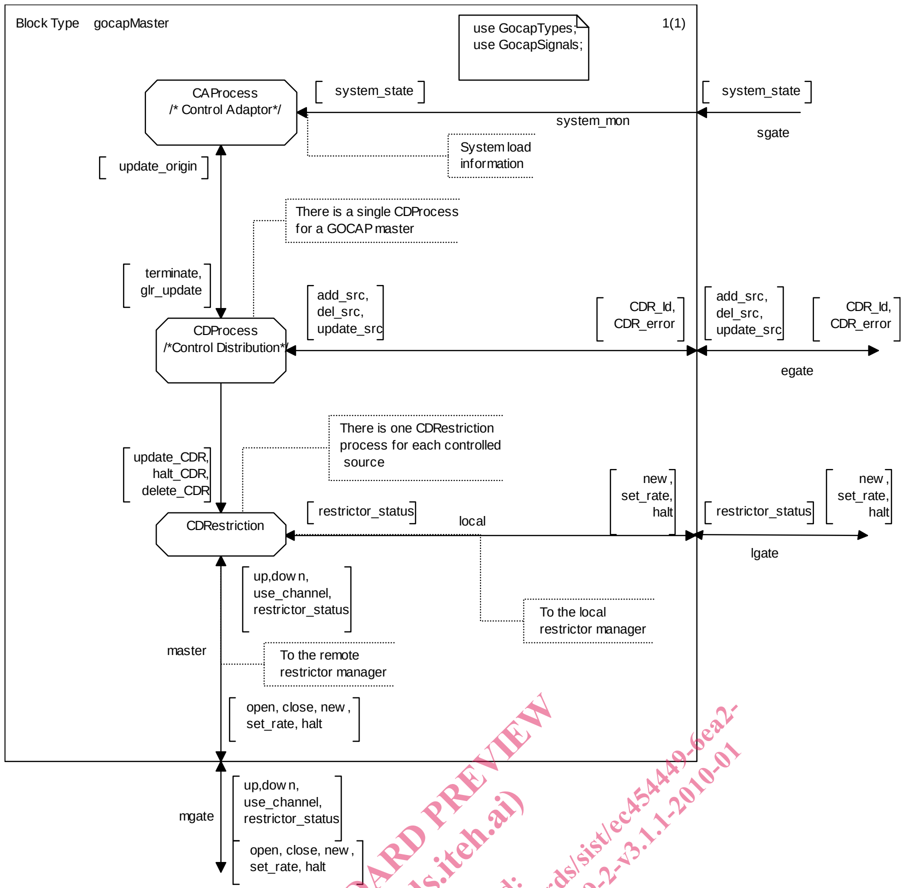

Figure 4: Signals and channels for a GOCAP Master. A node that hosts a GOCAP Master will usually host a GOCAP Slave too

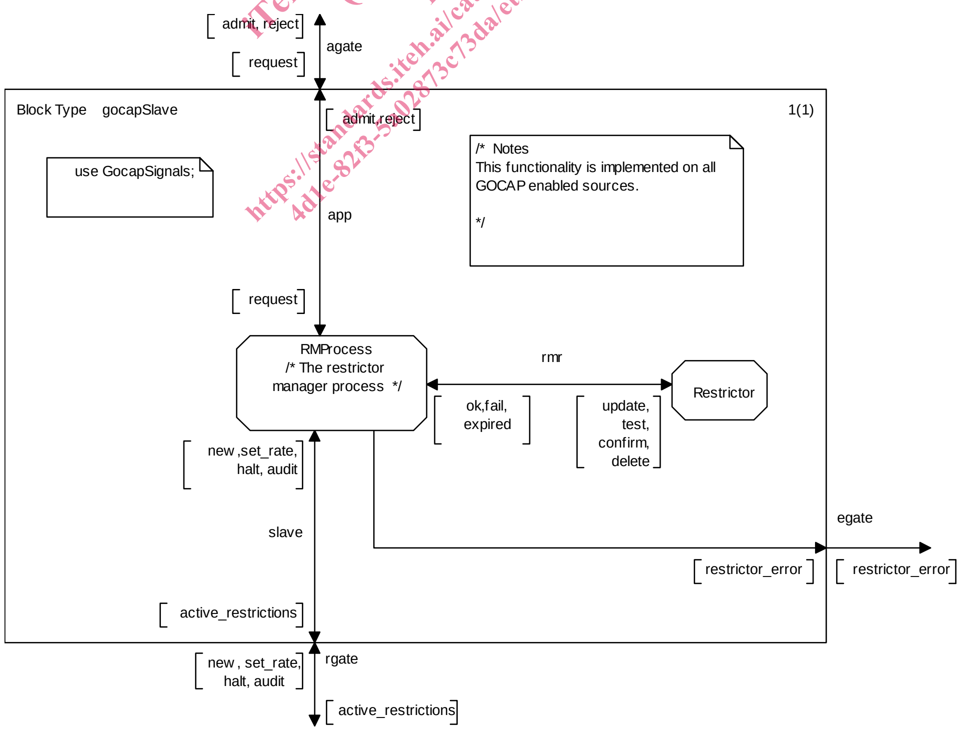

Figure 5: A GOCAP Slave, showing restrictor management