ETSI TR 104 073 V2.1.1 (2025-07)

TECHNICAL REPORT

Intelligent Transport Systems (ITS);
Facilities layer;
Radio Resource Management Study;
Release 2

Reference

DTR/ITS-246

Keywords

congestion control, ITS, radio resource management

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 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are 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 IPR online database.

Pursuant to the ETSI Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, 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.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™**, **LTE™** and **5G™** logo are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM**® and the GSM logo are trademarks registered and owned by the GSM Association.

Foreword

This Technical Report (TR) has been produced by ETSI Technical Committee Intelligent Transport Systems (ITS).

Modal verbs terminology

In the present document "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Introduction

Cooperative Intelligent Transportation Systems (C-ITS) Release 1 standards have been developed from out of the perspective of static information dissemination in which the applications and facilities layer services (ITS-S services) have no knowledge of the transmission capabilities in the available radio channels.

As the number of the applications and services provided by the facilities layer in Release 1 is limited, validations have shown that a predictable performance of these applications and ITS-S services can be realized as long as there is a radio channel usage agreement among stakeholders establishing an acceptable system/radio channel(s) behaviour. This is especially true since the Release 1 ITS services have a limited safety impact and by the C-ITS trust established by these stakeholder agreement and related C-ITS security.

In Release 2 ITS services with a higher safety impact are being introduced. The introduction of these ITS services lead to significant higher Quality of Service (QoS) requirements which include higher dynamic data-frame transmission needs. Since C-ITS data transmissions are mostly non-deterministic, the management of QoS in the network is limited, and the applications and ITS-S services are the only functionalities which can know the importance and impact of being able and being not able to exchange relevant information, further improvement of the QoS should be handled by the applications and ITS-S services themselves by being aware of the underlying transmission capabilities. Such knowledge enables the applications and ITS-S services to take appropriate measures in occurring communication situations.

As applications and ITS-S services do not know about the existence of other active applications and ITS-S services a general functionality which can keep track of all active applications and ITS-S services as well as being aware of all the available ITS-S lower layer functionalities and channel(s) should enable to increase the QoS. This functionality should be technology neutral and reside at the Facilities Layer (FL). Such functionality will manage the resources among the applications and ITS-S services and therefore called Resource Management (RM).

The present document evaluates possible RM approaches such that the most favourable version or versions can be derived in a future update of the present document. As a baseline, earlier results from the Multi-Channel Operation (MCO) study for Release 2 [i.1] have been used.

Any technology agnostic RM relies on proper information from lower layers about the dissemination possibilities and should ensure that no harmful interference can be caused by adjacent channel operation. The MCO study [i.1] showed that there are congestion requirements to comply to realize required non-interference.

The present document therefore also studies the related requirements which should be set for lower layers and how these should interface to higher layers. The maximization of technology agnostic parameter exchange between lower layers and the FL are also investigated. These parameters relate mostly to Congestion Control and its optimization.

Since Congestion Control handling at MAC/PHY layers can be handled differently for different radio access technologies, these aspects have to be investigated for the technologies that are considered.

While for the higher layer RM related requirements and parameters in the present document are intended technology agnostic, the present document with respect of addressing the lower technology specific layer aspects only the ITS-G5 technology is addressed and therefore also only the ITS-G5 related adjacent channel operations aspects, as identified in ETSI TR 1013 439 [i.1]. For the operation with other lower layer technologies such as LTE-V2X or NR-V2X further studies might be desirable.

There are several aspects to be investigated before a Resource Management solution can be defined. The present document addresses three specific aspects that need to be addressed for the definition of ITS resource management solutions: the analyses of the Release 2 ITS-S services, the identification of Congestion Control (CC) optimization possibilities and the identification of the adjacent channel interference management possibilities.

Other aspects such as CBR Management Method Selections and the selection of a CC Approach and Resource Management Process Selection will need to be considered to come to a full definition for a Release 2 RM specification, compatible with Release 1, but are not covered by the present document.

1 Scope

The present document identifies Radio Resource Management (RRM) at the Facilities Layer in relation to critical control requirements, capabilities, principles and parameters which could enable the definition of a mechanism supporting highly time and size dynamic data exchanging services to operate robust, interoperable and backward compatible with existing ITS Release 1 and upcoming Release 2 ITS services in the 5,9 GHz ITS allocated band.

2 References

2.1 Normative references

Normative references are not applicable in the present document.

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

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

The following referenced documents may be useful in implementing an ETSI deliverable or add to the reader's understanding, but are not required for conformance to the present document.

[i.1] ETSI TR 103 439: "Intelligent Transport Systems (ITS); Multi-Channel Operation Study; Release 2".

[i.2] ETSI TR 101 607: "Intelligent Transport Systems (ITS); Cooperative ITS (C-ITS); Release 1".

[i.3] ISO 26262-9:2018: "Road vehicles — Functional safety, "Part 9: Automotive safety integrity level (ASIL)-oriented and safety-oriented analyses".

[i.4] ETSI TR 103 903: "Intelligent Transport Systems (ITS); Specification Framework Release 2".

[i.5] ETSI TS 103 141: "Intelligent Transport Systems (ITS); Facilities layer function; Multi-Channel Operation (MCO) for Cooperative ITS (C-ITS); Release 2".

[i.6] ETSI TS 103 697: "Intelligent Transport Systems (ITS); Architecture; Multi-Channel Operation (MCO) for Cooperative ITS (C-ITS); Release 2".

[i.7] C2C-CC Profile Release 1: "Car2Car Communication Consortium - Basic System Profile (BSP)".

[i.8] C-Roads Profile Release 2: "C-ROADS Harmonised Communication Profile for C-ITS".

[i.9] IEEE 802.11™ (2024): "IEEE Approved Draft Standard for Information Technology -- Telecommunications and Information Exchange Between Systems Local and Metropolitan Area Networks -- Specific Requirements - Part 11: Wireless Local Area Network (LAN) Medium Access Control (MAC) and Physical Layer (PHY) Specifications".

[i.10] ETSI TS 103 836-4-1: "Intelligent Transport Systems (ITS); Vehicular Communications; GeoNetworking; Part 4: Geographical addressing and forwarding for point-to-point and point-to-multipoint communications; Sub-part 1: Media-Independent Functionality; Release 2".

[i.11] ETSI TS 103 695: "Intelligent Transport Systems (ITS); Access layer specification in the 5 GHz frequency band; Multi-Channel Operation (MCO) for Cooperative ITS (C-ITS); Release 2".

[i.12] ETSI TR 101 612: "Intelligent Transport Systems (ITS); Cross Layer DCC Management Entity for operation in the ITS G5A and ITS G5B medium; Report on Cross layer DCC algorithms and performance Evaluation".

[i.13] ETSI TS 102 636-4-2: "Intelligent Transport Systems (ITS); Vehicular Communications; GeoNetworking; Part 4: Geographical addressing and forwarding for point-to-point and point-to-multipoint communications; Sub-part 2: Media-dependent functionalities for ITS-G5".

[i.14] ETSI TS 102 687 (V1.2.1): "Intelligent Transport Systems (ITS); Decentralized Congestion Control Mechanisms for Intelligent Transport Systems operating in the 5 GHz range; Access layer part".

[i.15] ETSI TS 103 175: "Intelligent Transport Systems (ITS); Cross Layer DCC Management Entity for operation in the ITS G5A and ITS G5B medium".

[i.16] ETSI EN 302 571: "Intelligent Transport Systems (ITS); Radiocommunications equipment operating in the 5 855 MHz to 5 925 MHz frequency band; Harmonised Standard covering the essential requirements of article 3.2 of Directive 2014/53/EU".

[i.17] V. Todisco, S. Bartoletti, C. Campolo, A. Molinaro, A. O. Berthet and A. Bazzi: "WiLabV2Xsim: Performance Analysis of Sidelink 5G-V2X Mode 2 Through an Open-Source Simulator," in IEEE™ Access, vol. 9, pp. 145648-145661, 2021.

[i.18] Wu Z., Bartoletti S., Martinez V., Bazzi A.: "A Methodology for Abstracting the Physical Layer of Direct V2X Communications Technologies". Sensors 2022, 22, 9330.

[i.19] G. Bansal, B. Cheng, A. Rostami, K. Sjoberg, J. B. Kenney and M. Gruteser: Comparing LIMERIC and DCC approaches for VANET channel congestion control," 2014 IEEE™ 6th International Symposium on Wireless Vehicular Communications (WiVeC 2014), Vancouver, BC, Canada, 2014, pp. 1-7, doi: 10.1109/WIVEC.2014.6953217.

[i.20] M. Sepulcre, J. Mira, G. Thandavarayan and J. Gozalvez: "Is Packet Dropping a Suitable Congestion Control Mechanism for Vehicular Networks?," 2020 IEEE™ 91st Vehicular Technology Conference (VTC2020-Spring), Antwerp, Belgium, 2020, pp. 1-5, doi: 10.1109/VTC2020-Spring48590.2020.9128822.

[i.21] G. Thandavarayan, M. Sepulcre and J. Gozálvez: "Cooperative Perception for Connected and Automated Vehicles: Evaluation and Impact of Congestion Control", IEEE™ Access, vol. 8, pp. 197665-197683, October 2020, DOI: 10.1109/ACCESS.2020.3035119.

[i.22] F. Marzouk, R. Zagrouba, A. Laouiti, and P. Mühlethaler: "An empirical study of Unfairness and Oscillation in ETSI DCC", In International Conference on Performance Evalutaion and Modeling in Wired and Wireless networks PEMWN'2015, November 2015.

[i.23] A. Bazzi: "Congestion Control Mechanisms in IEEE 802.11p™ and Sidelink C-V2X", 53rd Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 2019, pp. 1125-1130, doi: 10.1109/IEEECONF44664.2019.9048738.

[i.24] E. Xhoxhi and F. Schiegg: "Benefits of DCC Facilities in ITS-G5 Networks - First Simulated Results", IEEE™ 95th Vehicular Technology Conference: (VTC2022-Spring), Helsinki, Finland, 2022, pp. 1-7, doi: 10.1109/VTC2022-Spring54318.2022.9860904.

[i.25] G. Thandavarayan, M. Sepulcre, J. Gozálvez, and B. Coll-Perales: "Scalable Cooperative Perception for Connected and Automated Driving", Journal of Network and Computer Applications, vol. 216, pp. 103655, May 2023.

[i.26] A. H. Sakr: "Evaluation of Redundancy Mitigation Rules in V2X Networks for Enhanced Collective Perception Services", in IEEE™ Access, vol. 12, pp. 137696-137711, 2024.

[i.27] Q. Delooz et al: "Analysis and Evaluation of Information Redundancy Mitigation for V2X Collective Perception", in IEEE™ Access, vol. 10, pp. 47076-47093, 2022.

[i.28] A. Figueiredo, P. Rito, M. Luís and S. Sargento: "Enhancing Vehicular Network Efficiency: The Impact of Object Data Inclusion in the Collective Perception Service", in IEEE Open Journal of Intelligent Transportation Systems, vol. 5, pp. 454-468, 2024.

[i.29] F. Schiegg et al.: "Automated Vehicle Marshalling: The First Functionally Safe V2X Service for Connected Automated Driving" in IEEE™ Open Journal of Vehicular Technology, vol. 6, pp. 927-947, 2025.

[i.30] G. Bansal and J. B. Kenney: "Achieving weighted-fairness in message rate-based congestion control for DSRC systems" Proceedings IEEE™ 5th WiVeC, Dresden, Germany, 2013.

[i.31] Directive 2010/40/EU of the European Parliament and of the Council of 7 July 2010 on the framework for the deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport.

[i.32] ETSI EN 303 797 (V2.1.1): "Intelligent Transport Systems (ITS); ITS-G5 Access layer in the 5 GHz frequency band; Release 2".

[i.33] Claudia Campolo, Antonella Molinaro, Riccardo Scopigno: "Vehicular ad hoc Networks, Standards, Solutions, and Research", Springer, ISBN: 978-3-319-15496-1 (Print), 978-3-319-15497-8 (Online).

[i.34] C2C-CC: "Vehicle C-ITS station profile #2037", RS_BSP_240.

3 Definition of terms, symbols and abbreviations

3.1 Terms

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

application: computer software functionality that performs a specific function directly for an or for another application

automated system: computer (ICT) system that collects information and can react and perform tasks based on the data

NOTE: Automated systems are made using three entities:

- Sensors.
- Microprocessors.
- Actuators.

communication flows: data path to be followed by the information through an ITS from the source ITS-S to the sinking ITS-Ss

direct communication: message exchange between data sourcing ITS-Ss and data sinking ITS-Ss, without involvement of intermediary networking routing functions

NOTE: Messages transported via Indirect Communication can be considered as Direct communicated when the messages cannot be processed (carried over in a secured envelope) along the indirect communication flow.

equivalent bandwidth: computed value coming from a computation of relevant statistical properties of the traffic source (e.g. Markov-modulated arrivals, moment generating functions, or large-deviation techniques), ensuring that the probability of exceeding resource limits remains below a defined threshold

NOTE: Equivalent bandwidth originates from wired networks and refers to the bandwidth that a given service effectively requires to meet its QoS constraints.

extra ITS: external ITS-S communication using a non-interoperable ITS or interference from non-ITS sources, also inter ITS

indirect communication: message exchange between data sourcing ITS-Ss and data sinking ITS-Ss, with the involvement of intermediary networking routing functions such as the Internet

intra ITS: internal ITS-S communication using interoperable ITS

ITS constellation: group of ITS-Ss which can exchange information

ITS service: service provided by an application to an end user or automated function

ITS-S service: Facilities Layer service which generates or processes data, triggered by an application or other ITS-S Service, or a Facilities Layer service which executes system control functions for the purpose of correct operation of other ITS-S services or applications

3.2 Symbols

Void.

3.3 Abbreviations

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

3GPP	3rd Generation Partnership Project
5G-NR	5th Generation New Radio
AI	Artificial Intelligence
AIFS	Arbitration Inter-Frame
AL	Access Layer
ALI	Access Layer Instance
ASIL	Automation Safety Integrity Level
AVM	Automated Vehicle Marshalling
BE	Best Effort
BME	Bandwidth Management Entity
CA	Cooperative Awareness
CAM	Cooperative Awareness Message
CAS	Cooperative Awareness Service
CBR	Channel Busy Ratio
CC	Congestion Control
CDF	Cumulative Distribution Function
C-ITS	Cooperative Intelligent Transportation Systems
CL	Channel Load
CP	Collective Perception
CPM	Collective Perception Message
CPS	Collective Perception Service
CSMA	Carrier Sense Multiple Access
CW	Continues Wave
DCC	Decentralized Congestion Control
DEN	Decentralized Environmental Notification
DENM	Decentralized Environmental Notification Message
DENS	Decentralized Environmental Notification Service
DoS	Denial of Service
ECC	Electronic Communications Committee
EIRP	Effective Ideal Radiated Power
EU	European Union
FAC	Facility
FL	Facilities Layer
FUSA	Functional Safety
GCBR	Global Channel Busy Ratio
GCRA	Generic Cell Rate Algorithm
I2V	Infrastructure to Vehicle
I2X	Infrastructure to Everything
ICT	Internet Communication Technology
ITS	Intelligent Transportation Systems
ITS-S	Intelligent Transportation Systems - Station
IVI	Infrastructure to Vehicle Information
IVIM	In Vehicle Information Message
LTE	Long-Term Evolution
MAC	Medium Access Control
MAP	Road topology
MAPEM	MAP (topology) Extended Message
MCM	Manoeuver Coordination Message

MCO	Multi-Channel Operation
MCS	MCS in radio communication
MHE	Message Handling Entity
MIM	Marshalling Infrastructure Message
MVM	Marshalling Vehicle Message
NL	Networking and transport Layer
NR	New Radio
OFDM	Orthogonal Frequency Division Multiplexing
PER	Packet Error Rate
PHY	Physical layer
PLCP	Physical Layer Convergence Protoco
PRR	Packet Reception Ratio
QAM	Quadrature Amplitude Modulation
QM	Quality Management
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RM	Resource Management
RRM	Radio Resource Management
RSSI	Received Signal-Strength Indicator
RSU	Roadside Unit
RWW	Roads Works Warning
SAM	Service Announcement Message
SDV	Slow Driving Vehicle
SINR	Signal to Noise Interference Ratio
SP	System Profile
SPATEM	Signal Phase And Timing Extended Message
SRTI	Safety Related Traffic Information
TC	Traffic Class
TR	Technical Report
TS	Technical Specification
V2V	Vehicle to Vehicle
V2X	Vehicle to Everything
VAM	Vulnerable road user Awareness Message
VoI	Value of Information
VRU	Vulnerable Road User
X2N2X	Anything to Network to Anything
X2X	Evertything to Everything

4 Considerations

4.1 Introduction

The ITS Release 1 basic Cooperative use cases, supported by the standards listed in ETSI TR 101 607 [i.2], are awareness-oriented. When considered from an Automotive Safety Integrity Level (ASIL) perspective, according to ISO 26262-9 [i.3], ITS Release 1 supports some use cases of the Quality Management (QM) ASIL. For this level, there are no liability or other quality requirements defined, as identified in the ITS Framework Release 2 ETSI TR 103 903 [i.4].

The ITS Framework Release 2 (addressing both Cooperative-ITS (C-ITS) and non-C-ITS use cases), detailed in ETSI TR 103 903 [i.4], identifies specific Release 2 ITS services that are expected to support ASIL C and potentially ASIL D.

The support of higher ASIL levels results in additional and more stringent data quality and system performance requirements, including additional QoS requirements.

The present document focusses on two QoS related system improvement aspects:

1) how the ITS Service information dissemination could be better scheduled statistically; and

2) at the ITS Station (ITS-S) system level, how the information dissemination could be better managed dynamically.

Both aspects are system aspects from which the first should be specified in System Profiles (SPs). For Release 1 the C2C-CC Profiles [i.7] and the C-Roads Profiles [i.8] have specified this to enable a predictable static message exchange behaviour. Both profiles refer to the ETSI standards listed in the ETSI TR 101 607 [i.2] for Release 1.

With regards to the second aspect, aspects related to the influence of ITS Services information exchange are identified in clause 4.2. Aspects related to the information dissemination QoS are clarified in clause 4.3. Finally, in clause 4.4, the system context is illustrated.

4.2 Message Handling

Message handling is the static setting of C-ITS information exchange in a specific environmental context. As stated in clause 4.1, C2C-CC and C-Roads limit the Release 1 information dissemination to V2V, I2V message exchange in a single direct communication specific 10 MHz channel. C-Roads identifies other I2X/X2I direct communication into other channels for early C-Roads Release 2 within the same C-ITS Ecosystem QoS.

In this context the C-Roads hybrid profile [i.8] (Internet based) is not considered at present for this study as it does not provide dissemination QoS mechanisms for higher levels of safety and automation as expected for Release 2.

At present according to current Functional Safety Assessments (FUSAs), Internet information exchange does not provide sufficient QoS such that it can be used for safety related information exchange to satisfy higher ASIL levels. As such related information exchange is not considered in this study.

At the system level QoS are realized by means of stakeholder agreements and technical profiles such as the one from C2C-CC and C-Roads. As such they cannot be considered in this study.

4.3 Quality of Service improvement

In Information and Communication Technology (ICT), QoS cannot be fully guaranteed. The main three areas, in which QoS mechanisms are realized:

1) From a functional perspective internet communication capabilities are business driven statistically estimated and adjusted ensuring a general level of QoS. Initially this was only statically realized and later extended to support exceptional (large) events such as a football game with over 50 000 spectators and based on related statistics. For instance, when there is a large sport event, additional equipment is temporarily installed. This can be seen as scenario management.

2) Internet protocol communication in general is handshaking based which allows the disseminating station to verify whether information is received. Direct communications are sensor networked (broadcast/unicast) based in which this verification cannot take place. QoS is reached by system agreements and resource assessment.

3) From a technical perspective information can for example be given different priorities or be offered different communication capabilities fitting to the information behaviour. This however always goes with a cost in which lower priorities will have to pay for the burden of the higher priorities.

In an open system such as the internet the large number of users (millions of users) damp the dynamic behaviour in general which has a positive effect on the QoS, by which the above three mechanisms are sufficient to realize an acceptable QoS for most internet services.

Information exchange in the domain of transport very quickly has safety impact and requires a higher level of QoS. While some of the ITS Release 1 use cases can still be realized with an internet level of QoS, extended ITS Release 2 use case QoS requirements are more demanding.

One aspect of satisfying automated mobility systems is to provide higher levels of QoS to increase the probability so that information can be disseminated and reaches the users of that information under time environmental critical conditions while meeting other, for instance time, requirements. The certainty depends largely on the openness and accessibility of the network. This includes ensuring that there is enough bandwidth available as to aspects of reducing the sensitiveness to Denial-of Service (DoS, a typical cyber-attacks), attacks which make use of the bandwidth limitations of the network. Only when common agreements in a closed communication environment are guaranteed, a high level of QoS can be reached.

In case of safety related services, specific levels of QoS are required as otherwise the safety related operation cannot be guaranteed. As QoS are coming with a price, discussions are started to what level of costs the safety should be guaranteed. This is a society cultural question and generally out of scope of standardization. However, C-ITS is of interest for a specific group of stakeholders and not to the society as a hole, which allowed the user to come to an agreement about the requirement levels for QoS.

For advanced C-ITS services and automated systems several approaches are being developed to realize a controlled system environment for extended trustworthiness and QoS needed for Release 2.

In the C-ITS ecosystem the message behaviour depends on mostly non-predictable events and circumstances. Within this ecosystem various parties are interested in different C-ITS services and make use of the communications capabilities in different ways. To realize a trustworthy communication behaviour, organizational agreements (regulations and/or consortia) for establishing a stable environment are required. As the C-ITS information dissemination is very dynamic in time and size, additional technical measures are needed to enable applications and ITS-S services to be aware of the communication capabilities and to allow them to take the appropriate measures in any traffic scenario. As a result, for C-ITS but also for other ITS services, resource management methods will increase the dissemination trustworthiness. As this is a system related trustworthiness, each participating ITS-S needs to address the trustworthiness and QoS in a standardized manner as otherwise the resulting behaviour of a group of ITS-S's cannot be predicted.

A basic trustworthiness and QoS improvement method is to increase the media (radio) access possibilities by e.g. increasing the communication bandwidth or by realizing agreements among the stakeholders on how to use the available media. This is useful for many of the non-safety or limited safety relevant ITS services.

As spectrum is always limited, the available bandwidth for exchanging information is limited. For more critical safety requirements, (ASILs C-D) this limitation can be an issue, and therefore more stringent measures are needed to realize a better QoS.

The way to realize an acceptable QoS depends on the network configuration, it is in a direct ITS-S - ITS-S (X2X) or in a ITS-S - Service Provider - ITS-S (X2N2X) network (see ETSI TR 103 903 [i.4] and clause 4.4.5 for details on the Networking & Transport layer involvement).

It could be that all stakeholders agree for a specific use case only X2N2X networks are used.

It could be what information should be disseminated, like it is currently realized based on agreements (and associated profiles) between Car2Car Communication Consortium and C-Roads in Europe, where it has been agreed what Release 1 messages is at least disseminated in the single 10 MHz channel based on the IEEE 802.11p [i.9] access layer technology but can be, in parallel, be distributed via X2N2X networks.

Practice has shown that for basic Safety Related Traffic Information (SRTI) information exchange both methods are useful.

When moving to more critical safety and automation Release 2 use cases, where more stringent QoS requirements surface, X2N2X QoS capabilities are expected not to grow with these needs, and direct data exchange between data source and data sink (ITS-S to ITS-S) without involvement of networks may therefore be the only possible solution for use cases with higher QoS needs.

The present document intends to introduce a method which further increase the QoS enabling the applications and ITS-S services to be aware of the dissemination capabilities and whether they are influenced by other applications in the same ITS-S or in other ITS-S, being these applications disseminated in the same or other channels.

The present document therefore focuses on aspects which are relevant for the definition of a Resource Management (RM) method which supports an QoS improvement.

4.4 The system context

4.4.1 Overview

Previous work related to resource management was realized by the Multi-Channel Operation (MCO) work resulting in the MCO architecture for C-ITS as specified in ETSI TS 103 697 [i.6]. Present MCO functionalities are specified in the ETSI Release 2 standards, ETSI TS 103 141 [i.5], ETSI TS 103 836-4-1 [i.10] and ETSI TS 103 103 695 [i.11].

Figure 1 represents the MCO architecture as specified in ETSI TS 103 697 [i.6]. The RM has effect on the aspects at the various layers, as clarified in the following clauses.

The MCO specifications focussed on the realization of the multi-channel use, while the Resource Management (RM) extends and generalizes this concept also for use by single channel systems and is focussed to increase the QoS. The MCO architecture and related functionalities are therefore the basis for a wider resource management specification.

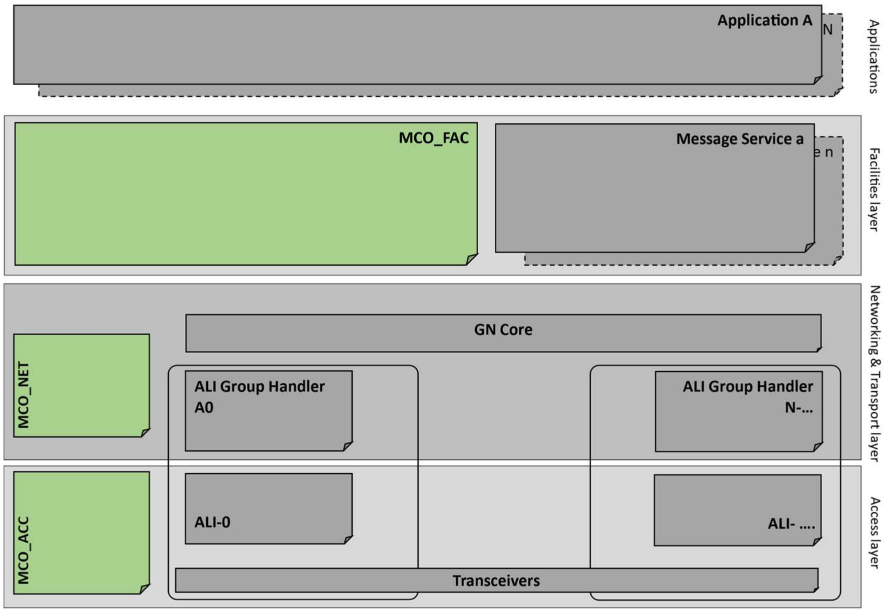


Figure 1: MCO Architecture
(Source: ETSI TS 103 697 [i.6])

With RM at the Facilities layer, a functionality maintaining dynamic knowledge about the lower layer capabilities, it realized higher levels of QoS by advising and directing applications about their disseminating possibilities by which the applications are enable to take appropriate functional and technical measures as required for the use case or use cases it realizes. To allow the RM to fulfil its tasks, it communicates to the applications as well as to relevant lower layer functionalities as required. While the RM itself is technology agnostic, it can interface with technology specific entities as lower layers, although it expects technology agnostic parameter values. For the realization of the RM processes, RM related technology agnostic and technology specific functions can be expected at all layers. From a lower layer perspective, the RM is especially dependent on information about the Channel Load (CL), which is an aspect related to the Access Layer (AL) to be able to realize QoS improvements. In the following clause, relevant RM specific functional and technical aspects at the different layers are recognized, and relevant studies are identified.

4.4.2 Channel Load

The bandwidth of a channel is always limited, and it needs to be used efficiently. This means maximizing the data that can be exchanged in such channel. At the same time, overusing the channel may cause large interference and high error probability at the receivers. As a result, there is a maximum possible use of the channel and approaching it allows to optimally use the resources. How this is handled, depends on the channel management principle that is used.

The specific split of radio resources, which can be in the frequency, time, space or code or a combination of these splitting principles, is technology specific.

In ITS-G5 based on IEEE 802.11 [i.9] technologies resources are split in the time domain with a listen-before-talk mechanism, whereas in LTE-V2X and 5G-NR-V2X sidelink there is a grid of time and frequency slots with synchronous access and scheduling procedures. In ITS Release 1, there are separate mechanisms identified for ITS-G5 and for LTE-V2X sidelink. For ITS-G5 the mechanism is specified in several ETSI specifications as identified in ETSI TR 101 607 [i.2] and for LTE-V2X sidelink by ETSI and 3GPP specifications.

In Release 2, an access layer technology agnostic solution at the higher layers needs to be defined to facilitate the use of various access layer technologies in the ITS spectrum bands. This implies the need to abstract the definition of used and available resources from the access layer to the layers above.

The measurement of the channel occupation and the evaluation of the available resources remain anyway technology dependent and is therefore performed in the access layer, defined separately per each technology.

When looking at the single channel and access technology, it is important to note that besides the inter ITS interference coming from ITS frames using the reference ITS access technology (ITS access technology used by the measuring ITS-S), extra ITS interference needs to be taken into account. When this external interference is above a given threshold it might also contribute to the channel load. Interference generated by a competing ITS technology can be also part of this extra (inter) ITS interference since it cannot be recognized by the receiver as useful signal. In the future, a differentiation between intra and extra ITS signal and interference might be required having in mind that ITS is operating in a shared environment.

4.4.3 Access Layer operation

The Access Layer (AL) is composed of the Medium Access Control (MAC) and Physical sub-layers and handles the radio spectrum access for the transmission of frames. Higher layers can provide data packets to the access layer and the access layer will decide whether or not to send frames, and when to transmit them by the mechanisms of the media access protocol.

With the aim to identify the occupation of the channel, in many systems including ITS-G5, the Channel Busy Ratio (CBR) is a metric derived from measurements which gather information about the load in the channel. This information can be provided to higher layers to inform them about the state of the channel from the local perspective of the ITS-S specific transceiver. The measurement performed locally can also be shared through higher layers with the neighbouring stations to improve awareness of channel use, as discussed in clause 4.4.5 in more detail. In Release 1, the CBR was directly related to a system CBR limit to control the amount of data that could be sent. The validity of such limit and the solutions defined in Release 1 to guarantee the respect of such limits are further elaborated in clause 5.

4.4.4 Adjacent channel

The MCO report ETSI TR 103 439 [i.1] investigated the influence of operating information dissemination in adjacent channels and concluded that the traffic in one channel affects the resources available in the adjacent ones. The result also shows that the second adjacent channel can be neglected. This means that the maximum occupation allowed in one channel may affect the performance and thus the maximum occupation of the adjacent ones. These aspects are further investigated in clause 5.4.

4.4.5 Networking & Transport Layer operation

The basic operation of the Networking & Transport Layer (NL) is responsible for directing packets to the appropriate radio configuration at the AL. In standard ICT it is the layer where it is decided what AL configuration is used to route packets to. For some ITS ecosystems (see ETSI TR 103 903 [i.4]) there is no routing at the NL. According to Release 1, the NL can only be used for forwarding packets within a specific area. In addition, Release 1 defines mechanism to share the CBR measured locally by the ITS-G5 with neighbouring ITS-Ss, with the scope to have a better estimation of what is called Global CBR (GCBR). For other access technologies this could be realized differently. As any possible improvement of NL related standards depend on the improvements realized on other layers, at present no investigations are realized at the NL, but may be added in an update of the present document.

4.4.6 Applications, Facilities Layer and Resource Management

At the Facilities Layer (FL), the timely varying spectral resource capabilities come together with the timely changing functional requirements of applications and ITS-S services. As identified in the Multi-Channel Operation (MCO) concept study (ETSI TR 103 439 [i.1]), applications do not have any notion of the existence of other applications and their communication needs, therefore the Resource Management (RM) (MCO management in ETSI TR 103 439 [i.1]) in the FL intends to facilitate a predictable use of the various available channels.

The RM functionality in the FL not only selects a specific channel for the use by a service or application but also supports the use of multiple channels including the resource distribution among the requestors or resources. It appears preferable to rename the MCO FL functionality as defined in ETSI TR 103 439 [i.1], into a more general RM.

The RM is a FL functionality which should replace what is currently specified as the MCO-Facilities (MCO-FAC) in ETSI TS 103 141 [i.5].

The RM operation depends on the capabilities of the lower layers and on the proper operation of the applications and the message generating and manipulating ITS-S services. To allow the RM to provide robust operation of applications and services, it is required that all dataflow functionalities at all layers, from applications to AL functionalities, operate with similar trustworthiness and QoS.

In clause 5, relevant lower layer robustness improvements are depicted; in clause 6, the application and ITS-S service parameters and possibilities are considered; and in clause 7, the functional layer mechanisms are investigated and considered.

5 Congestion Control Optimization

5.1 Introduction

In Release 1, congestion control is managed separately for different ALI groups. The extended concept as presented in the present document is based on ITS-G5 technology Access layer analysis but is not limited to this specific Access layer technology. For use with other technologies additional studies might be needed.

This clause focusses on Decentralized Congestion Control (DCC) in a single channel, reviewing the standards of Release 1 and identifying the aspects that in Release 2 can benefit from modifications or require modifications to comply with the MCO framework.

5.2 Decentralized congestion control in Release 1

5.2.1 Introduction

This clause provides a synthetic review of DCC as it is defined in Release 1. First of all, it recalls the definition of the main metrics used for congestion control and then details the limits that apply and the constrains imposed to the congestion control algorithm. The congestion control algorithm to be used is not specified by the standards but needs to respect the given constraints.

5.2.2 Overview of Release 1 DCC

The general description of Release 1 DCC is provided in ETSI TS 103 175 [i.15], which details the Entities at the various layers of the C-ITS protocol stack. In Release 1, DCC has Entities at the access layer, detailed in ETSI TS 102 687 [i.14], and at the networking & transport layer, detailed in ETSI TS 102 636-4-2 [i.13].

5.2.3 Metrics

DCC is based on the concept of Channel Busy Ratio (CBR), as defined in ETSI EN 302 571 [i.16], which determines the level of load of the channel. More specifically, the receiver determines every T_{CBR}, set to 100 ms, the time when the strength of the received signal exceeds -85 dBm, called T_{busy}. The CBR is calculated every T_{CBR} as:

$$CBR = \frac{T_{busy}}{T_{CBR}} \tag{1}$$

NOTE: It is common use to consider the own transmission time as part of T_{busy}, even if this is not explicitly written in the current specifications.

The measured CBR can be elaborated at the networking & transport layer and exchanged with the neighbouring nodes up to the second hop. This allows, as detailed in ETSI TS 102 636-4-2 [i.13], to obtain average values of the CBR in a larger area than the one directly observed by the station. This aspect is further discussed in clause 5.2.6.

Other metrics that are relevant for the DCC process are:

- The duration of a single transmission, denoted as T_{on}.
- The time between the end of a transmission and the beginning of the following one, denoted as T_{off}.

- The portion of time within a reference interval (set to 1 second) when the station is transmitting in the given channel, called *duty cycle*.

5.2.4 Limit imposed to the channel load

The limit for the channel load of each station is indicated in the ETSI EN 302 571 [i.16], which in turn is based on ETSI TS 103 175 [i.15] and ETSI TR 101 612 [i.12]. The limit provides a minimum time between two consecutive transmissions where the transceiver is not allowed to generate a signal (minimum T_{off}, as defined in clause 5.2.3), based on the duration of the last transmission (T_{on}, as defined in clause 5.2.3) and the measured CBR.

For any value of the CBR, T_{off} needs to be larger than 25 ms. If the CBR is equal or above to 0,62, the T_{off} needs also to comply with equation (2):

$$T_{off} \geq \min\left\{1\ 000, T_{on} \times \left(4\ 000 \times \frac{CBR - 0{,}62}{CBR} - 1\right)\right\} \tag{2}$$

with T_{off} expressed in milliseconds. The rationale behind the equation above is discussed in Annex A.

In addition, ETSI EN 302 571 [i.16] indicates a maximum T_{on} of 4 ms and a maximum duty cycle of 3 %.

5.2.5 Release 1 DCC at the access layer and constraints to the algorithm

ETSI TS 102 687 [i.14] details the constraints for the algorithm to be implemented at the access layer. The specific algorithm is not defined but left to the implementer. The constraints are as follows:

- The algorithm has to run on each frequency channel specified in ETSI EN 302 571 [i.16] independently.
- The algorithm has to run in an infinite loop.
- The algorithm has to be activated at least every 200 ms.
- The algorithm does not have to exceed the limits discussed in clause 5.2.4.

In the same document, two possible classes and an example for each of them are reported.

The first class is called reactive and consists in the use of a number of states. The station moves from one state to another based on the measured CBR; the state defines the value of the minimum T_{off}. The minimum T_{off} that follows from each state can be dependent on T_{on} (for example, two different values can be defined if the packet is small or large). ETSI TS 102 687 [i.14] does not specify the number of states or the values to be used per each state, but gives an example in its Annex A. In the reactive approach, at each step the station measures the CBR, identifies the state, and sets the minimum T_{off} accordingly. The minimum T_{off} has thus a granularity that is defined by the number of states and the associated values.

The second class is called proactive and increases or reduces the minimum T_{off} based on the measured CBR. In this case, ETSI TS 102 687 [i.14] specifies the calculations that need to be performed from the measured CBR to the minimum T_{off}, also including the values of the several parameters used by the calculations. ETSI TS 102 687 [i.14] also describes, in its Annex B, a possible implementation of the "gate keeping" function at the interface to the networking & transport layer to cope with messages of variable size.

5.2.6 Release 1 DCC at the networking & transport layer

ETSI TS 102 636-4-2 [i.13] describes the media-dependent functionalities for ITS-G5 at the networking & transport layer, also called DCC_NET. If GeoNetworking is implemented, DCC_NET is mandatory. If it is implemented:

- It maintains a number of DCC state variables (CBR_L_0_Hop, CBR_L_1_Hop, CBR_L_2_Hop, CBR_R_0_Hop, CBR_R_1_Hop, CBR_G, and CBR_Target), also using the Location Table Entry Extension for ITS-G5 (LocTEX-G5); LocTEX-G5 is an extension of the location table of the GeoAdhoc router.
- It periodically calculates the global Channel Busy Ratio, CBR_G.
- It processes and provides DCC-related information from/to DCC_CROSS.

- It transmits and receives DCC-related information to other GeoNetworking routers using the extensions for GeoNetworking packet handling.
- It optionally sets the transmission power limits, as specified in ETSI TS 103 175 [i.15], clause 6.2.

Regarding the state variables CBR_L_0_Hop, CBR_L_1_Hop, and CBR_L_2_Hop, they are the CBR measured by the station (0 to indicate no hops), the CBR received by its neighbours (1 hop), and the CBR that its neighbours have forwarded as they have in turn received from them their neighbours (2 hops), respectively. CBR_L_1_Hop and CBR_L_2_Hop are conservative values derived from the possibly several CBR values received from the neighbours, with the specific calculations described in ETSI TS 102 636-4-2 [i.13], clause 5.3.

CBR_R_0_Hop and CBR_R_1_Hop are the measured CBR shared by the station and the received CBR shared by the station, calculated as an average of those received from the neighbours.

Per each received packet, each station stores the CBR_R_0_Hop and CBR_R_1_Hop received from all the neighbours in the LocTEX-G5, together with information about transmit power and Received Signal-Strength Indicator RSSI.

Each station then calculates the *global CBR* CBR_G based on CBR_L_0_Hop, CBR_L_1_Hop, and CBR_L_2_Hop, which are in turn evaluated as a function of all received CBR_R_0_Hop and CBR_R_1_Hop. CBR_G is conservatively calculated, as detailed in ETSI TS 102 636-4-2 [i.13], clause 5.3. These calculations are performed every 100 ms.

CBR_Target is finally the value used by DCC as target, which needs to be the same as the one used at the access layer; currently, it is set to 0,62. The value of CBR_Target is used during the calculation of CBR_G.

5.2.7 Release 1 DCC at the facilities layer

The overall DCC architecture detailed in ETSI TS 103 175 [i.15], includes at the facilities layer a DCC_FAC function. However, DCC_FAC was not eventually defined in Release 1.

5.3 Critical analysis of Release 1 DCC

5.3.1 Introduction

In this clause, an analysis is conducted of the main aspects that could be revised in Release 2.

5.3.2 Analysis of the limit

When looking at the single generic channel, the limit imposed in Release 1 implies restrictions when the CBR exceeds 0,62. This number, which was chosen having a few services and relatively small messages in mind, appears lower than the maximum channel load tolerated in ITS-G5. It appears therefore reasonable to verify if such value is still a good choice or it should be revised once moving to Release 2. It can be noted that an increase of the limit would mean that the already enrolled devices still comply, while a reduction of the limit would need to include exceptions for the devices implemented following the specifications in Release 1.

With the aim to investigate this aspect, a simulation campaign was performed using the open-source WiLabV2Xsim [i.17]. Selected results are shown hereafter assuming a variable number of cars distributed on a highway segment of 8 km with 3 lanes per direction. The simulations are intended to obtain general considerations without focusing on a specific service; for this reason, each vehicle is assumed to transmit 10 packets per second, and all packets are of the same size; different values for the packet size are assumed in different simulations . In particular, 350, 550, and 1 000 bytes are considered, where the first value is a reasonable reference for Release 1 messages and the last one for larger Release 2 messages. These three values correspond to a duty cycle of approximately 0,58 %, 0,84 %, or 1,44 %.

ITS-G5 nodes are assumed to transmit at 23 dBm Effective Isotropic Radiated Power (EIRP), with 3 dBi antenna gain at the receiver, and 6 dB receiver noise figure. The access category Best Effort (BE) is adopted, corresponding to Arbitration Inter-Frame Spacing (AIFS) equal to 110 µs, Congestion Window (CW) equal to 15. The sensing threshold, when a preamble is detected, is set to -85 dBm, whereas it is set to -65 dBm otherwise. The modulation and coding scheme corresponding to Quadrature Phase Shift Keying (QPSK) and 1/2 coding rate is assumed. The modified Electronic Communications Committee (ECC) rural model from ETSI TR 103 439 [i.1] is used for the propagation, with correlated shadowing with standard deviation 3 dB. The model implies a loss exponent of 2 up to 128 m, then 2,8 up to 512 m, and 3,3 m above 512 m.

To evaluate the correctness of the decoding of each received signal, the average Signal to Noise and Interference Ratio (SINR) is calculated, where both noise and interference are assumed Gaussian and White. The reception is identified as correct when the SINR is above a given threshold as elaborated in "Methodology for Abstracting the Physical Layer of Direct V2X Communications Technologies" [i.18] Following the approach detailed in the same reference, the threshold is set to 1,2, 1,9, and 2,4, for 350, 550, and 1 000 bytes, respectively.

Results are provided in terms of measured CBR, Packet Reception Ratio (PRR) and Packet Error Ratio (PER). The PRR is obtained dividing the number of correctly decoded packets to the number of packets attempted to be decoded. The PER is the complementary value, i.e. the number of packets not correctly decoded divided by the number of packets attempted to be decoded.

Figure 2 represents the CBR varying the density and with different packet size. More specifically, it shows the median (i.e. the 50th percentile) of the distribution of the CBR measured in each interval by each station. As expected, the CBR increases almost linearly for low densities where collisions are limited and then slower when collisions become frequent. In addition, as expected, the increase is faster with larger packets. What is instead less obvious and therefore relevant to remark is that slower increase in CBR occurs at larger CBR values when the packets are larger; this effect is due to the lower relative impact of the inter-frame spaces when the packets are larger.

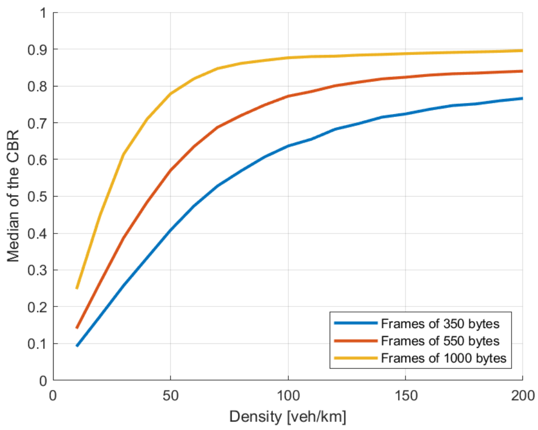


Figure 2: Median of the CBR varying the density

In Figure 3, the median CBR is shown in the x-axis and the PRR is plotted in the y-axis, when implicitly varying the density. Specifically, the PRR is calculated as an average at a distance of 400 m between transmitter and receiver. These results further show that the maximum CBR that can be tolerated changes when different packet size is assumed, with a higher CBR corresponding to a larger packet if the same PRR is assumed. This means that the larger packets expected in Release 2 may allow to increase the limit for congestion control. However, the increase appears limited.

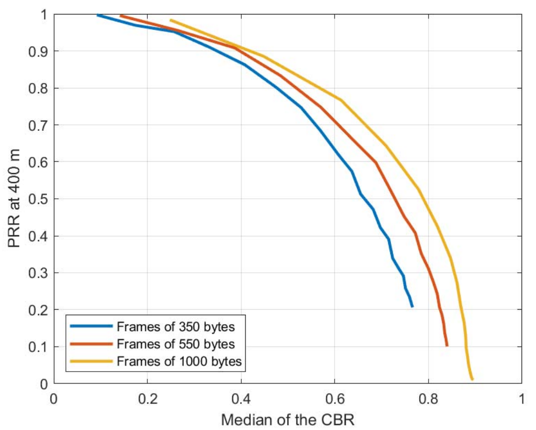


Figure 3: Average PRR at 400 m varying the median CBR

In Figure 4, the median CBR and a measure of the average number of messages that are correctly decoded at 400 m per km and per second are shown varying the average messages sent per km and per second. Also in this case, the results are obtained by implicitly varying the vehicle density. Looking at the blue curves, corresponding to the median CBR, it can be noted that trends are consistent to those shown with the density in the x-axis in Figure 2; indeed, since 10 packets per second are sent by each vehicle, the average sent messages per km and second is proportional to the density. Looking at the red curves, which more specifically shows the product between the average number of messages sent per km per second and the PRR at 400 m, it can be noted that they increase for small values of the sent packets (i.e. for low vehicle density) and then decrease after a maximum is reached. In each of the three red curves, a star is added when the median CBR corresponds to 0,62; as observable, the marked values are close to the maximum of the curves, meaning that a load on the channel higher than the limit imply an excessive increase in the collision probability.

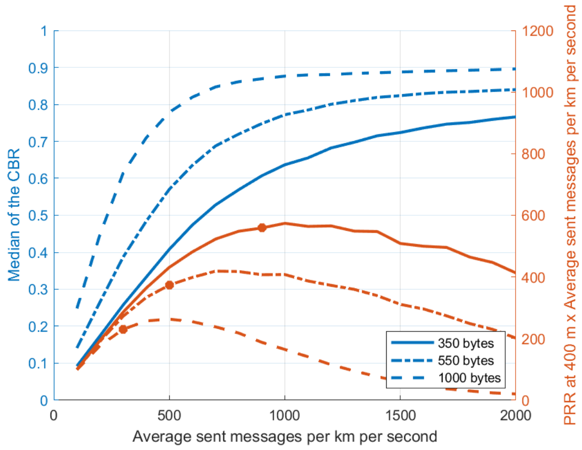


Figure 4: Median CBR and measure of the average number of messages correctly received at 400 m per km per second

Finally, in Figure 5, the PER at 400 m is shown together with the CBR varying the average messages sent per km and per second. Looking at the PER, it increases with larger packets, both due to a higher occupation of the channel and a higher SINR required to correctly decode the packet. Also in this case, the values corresponding to a median CBR equal to 0,62 are marked with a star, showing that the PER at 400 m is between 0,2 and 0,4 when the channel occupation is close to the limit imposed by current specifications. By looking jointly at Figure 4 and Figure 5, it can be noted that the maximum average number of correctly decoded messages at 400 m is achieved close to the channel occupation limit, despite the PER is larger than 0,2.

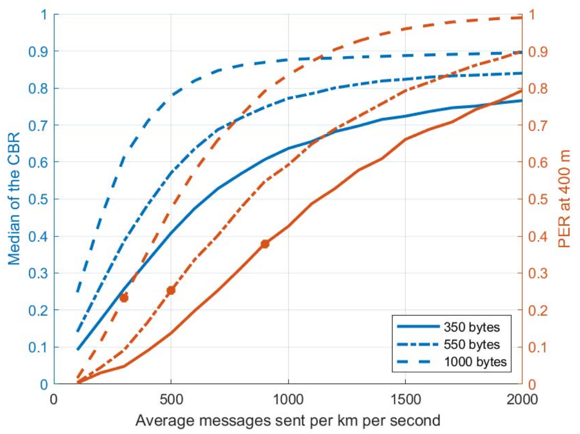


Figure 5: Median CBR and frame error probability at 400 m, varying the messages sent per km per second

The overall conclusion is that, assuming packets all of the same size, a different size could lead to a different maximum congestion threshold. Similar considerations apply also to other cases, such as if the packets are all with the same priority, but the priority is varied, because of different durations of the inter-frame spaces; or if the reference distance is changed. This means that it is not possible to give a specific number as the optimum limit, but this depends on a large number of variables which may change in time, in space, and depending on the use case.

These results are consistent with what discussed in the reference literature and the main deriving observations are as follows.

NOTE 1: A specific optimum value does not appear possible to derive, as it depends on a very large number of factors, which can also depend on the application and vary in time and space.

NOTE 2: The threshold 0,62 currently in the standards appears reasonable also for Release 2.

NOTE 3: A mechanism allowing the infrastructure to periodically update the limit should be introduced. Such periodicity may be in the order of months or years and should introduce small variations based on measurements from the real implementations. The mechanism should be defined considering fairness between vehicles that may have temporarily different values during the update; this may be granted by only allowing small variations but needs further analysis. As an example, parameter updates could be distributed via the certificates.

5.3.3 Analysis of algorithms

The current limits imposed to the algorithm that an implementation can use may not be sufficient. The current limits imposed by the standard ETSI TS 102 687 [i.14], in fact, do not guarantee fairness and may cause different implementations to have different levels of access to the channel in congested situations.

With the objective to elaborate on this statement, hereafter some results are shown using the same simulator and settings of clause 5.3.2, adding the reactive and adaptive approaches described in ETSI TS 102 687 [i.14], without GeoNetworking. These results are not meant to comprehensively compare specific solutions, for which the reader can refer to the following references [i.19], [i.20], [i.21] and [i.22], but only to remark the different impact that different algorithms can have.

In particular, hereafter either 200 vehicles per km are assumed to transmit messages of 350 bytes (scenario A), or 100 vehicles per km are assumed to transmit messages of 1 000 bytes (scenario B). These settings allow to investigate a scenario where the channel is slightly overloaded if no congestion control is implemented. Each vehicle generates one packet every T_{on}+T_{off}, where T_{off} is set by the congestion control algorithm, based on the measured CBR. A new packet is generated at least every 1 second and at most every 100 ms. Each vehicle generates one packet every 100 ms if DCC is not active.

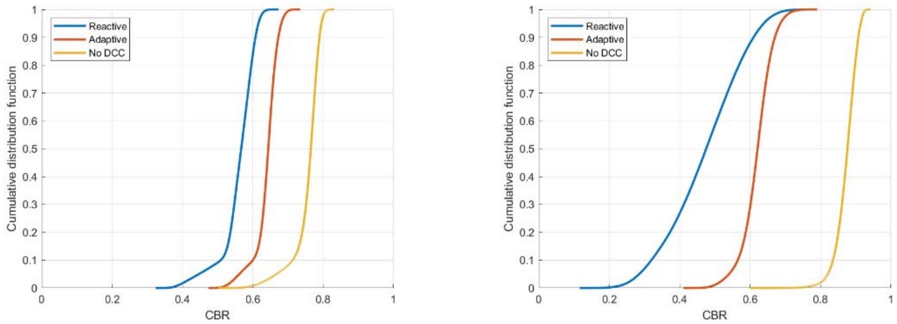
When the reactive approach is used, the profile defined by the CAR 2 CAR Communication Consortium in the Vehicle C-ITS station profile, Release 1.6.7 is used. The profile requires that the CBR is updated making an average of the last measured CBR and the one calculated in the previous measurement interval. The profile also sets the values of the minimum T_{off} for CAM messages to those of Table A.2 in ETSI TS 102 687 [i.14], which are therefore used for the 350 bytes packets; consistently, for the 1 000 bytes packets those of Table A.1 in ETSI TS 102 687 [i.14] are used. These values are also reported in Table 1.

Table 1: Settings of reactive DCC

State	CBR	Minimum T_{off} with packets of 350 bytes	Minimum T_{off} with packets of 1 000 bytes
Relaxed	< 30 %	50 ms	100 ms
Active 1	30 % to 39,99 %	100 ms	200 ms
Active 2	40 % to 49,99 %	200 ms	400 ms
Active 3	50 % to 65 %	250 ms	500 ms
Restrictive	> 65 %	1 000 ms	1 000 ms

When the adaptive approach is used, the settings defined in ETSI TS 102 687 [i.14] are used.

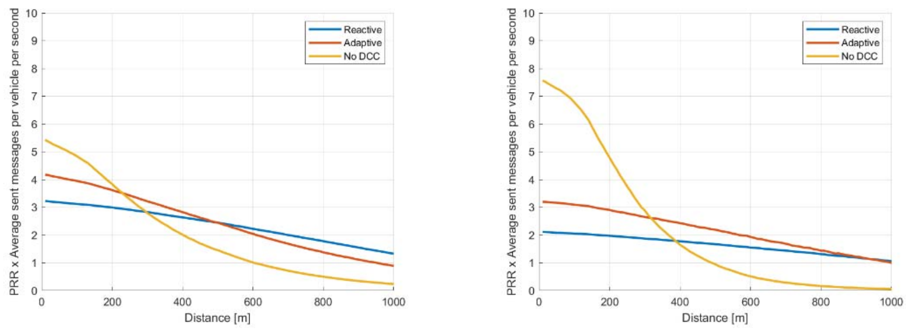
Figure 6 and Figure 7 refer to scenarios where all the vehicles implement the same algorithm, reactive or adaptive, or do not implement any algorithm. Figure 6, in particular, shows the Cumulative Distribution Function (CDF) of the CBR measured by every vehicle in every measurement interval T_{CBR}. The two subfigures refer to the two scenarios. The yellow curve, referring to no DCC implemented, provides a reference in the case all vehicles transmit exactly 10 packets per second. What can be observed is that both the reactive and the adaptive approaches reduce the measured CBR, which remains in most of the cases below or slightly above 0,6 as required by the limit imposed. The reactive approach implies a lower CBR than the adaptive approach, meaning that it is able to reduce the average packet generation frequency more than the adaptive approach. A last observation that appears relevant is that in most of the cases the CDF tends to remain around a small range of the CBR; this is consistent with the stationarity of the scenarios and means that the algorithm tends to reach an equilibrium at a certain T_{on}+T_{off}. However, in Scenario B, where the channel is more congested, the CDF related to the reactive approach spans over a larger range of values of the CBR, which suggests that the algorithm does not converge to an equilibrium.



(a) 200 vehicles per km, 350 bytes (b) 100 vehicles per km, 1 000 bytes

Figure 6: Cumulative distribution function of the measured CBR

The impact on the quality of the communication is shown in Figure 7, where the PRR multiplied by the average messages sent per vehicle per second is shown varying the transmitter-receiver distance. The metric observed corresponds to the average number of packets that can be correctly decoded by a receiver at the given distance and makes the performance of the various cases comparable, even if the number of generated packets is not the same. As observable, in both scenarios, when no DCC is implemented, the communication becomes unreliable at a shorter distance compared to the other cases. For example, looking at scenario A (Figure 7(a)), only one average packet can be correctly decoded at 600 m if there is no DCC, whereas more than two can be correctly decoded if a DCC algorithm is implemented. Comparing the two algorithms, the adaptive approach appears to provide a slightly higher number of correctly decoded packets, at least up to a certain distance. Overall, a higher transmission rate privileges the reception at shorter distance, whereas a lower transmission rate does the opposite.

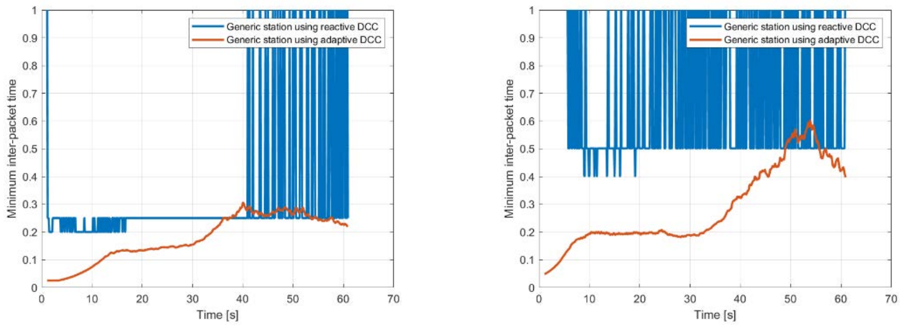


(a) 200 vehicles per km, 350 bytes (b) 100 vehicles per km, 1 000 bytes

Figure 7: Average number of messages per second correctly received from a station at a given distance

The results shown above suggest that the adaptive approach is able to maintain the channel occupation closer to the limit imposed, with a more stable value of the selected T_{on}+T_{off} and with slightly higher number of correctly decoded packets at short-to-medium distances. However, the difference between adaptive and reactive may not appear significant.

In order to verify what happens if not all the vehicles implement the same algorithm, Figure 8 refers to a case where 50 % of the vehicles implement the reactive approach and the remaining 50 % implement the adaptive approach. In Figure 8, in particular, the minimum inter-packet time (i.e. T_{on}+T_{off}) of two stations are shown varying the simulation time. The two stations are randomly selected, one among those implementing the reactive approach and the other among those implementing the adaptive approach. What can be observed is that the two stations have significantly different values of the minimum inter-packet time; the reactive approach, in particular, tends to use a larger gap, which means a lower frequency of packets sent. Having stations that generate packets at different frequencies because they are using different algorithms appears as a possible issue.



(a) 200 vehicles per km, 350 bytes (b) 100 vehicles per km, 1 000 bytes

Figure 8: Variation of the minimum inter-packet generation interval for randomly selected stations varying time

The results shown are consistent with what has been discussed in the literature. The following observations can be made.

NOTE 1: The specific algorithm that is adopted impacts on the system performance. The algorithm acts on the trade-off between the number of packets sent and the probability of collisions; reducing the packets sent may reduce the information available at shorter distance but increase that available at longer distance. This means that the distance from the limit at which the congestion control algorithm works has a significant impact on the system.

NOTE 2: To comply with the limit described in clause 5.2.4 may not be sufficient to guarantee a certain performance and fairness among stations. Different algorithms may imply, in fact, that different amount of information is available at the receivers, which means that the principle of fair access to the channel may not be guaranteed. A solution may be to have a single algorithm or at least stricter rules for the definition of the algorithm.

5.3.4 Analysis of possible variables impacting on congestion control

Different options to cope with congested situations are suggested in Release 1 specifications, including Transmit Power Control, Transmit Rate Control, and Transmit Data-Rate Control. For more details see the references [i.20], [i.23], and [i.24].

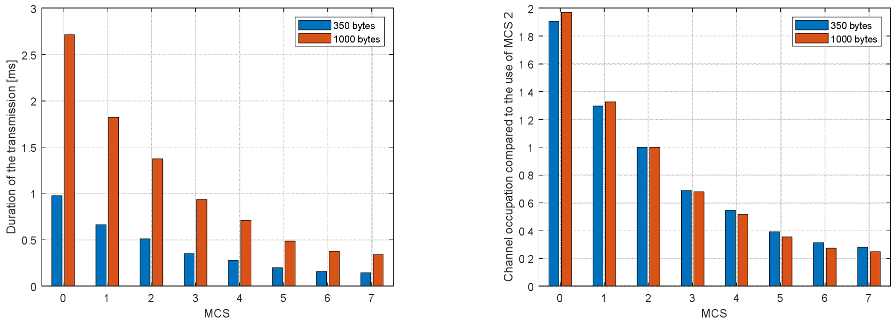
Variations of the power appear complex when looking at control at the facilities layer and also questionably effective: power variations, indeed, affect communication range but do not alter the CBR measurement close to the transmitting vehicle; this causes in turn variable power level and range in the spatial domain, which affects fairness and causes lower overall system performance. Additionally, if all stations uniformly reduce the power consistently, the effect is only an average reduction of the ratio between the useful signal and the noise power levels, without an improvement of the ratio between the useful signal and the interference power levels [i.23]. Power control may still remain an option when dealing with unicast communications; in such case, in fact, the transmitter may control the power level in order to guarantee a sufficient reliability to the communication while minimizing the interference generated to the others.

Data-rate control appears easier to control at the facilities layer and possibly more effective to control the trade-off between reliability and channel occupation. This means at the access layer to select a different MCS, which corresponds at the facilities layer to the selection of a different ALI. However, the choice of the MCS (or ALI) has also an impact on the range, since higher MCS implies lower protection to noise and interference. The use of data-rate control therefore requires careful considerations about the use cases that need to be supported.

Still discussing the data-rate control, it is also worth observing that the impact may be limited. This is further elaborated hereafter, by assuming the eight MCSs of the basic version of ITS-G5 and calculating the duration of packet transmissions. Based on the specifications of ITS-G5, the duration t_{pck} of a packet transmission with payload of N_{byte} bytes can be approximated as follows:

$$t_{pck} \approx t_h + \frac{\lceil 8 \times N_{byte} \rceil}{n_{bps}} T_{OFDM} \quad (3)$$

where $t_h = 40$ µs is the duration of the Physical Layer Convergence Protocol (PLCP) preamble and header, n_{bps} is the number of useful bits carried by an Orthogonal Frequency Division Multiplexing (OFDM) symbol, and $T_{OFDM} = 8$ µs is the duration of an OFDM symbol. The value of n_{bps} depends on the specific MCS and ranges between a minimum of 24 and a maximum of 216, with 52 used as default (normally known as MCS 2 and corresponding to 4-QAM, coding rate 1/2). The duration deriving from this equation is shown in Figure 9 for packets of 350 or 1 000 bytes. As observable, compared to the use of MCS 2 (which is the default value in ITS-G5), a data-rate control can at most reduce the channel occupation by a factor close to 4; this however requires moving to the less reliable MCS, which strongly impacts on the range. Small variations of the MCS have limited impact on the channel occupation.



(a) Duration of a transmission (b) Channel occupation compared to MCS 2

Figure 9: Impact on channel occupation of data-rate control in ITS-G5

The last option discussed is the transmit rate control, which means controlling the average number of packets sent and can be implemented in two ways. One way is discarding a portion of the packets before they are transmitted, and the other way is to control the generation rate. Looking at the former option, it can in turn be implemented at the facilities or at the access layer. If it is performed at the access layer, there is no possibility to differentiate among packets based on their content and the withdrawal is unavoidably performed in a random way. Among the various options, the control of the generation rate appears clearly preferable and the interaction between the facilities layer and the running applications and services can further help to optimally identify the information to be sent. When referring to Release 1, a withdrawal at the access layer appears as the only viable solution.

Overall, the transmit rate control is the one that can impact more significantly on the channel occupation without affecting communication range. Its main drawback is a reduction of the information update, which is however balanced by a reduction of the packet error rate if the congestion control algorithm works properly.

NOTE: It is observed that transmit rate control is the most effective approach to control congestion; it is also noted that it would be better implemented with a control made by the application rather than by discarding part of the packets. Data-rate control can also help in some cases, if the communication range can be traded-off. Transmit power control seems helpful only in specific cases, such as for unicast transmissions.

5.4 Interference from adjacent channels

5.4.1 Introduction

The congestion status of a given channel is not only dependent on the actual usage of the channel itself but also on the usage of the spectrum that is adjacent to it. This spectrum might be occupied by interoperable ITS services using the same protocol or by systems which are not interoperable or compatible. Interference management operations are basically different for the two cases. In the first case, it can be handled by internal RM control operations and functions that are part of the RM specifications. These operations can be called cross-channel RM functions, which include the cross-channel CBR evaluation and the cross-channel load control.

In the non-interoperable case, the cross-channel RM control is not possible due to the lack of interoperability. These effects can only be considered as general interference from adjacent channels similar to co-channel interference from non-interoperable sources.

All non-interoperable interference effects are included in the RM monitoring/evaluation process based on the CBR measurement at the access layer and thus contribute to the channel load in general. It has to be noted that these non-interoperable interference effects can lead to a significant increase of the channel load and to a reduced transmission probability/capability of the ITS systems due to the clear channel assessment performed by CSMA/CA before transmitting. In the RM and the MAC procedure a differentiation between the two different kinds of channel loading effect would be beneficial and could contribute to the proper operation of the ITS.

5.4.2 Impact of interference from interoperable C-ITS in adjacent channels

The impact of interference caused by stations using the same protocols in adjacent channels is extensively studied in ETSI TR 103 439 [i.1].

In Figure 10 the basics of the interfering effects in an MCO operation are depicted with the focus onto the direct adjacent channel as considered in ETSI TR 103 439 [i.1].

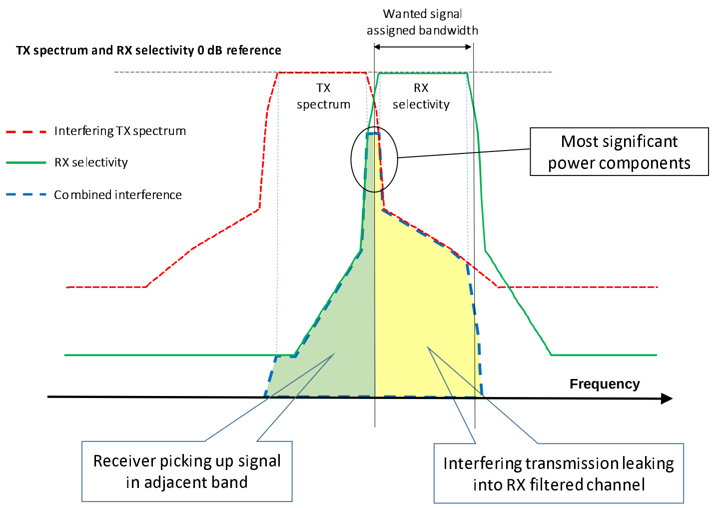


Figure 10: Impact of an interfering transmitter and a victim receiver on the reception of wanted signals

The study in ETSI TR 103 439 [i.1] was performed assuming a highway scenario with multiple lanes per direction and different levels of road traffic. Furthermore, different settings were assumed in terms of which channels were used by the vehicles moving in the scenario. Both cases where similar data traffic was generated on average in the various channels (balanced load) and cases where the load was imbalanced, were considered.

The main conclusions of the study can be summarized as follows:

- When transmissions are performed in channels that are not directly adjacent to each other, the interference that they cause reciprocally only slightly reduces the probability to correctly receive the packets. This also holds even when the channels are congested. This implies that there is no need to take the congestion status of the channels beyond the first adjacent channels into account in the internal RM operations.
- Focusing on transmissions in the first adjacent channel, they do not impact relevantly on the clear channel assessment performed by CSMA/CA before transmitting, but impact on the calculation of the CBR. As a consequence, rarely transmissions are deferred due to interference from adjacent channels, but the interference may increase the CBR and therefore reduce the number of transmissions overall performed in the given channel. When two channels that are directly adjacent to each other are highly loaded, the effect is a reduction of the average number of transmissions due to congestion control, thus a reduction of the collision probability in each of the two channels. This in part counterbalances the increase of errors at the receivers.
- A highly loaded first adjacent channel can cause a reduction of the probability of correct reception of the packets which leads to an estimated loss of range that can reach 25 % - 30 % (where the range is defined as the maximum distance at which a certain minimum reception probability is obtained).
- Distributing the data traffic over two channels, even directly adjacent to each other, is always preferable to having all the data traffic in a single channel. This means that the reciprocal interference caused by transmissions in adjacent channels has an impact lower than the advantage obtained by halving the data load in a single channel.
- Due to the current constraints imposed to the transmission mask, reducing the transmission power in one channel may not significantly reduce the interference it causes to the adjacent channels; even if this also depends on the specific implementation, the observation is that using power control to reduce the interference between adjacent channels may not be effective.
- Reducing the congestion control limit in one channel may reduce the maximum interference that the transmissions in that channel causes to those in its first adjacent channels; reducing the limit in one channel may therefore be used to protect its first adjacent channels.

Given these considerations, interference from stations using the same protocols potentially reduces communication performance in an adjacent channel, but in practice performance degradation does not appear to be critical. It may be helpful to reduce the congestion control limit in some channel ALI Groups in order to prevent the communication performance degradation in their first adjacent channels, including to available capacity and communication range. This would give priority to channels with less stringent requirements.

In all the cases, distributing the traffic over multiple channels appears beneficial.

5.4.3 Adjacent channel monitoring

Different approaches for the monitoring operations of adjacent channel interference have to be taken for single channel and multichannel ITS-S. A single channel device cannot differentiate between interoperable or non-interoperable deployments in the adjacent channels without switching to these channels, thus a single channel ITS-S can only passively monitor the interference increase as part of the channel load measurement, independently from the nature. In contrast, multichannel devices can perform this differentiation and can take the required control decisions.

5.4.4 Cross-channel RM control

In a single channel device, the adjacent channel interference will be treated as normal unspecified channel load element together with the co-channel load elements. The resulting levels of the load estimation in the access layer will be taken into account in the management algorithms. The co-channel load originating from an interoperable ITS system could be treated differently from the non-interoperable and adjacent channel interference.

In a multichannel device a more detailed view of the interference characteristics in the adjacent channel can be evaluated. This information can be used for an optimization of the RM control algorithm. In order to differentiate between load originating from interoperable systems and load originating from non-interoperable systems, the channel load information from the co-channel and adjacent channel would need to be reported in a structured manner.

5.5 Modifications towards Release 2

5.5.1 Introduction

This clause focuses on the modifications that are required to comply with the RM approach in Release 2, where congestion control needs to indicate the available resources in an abstract way at the facilities layer. Modifications in channels that are already used today do not need to affect the performance of devices that are already on the road.

In Release 1, the congestion control mechanism is fundamentally performed at the access layer, in which case it is technology-dependent and cannot affect the message generation. In Release 2, the RM functionality needs to be aware of the available resources and to manage them. The following two approaches appear possible:

1) the access layer provides the information about the channel load to the facilities layer, and congestion control is performed at the facilities layer; or

2) congestion control is performed at the access layer and the available resources are abstracted to the facilities layer.

The approach one listed above foresees that per each channel either:

i) the facilities layer implements a technology specific algorithm per each technology that is supported; or

ii) it implements an agnostic algorithm, which in turn requires the definition of a technology-agnostic channel load measure derived from the measurements performed differently by each technology.

Having a technology specific algorithm appears undesirable as it complicates the design of the RM, with the risk to duplicate some functions at both the access layer and the facilities layer. At the same time, having a technology-agnostic channel load metric and a translation from a technology specific measurement to this metric appears to be very challenging.

For these reasons, the approach two above appears preferable and therefore only that approach is further discussed in the following.

5.5.2 Congestion control in Release 2

The approach proposed for Release 2 congestion control is to maintain it mainly at the access layer, but abstracting to the facilities layer the resources that are available. In the case of ITS-G5, the available resource may correspond to $T_{on}/(T_{on}+T_{off})$, with values calculated using a reactive or adaptive approach. Other access technologies may implement different procedures.

This approach requires the definition of resources at the facilities layer that are agnostic of the specific access technology, which might be, for example, a bandwidth-time product in each interval of time. The available resources can be mapped in the facilities layer to each ALI group. When the facilities layer is aware of the resources that are available per each ALI group and is able to map the messages that need to be delivered to a number of resources that they would consume, the RM can make the appropriate decisions.

Concerning the congestion control algorithm, this approach does not require relevant modifications to the algorithms used in Release 1.

In addition to the provision of the available resources to the facilities layer, the access layer needs to inform it of relevant events, such as queue overflow, drop of packets, and time out of packets; their identification and notification by the access layer need therefore to be specified in Release 2.

5.5.3 Resources at the facilities layer

One issue that needs to be solved is how to abstract the radio resources from the access layer to the facilities layer.

The main difficulty is that the resources available and the resources used do not only depend on the average number of bits generated but also:

i) on the number of packets, since each packet has its own headers adding overhead; and

ii) on the specific settings at the access layer, including for example the MCS.

The facilities layer should have, as much as possible, a precise and complete understanding of the options available. It should have for example all the information to decide, under congested situations, if for a given message it is better to use one or the other channel, one or the other MCS, or even discard it, if necessary.

The knowledge of the available channels and MCSs is already abstracted using the concepts of ALI and ALI group. What still needs work is the definition of the resources, the calculation of the available resources, and the calculation of the resource used to transmit a message in a given ALI, possibly with some knowledge about the expected performance corresponding to the use of that specific ALI. The proposed definition of the technology-agnostic resources is as follows:

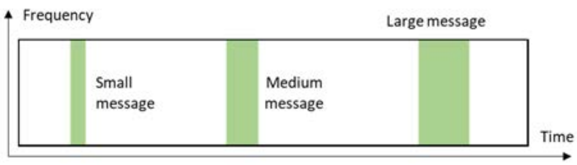
- Portion of time and bandwidth. This definition is unitless and is valid for any bandwidth or time period. As an example, the resources corresponding to 0,005 are those equivalent to the transmission of either:

    a) messages that occupy the whole bandwidth during 0,5 % of the time; or

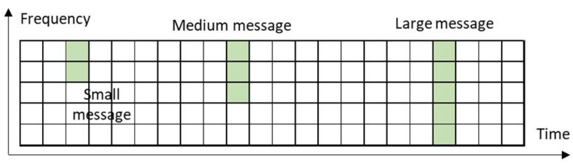
    b) messages that occupy the half of the bandwidth during 1 % of the time; or

    c) messages that occupy one fifth of the bandwidth during 2,5 % of the time.

This definition is valid for radio access technologies both with and without resource grids. As an example, ITS-G5 (Figure 11a) makes use of the whole bandwidth in all transmissions, providing a high granularity in time. On the other hand, LTE-V2X and NR V2X (Figure 11b) allow transmitting using part of the bandwidth but the time duration is fixed.



(a) ITS-G5



(b) LTE-V2X or NR V2X

Figure 11: Resources in time and frequency

This definition can be also used for the resources needed to transmit a single message, although this requires mapping the message size with the resource use. Given that the channel used for a transmission depends on the message size and on parameters known at the access layer, one proposed solution is to have at the facilities layer one table per each ALI indicating the resources required to transmit a message of a given size. The table should have at least:

i) one column indicating the message size; and

ii) one column indicating the resources required for the transmission of a message of that size.

Using this table, the RM can calculate the resources it is using and apply procedures that require further work. As a simple example, if the facilities layer has a budget of 0,005 per second for an ALI group, and messages are generated every 100 ms on average, consuming in the selected ALI of that ALI group 0,0001, the RM knows that this flow consumes overall 10 × 0,0001 = 0,001, which is 1/5 of the available resources of the ALI group.

An additional aspect to be considered is that the RM also needs indications on the performance expected when choosing among ALIs that imply different uses of resources. For example, an ALI corresponding to a lower MCS will use more resources than another ALI corresponding to a higher MCS, but the advantage is that the transmission is more robust to noise and interference, which means that it is expected to provide more range. For this scope, one possibility is to store per each ALI a reference range, which may be for example calculated as the maximum distance at which the error rate is lower than 10 %, assuming a Gaussian channel, absence of interference, and predefined settings for the transmission power and receiver characteristics. As an example, it is assumed that there is one ALI group with two ALIs, ALI A and ALI B; the range of ALI A is 1 km and it requires 0,001 resources (defined as above discussed) to transmit a message of 800 bytes; the range of ALI B is 500 m and it requires 0,0005 resources to transmit the same message; given this information, the RM can decide if it is better to have a longer range but using more resources or the opposite.

Even if further work is required, the definition of the described information per each ALI appears a relatively simple approach to abstract at the facilities layer the resources used by a transmission with a specific access layer configuration and the performance expected. One point of attention that needs further elaboration is how to include in the calculations the impact of the overhead introduced by GeoNetworking and security.

5.5.4 Introducing a mechanism for the update of parameters

One feature that is not available in Release 1 and might be worth introducing in Release 2 is the possibility to update the specific value for a limited number of parameters, such as the congestion control limit in each channel. Such updates would require the control from a central entity and a distribution to all ITS-S in a relatively short time, which in turn means that it would require the use of I2V links.

EXAMPLE: One option to be explored is to exploit the distribution of the certificates to also include possible parameter updates.

6 Application Requirements

6.1 Introduction

To ensure that active applications in an ITS-S are able to make appropriate decisions about when and how to exchange information with other ITS-Ss, they need to be aware of the available communication resources at any given time. In general, more than one application can be active in an ITS-S and each application does not know the static and dynamic communication requirements from the other applications. The number of active applications in Release 1 is limited but for Release 2 they are significantly extended and therefore some resource management should be considered.

The message behaviour is strongly dependent on the way the dissemination of messages is triggered and the amount of information it needs to disseminate.

Furthermore, dissemination of a specific message type could be triggered by many applications, while other messages may be triggered by a single application or related to a specific use case. The DENM is typically a message type for which the message dissemination can be triggered by multiple applications, while the CAM is a typical message type which is related to a single use case, including message generating rules. The CAM is not triggered by applications but is a stand-alone functionality which has no interface to an application. Any resource management functionality most probably will only have to gather relevant management information for CAM dissemination directly from the CAS but as the DENS only knows that it has to disseminate a message when triggered, any resource management should in the first place get information from the triggering applications and possibly secondary from the DENS. In the following, for relevant dissemination information, reference is only made to applications while for several cases this could be also an ITS-S service.

Since in Release 2 implementations the number of active applications is expected to significantly increase, it could be not sufficient to just have the static knowledge about the available resources. Having a generalized knowledge of the needs of all applications could allow a resource management functionality to distribute the available resources between all applications. This could be detailed with different granularity of the resource distribution, which could range from a binary switch-on/off of an application to a very fine allocation of resources to the applications active at any given time.

Different methods exist in the initial Release 2 ETSI MCO concept. For the management of the resource allocation to the active applications in the ITS-S several aspects have to be considered:

- the individual application or ITS-S service communication requirements and thus resource needs;
- the application priority;
- available resource management mechanisms and granularity of the resource allocation;
- external resource limitation based on legal and regulatory restrictions.

The first three aspects are ITS-S internal (intra ITS) whereas the last one is external to the ITS-S (inter ITS).

The inter ITS aspect is an ITS intercommunity aspect which might require legal agreements or technical specifications agreed between stakeholders (profiles) to ensure which applications are active when and where in the available spectrum. It will need a secure control mechanism to allow for an external input of the required control information.

To ensure the robust operation of C-ITS as defined by the European Union (EU) in the Directive 2010/40/EU [i.31] and its amendments, this EU Directive and related specifications should ensure robust operation of all C-ITS applications in the assigned C-ITS spectrum.

In the following clauses, the intra ITS aspects related to the way the dissemination of messages is triggered and the amount of information it needs to disseminate are considered for the known ITS applications and ITS-S services.

6.2 Applications and ITS-S services

6.2.1 Introduction

There are applications which trigger event message generation and dissemination at the Facilities Layer by ITS-S services such as the Decentralized Environmental Notification Service (DENS). There are also ITS-S Message Services like Cooperative Awareness Service (CAS) which include message generation rules themselves, by which they trigger the generation and dissemination of messages. However, in future it cannot be excluded that such message services could be triggered as well. In general, message dissemination can be initiated by any application or message service with included message generation rules.

As message services disseminate messages in the same channel as other message service, the dissemination of one message service can be influenced by the message dissemination of one other message service. This dependency does not occur when the channel occupation is not near to the set congestion limits, however, already ahead of the possibility of entering such state, it could be of interest to applications and message services to be aware of this and make different choices when this would apply.

In principle, the understanding of the possible application message triggering and dissemination intensions are required for the realization of resource management. Considering that applications communicate with the ITS-Services and that therefore the message service knows the real disseminations requirements, currently it is considered sufficient that the dynamic behaviour expectations of the message service is communicated to the RM. The RM may also communicate with the applications for more general information, but this is not considered for the time being.

In general, ITS-S-MSs dissemination behaviour can be categorized in three types as follows:

- Broadcasted Event type, such as Decentralized Environmental Notification message (DENM) triggered by Applications (such as the Triggering Conditions as specified by the C2C-CC [i.7]).
- Broadcasted Awareness types such as:
  - Fixed Repetitive type, with fixed message size, such as MAPEM, IVI and SAM.
  - Adaptable Repetitive type, with predictable but no predefined message size, such as CAM, CPM, SPATEM and MIM/MVM.
- Broadcasted Streaming type, with fixed rate, fixed size and continuous, such as video streaming.

In the following clauses the message services identified are further detailed with regards to their communication behaviour.

NOTE: The mentioned ITS-S-MS are specified by ETSI and ISO but their references are not provided here as their reference number can differ for difference releases.

6.2.2 Decentralized Environmental Notification (DEN) Message

The DEN is an event message (as identified in clause 6.2.1). Especially the dissemination of DENMs is triggered by one or more applications which can support the realization of different use cases and environmental scenarios. It can be expected that the number of use cases and scenarios will increase in the future.

Depending on the safety impact of the specific use case, DENM dissemination via direct communication requires a specific priority over other message disseminations. In general, direct communications is direct safety impact oriented and Internet communications indirect safety impact oriented (more details in clause Ecosystems ETSI TR 103 903 [i.4]). While there are ITS use cases best serviced by either one of these communication solutions, there are also ITS use cases which can be serviced by both (see the Hybrid Solution by C-ROADs [i.8]).

One example is the "End of Queue" use case. For this use case the information disseminated via direct communications can be used for a direct stop initiated by a driver of a vehicle or by the automated vehicle itself, while disseminated information via Internet communication will generally be used for rerouting of a trajectory and not predictably for a direct stop, thus they are the same use case but seen as different scenarios. Information disseminated via direct communication (as there are more stringent requirements) can also be used for those use cases which require the same information but can also be serviced by Internet communications. The communication requirements for the indirect scenario are out of scope of the present document.

In Release 1, it was identified that high priority should be provided to these types of messages. It was recognized that events, as they have immediate impact, should be handled first compared to any other kind of other information dissemination. While for Release 1 only two levels were identified, it could happen that for later releases (at present it has not been defined for Release 2) the number of priorities should be extended.

NOTE 1: Extension of the priority levels is not only something of relevance for DEN but also for other services and therefore a generic extension should be considered.

NOTE 2: Release 1 operational equipment is based on IEEE 802.11 [i.9] which includes four QoS Traffic Classes (TCs). These TCs are directly coupled to the four priorities used at the Facilities Layer in C-ITS.

At present two event types can be recognized:

- An event which was not planned, mostly identified by an automated functionality in the vehicle, such as a Slow Driving Vehicle (SDV). This is a type of event which may just popup.
- An event which was planned, typically managed by authorities and switched on and off by humans, such as a Roads Works Warning (RWW).

While with the first type the dissemination of DENMs could take only a very short time, with the second type the disseminations of DENMs could take place for days or even month. This time related aspect is the only difference. Since at present both of the event types can be detected, activated and deactivated, the DENMs will be disseminated with a repetition rate fitting the environment and the use case as required.

From an ITS-S communication perspective, DENM dissemination is not application but use case dependent. In general, an application does not statically know in advance what its communication needs are. Only dynamically, when a use case determines that it needs to disseminate DENMs, the application can notify its needs in terms of resources. This is valid for both the managed and not managed type of events.

Possible parameter consideration for DENM:

- Application statistically (at the time of application activation (application registration)):
  - Application indicator (internal for ITS-S).
  - Number of use cases supported (internal for ITS-S).
  - Expected Priority level(s) to be supported.
  - Expected Access Technology and spectrum requirements (possibly including primary and secondary options).
  - Max message generation rate when activated.
  - Max message size when activated.
- Application dynamically (based on activation of DENM dissemination cycle (application registration)):
  - Application indicator.
  - Use case indicator.

- Dissemination initiation (Request):
    - Actual Priority of the cycle of DENMs to be disseminated.
    - Actual Technology and spectrum requirements (possibly including primary and secondary options).
    - Real message generation rate.
    - Real message size.
    - Expected start time of dissemination.
    - Expected termination time.
- Dissemination termination:
    - Termination indicator.

NOTE 3: A dissemination cycle is considered to be the period in which a sequence of DENM with a specific rate are actively disseminated.

NOTE 4: At present, a constant rate and constant message size is considered for DENM. Further it is considered that applications realize use cases with a similar message dissemination behaviour.

NOTE 5: To allow any management of the DENM message as provided by several applications and their use cases, it should be clear to the resource management which application/use case initiate DENM dissemination and therefore it is required to provide related DENMs with indicators about which application/use case initiated the dissemination. As a result, each DENM from a cycle needs to include all application dynamic information.

6.2.3 Cooperative Awareness (CA) Message

The CAM is a broadcast awareness message (as identified in clause 6.2.1) disseminated by an ITS-S which represents a road user e.g. vehicle, truck, motorbike, bicycle or pedestrian. The CAM provides information about the dynamic state of the represented road user. This information includes parameters such as location, time and dynamic parameters as speed and direction.

Since a CAM only provides information about its own identified road user, it is considered as a single use case only.

As specified in the CA ITS-S-MS specifications, the dissemination has two specific characteristics:

- The message dissemination rate depends on the dynamic behaviour of the station. In principle, it depends on the speed, acceleration and movement. The dynamic dissemination behaviour is expected to depend on the type of represented road user and can differ for a vehicle compared to a bike.
- The information to be shared as part of the CAM includes static as well as very dynamic parameters. As it is not that relevant to exchange the static parameters too often, related information is shared not as often as for the dynamic parameters, with the result that the disseminated message has a regular but not constant message size.

CAM transmission can therefore be predictably estimated but depend on the environment (for vehicles, it differs for urban, sub-urban and highway). CAMs predictability depends on the intelligence of the system which disseminates these messages. This can be statically performed or done in various dynamic ways, including the use of Artificial Intelligence (AI).

Possible parameter consideration for CAM:

- Application statistically (at the time of application activation (application registration)):
    - Application indicator (internal for ITS-S).
    - Number of use cases supported (fixed = 1 for the time being) (internal for ITS-S).
    - Expected Priority level(s) to be supported.
    - Expected Technology and spectrum requirements (possibly including primary and secondary options).

    - Maximum message generation rate.
    - Minimum message generation rate.
    - Maximum message size.
    - Minimum message size.
    - Distribution type.

- Application dynamically (based on activation of CAM dissemination (application registration)):
    - Application indicator.
    - Use case indicator (fixed = 1 for the time being).
    - Dissemination initiation (repeatedly based on environmental changes) can be repeated as long as needed:
        - Actual Priority of the cycle of CAMs to be disseminated.
        - Actual Technology and spectrum requirements (possibly including primary and secondary options).
        - Expected average message generation rate.
        - Expected maximum message generation rate.
        - Expected minimum message generation rate.
        - Expected average message size.
        - Expected maximum message size.
        - Expected minimum message size.
        - Required operation limit message rate.
        - Expected start time of dissemination.
        - Expected termination time.
        - Distribution type.
    - Dissemination termination (for instance when the vehicle is parked).

6.2.4 Collective Perception (CP) Message

The CPM is a broadcasted awareness message (as identified in clause 6.2.1) that is continuously generated with variable interval and message size. The variable interval comes from the fact that the CPM is only generated when certain rules are satisfied and not based on a predefined interval. The CPM includes information recognized by the equipment from which the ITS-S is part off. The disseminate CPM can include (a subset of) the objects perceived by the station, information about the sensing capabilities, and information about the perception regions.

The present Collected Perception Service (CPS) specification specifies a set of perceived object inclusion rules that significantly and dynamically influence the CPM size and generation rate (or interval). These rules are mainly based on the dynamics (e.g. position and speed) and type (e.g. VRUs vs vehicles) of the perceived objects. These object inclusion rules were extensively studied in related ETSI studies and in the scientific literature (see [i.25] and [i.28]). The CPM generation rules also include the possibility to dynamically include in each CPM a variable number of objects taking into account their value or utility, which is referred to as Value of Information (VoI) which refers to as redundancy in the scientific literature [i.26] and [i.27]. In addition, the object inclusion rules defined in the CPS standard allow the adaptation of the CPM size (or number of perceived objects) and interval based on inputs from RM. This ensures that the resources are used efficiently, and the amount of information sent by the collective perception service fits into the radio channel.

As specified, the sender can design its own rules about the inclusion of perceived objects (with *ObjectInclusionConfig* flag set to "0"), which affect the CPM size and rate. The CPM is considered to support multiple use cases as the perceived objects represent different traffic participant types and other types such as empty road slots. As different traffic participant types could represent different message sizes and generation rules, one CPM can include objects representing of traffic participant types of one kind while other CPMs include objects representing of traffic participant types of one or more other kind. Also, for other reasons linking specific use cases to specific CPMs is advisable. This means not to disseminate object information of all objects recognizable by the sensor but select those object information relevant for the use cases to be supported . For system flexibility and robustness, it is advice not to include all the perceived objects in a single CPM. One of the main reasons is to avoid reaching the maximum message size and requiring the allocation of more than 10 MHz bandwidth. In general, from a system perspective it is better to keep awareness messages small so there is flexibility of making system choices.

For the rest for CPM the same applies as for CAM.

6.2.5 MIM and MVM

The Marshalling Infrastructure Message (MIM) and the Marshalling Vehicle Message (MVM) are used by the Automated Vehicle Marshalling (AVM) service or low-speed remote controlled automated driving (e.g. in parking areas or factories). The MIMs are disseminated by the infrastructure and the MVM are disseminated by vehicles.

Each MIM sent by the infrastructure can target up to 32 vehicles and therefore its size can significantly change over time depending on the number of vehicles being remotely controlled. The MVM sent by vehicles have optional elements that also generate messages with variable size, but its variation is significantly lower than the MIM.

A sequence of MIMs and MVMs are exchanged during the initialization of the AVM service. During the driving mode, the infrastructure and the vehicle have to periodically exchange MIMs and MVMs for the correct operation of the service. To this aim, the AVM service introduces a message generation based on a mix of periodic and event-driven messages. By default, AVM messages are generated continuously at a recommended rate of 10 Hz. Additionally, the generation of MIMs and MVMs may be triggered by events, such as emergency stops. These events may cause the generation of one or more new messages and thus disrupt the periodic pattern.

With regards to message dissemination behaviour, it can be expected that it has a more static behaviour as CAM but at high rates. It can be expected that the parameter set is quite similar to CAM/CPM.

6.2.6 Signal Phase and Timing Message (SPATEM)

SPATEMs are semi-static messages which means that their size can slightly change. The timing of the SPATEM depends on the changes in the traffic-light behaviour. Today's dynamically SPATEM assigning systems could update the sequence about once every 0,1 second. So, updates of 10 Hz are not an exception. The packages however are not large and both size as update rates although dynamic are still quite predictable. With regards to channel use, only Roadside Units (RSUs) will transmit SPATEMs and therefore it can be expected that in a range of 200 m - 400 m only a very few RSU will transmit. After analysing the SPATEM, it can be recognized that the SPATEM can use the same parameter structure as the CAM and only will have different values.

6.2.7 MAP Message (MAPEM)

The MAPEM messages are static messages which provide an overview over the road topology with all lane descriptions and stop lines etc. The size always stays the same and it is disseminated with a fixed lower frequency such as 1-2 Hz by RSUs. Like for SPATEM, the same parameter structure as for CAM can be used.

6.2.8 In Vehicle Information Message (IVIM)

The IVIM is a message which is intended to represent for instance a sign. A sign is expected to be static, like a speed limit sign. However, road operators do change the prelimits depending on the situation on the road. Signs can therefore be static or semi-static (changing ones in a while, in intervals of minimal 30 seconds).

IVIMs are awareness messages which can only be disseminated by authorities which could be RSUs but also special vehicles. Like SPATEM and DENM they are generated by authority controlled ITS-Ss. The number of ITS-Ss within a certain vicinity will be limited. It can be assumed not to be more then 4. The impact on the channel use can be assumed minor.

The size is known and static while the rate may be of a few Hz and it can slightly change depending on the road it is active. The dissemination is timely predictable to allow a resource management functionality to fulfil its task. The same parameter structure as for CAM can be used.

6.2.9 Service Announcement Message (SAM)

At present from a message dissemination perspective, the SAM can be seen similar to the IVIM. The SAM is not yet considered to be used for the announcement of dynamic safety use cases, which could result in some additional dissemination requirements (at present this is not foreseen). SAM dissemination can be predicted sufficiently to allow resource management to fulfil its task. The same parameter structure as for CAM can be used.

6.2.10 Vulnerable road user Awareness Message (VAM)

The VAM is an awareness message similar as the CAM its size is smaller and rate more predictable and lower. Considering road scenarios, the number of present VRUs in an area could however be much more then Vehicles in the same area. The dynamic behaviour of VRUs is much slower than that of Vehicles and therefore the channel use behaviour is likely to be more predictable for the RM than that of the dissemination of CAMs. The same parameter structure as for CAM can be used.

6.3 Types of ITS-S services from RM perspective

The envisioned RM solution will provide indications to the ITS-S services so that they can dynamically adapt the messages they generate in real time. To this aim, the ITS-S services could reduce the number of messages and/or their size to follow the indications provided by the RM, while at the same time send additional messages to be offloaded to alternative channels. In this context, the following types of ITS-S services are identified:

- Size adaptation. ITS-S services that adapt the message size (dynamic inclusion of objects in CPMs or vehicles in MIM, or optional elements in any other service) but keep the message interval fixed.
- Interval adaptation. ITS-S services that adapt the interval of the message generation instead of adapting their size.
- Size and interval adaptation. ITS-S services that have the flexibility and intelligence to adapt both the message interval and size following the indications of RM.

According to how the ITS-S services generate their messages, the following types are also envisioned:

- Predefined rules. ITS-S services that have their own message generation rules that trigger the generation of new messages. When enough resources are available, they generate the necessary messages following these rules. If the available resources are higher than the resources needed by these ITS-S services, they simply follow they predefined message generation rules. Examples are the CA and DEN services (see Figure 12a).
- Adaptive rules. ITS-S services that adapt the messages they generate to the resources available. The more resources available the more messages they can generate, up to a certain limit that can be high. One example is the CP service, that could adapt the number of perceived objects and regions following the instructions of the RM (see Figure 12b).

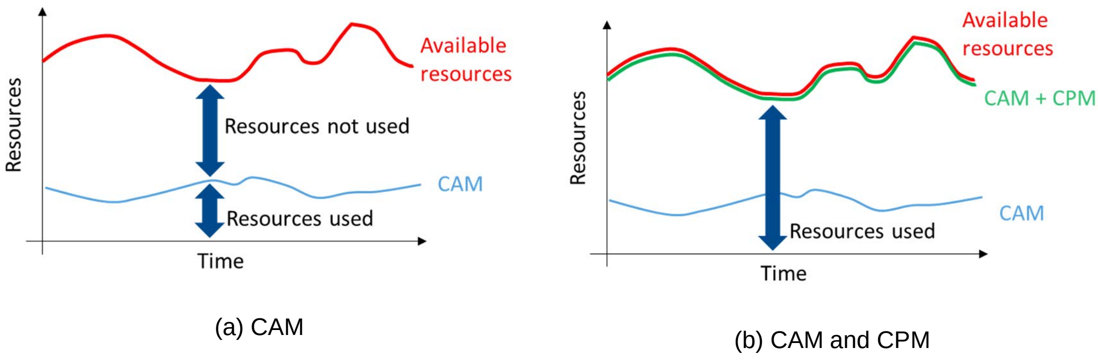


Figure 12: Illustration of resources consumed by CAM and CPM

6.4 Heterogeneous resource requirements

6.4.1 Same services but different resource needs

One important aspect to consider in the design of the RM is the fact that two (or more) ITS-S could be running exactly the same ITS services, but the radio resources each one consumes can differ widely. Some examples are described below:

- **Cooperative Perception Service (CPS).** An ITS-S on a vehicle typically detects less objects than an ITS-S on an RSU, so that the amount of information they need to generate is different. Also, two nearby vehicles can also detect a very different number of objects based on their specific location and sensors. The quality of the sensors is also an important factor, since the quality of the detections has an impact on the amount of information that needs to be transmitted.

- **Automated Vehicle Marshalling (AVM).** Vehicles participating in an AVM system transmit Marshalling Vehicle Messages (MVMs). The infrastructure controlling these vehicles generates Marshalling Infrastructure Messages (MIMs) that are larger in size and have to be more frequently transmitted (see for more information the publication "Automated Vehicle Marshalling" [i.29]). Even though both participate in AVM, their resource needs are intrinsically asymmetric.

- **Cooperative Awareness (CA).** Two implementations may have different resource requirements depending on the optional elements implemented, and the specific driving conditions. As an example, a vehicle could be in a traffic jam in one driving direction and a nearby vehicle could experience free flow conditions in the other direction. The CA service of the vehicle stopped would generate CAMs at 1 Hz, while the other could require the transmission of CAMs at 10 Hz depending mainly on its speed.

The design of the RM should be able to handle this heterogeneity of resource needs to optimize the bandwidth efficiency and system performance.

6.4.2 Different releases with different needs

The resources needed by different ITS-S also depend on their Release, since they are expected to implement different ITS-S services in each release. Some examples are shown below:

- A Release 1 ITS-S generates essentially CAMs and only occasional DENMs.

- A Release 2 ITS-S also implements collective perception, so it generates CPMs in addition to CAMs and sporadic DENMs.

- A Release 3 ITS-S could also implement manoeuvre coordination, generating extended Manoeuver Coordination Messages (MCMs) in addition to CPMs, CAMs and DENMs.

On any given road different ITS-S implementing different Releases can be intermixed, each with its own resource needs-ranging from a lightweight CAM-only sender to a fully-featured node that generates CAM, CPM, MCM and DENM concurrently. The design of the RM should ideally take into account these aspects.

6.5 Indications to ITS-S services

ITS-S services will dynamically adapt the messages they generate (size and/or interval) following the indications of RM. Different options are possible with different levels of abstractions for the RM to inform the ITS-S about the messages they can generate:

- **Bits/s.** The RM could limit the amount of bits/s that each ITS-S can generate in the default channel (or alternative channels). With this approach the ITS-S services could adapt e.g. the message interval for a given size.

- **Bits.** The RM could assign a certain number of bits to the ITS-S services to indicate them that they are allowed to generate one or more messages so that the total amount of bits does not exceed the assigned one, irrespective of the time. In the next update, the RM could take into account if the ITS-S services consumed all the bits assigned or not, and re-assign them accordingly. This approach is similar to the previous one but avoids the complexity of time management in the services, since they only have to check the number of assigned bits left to generate new messages.

- **T_{on}/T_{off}.** This approach was used in Release 1 as detailed in clause 5.2. Radio access technologies like LTE-V2X and NR V2X would not support this approach because they have a fixed T_{on}, except in those cases where one facility-layer message has to be segmented in multiple packets at the access layer.

7 Resource Management concepts

7.1 Introduction

Resource management plays a pivotal role in ensuring efficient operation and Quality of Service (QoS) in communication systems. The Facilities Layer (FL) is particularly well-suited for this task, as it can integrate the capabilities of lower layers with the dynamic functional requirements of ITS-S services. As highlighted in the Multi-Channel Operation (MCO) concept study in ETSI TR 103 439 [i.1], ITS-S-MS at this layer are unaware of the communication needs of other ITS-S-MS. This necessitates a robust Resource Management (RM) functionality to ensure a consistent and efficient use of available channels.

The RM ITS-S functionality should be seen as an improvement of the original MCO functionalities as specified by the set of MCO specifications. The RM operates by harmonizing dataflows between message handling services to realize improved trustworthiness and QoS while making an efficient use of the channels. It leverages the capabilities of lower layers while supporting technology agnostic operation of message-disseminating ITS-S services. This layered approach ensures that resource allocation and communication management remain dynamic, adaptive, and aligned with real-time network conditions. The following clauses outline key concepts and mechanisms for resource management, drawing analogies to wired network practices while addressing the unique requirements of vehicular networks.

7.2 Analogy with Wired Networks

Resource management in vehicular networks shares conceptual similarities with wired networks, particularly in the context of resource reservation and QoS. In wired networks, clients request resources for specific flows, defining the flow parameters and QoS requirements. The network evaluates these requests and accepts them only if all nodes along the path can support the resource demands. Nodes include edge and inner nodes ensuring seamless communication. Figure 13 illustrates resource reservation for QoS flows in wired networks.

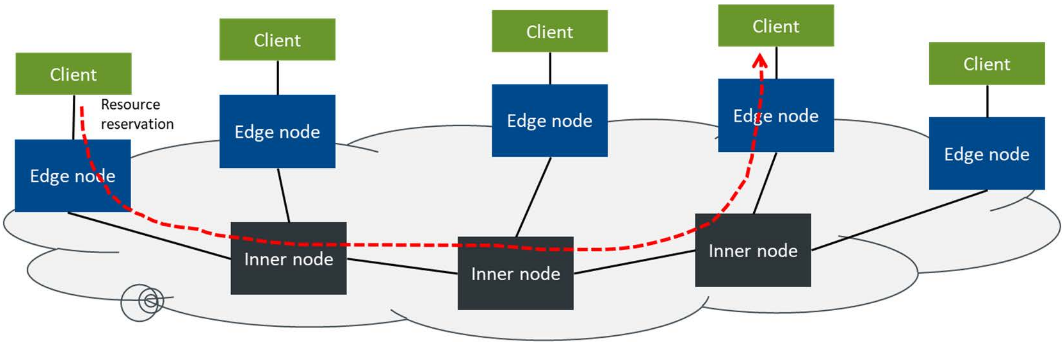


Figure 13: Resource Reservation in Wired Networks

In vehicular networks, this analogy holds with certain adjustments. Here, applications or ITS-S services act as clients, requesting resources for message dissemination based on their QoS needs. The RM evaluates these requests and grants or denies them based on channel availability and service priority. Unlike wired networks, vehicular networks lack intermediate nodes for resource validation. Instead, the system uses channel load measurements for resource management. This model supports dynamic and adaptive resource allocation tailored to the real-time demands of vehicular communication systems. Figure 14 shows how the concept of resource reservation can be adapted to vehicular networks.

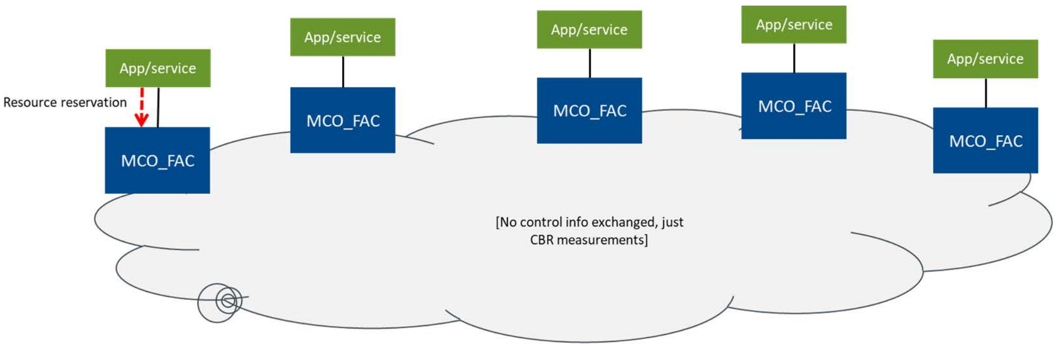


Figure 14: Resource Reservation Analogy in Vehicular Networks

The trustworthiness and QoS of the vehicular network are not guaranteed by RM. RM is only able to manage and supervise the dissemination of the messages. In wired vehicular networks the trustworthiness and QoS are fixed by the system design, which is a closed box approach in which it is known what network can be expected at least in terms of congestion.

In the ITS, this should also be managed. This management is handled by an agreement between the stakeholders on what use cases and services are allowed to exchange their message in the specific available radio channel or channels.

7.3 Resource Management architecture and its mechanisms

7.3.1 Overview

This clause describes mechanisms for resource management that can be part of the RM functionality, in particular mechanisms that can be used in Bandwidth Management Entity (BME) and Message Handling Entity (MHE), as defined in the MCO architecture.

For the RM, the BME can include functionalities such as Admission Control and Bandwidth Management. These functionalities require the BME to collect application requirements, monitor channel conditions, and configure radio interfaces:

- The Admission Control functionality is critical in regulating resource usage when a service is activated. It evaluates resource availability and prevents the activation of services that exceed current bandwidth or violate predefined priorities or regulatory constraints. See clause 7.3.2 for details on different admission control techniques.

- The Bandwidth Management functionality ensures the adaptive allocation of radio resources to meet varying traffic demands. This mechanism is applied exclusively to services that have been admitted through admission control, ensuring that only authorized services utilize the available bandwidth. Additionally, the priorities of the messages are a key factor in bandwidth management. Higher-priority messages, such as safety-critical notifications (e.g. DENMs), are allocated bandwidth preferentially to ensure timely delivery. By dynamically adjusting resource allocation in real-time based on priorities and current conditions, this approach optimizes performance while adhering to predefined admission policies. This approach adjusts bandwidth allocation in real-time based on service requirements and current conditions. See clause 7.3.2 for examples of bandwidth management techniques.

The MHE can implement Traffic Shaping and Traffic Policing policies to manage the transmission of messages effectively, ensuring compliance with the configuration limits set by the BME:

- The Traffic Shaping policies can be used to smooth and regulate the traffic generated by each service, buffering or delaying excess traffic to ensure long-term compliance with traffic limits. The MHE can be responsible for performing this task to ensure that each service adheres to its allocated bandwidth and complies with the resource management policies defined by the BME. For common algorithms see clause 7.3.4.

- Traffic Policing ensures strict traffic limits by marking or discarding excess traffic without buffering. This enforcement guarantees that the aggregation of all the messages generated by the services and sent down to the lower layers operate within their permitted limits, maintaining overall network stability and fairness. See for more details clause 7.3.5.

These entities work collaboratively to optimize resource allocation and maintain QoS in vehicular communication systems. Figure 15 illustrates how these mechanisms could be integrated into the BME and MHE of RM.

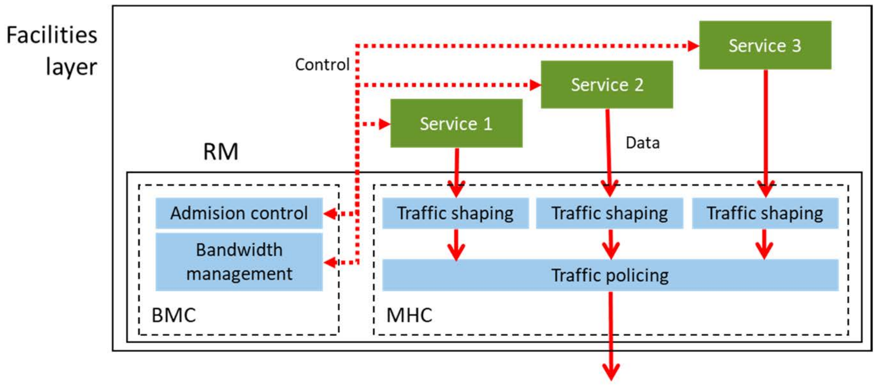


Figure 15: Architecture and mechanisms for Resource Management (RM) at Facilities layer

7.3.2 Admission Control Techniques

Peak Resource-Based Admission Control: Limits resource allocation based on the peak resource demands of a service. It avoids allocating resources that might exceed network capacity during peak usage periods. A practical example is restricting the activation of multiple service with a high peak message generation rate during a traffic jam.

Average Resources Admission Control: Ensures that the cumulative resource usage remains within bounds by summing up the average resource demands of all active services. This typically involves adding average resource demands, rather than peak values that could be rarely produced. For instance, this method might deactivate a service when the sum of average resources is higher than the available resources.

Equivalent Bandwidth Admission Control: Allocates resources based on the equivalent bandwidth required for a service, considering both average and peak demands. The equivalent bandwidth is calculated by combining the statistical distribution of traffic loads with the desired QoS parameters, such as delay, jitter, and packet loss rate. For example, if a service exhibits bursty traffic patterns, the equivalent bandwidth will factor in both the average rate and a margin to accommodate bursts, ensuring reliable operation. This approach is well-suited for services where peak demands need to be balanced against average usage to optimize resource allocation. This approach provides a more realistic estimation of resource requirements compared to simple peak or average calculations.

Statistical Admission Control: Utilizes probabilistic models to estimate resource requirements and allocate them based on expected traffic patterns. Unlike equivalent bandwidth control, statistical admission control focuses on the likelihood of multiple services requiring peak resources simultaneously. This approach uses historical data and probability distributions to anticipate resource needs, ensuring that resources are not over-allocated based on rare peak scenarios. In practice, it might prioritize safety-critical messages during periods of high traffic density, while reserving capacity for unexpected spikes in demand.

Policy-Based Admission Control: Implements resource constraints based on pre-defined policies, such as user priority levels or regulatory requirements, ensuring that critical services receive necessary resources. This method can be combined with other admission control mechanisms. For instance, policy-based rules can complement Statistical Admission Control by defining thresholds for resource allocation under specific conditions, such as prioritizing certain messages during congestion. Similarly, it can enhance Equivalent Bandwidth Admission Control by incorporating policies that adjust the equivalent bandwidth calculations based on application-specific or regulatory priorities, ensuring more granular and adaptive resource allocation. For example, a policy might specify that safety-critical messages always have precedence over awareness messages, regardless of statistical or bandwidth estimates.

7.3.3 Dynamic Bandwidth Management Techniques

Proportional Fairness: Balances resource distribution equitably among applications and services by ensuring that each of them receives a fair share relative to its needs, promoting overall network efficiency. Message priorities play a critical role in this mechanism. Messages are grouped by their priority levels, and resources are allocated proportionally within each group. For multiple priority levels, higher-priority groups are allocated resources first, and any remaining bandwidth is distributed among lower-priority groups using proportional fairness. For instance, DENMs (high-priority messages) are served first to ensure safety-critical operations, while CAMs and CPMs share the remaining bandwidth in proportion to their demands (assuming that they have the same priority).

Max-Min Fairness: Prioritizes applications and services with the least resources by maximizing their resource allocation without significantly impacting others, ensuring minimum fairness for all users. Message priorities are also important in this context. High-priority messages are allocated resources first, ensuring their timely delivery. Once higher-priority demands are satisfied, the remaining resources are distributed among lower-priority messages in a way that maximizes the minimum allocation, ensuring no service is entirely starved of resources.

Message Generation Scheduling: This advanced technique involves the BME directly controlling the scheduling of message generation by services. The BME considers the available bandwidth, current channel conditions, and message priorities to decide when each service can generate a new message. For instance, higher-priority messages might be scheduled for immediate generation, while lower-priority messages are deferred to avoid congestion. This approach ensures an optimal balance between resource utilization and the timely delivery of high-priority messages.

Machine Learning-Based Management: Applies predictive analytics to forecast traffic patterns and optimize bandwidth allocation proactively. Machine learning algorithms can also account for message priorities by learning from historical traffic patterns and adapting allocation strategies. For example, the system might predict an increase in high-priority messages and pre-emptively allocate bandwidth to accommodate this demand, ensuring timely message delivery across all priority levels.

7.3.4 Traffic Shaping Policies

Token Bucket Algorithm: Controls the flow of traffic by allowing bursts within a limit, regulated by token generation rates. The algorithm internally uses tokens, which are generated at a constant rate, to authorize the sending of packets. Each packet consumes a token, and traffic exceeding the token rate is delayed until tokens are available again. This mechanism ensures compliance with average traffic rates while permitting flexibility. For instance, it can handle sudden bursts of messages without disrupting other services.

Leaky Bucket Algorithm: Smooths traffic by enforcing a constant output rate, discarding excess data beyond the bucket's capacity. The algorithm uses a fixed-size bucket where packets are added at any rate but are released at a steady, predetermined rate. If the bucket overflows, excess packets are dropped. This ensures a steady flow of traffic and can be used to regulate bursts of messages to avoid network saturation.

Dual Token Bucket Algorithm: Combines two token buckets to manage multiple traffic priorities. The primary bucket regulates high-priority traffic, while the secondary bucket manages lower-priority traffic. High-priority traffic consumes tokens from the primary bucket, ensuring prompt transmission, while lower-priority traffic waits until both buckets have sufficient tokens.

Virtual Scheduling Algorithm: Simulates scheduling in a virtual timeline, ensuring fair distribution of resources by assigning each packet a virtual departure time. Packets are sent in the order of their virtual departure times, maintaining compliance with predefined traffic limits. This could prioritize transmission of event-driven messages over general status updates, ensuring timely delivery of essential information.

Generic Cell Rate Algorithm (GCRA): Monitors and regulates traffic based on cell arrival times to ensure adherence to the specified rate and burst tolerance. GCRA uses a virtual scheduling mechanism to check if incoming packets comply with the configured traffic profile. Non-compliant packets are either dropped or marked.

Sliding Window Algorithm: Controls the rate of data transmission by monitoring traffic over a moving time window. It calculates the volume of data sent during the window and ensures it does not exceed a predefined threshold. If traffic exceeds the limit, packets are delayed or dropped. This technique can be employed to manage the steady flow of messages from vehicles during a convoy, maintaining a balance between throughput and compliance with traffic limits.

7.3.5 Traffic Policing Policies

Unlike traffic shaping, which smooths traffic over time by buffering excess data, traffic policing focuses on immediate compliance by dropping or marking traffic that exceeds the defined limits. This ensures that the system adheres strictly to resource policies without introducing delays caused by buffering. Traffic policing mechanisms often include techniques similar to those used in traffic shaping, such as the Token Bucket and Leaky Bucket algorithms. However, in traffic policing, these mechanisms enforce strict limits by discarding non-compliant packets instead of buffering them. Figure 16 illustrates the differences between traffic policing and shaping when they are applied to the same message flow. In the figure, the vertical axes represent the message rate, and the horizontal axes represent the time.

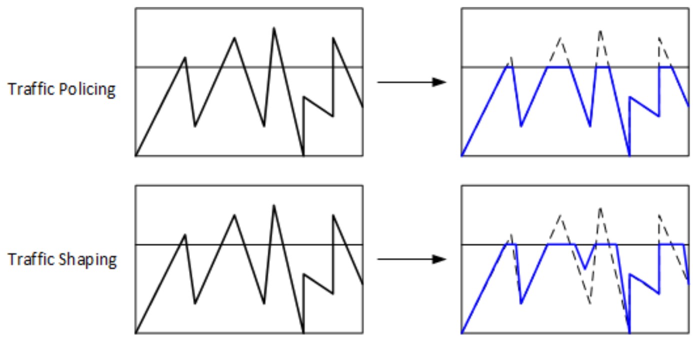


Figure 16: Effect of traffic Policing and Shaping on a message flow

7.3.6 A Basic solution

While RM can become highly complex when all functionalities are implemented in detail, a basic yet effective approach is achievable by focusing on essential requirements. The key idea is to implement a simple RM mechanism that still fulfils necessary operational and regulatory needs without the overhead of full complexity.

At the core of a basic solution is the division of functions between two primary entities: Admission Control and Bandwidth Management. Admission Control serves as the initial gatekeeper by applying straightforward policies to regulate service activation. For example, it may restrict the activation of certain services in some regions, or the simultaneous activation of multiple services in scenarios where their combined transmission rates might exceed system limits. This regulation ensures that certain services are either disallowed in specific regions or managed to avoid extreme cases, such as preventing three services transmitting at high frequencies when the overall system can only support a lower cumulative rate.

Once services pass through Admission Control, the BME could allocate the available resources based on proportional fairness. In this approach, the system calculates (at the Access or Facilities layer, see clause 5.5) the resources available for the ITS-S according to the current channel load and distributes it among the active services according to their priority and recent resource usage. The distribution continues in a tiered fashion, starting with the highest priority services, and if any resources remain, they are assigned to lower priority services. Each service then receives an allocation expressed as e.g. bits per second, which is derived from translating the computed resources and the ALI with the table discussed in clause 5.5. While a fixed ALI might be sufficient for basic implementations, more advanced setups could dynamically adjust the ALI based on factors like available bandwidth, radio access technology, or modulation and coding schemes.

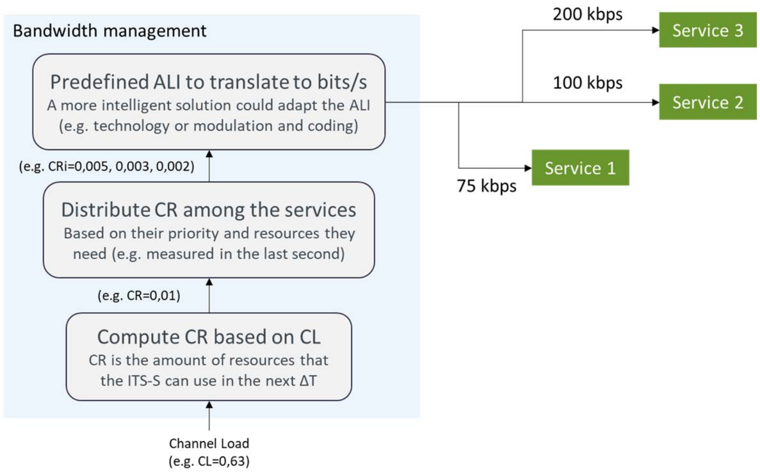


Figure 17: Example of basic bandwidth management

A more detailed operation of the bandwidth management performed by RM is illustrated in Figure 18. As can be observed, during the first RM interval, $t = [0, \Delta T)$, there is no information yet available about the current channel load, so all services operate in an open-loop mode—each one generating message according to its default behavior or immediate needs. At the end of this first interval, $t = \Delta T$, the RM module obtains a new channel load measurement and collects the individual resource needs R_i from each active service i. Based on the aggregate demand $R = \sum_i R_i$, and the channel load, RM computes the available resources for the station and for each service, referred to as CR_i for service i. In the second RM interval, $t = [\Delta T, 2\Delta T)$, services adjust their message generation rates according to the assigned resources, thereby entering a closed-loop control phase. At the end of this second interval, the process is repeated: a new measurement of the channel load is taken, updated service demands are reported, and RM recomputes the available resources for the station and for each service.

This approach that computes the available resources for the station at the Facilities layer opens the door to handle scenarios with heterogeneous resource needs (see clause 6.4). Congestion control mechanisms such as achieving weighted-fairness [i.30] could be applied to allow that different ITS-S experiencing the same channel load have a different number of available resources depending on their service needs.

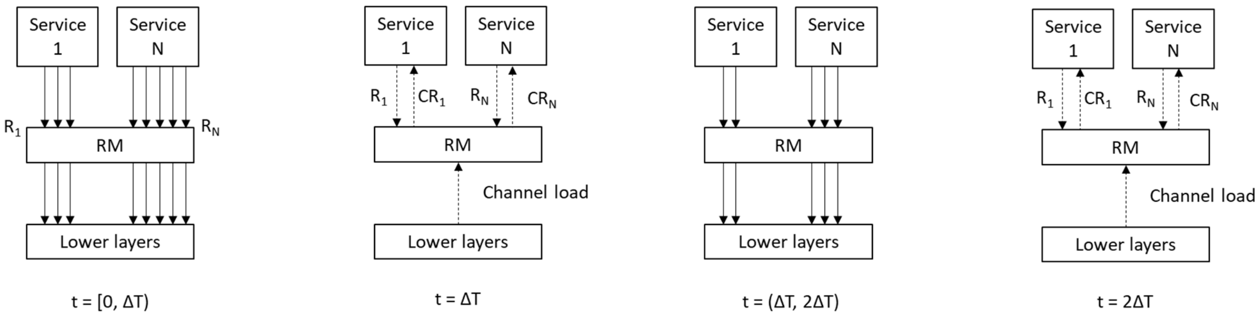


Figure 18: Interaction between RM and services for bandwidth management

In addition to these primary functions, the RM process includes a message handling entity. In the basic solution, the system does not actively enforce traffic shaping or policing. Instead, it assumes that services will adhere to their resource assignments and that any temporary excesses will not significantly disrupt overall performance. Continuous monitoring of resource consumption - for instance, measuring the number of bits generated by each service in the last second provides feedback to adjust future resource distributions.

Overall, this basic RM mechanism illustrates a balanced approach where simplicity meets functionality. It relies on policy-based restrictions to manage service admission and uses a fair, priority-driven method to allocate bandwidth, while also keeping the system flexible enough to accommodate service behaviour without overly rigid controls. This foundation can be enhanced with more sophisticated techniques as needed, but it provides an effective baseline for managing resources in a controlled environment. An evaluation showing the feasibility of this basic solution is presented in Annex E.

Annex A: Rationale for the limits imposed by congestion control

The rationale for equation (2) in clause 5.2.4, which is derived from ETSI EN 302 571 [i.16], is described in ETSI TS 103 175 [i.15] and hereafter elaborated.

The starting point is the derivation of the maximum channel occupation of the generic station, hereafter called CR_{Limit}, as a function of the overall channel load, which is measured by the CBR. The maximum channel occupation can also be written as a function of the minimum time between two consecutive transmissions where the transceiver is not allowed to generate a signal, denoted as $T_{offLimit}$ and the duration of the last transmission T_{on} as:

$$CR_{Limit} = \frac{T_{on}}{T_{on}+T_{offLimit}} \quad \text{(A.1)}$$

Since the maximum channel occupation of the generic station needs to depend to the number of stations, but this information is not explicitly available, an assumption is made on the CBR. In particular, it is assumed that the maximum CBR, denoted as CBR_{Limit}, is a function of the number of stations concurring to access the same channel in the same area, denoted as N_{sta}. This assumption is introduced to allow an implicit derivation of the number of station directly from the measured CBR. This assumption can also be written as:

$$CBR_{Limit} = N_{sta} \times a + b \quad \text{(A.2)}$$

where a and b are two constants, and thus:

$$N_{sta} = \frac{CBR_{Limit}-b}{a} \quad \text{(A.3)}$$

Given the maximum CBR and the number of stations, the maximum channel occupation of the generic station can be calculated as:

$$CR_{Limit} = \frac{CBR_{Limit}}{N_{sta}} = \frac{CBR_{Limit}}{\left(\frac{CBR_{Limit}-b}{a}\right)} = \frac{a \times CBR_{Limit}}{CBR_{Limit}-b} \quad \text{(A.4)}$$

Using equations (A.1) and (A.4), by first writing $T_{offLimit}$ as a function of CR_{Limit} and then substituting CR_{Limit} with its expression as a function of CBR_{Limit}, resulting in:

$$T_{offLimit} = T_{on} \times \left(\frac{1-CR_{Limit}}{CR_{Limit}}\right) = T_{on} \times \left(\frac{1}{CR_{Limit}} - 1\right) = T_{on} \times \left(\frac{1}{a}\frac{CBR_{Limit}-b}{CBR_{Limit}} - 1\right) \quad \text{(A.5)}$$

The final step is that the measured CBR, indicated as CBR, is assumed equal to CBR_{Limit}, which brings to:

$$T_{offLimit} = T_{on} \times \left(\frac{1}{a}\frac{CBR-b}{CBR} - 1\right) \quad \text{(A.6)}$$

Which corresponds to equation (2) in clause 5.2.4 (and thus the limit indicated by ETSI EN 302 571 [i.16]) when $a = 1/4\,000$ and $b = 0{,}62$ are used. These two parameters were empirically determined as a good trade-off between convergence speed and stability.

It can be observed that the assumption that the measured CBR is equal to the limit implies that each station tends to consider the channel always congested, which may not be true. However, if the channel is not congested the effect is that the station under observation underestimates the number of contending stations and overestimates the portion of resources it can use. Given that the channel is not really congested, this may only cause an increase of the CBR until it actually reaches the level of congestion.

Annex B: Methods to characterize congestion control algorithms

B.1 Modelling of a rate control loop

B.1.1 Definition of the channel busy ratio limit

Several independent transmitters that want to share a radio channel without coordination by a central station need a mechanism that avoids data packet collisions and that takes care of the available channel resources. For simplification of the derived model of the congestion control it is assumed that data packet collisions are effectively avoided by the CSMA/CA algorithm when the radio channel is not overloaded and that the load is only controlled by the rate and duration of the data packets. This simplifies the description and makes an analytic evaluation possible. Anyhow, the congestion control is designed to avoid packet collisions by keeping the channel load reasonably low.

Obviously, all transmitters should not try to put more packets on the channel than the channel capacity allows. It is even so that the full channel capacity can only be reached at the cost of massive packet collisions (packets of different transmitters overlap in time). Most of these overlapping packets cannot be decoded in the receivers and are therefore waste of channel resources. Hence, the optimum channel load is well below the channel capacity. A congestion control should allow a system of multiple transmitters to use the channel resources up to, or close to, the optimum channel load, and avoid an operation above this load limit.

The channel resources R can be seen as percentage of the total available transmit time for all transmitting nodes ($0 < R < 1$). Whereas the channel utilization u is the percentage of the total available time each node can transmit. Assuming that all nodes N share the available resources equally, u can be found by dividing R by N (equation (B.1)):

$$u = \frac{R}{N} \qquad \text{(B.1)}$$

When the channel load is the only input parameter to the control algorithm, the transmitters do not know the number of other transmitters contributing to the channel load, and the distribution of the channel load cannot be done by just dividing the available resources by the number of nodes. The control algorithm can only inherently divide the resources equally from only knowing the channel load. This is possible when the channel load has a one-to-one relation to the number of nodes.

ETSI specifies in ETSI TS 103 175 [i.15] a linear relation between the channel load limit and the number of contributing nodes (see also Annex A and equation (B.2)):

$$CBR_{Limit} = N \times a + b \qquad \text{(B.2)}$$

Where CBR_{Limit} is the upper channel busy ratio limit that should not be exceeded by the measured channel load CL and a and b are parameters. An example for a = 1/4 000 and b = 0,62 based on ETSI EN 303 797 [i.32] is shown in Figure B.1.

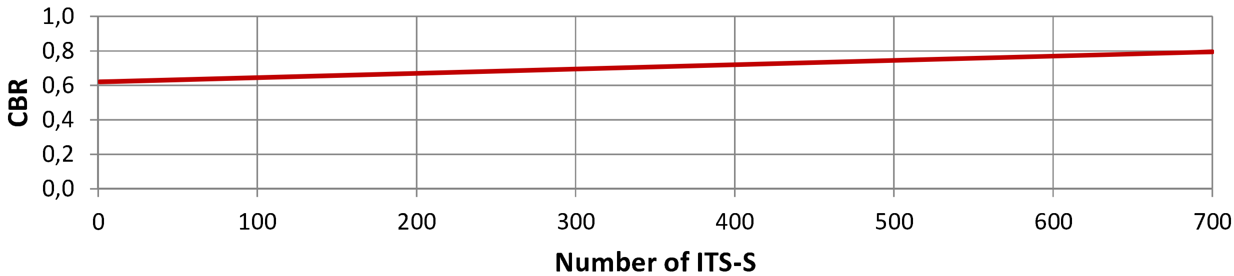


Figure B.1: Example of a CBR limit according to equation (B.2)

The measured channel load results from the sum of the channel utilizations u of all nodes (equation (B.3)):

$$CL = \sum_{k=1}^{N} u_k \qquad \text{(B.3)}$$

Figure B.2 shows the contribution of each rate controller to the channel load. Other types of congestion control are not considered in the present document since they are less effective and more complicated to implement.

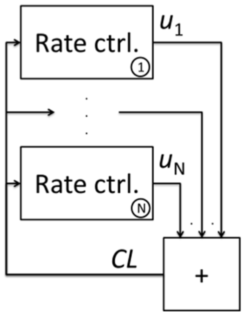


Figure B.2: Contribution of each rate control to the channel load

When all rate controllers are the same, all channel utilizations u_k are equal to u and the resulting channel load CL is given by equation (B.4):

$$CL = N \times u \qquad \text{(B.4)}$$

When these equal rate controllers are working properly and equilibrium is reached, all u_k are equal $\tilde{u}$ and the resulting steady state channel load $\widetilde{CL}$ is given by equation (B.5).

$$\widetilde{CL} = N \times \tilde{u} \qquad \text{(B.5)}$$

B.1.2 Function of the rate controller

The channel utilization u of each node is controlled by adjusting the time T_{off} in between two transmitted packages according to the given duration T_{on} of the package (equation B.6), so that the resulting utilization u does not exceed the limit U_{max} given by (equation (B.7) and equation (B.8)):

$$u = \frac{T_{on}}{T_{on}+T_{off}} \qquad \text{(B.6)}$$

$$U_{max} = \frac{CBR_{Limit}}{N} \qquad \text{(B.7)}$$

$$u \leq U_{max} \qquad \text{(B.8)}$$

Hence, the rate controller is working in discrete time steps with a variable length of $T_{off} + T_{on}$.

The number of nodes N is unknown to each node, but an upper bound for the number of nodes N_{est} can be estimated from the measured channel load CL when assuming that $CL \leq CBR_{\text{Limit}}$ and by substituting CBR_{Limit} by CL in equation (B.2). Since CL can be smaller than b, the minimum number of estimated nodes N_{est} is fixed to at least one in equation (B.9):

$$N_{est} = max\left(1; \frac{CL-b}{a}\right) \qquad \text{(B.9)}$$

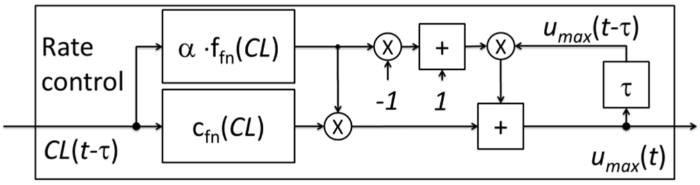
From this a lower bound of U_{max} can be calculated by use of equation (B.7) and equation (B.2) when substituting N with N_{est} (see equation (B.10)):

$$U_{max} = \frac{N_{est} \cdot a+b}{N_{est}} \qquad \text{(B.10)}$$

The distributed rate controllers are not synchronized and therefore the measurement period τ for the channel load determination should be long enough. Either 100 ms or a duration of T_{off} +T_{on} are considered. Also, a dissemination of the CL values between the nodes can be foreseen to increase the robustness of the measurement.

B.1.3 Structure of a feedback controller

Figure B.3 shows a common basic structure how a discrete time rate control in each node can be implemented. The channel load $CL(t\text{-}\tau)$ measured in the previous time step at $t\text{-}\tau$ and the previous dynamic upper bound of the channel utilization $u_{max}(t\text{-}\tau)$ is used by the control function c_{fn} and the filter function $\alpha \times f_{fn}$ to determine the next dynamic channel utilization limit $u_{max}(t)$.



NOTE: The filter function f_{fn} might not be used, and even the factor α is set to one in most implementations.

Figure B.3: Block diagram of the rate control

Equation (B.11) relates to the rate controller shown in Figure B.3 and results in the dynamic channel utilization limit $u_{max}(t)$:

$$u_{max}(t) = \alpha \times f_{fn}\big(CL(t-\tau)\big) \times c_{fn}\big(CL(t-\tau)\big) + \Big(1 - \alpha \times f_{fn}\big(CL(t-\tau)\big)\Big) \times u_{max}(t-\tau) \quad \text{(B.11)}$$

When no filtering to the input signal $CL(t\text{-}\tau)$ is used in the controller, the filter function is constant and equal to one ($\alpha \times f_{fn} = 1$). In this case the dynamic channel utilization $u_{max}(t)$ is just given by the control function $c_{fn}(CL(t\text{-}\tau)) = u_{max}(t)$. This type of controller implementation is the simplest one, but it is prone to instabilities, since the stability and the control equilibrium are both given by the control function. Using a filtering function that differs from one ($\alpha \times f_{fn} \neq 1$), offers the possibility to decouple the control equilibrium from the controller stability as will be shown in clause B.2.

B.2 Control equilibrium

The control equilibrium for a control function f_{fn} is the steady state channel load $CL(t\text{-}\tau) = CL(t) = CL$ produced by a given number of nodes N that are all utilizing the channel with the same constant $u_{max} = u_1 = u_2 = \ldots = u_N = U_{max}$. For the steady state equilibrium equation (B.11) simplifies to equation (B.12):

$$U_{max} = \alpha \times f_{fn}(CL) \times c_{fn}(CL) + \big(1 - \alpha \times f_{fn}(CL)\big) \times U_{max} \quad \text{(B.12)}$$

From equation B.12 the channel utilization limit U_{max} as function of the channel load CL can be determined:

$$U_{max} = c_{fn}(CL) \quad \text{(B.13)}$$

Equation (B.13) shows that the filter function $\alpha \times f_{fn}$ has no influence on the control equilibrium CL and on the steady state channel utilization U_{max}. Only the control function c_{fn} determines the steady state characteristics of the controller shown in Figure B.3.

Clause B.3 and clause B.4 will show that the controller stability and the controller dynamics are not only influenced by the control function c_{fn} but also by filter function $\alpha \times f_{fn}$.

Finally, when substituting $U_{max}=u$ into equation (B.4), the steady state control equilibrium $\widetilde{CL}$ can be calculated from the control function c_{fn} for a given number of nodes N when solving equation (B.14) for $\widetilde{CL}$:

$$\widetilde{CL} = N \times c_{fn}\big(\widetilde{CL}\big) \quad \text{(B.14)}$$

B.3 Stability

B.3.1 Different types of stability

Congestion control algorithms are implemented as discrete-time controllers. That means that they will change the control value (channel utilization u) not at any arbitrary point in time, but only after discrete time intervals τ. This implies that a stability evaluation cannot be done by inspecting the continuous time differential equation of the control loop alone. In addition, discrete time oscillations, caused by too long intervals τ, need to be studied.

Controller stability is given when within a given range of node numbers N the control loop converges independently of the initial channel load $CL(t_0)$ to a channel load $CL(t)$ that is bounded within a defined small range (bounded stability). Such a bounded stability criteria is necessary to account for quantization steps in time and channel utilization.

B.3.2 Description of the controller by a differential equation

To analyse the dynamic behaviour of the distributed controllers the discrete time control equation (B.11) is converted into the differential equation (B.17). This is done in a first step by dividing equation (B.11) by the discrete time step size τ and rewriting it in such way that the difference equation (B.15) is obtained.

The left side of equation (B.15) is the ratio between the utilization difference $\Delta u_{max} = u_{max}(t)$-$u_{max}(t$-$\tau)$ and the time difference $\Delta t = \tau$ between $u_{max}(t)$ and $u_{max}(t$-$\tau)$. When additionally substituting t-τ by t-Δt on the right side of equation (B.15), equation (B.16) is obtained.

Under the assumption that the control function $\mathrm{c_{fn}}$ and the filtering function $\mathrm{f_{fn}}$ are both analytic, the difference equation (B.16) can be converted by a limiting process into the differential equation (B.17). Where $u(t)$ is the continuous time function corresponding to the discrete time function $u_{max}(t)$:

$$\frac{u_{max}(t)-u_{max}(t-\tau)}{\tau} = \frac{\alpha\times \mathrm{f_{fn}}\big(CL(t-\tau)\big)}{\tau} \times \Big(\mathrm{c_{fn}}\big(CL(t-\tau)\big) - u_{max}(t-\tau)\Big) \quad \text{(B.15)}$$

$$\frac{\Delta u_{max}}{\Delta t} = \frac{\alpha\times \mathrm{f_{fn}}\big(CL(t-\Delta t)\big)}{\tau} \times \Big(\mathrm{c_{fn}}\big(CL(t-\Delta t)\big) - u_{max}(t-\Delta t)\Big) \Big| \lim_{\Delta t\to 0} \quad \text{(B.16)}$$

$$u' = \frac{du}{dt} = \frac{\alpha\times \mathrm{f_{fn}}\big(CL(t)\big)}{\tau} \times \big(\mathrm{c_{fn}}\big(CL(t)\big) - u(t)\big) \quad \text{(B.17)}$$

For N identical controllers, the time behaviour of the channel utilization $u(t)$ as function of the number of nodes N results from solving differential equation (B.18):

$$u' = \frac{\alpha\times \mathrm{f_{fn}}(N\times u)}{\tau} \times (\mathrm{c_{fn}}(N\times u) - u) \quad \text{(B.18)}$$

By use of equation (B.4) also the time behaviour of the channel load $CL(t)$ can be obtained from equation (B.18).

B.3.3 Convergence

Even when equation (B.14) has a solution for the equilibrium channel load $\widetilde{CL}$ that lies within 0 and 1 as given by the definition of the channel load, the controller might not converge to this equilibrium. This can be the case when the slope of the control characteristic is not monotonic towards the point of equilibrium. This could be caused by an ill formed control function with ripples in the slope, or when equation (B.14) has more than one solution for $\widetilde{CL}$.

For convergence following criteria based on equation (B.16) should be met:

$$(\mathrm{c_{fn}}(N\times u) - u < 0 \text{ for } 1 > u > \tilde{u}) \text{ and } (\mathrm{c_{fn}}(N\times u) - u > 0 \text{ for } 0 < u < \tilde{u}) \quad \text{(B.19)}$$

Where $\tilde{u}$ is the controller equilibrium as calculated from equation (B.5) an equation (B.14).

B.3.4 Continuous-time stability

The convergence criteria given by equation (B.19) are not sufficient to guarantee a stable control loop. Only when in addition the solution of differential equation (B.18) shows at least a decaying oscillation, or even no oscillation, for the channel utilization $u(t)$ within a given range of node numbers N and for all starting values $0<u(t_0)<1$, the continuous-time control loop is stable for this range of N.

B.3.5 Discrete-time Stability

Since the congestion control is implemented as controller that measures the channel load CL and changes the channel utilization u_{max} repeatedly after discrete-time intervals τ, the continuous-time stability as given in clause B.3.4 is a necessary ,but not sufficient criteria for the stability of the control loop.

When the continuous time solution converges for all stations to the same stable value $\tilde{u}$ (see equation (B.5)), the discrete-time controller can be unstable when $u(t)$ changes in one time step within an interval of τ by at least twice the distance to $\tilde{u}$ (overshoot). The $u(t)$ change over one time step with a duration of τ equals $u'(t) \times \tau$. Where $u'(t)$ is the slope of the $u(t)$ function at the time t. From this, the discrete time realization for a controller that overshoots $\tilde{u}$ is stable when equation (B.20) is fulfilled for all $t \geq 0$ within a given range of node numbers N and for all starting values $0 < u(t_0) < 1$:

$$\tau \times |u'(t)| < 2 \times |\tilde{u} - u(t)| \quad \text{(B.20)}$$

This implies that a discrete time controller is also stable when the channel utilization does not overshoot the controller equilibrium $\tilde{u}$ as expressed by equation B.21. When equation B.21 is met, the controller shows a favourable behaviour without discrete time oscillations:

$$\tau \times |u'(t)| \leq |\tilde{u} - u(t)| \quad \text{(B.21)}$$

Since τ is always positive, equation (B.20) and equation (B.18) can be combined to form the discrete-time stability criterium given in equation (B.22):

$$\left|\alpha \times \mathrm{f_{fn}}\big(N \times u(t)\big) \times \Big(\mathrm{c_{fn}}\big(N \times u(t)\big) - u(t)\Big)\right| < 2 \times |\tilde{u} - u(t)| \quad \text{(B.22)}$$

This criterium is a function of the channel utilization $u(t)$ and the node number N. It should be fulfilled within the whole range of $0 < u(t) <1$ and within the given range of node numbers N where the controller should be stable.

B.4 Control dynamics

The convergence speed of the controller can be characterized with a channel load step function. Since the equilibrium might not be exactly reached because of quantization effects and measurement noise, a definition of the convergence speed by the time until a certain percentage of the channel load equilibrium is reached can be used. In addition, it makes sense to characterize the decay time of the controller oscillations, since it can be much longer that the time the channel utilization reaches a certain percentage of the equilibrium. Another criterium to characterize the control dynamics is the overshoot and undershoot relative to the equilibrium channel utilization when a channel load step function is applied. Also, the step size of the channel utilization quantization is an important parameter when defining the bounded stability criteria.

To get a first impression of the convergence speed and the decay time of the controller oscillations an inspection of the result of the continuous time differential equation (B.18) is helpful. But this result is just a guide value for the discrete time controller as given by equation (B.11). The discrete time implementation can converge even faster than the continuous time controller when $\tau \times |u'(t)| = |\tilde{u} - u(t)|$, or much slower when only equation (B.22) is fulfilled.

The control dynamics of the discrete time controller can be calculated iteratively from equation (B.11). Only for very simple control functions $\mathrm{c_{fn}}$ and filtering functions $\mathrm{f_{fn}}$ exact closed form analytic solutions are possible for the dynamics of a discrete time controller.

Annex C: Characterization of congestion control algorithms

C.1 Reactive congestion control

C.1.1 Control equilibrium of the reactive congestion control

The reactive congestion control algorithm is based on a control function for T_{off} (equation (C.1)):

$$T_{off}(t) = CBR(t-\tau) \times T_a + T_b \tag{C.1}$$

Where T_a and T_b are parameters and the channel busy ratio *CBR* is the measured channel load. This implies that the equilibrium channel load *CL* is not only depending on the number of nodes *N*, but also on the duration of the transmissions T_{on}. This follows from equation (B.4) and equation (B.6) when assuming that *CL* is equal *CBR* (*CL* = *CBR*):

$$CL = N \times \frac{T_{on}}{T_{on}+T_{off}} \tag{C.2}$$

When combining equation (C.1) with equation (C.2) the T_{off} control function can be rewritten to:

$$T_{off}(t) = N \times \frac{T_{on}(t-\tau)}{T_{on}(t-\tau)+T_{off}(t-\tau)} \times T_a + T_b \tag{C.3}$$

To simplify the calculation, it is assumed that all nodes use the same time independent T_{on} = const. With this assumption and based on equation (C.3) the equilibrium $\tilde{T}_{off}$ can be calculated from equation (C.4):

$$\tilde{T}_{off} = N \times \frac{T_{on}}{T_{on}+\tilde{T}_{off}} \times T_a + T_b \tag{C.4}$$

What leads to a quadratic equation for $\tilde{T}_{off}$:

$$\tilde{T}_{off}^2 + \tilde{T}_{off} \times (T_{on} - T_b) - T_{on} \times (T_b + N \times T_a) = 0 \tag{C.5}$$

Equation (C.5) has following two solutions for $\tilde{T}_{off}$:

$$\tilde{T}_{off}\Big|_2^1 = \frac{1}{2}\left((T_b - T_{on}) \pm \sqrt{(T_{on} + T_b)^2 + 4 \times T_{on} \times N \times T_a}\right) \tag{C.6}$$

The channel load equilibrium $\widetilde{CL}$ results from equation (C.2) when substituting T_{off} by $\tilde{T}_{off}$ obtained from equation (C.6).

$$\widetilde{CL}\Big|_2^1 = \frac{2 \times N \times T_{on}}{T_b + T_{on} \pm \sqrt{(T_{on}+T_b)^2 + 4 \times T_{on} \times N \times T_a}} \tag{C.7}$$

As a result of the quadratic equation (C.5) there are two channel load equilibrium $\widetilde{CL}$ solutions given by equation (C.7). For the control stability this requires that only one $\widetilde{CL}$ solution lies within the channel load boundaries of $0<CL<1$ and an analysis of the convergence within this channel load range is necessary.

As an example, from Table A.1 and Table A.2 in ETSI TS 102 687 [i.14] the parameters T_a and T_b that fit best to the table entries can be calculated by a linear regression as shown in Figure C.1 and Figure C.2. The results are listed in Table C.1.

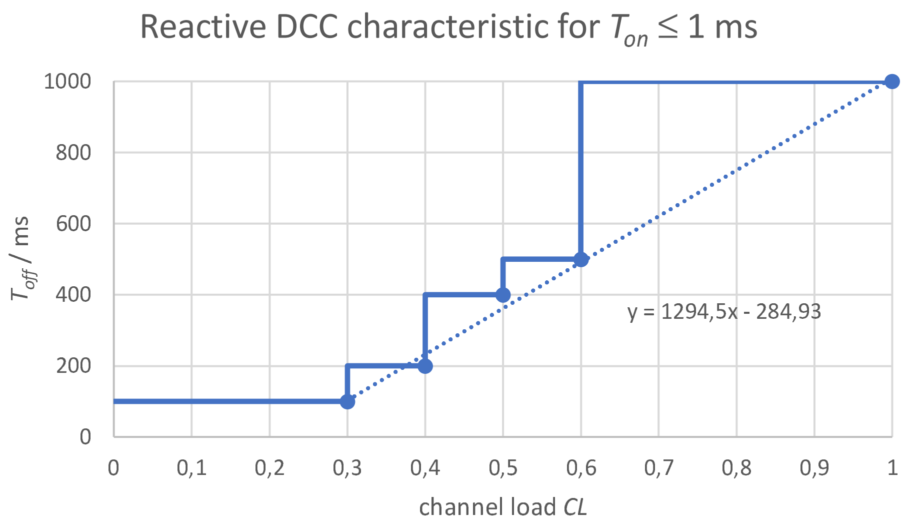


Figure C.1: Linear regression of the DCC function given in Table A.1 of ETSI TS 102 687 [i.14]

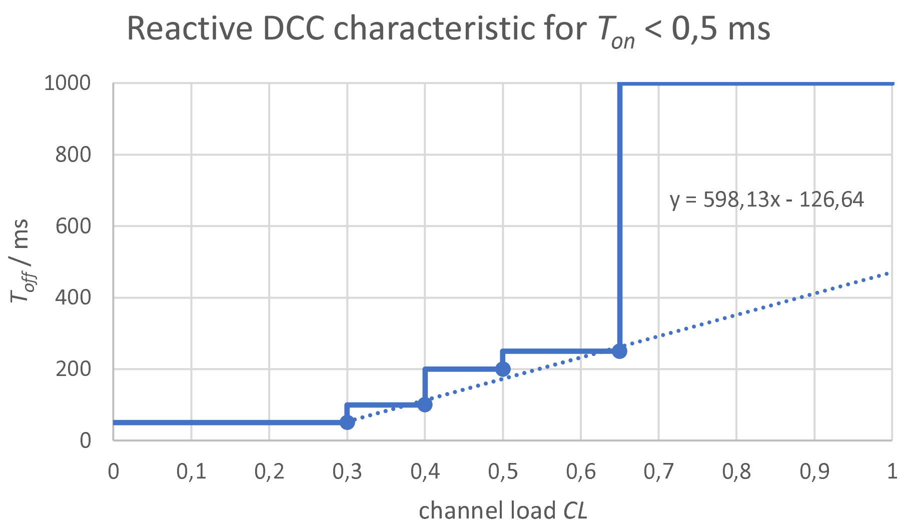


Figure C.2: Linear regression of the DCC function given in Table A.2 of ETSI TS 102 687 [i.14]

Table C.1: Parameters T_a and T_b derived from ETSI TS 102 687 [i.14]

Parameter	T_{on} < 0,5 ms	0,5 ms ≤ T_{on} ≤ 1 ms
T_a	598 ms	1 295 ms
T_b	-127 ms	-285 ms

Figure C.3 and Figure C.4 show the positive channel load equilibria calculated from equation (C.7) as function of the node number N and the parameters given in Table C.1. The dashed lines show the results outside the T_{on} range to show that these results either over utilize or underutilize the channel for large node numbers. The dotted line is the CBR limit for the example based on ETSI EN 303 797 [i.32] shown in Figure C.3.

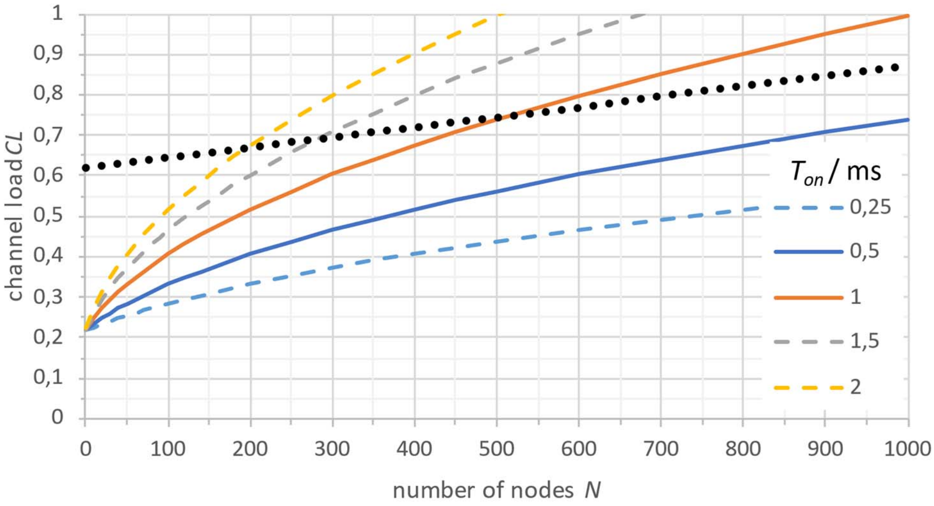


Figure C.3: Equilibrium channel load of reactive congestion control for T_a and T_b from Table C.1 for 0,5 ms <T_{on} ≤ 1 ms

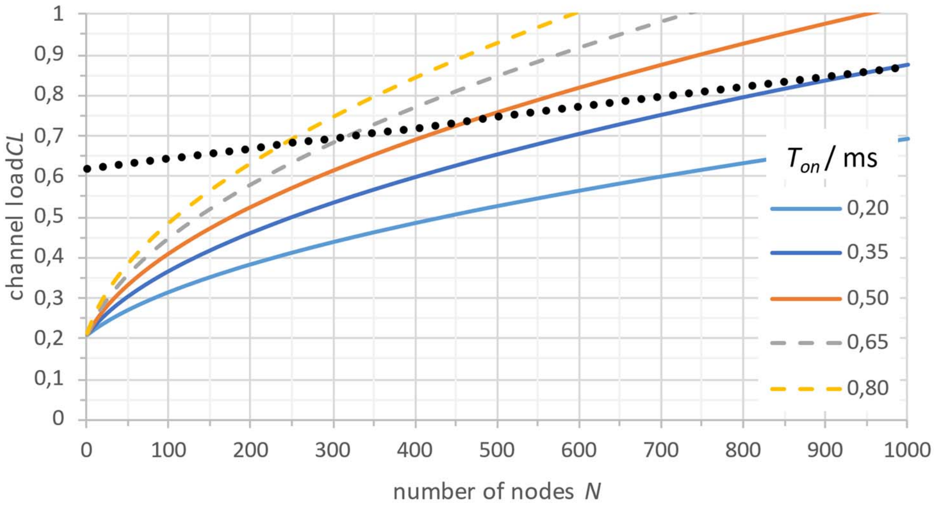


Figure C.4: Equilibrium channel load of reactive congestion control for T_a and T_b from Table C.1 for T_{on} ≤ 0,5 ms

C.1.2 Convergence of the reactive congestion control

To analyse the convergence, the channel load control function c_{fn} of the reactive controller is be determined by first substituting equation (C.1) in equation (C.2) under the assumption that *CL=CBR*:

$$CL = N \times \frac{T_{on}}{T_{on} + CL \times T_a + T_b} \quad \text{(C.8)}$$

With the help of equation (B.14) the channel load control function c_{fn} can be found from equation (C.8):

$$c_{fn}(CL) = \frac{T_{on}}{CL \times T_a + T_b + T_{on}} \quad \text{(C.9)}$$

To check the convergence criteria given by equation (B.19), the channel load control function c_{fn} (equation (C.9)) is used:

$$\left(\frac{T_{on}}{N \times u \times T_a + T_b + T_{on}} - u < 0 \text{ for } 1 > u > \tilde{u}|_2^1\right) \text{ and } \left(\frac{T_{on}}{N \times u \times T_a + T_b + T_{on}} - u > 0 \text{ for } 0 < u < \tilde{u}|_2^1\right) \quad \text{(C.10)}$$

Where $\tilde{u}|_2^1$ represents the solution of the equilibrium channel utilization that is within the allowed range of $0 < \tilde{u} < 1$ obtained from equation (B.5) and equation (C.7). In case equation (C.7) has two solutions within this range, convergence is only given when the range of u is limited in such a way that only one solution is within this range and equation (C.10) is fulfilled there.

Or in terms of CL equation (C.10) can be written as equation (C.11):

$$\left(\mathrm{c_{fn}}(CL) - CL < 0 \text{ for } 1 > CL > \widetilde{CL}\Big|_2^1\right) \text{ and } \left(\mathrm{c_{fn}}(CL) - CL > 0 \text{ for } 0 < CL < \widetilde{CL}\Big|_2^1\right) \qquad \text{(C.11)}$$

When applying the convergence criteria to the examples based on ETSI TS 102 687 [i.14] given in in Table C.1 it can be seen form Figure C.5 and Figure C.6 that the reactive congestion control will converge to the negative equilibrium when *CL* is somewhere below 0,2. The correct convergence area above a *CL* of around 0,2 to the positive *CL* equilibrium is highlighted in green, the wrong convergence to the negative *CL* equilibrium is marked in red, and the negative (impossible) *CL* area is marked in white. This is the reason why this control algorithm needs to be implemented by a table that has no entries below a *CL* of 0,3 to avoid a convergence to the impossible negative *CL* equilibrium.

The convergence direction is given by $\mathrm{c_{fn}}(CL)$ which has a pole when $CL \times T_a + T_b + T_{on}$ gets zero. Since T_a and T_{on} are always positive, a pole for a positive *CL* can only happen when T_b is negative. For the example given in in Table C.1 and when assuming that the T_{on} values are below 2 ms the pole is at a $CL = (285\ ms - T_{on})/1\ 295\ ms \approx 0{,}22$ or at a $CL = (127\ ms - T_{on})/598\ ms \approx 0{,}21$ as shown in Figure C.5 and Figure C.6.

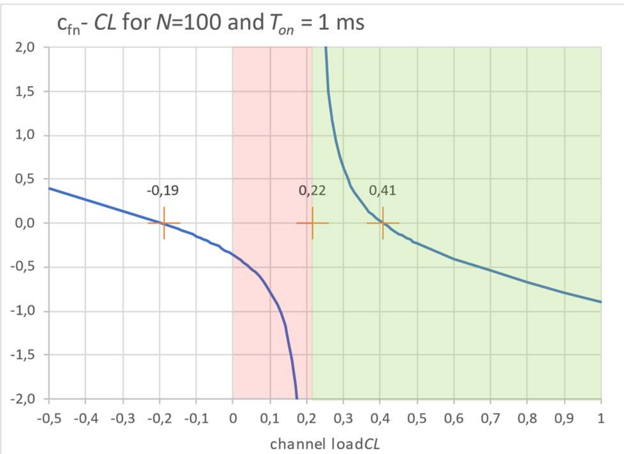


Figure C.5: Convergence criteria function for *N* = 100, *T*$_{on}$ = 1 ms, and the parameters given in Table C.1

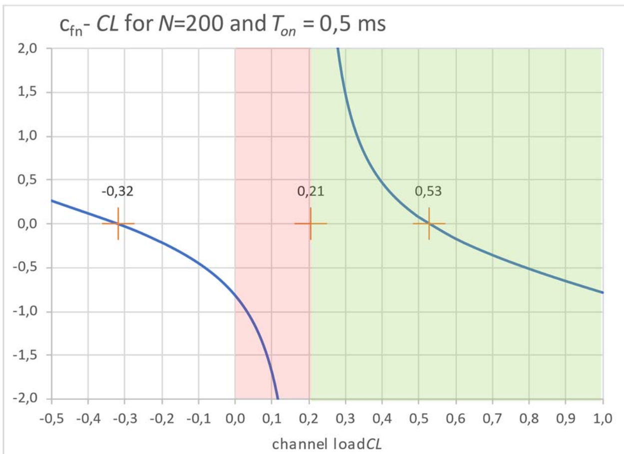


Figure C.6: Convergence criteria function for *N* = 100, *T*$_{on}$ = 0,5 ms, and the parameters given in Table C.1

C.1.3 Stability of the reactive congestion control

C.1.3.1 Stability of the reactive congestion control without filtering according to ETSI TS 102 687

C.1.3.1.1 Continuous time stability

According to clause B.1.3 the control equation without filtering for a reactive congestion control is:

$$u_{max}(t) = \frac{T_{on}}{CL(t-\tau) \times T_a + T_b + T_{on}} \qquad \text{(C.12)}$$

Equation (C.12) leads with equation (B.17) and equation (B.4) to differential equation (C.13) hat can be solved by separation of the variables:

$$\frac{dCL}{dt} = \frac{1}{\tau} \times \left(\frac{N \times T_{on}}{CL \times T_a + T_b + T_{on}} - CL\right) \qquad \text{(C.13)}$$

$$\tau \times \int \frac{1}{\left(\frac{N \times T_{on}}{CL \times T_a + T_b + T_{on}} - CL\right)} dCL = \int 1\, dt \quad \text{(C.14)}$$

After some calculations this leads to a solution for t in relation to the step size τ given by equation (C.15):

$$\frac{t+T_0}{\tau} = -\frac{1}{2} \times \left(\log(T_a \times CL^2 + B \times CL - N \times T_{\mathrm{on}}) - \frac{B}{A} \times \log\left(\frac{2 \times CL \times T_a + B - A}{2 \times CL \times T_a + B + A}\right)\right) \quad \text{(C.15)}$$

Where T_0 is an integration constant to adjust the time offset and A and B are constants given by:

$$\sqrt{{T_{\mathrm{on}}}^2 + (2 \times T_b + 4 \times N \times T_a) \times T_{\mathrm{on}} + {T_b}^2} = A \text{ and } T_b + T_{on} = B \quad \text{(C.16)}$$

The solution of the differential equation shows no oscillating term. Together with the convergence criteria and the initial value, this defines a parameter range and boundary conditions where a stable convergence to the control equilibrium of the continuous time reactive congestion control algorithm can be granted analytically.

Figure C.7 shows in red the solution given by equation (C.15) for N = 100, T_{on} = 1 ms, and the parameters given in Table C.1. Where the offset T_0 was adjusted so that the time t is zero for a channel load CL equal to one. The blue crosses show a numerical solution starting at a CL of one. The step size for this numeric solution is $\tau/50$. With this small step size, the difference to the analytic solution is neglectable. Furthermore, the analytic solution shows for a CL below 0,22 that the CL converges with increasing time t to decreasing channel load CL values and does not reach the positive equilibrium at CL = 0,405 as also shown in Figure C.5.

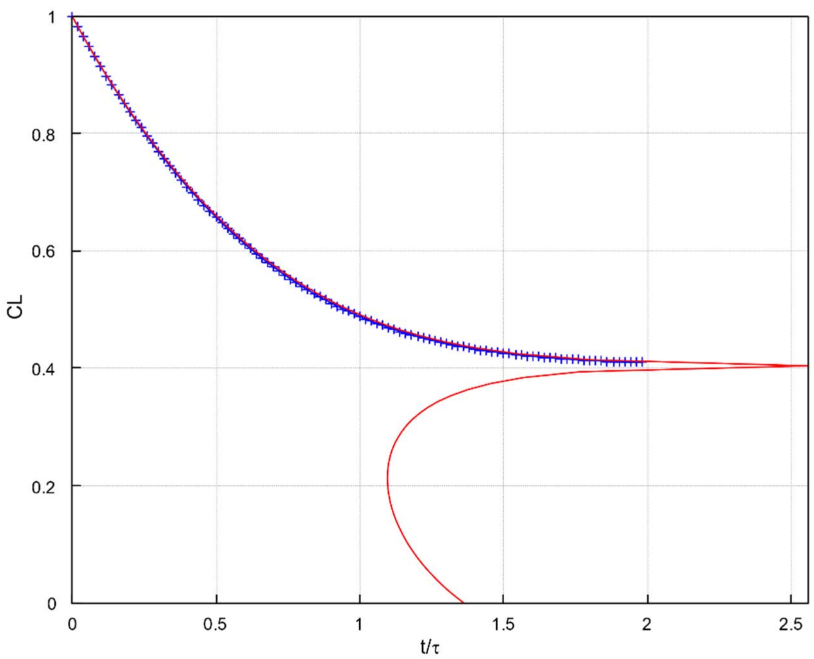


Figure C.7: Analytic and numeric solution of the continuous time reactive congestion control for N = 100, T_{on} = 1 ms, and the parameters given in Table C.1

C.1.3.1.2 Discrete T_{off} stability and convergence speed

Figure C.7 shows that a discrete time numeric solution of the reactive congestion control can reproduce the analytic solution when the time step size is small enough. This numeric solution uses 100 steps to converge to a value close to the equilibrium. Assuming a time step size of 100 ms, such an implementation could need around10 seconds to converge to a CL close to the control equilibrium, what is much longer than the time constants to be expected in mobile radio channels and road traffic scenarios.

The numeric solution shown in Figure C.7 is calculated iteratively using equation (C.17) and the parameters given in Table C.1 when setting $\alpha = 0{,}02$:

$$CL(t) = CL(t - \tau_n) + \alpha \times \left(\frac{N \times T_{on}}{CL(t-\tau_n) \times T_a + T_b + T_{on}} - CL(t - \tau_n)\right) \quad \text{(C.17)}$$

Equation (C.17) follows from equation (B.11) when using c_{fn} from equation (C.9), setting f_{fn} to one, and using equation (B.4) to calculate CL. α corresponds to the ratio $\frac{\tau_n}{\tau}$. Where τ_n is a fixed time step size used for the numeric solution. α defines how many iteration steps from zero to the time τ of the analytic solution are calculated. Hence, $\alpha = 0{,}02$ used for Figure C.7 corresponds to 50 iteration steps per τ. α is needed to scale the numeric solution in the same way as the analytic solution. It is obvious that the step size of the numeric solution needs to be small to resemble the analytic solution. For a discrete-time congestion control α together with f_{fn} is also essential to fulfil the stability criterium given in equation (B.22).

For a control algorithm it is not essential to resemble the analytic solution of the control function, it should be stable and converge quickly to the equilibrium. Therefore, in ETSI TS 102 687 [i.14] a state machine approach was proposed that discretises the *CL* and the T_{off} time into fife value pairs as shown in Figure C.1 and Figure C.2. To limit the T_{off} step size for each time step, the state machine will only transit from one state to the consecutive one. It is not allowed to step over a state, even when the measured *CL* is not in the range defined for the consecutive state. To determine the direction of the state transition, the *CL* range of the current state is compared with the measured *CL*. If the measured *CL* is within the *CL* range of the current state, then the state and the T_{off} time are not changed. Otherwise, depending on whether the measured *CL* exceeds or is below the *CL* range of the current state, the state transits to the next or the previous state with a longer or shorter T_{off} time.

Figure C.8 shows as solid-coloured lines the *CL* that results from the T_{off} values given in Table A.1 of ETSI TS 102 687 [i.14] for T_{on} = 1 ms. These are straight lines, what follows from equation (C.2). Each line corresponds to one of the states (① to ⑤) of the state machine. The dashed horizontal lines are the *CL* limits for a state transition that correspond to the state with the same colour. The red circles show where in the diagram the table entries for T_{on} = 1 ms are located. For comparison, the control equilibrium $\widetilde{CL}$ for continuous *CL* values according to equation (C.7) with T_a and T_b from Table C.1 is shown as black line.

NOTE: For other values of T_{on} the number of nodes where the state transition limit is reached would be different, as follows from equation (C.2).

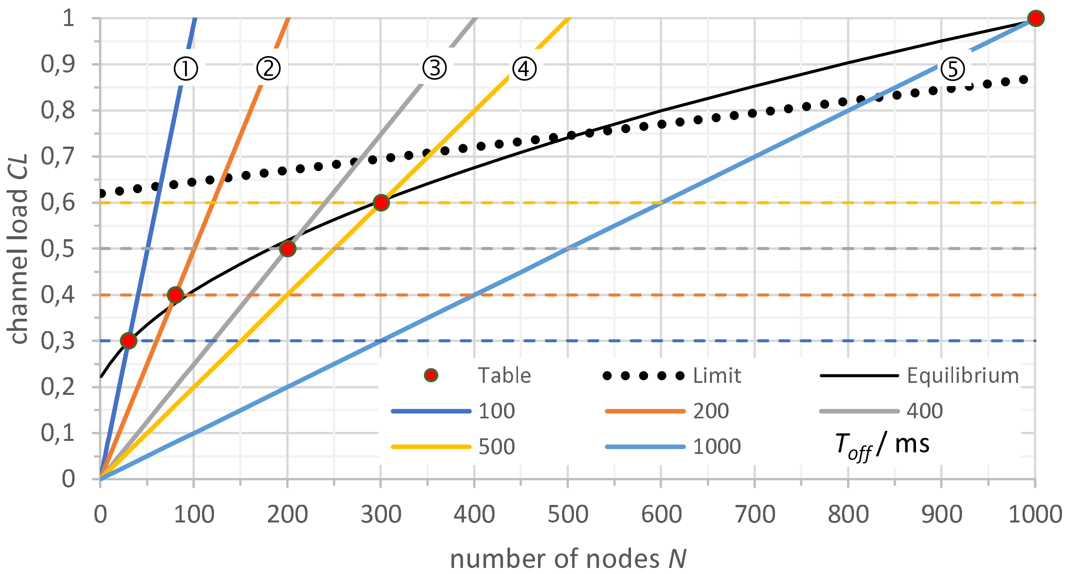


Figure C.8: *CL* resulting for the T_{off} values taken from the reactive DCC Table A.1 in ETSI TS 102 687 [i.14] for T_{on} = 1 ms

Figure C.9 takes a closer look on the control equilibrium of the table based DCC proposed in ETSI TS 102 687 [i.14]. The equilibrium is shown as bold black line. It can only be reached as average over time when the nodes are not synchronized and when only state transitions to the consecutive states are allowed.

The shaded areas (blue, orange, grey) are regions where not all nodes are in the same state so that in average the *CL* results to the respective state transition limit. Hence, in such a region the ratio between the number of nodes in one state and the number of nodes in another state depends on the total number of nodes *N*.

EXAMPLE: For node numbers below the blue shaded area in Figure C.9 all nodes will be in state ①. For a number of nodes within the blue shaded area the limit of *CL* = 0,3 would be exceeded when all nodes are in state ①. Since the nodes are assumed to be not synchronized, they measure the *CL* at random times within a 100 ms interval. Assuming that equilibrium is reached, all nodes measuring a *CL* of more or equal 0,3 will apply state ② all others will apply state ①. Since the nodes are not synchronized and are measuring the *CL* not at the same time, they will do the measurements one after the other with a random time difference. This allows them to get different measurements results and hence to transit to different states, so that in average in this example within the blue shaded area the *CL* will converge to the state transition limit of *CL* = 0,3.

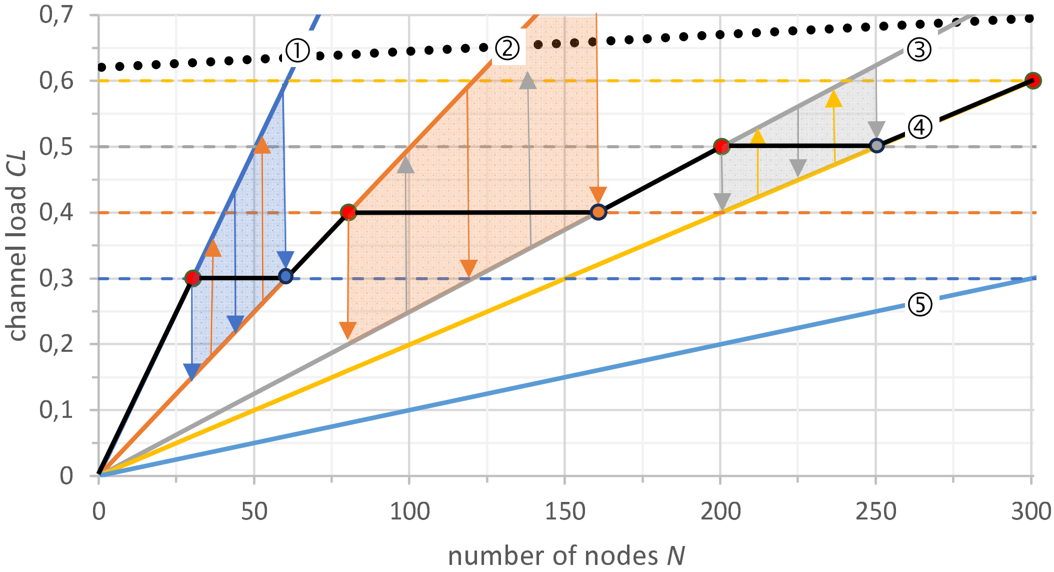


Figure C.9: *CL* resulting from the T_{off} values taken from the reactive DCC Table A.1 in ETSI TS 102 687 [i.14] for T_{on} = 1 ms and the bounded stability regions up to *N* = 300

From Figure C.9 it also gets clear why the nodes should not be synchronized and only transitions between consecutive states are allowed. For the orange shaded area in Figure C.9 it can be seen that when all nodes are synchronized and in state ③ the *CL* will be below the limit of 0,4 and all nodes would synchronously transit to state ②. As a result, the *CL* will jump to a much higher value in the next measurement period and all nodes will transit back to state ③. This results in a strong *CL* oscillation that can even overload the channel. For the same scenario it can be seen that when all nodes are synchronized and in state ③ the *CL* can be below 0,3. When it would be allowed to transit directly to state ① the channel would be heavily overloaded, and the controller gets immediately instable. For not synchronized nodes this issue is not so obvious, but it still exists and can lead to serious oscillations and channel overload.

C.1.3.2 Stability of the reactive congestion control with filtering according to C2C-CC: "Vehicle C-ITS station profile #2037", RS_BSP_240

In C2C-CC: "Vehicle C-ITS station profile #2037", RS_BSP_240 [i.34] a sliding average filter for the controller input signal is specified. This makes an analysis of the nonlinear control algorithm more complicated. Therefore, such an input filtering is not foreseen in clause B.1.3.

To model this input filtering, the channel load at $t - 2 \times \tau$ is needed. With equation C.8 this results in the dynamic channel load equation (C.18):

$$CL(t) = \frac{N \times T_{on}}{(CL(t-\tau)+CL(t-2\times\tau))\times\frac{T_a}{2}+T_b+T_{on}}. \quad (C.18)$$

This input filtering does not change the equilibrium channel load. Because it uses the average of two previous channel loads as input, and since in the steady state case all channel loads are considered to be equal to the equilibrium channel load, the average will also be the equilibrium channel load as given in equation (C.7).

Furthermore, when τ is close to zero, the average $\frac{CL(t-\tau)+CL(t-2\times\tau)}{2}$ will be $CL\left(t-\frac{3}{2}\times\tau\right)$:

$$\lim_{\tau\to 0}\frac{CL(t-\tau)+CL(t-2\times\tau)}{2} = CL\left(t-\frac{3}{2}\times\tau\right) \qquad \text{(C.19)}$$

From equation (C.19) follows that differential equation (C.13), its solution in equation (C.15), and all conclusions drawn for the continuous time behaviour of the reactive approach according to ETSI TS 102 687 [i.14] also apply for the DCC as specified in C2C-CC: "Vehicle C-ITS station profile #2037", RS_BSP_240 [i.34]. The difference lies in the behaviour of the discrete time / discrete T_{off} implementation.

C.2 Adaptive congestion control

C.2.1 Control equilibrium of the adaptive congestion control

The adaptive congestion control concept is based on the idea that the radio channel can support a given maximum total target message rate r_g. A Proportional Integral (PI) message rate control algorithm is used to adjust the transmission rate r_j of each network node according to equation (C.20):

$$r_j(t) = (1-\propto)\times r_j(t-\tau) + \beta_o \times \left(r_g - r(t-\tau)\right) \qquad \text{(C.20)}$$

$r(t-\tau)$ is the total rate contribution of all nodes N in the previous time step and results from equation (C.21). The time step duration is given by τ:

$$r(t-\tau) = N \times r_j(t-\tau) \qquad \text{(C.21)}$$

The target message rate r_g that can be supported by the radio channel depends on the message duration, therefore this concept works only fine when the message duration is (almost) fixed and known. Otherwise, a maximum message duration $T_{on\ max}$ needs to be estimated or considered. r_g can then be set according to equation (C.22) so that the channel load is limited to e.g. CL_{max} = 70 %, which is a good default value:

$$r_g = \frac{CL_{max}}{T_{on\ max}} \qquad \text{(C.22)}$$

Due to the shortcoming that r_g depends on T_{on}, equation (C.20) was reformulated in ETSI TS 102 687 [i.14] so that the estimated constant reciprocal message duration was put into the factor β_o and instead of the total rate contribution r the channel busy ratio *CBR* which is the measured channel load, was used. This results in equation (C.23):

$$r_j(t) = (1-\propto)\times r_j(t-\tau) + \beta \times \left(CBR_{target} - CBR(t-\tau)\right) \qquad \text{(C.23)}$$

This notation has the advantage that it directly uses the measured *CBR* and that the channel load is limited to CBR_{target} independent of T_{on}. Since the message rate is controlled, the channel utilization of each node still depends on T_{on}:

$$u = r_j \times T_{on} \qquad \text{(C.24)}$$

Consequently, from equation (C.24) follows in equation (C.4) with equation (B.4) the channel load:

$$CL_a = r_j \times N \times T_{on} \qquad \text{(C.4)}$$

Assuming that the *CL* is equal to the measured *CBR* and $\widetilde{r_j} = r_j(t) = r_j(t-\tau)$ the equilibrium rate $\widetilde{r_j}$ can be calculated from equation (C.25):

$$\widetilde{r_j} = (1-\propto)\times\widetilde{r_j} + \beta\times\left(CBR_{target} - \widetilde{r_j}\times N\times T_{on}\right) \qquad \text{(C.25)}$$

$$\widetilde{r_j} = \frac{\beta\times CBR_{target}}{\alpha+\beta\times N\times T_{on}} \qquad \text{(C.26)}$$

With equation (C.4) the channel load equilibrium as function of the number of nodes N follows from equation (C.26):

$$\widetilde{CL_a} = \frac{\beta\times N\times T_{on}}{\alpha+\beta\times N\times T_{on}}\times CBR_{target} \qquad \text{(C.27)}$$

The target channel load CBR_{target} is reached asymptotically for an infinite number of nodes N as follows from equation (C.27).

As an example, Figure C.10 shows $\widetilde{CL}_a$ as function of the node number N and the T_{on} time as coloured solid lines for the parameter values given in Table 3 of ETSI TS 102 687 [i.14]. $\widetilde{CL}_a$ does not exceed the CBR DCC limit specified in ETSI EN 303 797 [i.32] (dot ted line), but for a small number of nodes and short T_{on} times the channel is underutilized.

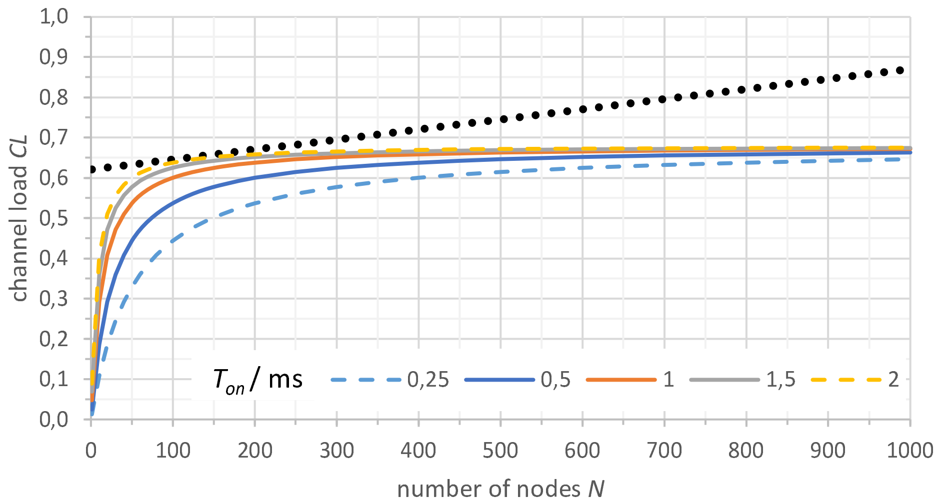


Figure C.10: Equilibrium channel load of adaptive congestion control for the parameter values given in Table 3 of ETSI TS 102 687 [i.14]

C.2.2 Convergence of the adaptive congestion control

C.2.2.1 Continuous time behaviour of the adaptive congestion control

From equation (C.4) the packet rate as function of the channel load CL_a, the number of nodes N, and the packet duration T_{on} can be derived:

$$r_j = \frac{CL_a}{N \times T_{on}} \tag{C.28}$$

Substituting equation (C.28) into equation (C.23) leads to equation (C.29):

$$\frac{CL_a(t)}{N \times T_{on}(t)} = (1-\propto) \times \frac{CL_a(t-\tau)}{N \times T_{on}(t-\tau)} + \beta \times \left(CBR_{target} - CL_a(t-\tau)\right) \tag{C.29}$$

For the assumption that the packet duration is time invariant and equal for all nodes $T_{on}(t) = T_{on}(t-\tau) = T_{on}$ equation (C.29) can be rewritten to:

$$\frac{CL_a(t) - CL_a(t-\tau)}{\tau} = \frac{1}{\tau} \times \left(N \times T_{on} \times \beta \times CBR_{target} - CL_a(t-\tau) \times (\propto + N \times T_{on} \times \beta)\right). \tag{C.30}$$

Equation (C.30) can be written as differential equation for small τ when $\tau \to 0$:

$$\frac{dCL_a}{dt} = \frac{1}{\tau} \times \left(N \times T_{on} \times \beta \times CBR_{target} - CL_a \times (\propto + N \times T_{on} \times \beta)\right) \tag{C.31}$$

When substituting $A = N \times T_{on} \times \beta \times CBR_{target}$ and $B = \propto + N \times T_{on} \times \beta$ into equation (C.31) it can be rewritten after separation of the variables into:

$$\int \frac{dCL_a}{A - B \times CL_a} = \int \frac{dt}{\tau} \tag{C.32}$$

When integrating equation (C.32) it results to:

$$\frac{t_0 - t}{\tau} = \frac{1}{B} log(|A - B \times CL_a|) \tag{C.33}$$

When undoing the substitution for A and B in equation (C.33) the continuous time solution for the adaptive congestion control results to:

$$\frac{t_0-t}{\tau} = \frac{1}{\propto+N\times T_{on}\times\beta} \times log\left(\left|N \times T_{on} \times \beta \times CBR_{target} - (\propto + N \times T_{on} \times \beta) \times CL_a\right|\right) \qquad \text{(C.34)}$$

Figure C.11 shows the analytic result calculated with equation (C.34) in red colour and a numeric result as blue crosses for the for N = 100, T_{on} = 1 ms, and the parameters given in in Table 3 of ETSI TS 102 687 [i.14]. The numeric result is obtained with a step size of t/τ= 1 as specified in of ETSI TS 102 687 [i.14]. In this example, for τ= 100 ms the channel load is close to the equilibrium after around 2 seconds, and in practice reaches it after 4 seconds. The convergence time of the discrete time adaptive congestion control is strongly depending on the term $\propto + N \times T_{on} \times \beta$ as will be shown in clause C.2.2.2.

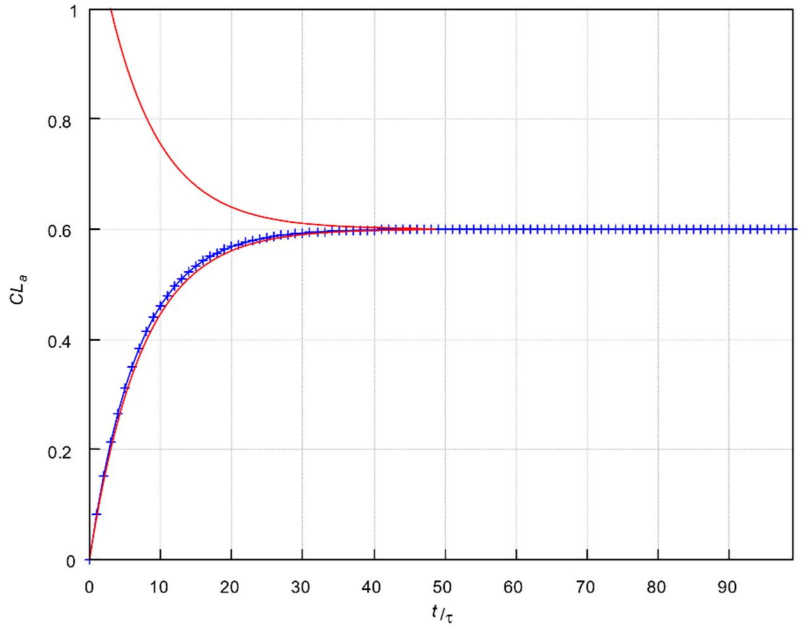


Figure C.11: Analytic and numeric solution of the continuous time adaptive congestion control for *N* = 100, *T_{on}* = 1 ms, and the parameters given in in Table 3 of ETSI TS 102 687 [i.14]

There are two analytic solutions for the continuous time behaviour of the adaptive congestion control depending on whether the initial channel load $CL_a(t_0)$ is above or below the equilibrium channel load $\widetilde{CL}_a$.

When $CL_a(t_0) > \widetilde{CL}_a$ then t_0^+ results from equation (C.35):

$$t_0^+ = \frac{\tau}{B} \times log(CL_a(t_0) \times B - A) \qquad \text{(C.35)}$$

$$t_0^+ = \frac{\tau}{\propto+N\times T_{on}\times\beta} \times log\left(CL_a(t_0) \times (\propto + N \times T_{on} \times \beta) - N \times T_{on} \times \beta \times CBR_{target}\right) \qquad \text{(C.36)}$$

With t_0^+ from equation (C.36) $CL_a(t)$ results for $CL_a(t_0) > \widetilde{CL}_a$ from equation (C.37):

$$CL_a(t) = \frac{A+e^{B\times\frac{t_0^+-t}{\tau}}}{B} = \frac{N\times T_{on}\times\beta\times CBR_{target}+e^{\frac{t_0^+-t}{\tau}\times(\propto+N\times T_{on}\times\beta)}}{\propto+N\times T_{on}\times\beta} \qquad \text{(C.37)}$$

When $CL_a(t_0) < \widetilde{CL}_a$ then t_0^- results from equation (C.38):

$$t_0^- = \frac{\tau}{B} \times log(A - CL_a(t_0) \times B) \qquad \text{(C.38)}$$

$$t_0^- = \frac{\tau}{\propto+N\times T_{on}\times\beta} \times log\left(N \times T_{on} \times \beta \times CBR_{target} - CL_a(t_0) \times (\propto + N \times T_{on} \times \beta)\right) \qquad \text{(C.39)}$$

With t_0^- from equation (C.39) $CL_a(t)$ results for $CL_a(t_0) < \widetilde{CL}_a$ from equation (C.40):

$$CL_a(t) = \frac{A-e^{B\times\frac{t_0^--t}{\tau}}}{B} = \frac{N\times T_{on}\times\beta\times CBR_{target}-e^{\frac{t_0^--t}{\tau}\times(\propto+N\times T_{on}\times\beta)}}{\propto+N\times T_{on}\times\beta} \qquad \text{(C.40)}$$

C.2.2.2 Stability of the discrete time adaptive congestion control

For calculating the stability criterium given in equation (B.22) the control function c_{fn}, the filtering function f_{fn}, and the equilibrium channel utilization $\tilde{u}$ of the adaptive congestion control is needed.

From equation (C.23) and equation (C.24) the control equation of the adaptive congestion control as function of the channel utilization u_a can be calculated:

$$u_{aj}(t) = (1-\propto) \times u_{aj}(t-\tau) + T_{on} \times \beta \times \left(CBR_{target} - CBR(t-\tau)\right) \quad \text{(C.41)}$$

From equation (B.11) and equation (C.41) it follows that the filter function of the adaptive congestion control $\mathrm{f_{fn}}_a$ is constant and equal to one:

$$\mathrm{f_{fn}}_a\big(CL(t-\tau)\big) = 1 \quad \text{(C.42)}$$

With equation (C.42) equation (B.11) simplifies to:

$$u_{aj}(t) = \alpha \times \mathrm{c_{fn}}_a\big(CL_a(t-\tau)\big) + (1-\alpha) \times u_{aj}(t-\tau) \quad \text{(C.43)}$$

From equation (C.41) and equation (C.43) the control function for the adaptive congestion control $\mathrm{c_{fn}}_a$ can be calculated:

$$\mathrm{c_{fn}}_a(CL_a) = T_{on} \times \frac{\beta}{\alpha} \times \left(CBR_{target} - CL_a\right) \quad \text{(C.44)}$$

Substituting equation (B.4), equation (C.27), equation (C.42), and equation (C.44) in equation (B.22) leads to equation (C.45):

$$\left|\alpha \times 1 \times \left(T_{on} \times \frac{\beta}{\alpha} \times \left(CBR_{target} - N \times u(t)\right) - u(t)\right)\right| < 2 \times \left|\frac{T_{on} \times \beta \times CBR_{target}}{\alpha + \beta \times N \times T_{on}} - u(t)\right| \quad \text{(C.45)}$$

When factorizing equation (C.45) to equation (C.46) it can be simplified to equation (C.47) which is in line with the result given in IEEE 802.11 [i.9] that was found via a series expansion approach:

$$(\alpha + \beta \times N \times T_{on}) \times \left|\left(T_{on} \times \beta \times CBR_{target} - T_{on} \times \beta \times N \times u(t) - \alpha \times u(t)\right)\right| <$$

$$< 2 \times \left|T_{on} \times \beta \times CBR_{target} - T_{on} \times \beta \times N \times u(t) - \alpha \times u(t)\right| \quad \text{(C.46)}$$

$$\propto + T_{on} \times N \times \beta < 2 \quad \text{(C.47)}$$

This result means that the discrete time adaptive algorithm converges to the $\widetilde{CL}_a$ value from equation (C.27) when the stability criterion of equation (C.47) is fulfilled.

The fastest convergence of the discrete time control is reached for $\propto + T_{on} \times N \times \beta = 1$ where the equilibrium is reached within one time step. For $\propto + T_{on} \times N \times \beta > 1$ the channel load overshoots the equilibrium and converges with an exponentially decaying oscillation. When $\propto + T_{on} \times N \times \beta \ll 1$ the discrete time control converges approximately like the analytic continuous time solution given by equation (C.34).

For the parameter values given in Table 3 of ETSI TS 102 687 [i.14] the stability criterion is given by $T_{on} \times N < 1653{,}3\ ms$.

Since this criterion cannot be guaranteed for large node numbers and long packet durations, the algorithm was enhanced in ETSI TS 102 687 [i.14] by a control loop gain saturation that limits the term $\beta \times \left(CBR_{target} - CL_a(t-\tau)\right)$ to a constant value G_{max} and by an input filtering function. While the filtering function has no influence on the control equilibrium function a saturation of the control loop gain by setting $\beta \times \left(CBR_{target} - CL_a(t-\tau)\right)$ to a constant G_{max} slightly reduces the channel utilization and has an influence on the convergence speed (see clause C.2.2.3). In ETSI TS 102 687 [i.14] different values for G_{max} are definied when $CBR_{target} - CL_a(t-\tau)$ is positive or negative. For the control equilibrium only G^{+}_{max} is relevant since the resulting new channel load equilibrium is below CBR_{target} and the gain saturation is realized by replacing $\beta \times \left(CBR_{target} - CL_a(t-\tau)\right)$ by G^{+}_{max} in the control equation C.25. From this follows the control equilibrium $\tilde{r}_{sj}$ in equation (C.48) for the gain saturated case:

$$\tilde{r}_{sj} = (1-\propto) \times \tilde{r}_{sj} + G^{+}_{max} \quad \text{(C.48)}$$

What results to a fixed equilibrium packet rate $\widetilde{r_{sj}}$ independent of the channel utilization:

$$\widetilde{r_{sj}} = \frac{G_{max}^{+}}{\propto}. \tag{C.49}$$

From equation (C.49) follows with equation (C.4) a linear relation between $N \times T_{on}$ and the equilibrium channel load $\widetilde{CL_{as}}$ as shown by equation (C.50) for cannel loads CL_a within the lower gain saturation area (see Figure C.12):

$$\widetilde{CL_{as}} = \frac{G_{max}^{+}}{\propto} \times N \times T_{on} \tag{C.50}$$

The gain saturation areas result from the gain saturation values G_{max}^{+} and G_{max}^{-} that are replacing $\beta \times \left(CBR_{target} - CL_a(t - \tau)\right)$ in equation (C.25). Based on this the gain saturation areas are given by an upper area with a lower bound CL_a^{-} as shown in equation (C.51) and a lower area with an upper bound CL_a^{+} as shown in equation (C.52):

$$CL_a^{-} = CBR_{target} - \frac{G_{max}^{-}}{\beta} \tag{C.51}$$

$$CL_a^{+} = CBR_{target} - \frac{G_{max}^{+}}{\beta} \tag{C.52}$$

Figure C.12 shows these areas and the impact of the gain saturation on the equilibrium channel load for the parameter values given in Table 3 of ETSI TS 102 687 [i.14]. It can be seen that in the gain saturation area below CL_a^{+} the equilibrium channel load follows the linear relation given by equation (C.50), while above CL_a^{+} it follows the non-linear equation (C.27). The upper gain saturation area is only reached in dynamic channel load scenarios. It has no influence on the equilibrium channel load.

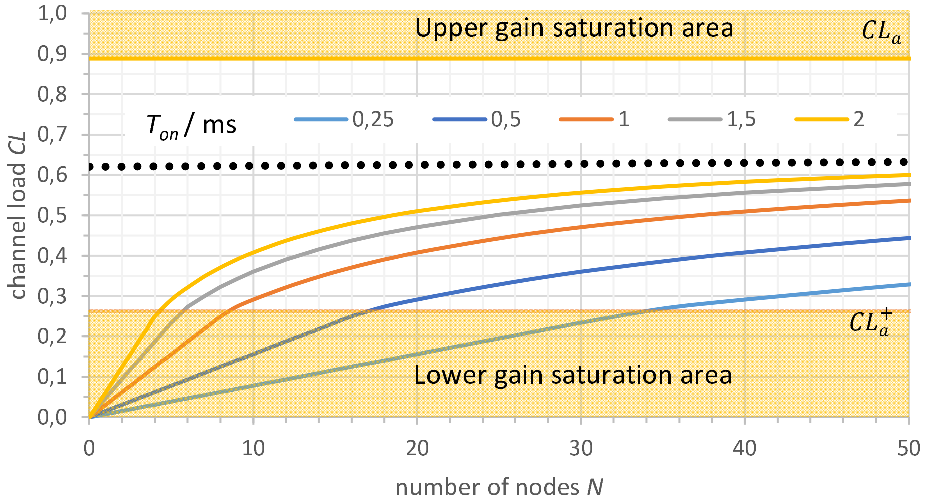


Figure C.12: Detail of equilibrium channel load of adaptive congestion control for the parameter values given in Table 3 of ETSI TS 102 687 [i.14] showing the effect of the gain saturation

The filtering as proposed in ETSI TS 102 687 [i.14] increases the stability range for $T_{on} \times N$ by a factor of two. Hence, the control loop can converge to an equilibrium for twice as many nodes or for doubled message sizes. Details about this need to be investigated.

C.2.2.3 Convergence speed of the discrete time adaptive congestion control

In [i.33] a series expansion approach was used to find the channel load values after each time step of the discrete time adaptive congestion control. From there, with $A = N \times T_{on} \times \beta \times CBR_{target}$ and $B = \propto + N \times T_{on} \times \beta$ the channel load CL_a after n time steps of the duration τ can be calculated according to equation (C.53) for the discrete time adaptive congestion control under the assumption that all nodes are synchronized. The more realistic case for unsynchronized nodes is discussed in Annex D:

$$CL_a(n \times \tau) = \frac{(1-B)^n \times (CL_a(0) \times B - A) + A}{B} \tag{C.53}$$

From equation (C.53) the convergence speed of the discrete time adaptive congestion control can be calculated. Where the convergence speed is defined by the time where the channel load crosses the boundaries of ±5 % deviation from the equilibrium value and afterwards stays within these boundaries when applying a channel load step from 0 % to 100 % or from 100 % to 0 %. Since the convergence times depend on the initial conditions at the time *t* = 0 s (see Figure C.11) these conditions of the network nodes need to be defined. An initial transmission rate $r_j = 0$ and a resulting $CL_a(0) = 0$ can be used to test an implementation of the control algorithm, while the example with $CL_a(0) = 1$ is more to demonstrate what happens when the channel is fully loaded and $r_j = \frac{1}{N \times T_{on}}$, what is a bit of an arbitrary assumption just to show the difference to $r_j = 0$.

For $1 - B > 0$ the power function $(1 - B)^n$ is monotonically increasing (see examples in Figure C.13), for $1 - B < 0$ it oscillates between positive results for even time steps values *n* and negative results for odd values of *n*. Since the factor $CL_a(0) \times B - A$ is negative for $CL_a(0) = 0$ this leads to an overshoot of the equilibrium $\frac{A}{B}$ (equation (C.27)) as can be seen as an example in Figure C.14.

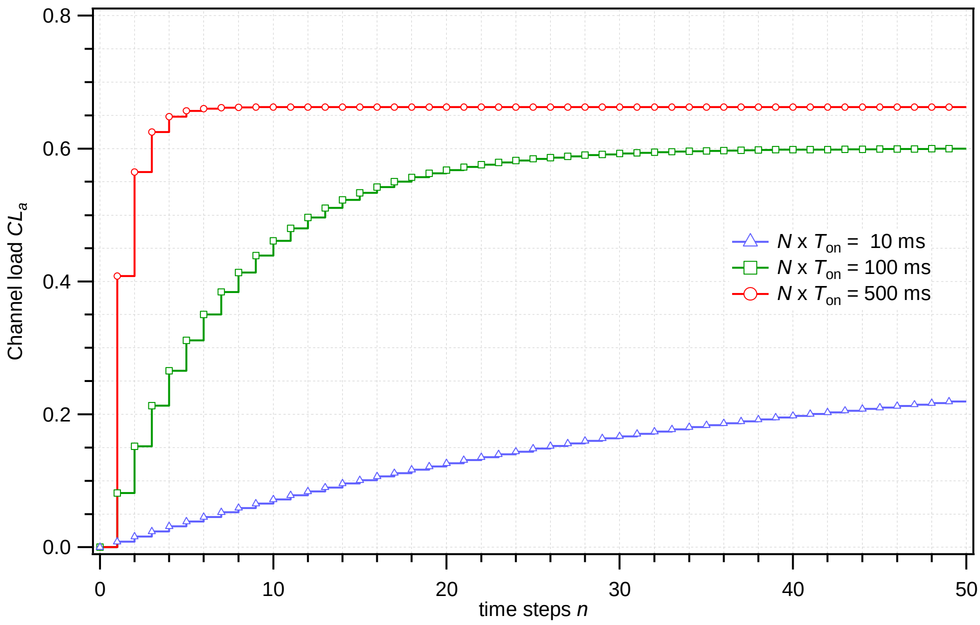


Figure C.13: Channel load of adaptive congestion control for different values of *N* x T_{on} resulting from equation (C.53) for the parameter values given in Table 3 of ETSI TS 102 687 [i.14]

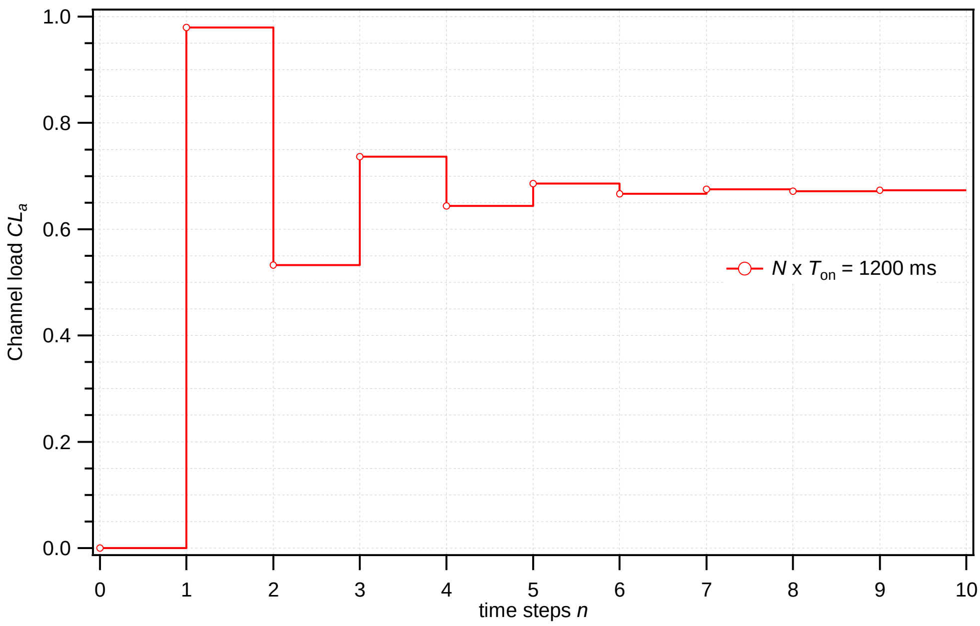


Figure C.14: Channel load of adaptive congestion control for N x T_{on} = 1 200 ms resulting from equation (C.53) for the parameter values given in Table 3 of ETSI TS 102 687 [i.14]

Thus, for $B < 1$ the convergence time $n_c \times \tau$ results from equation (C.53) when setting CL_a to 95 % or 105 % of the equilibrium $\frac{A}{B}$ (equation (C.27)). Which value to take depends on the initial channel load $CL_a(0)$ as shown in equation (C.54) and equation (C.55):

$$\text{For } B < 1 \text{ and } CL_a(0) = 0 \text{ follows } \frac{A}{B} \times 0{,}95 = \frac{(1-B)^{n_c^-} \times (-A) + A}{B} \tag{C.54}$$

$$\text{For } B < 1 \text{ and } CL_a(0) = 1 \text{ follows } \frac{A}{B} \times 1{,}05 = \frac{(1-B)^{n_c^+} \times (B-A) + A}{B} \tag{C.55}$$

The time steps n_c^- to reach 95 % of the channel load equilibrium for an initial channel load of 0 result from equation (C.56). For an initial message rate $r_j = \frac{1}{N \times T_{on}}$ the time steps n_c^+ to reach 105 % of the channel load equilibrium result from equation (C.57). Figure C.15 compares the results for n_c^- and n_c^+ from equation (C.56) and equation (C.57) for the initial cannel load $CL_a(0)$ equal to 0 or respectively equal to 1:

$$\text{For } B < 1 \text{ and } CL_a(0) = 0 \text{ follows } n_c^- = \frac{ln(0{,}05)}{ln(1-\text{B})} = \frac{ln(0{,}05)}{ln(1-\propto-N \times T_{on} \times \beta)} \tag{C.56}$$

$$\text{For } B < 1 \text{ and } CL_a(0) = 1 \text{ follows } n_c^+ = \frac{ln\left(\frac{A \times 0{,}05}{B-A}\right)}{ln(1-\text{B})} = \frac{ln\left(\frac{N \times T_{on} \times \beta \times CBR_{target} \times 0{,}05}{\propto + N \times T_{on} \times \beta \times \left(1-CBR_{target}\right)}\right)}{ln(1-\propto-N \times T_{on} \times \beta)} \tag{C.57}$$

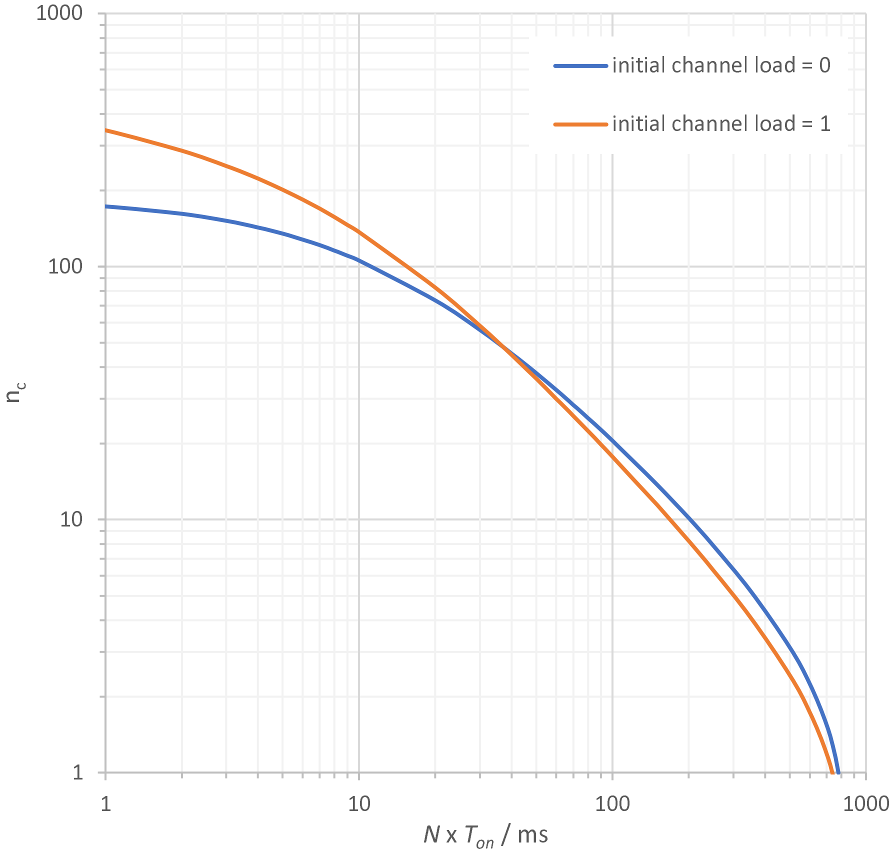


Figure C.15: Number of time steps to reach ±5 % deviation from the equilibrium channel load when starting from different initial channel loads for the adaptive congestion control with parameter values given in Table 3 of ETSI TS 102 687 [i.14]

For $B > 1$ the channel load will overshoot the equilibrium as has been shown in Figure C.14. This overshoot can overload the channel, and the convergence time consists of the channel overload recovery time plus the exponential channel load decay time. The linear channel load model will not hold for the channel load saturation at high loads. But it can give at least a rough estimation of the decay time as long as that CBR_{target} parameter is below a channel load of 0,8. The envelope of the channel load decay is given by equation (C.58):

$$\text{For } B > 1 \text{ follows } CL_{a\ envelope}(n) = \frac{A}{B} - \left(CL_a(0) - \frac{A}{B}\right) \times e^{ln(B-1) \times n} \tag{C.58}$$

With a gain saturation G_{max}^{+} the convergence time can increase, since below CL_a^+ the channel load CL_{as} follows equation (C.59) which is less steep in n compared to equation (C.53):

$$\text{For } CL_a < CL_a^+ \text{ follows } CL_{as}(n \times \tau) = \frac{(1-\alpha)^n \times (CL_a(0) \times \alpha - G_{max}^{+} \times N \times T_{on}) + G_{max}^{+} \times N \times T_{on}}{\alpha} \tag{C.59}$$

Depending on whether the channel load equilibrium is below or above CL_a^+, the convergence speed can either be determined from equation (C.59) or by piecewise use of equation (C.59) up to CL_a^+ and from there by use of equation (C.53) as shown in the beginning of the present clause.

Annex D:
Channel load measurement

In Annex C it was assumed that the channel load can be measured in any point of time. In practice it is measured by determining the total channel active time $T_{on\ total}$ and the total channel idle time $T_{off\ total}$ over a certain measurement period and then by calculating the measured channel load CL_m according to equation (D.1). This measured channel load is often called channel busy ratio CBR:

$$CL_m = CBR = \frac{T_{on\ total}}{T_{on\ total} + T_{off\ total}} \qquad \text{(D.1)}$$

When the measurement period is shorter than the idle time T_{off} between two consecutive messages transmitted by a certain node, the channel utilization of this node will be considered only in some of the channel load measurements. Hence, the measurement result will not be stable even when all nodes are not changing their channel utilization. Furthermore, such a sliding window approach to measure the channel load does exhibit a linear phase over frequency and has no flat frequency response. This has an influence on the control loop stability since some frequencies will be amplified in the control loop, leading to instabilities. Some authors even proposed to inject noise into the measurement results to avoid such a feedback effect in the control loop.

In addition, for an ITS-G5 access layer the measurement times are not synchronized what usually improves stability but complicates an analytic approach to determine stability. Even the unsynchronized measurement times will improve stability, they should not be taken for granted and the stability determination should be done for the worst case of synchronized nodes to guaranty stability for any random configuration.

First simulation results of the table based reactive algorithm show chaotic behaviour, even under static traffic and radio channel conditions. Such a chaotic behaviour is typical for nonlinear differential equations as given by the step function resulting from the table entries. Figure D.1 shows such a behaviour for a reactive DCC according to Table A.1 in ETSI TS 102 687 [i.14] for 401 nodes, T_{on} = 1 ms, and 100 ms CBR measurement time. The blue bars at the bottom of Figure D.1 show the transmission events. Even the scale is such that not all channel idle periods can be seen in this figure, it is obvious that the white spaces in between the transmissions are not deterministic. The coloured line shows in blue an underutilization and in red an over utilization of the channel. In average it fluctuates chaotically around the expected 60 % channel load. It is calculated by a sliding window of 1 second length. The CBR result obtained with a 100 ms sliding window duration is shown in black, it jumps chaotically between different values in the range from 0 to 100 %. This fluctuating value is the input to the reactive DCC.

To visualize these fluctuations for different number of nodes and thereby different T_{off} values the shaded region in Figure D.2 shows the range of channel load values measured with a sliding window duration of 1 second that were calculated for a time span of 20 seconds after bounded stability was reached. The green line shows the average of these channel load values. When comparing this to Figure C.9 the theory only fits in the far-left part of Figure D.2. This is because the measurement time span of 100 ms is much shorter than the T_{off} time of each node necessary to limit the channel load when there are more than 40 nodes in range.

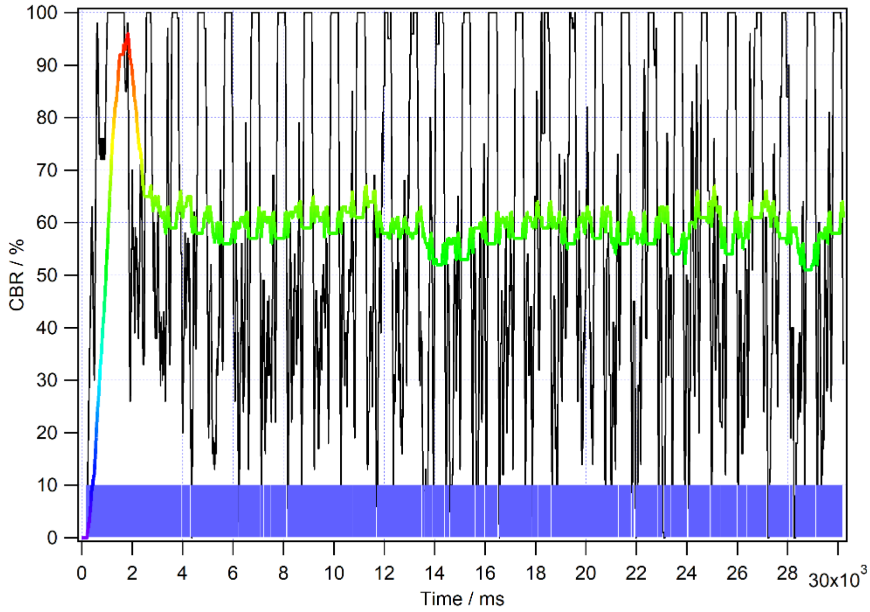


Figure D.1: Time behaviour for a reactive DCC according to Table A.1 in ETSI TS 102 687 [i.14] for 401 nodes, T_{on} = 1 ms, and 100 ms CBR measurement time

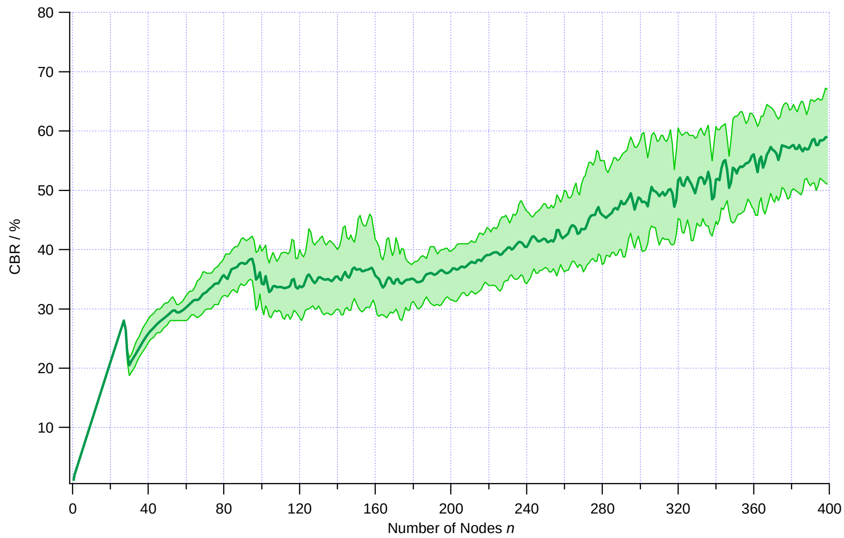


Figure D.2: Bounded stability for a reactive DCC according to Table A.1 in ETSI TS 102 687 [i.14] for T_{on} = 1 ms and 100 ms CBR measurement time

To reduce the chaotic behaviour of the measured channel load, the measurement period can be increase as specified in The C2C-CC Profile, where it was doubled to 200 ms. Figure D.3 shows the result for this doubled measurement period. As expected, the result is close to theory for around double the number of nodes compared to a measurement period of 100 ms. Still, for more than 80 nodes the behaviour gets again chaotic.

As a conclusion, the topic of the channel load measurement needs some further investigation to find the right way of doing it.

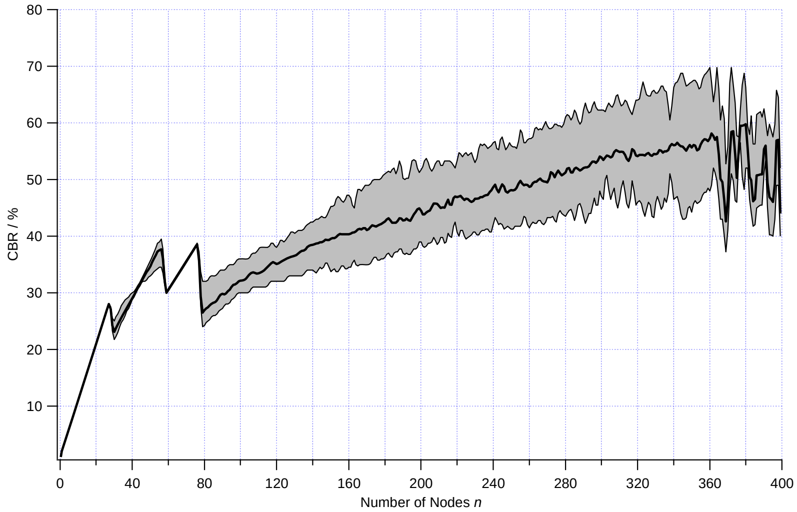


Figure D.3: Bounded stability for a reactive DCC according to C2C-CC: "Vehicle C-ITS station profile #2037", RS_BSP_240 [i.34] for T_{on} = 1 ms and 200 ms CBR measurement time

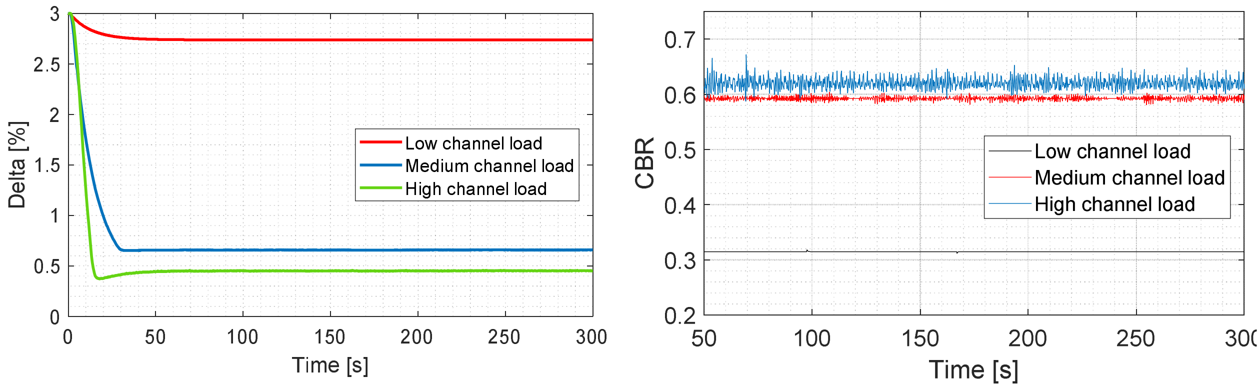
Annex E: Evaluation of basic RM solution

E.1 Evaluation set-up

The basic solution described in clause 7.3.6 is evaluated. The evaluation was performed using a virtualized testing environment to mimic real-world C-ITS operations. The RM module was integrated into the Vanetza C-ITS protocol stack at the Facilities layer with only minimal modifications - leveraging an extended version of the Socktapp tool - to support the additional congestion control functionalities at the facilities layer. To enable concurrent testing, the entire stack was encapsulated in a Docker container image, allowing multiple instances to run on a single hardware platform. Each container represented a distinct C-ITS station with five active V2X services, connected via a Docker virtual network. A scaling factor was applied to emulate low, medium, and high channel load scenarios by adjusting the effective channel load, evaluated through the Channel Busy Ratio (CBR). This approach also took into account realistic processing delays and hardware limitations, ensuring that the evaluation reflected operational conditions.

E.2 Results

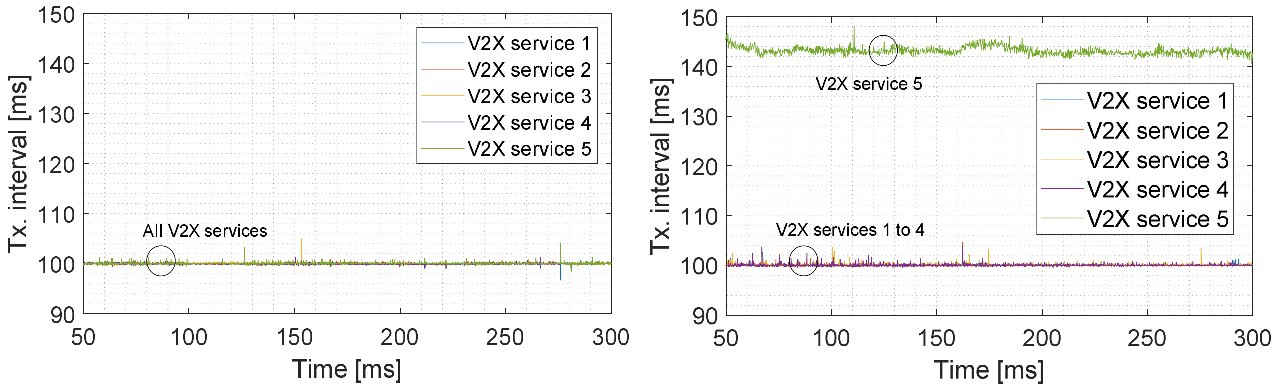
The experimental results confirmed that the basic solution achieves convergence and effective congestion control under varying channel loads. Under low channel load conditions, the system converged to a resource usage value (δ) of approximately 2,74 % (Figure E.1a). In contrast, for medium and high channel load scenarios, δ stabilized at around 0,66 % and 0,45 %, respectively. The evaluation showed that the implemented RM module could reduce the CBR from a potential 94 % - if all services transmitted at maximum rate - to about 62 % under high load (Figure E.1b). Additionally, while there were slight deviations between the configured and actual transmission intervals (attributable to real-time processing delays, see Figure E.2, these discrepancies did not impact the overall stability or robustness of the congestion control mechanism. The results validate that the integration of the RM within Vanetza and its execution in a Docker-based virtualized environment effectively bridge simulation-based studies and real-world testing.



(a) Available resources (delta)

(b) Channel load or CBR (Channel Busy Ratio)

Figure E.1: Stability and convergence of the implementation



(a) Low channel load (b) Medium channel load

Figure E.2: Time evolution of transmission interval of different services

History

Version	Date	Status
V2.1.1	July 2025	Publication