COntent Mediator architecture for content-aware nETworks

European Seventh Framework Project FP7-2010-ICT-248784-STREP

Deliverable D2.3
Global Architecture of the COMET System

The COMET Consortium

Telefónica Investigación y Desarrollo, TID, Spain
University College London, UCL, United Kingdom
University of Surrey, UniS, United Kingdom
PrimeTel PLC, PrIMETEL, Cyprus
Warsaw University of Technology, WUT, Poland
Intracom SA Telecom Solutions, INTRACOM TELECOM, Greece

For more information on this document or the COMET project, please contact:

Prof. George Pavlou
g.pavlou@ucl.ac.uk
Department of Electronic & Electrical Engineering
University College London
Torrington Place, London,
WC1E 7JE
UK

Document Control

Title: Global Architecture of the COMET System

Type: Public

Editor(s): Ioannis Psaras

E-mail: i.psaras@ucl.ac.uk

Author(s): David Florez Rodriguez (TID), Andrzej Beben, Piotr Wisniewski (WUT), George Kamel, Vahid Heydari Fami, Ali Norouzi, Ning Wang (UniS), Wei Koong Chai (UCL), George Petropoulos, Spiros Spirou (ICOM), Michael Georgiadis (PTL)

Doc ID: D2.3-v1.2.doc

AMENDMENT HISTORY

Version	Date	Author	Description/Comments
V0.1	06/12/2012	Wei Koong Chai	Release of ToC
V0.2	11/12/2012	Ioannis Psaras	First draft of all sections
V0.3	22/12/2012	David Florez	First update of Section 4
V0.4	29/12/2012	Ioannis Psaras	First draft of updated answers to reviewers' questions
V0.5	11/01/2013	Ning Wang	Input for the Coupled Approach and Evaluation
V0.6	20/01/2013	Andrzej Beben	Input for the Decoupled Approach
V0.7	25/01/2013	Ioannis Psaras	First complete draft
V0.8	05/02/2013	Ioannis Psaras	Second complete draft
V0.9	10/02/2013	David Florez	Second update of Section 4 and review of the document
V1.0	13/02/2013	Ioannis Psaras	Review and Finalisation of document
V1.1	16/02/2013	Ning Wang, Andrzej Beben, David Florez	Final comments and update of Annex
V1.2	19/02/2013	Ioannis Psaras	Final document released

Table of Contents

(This page is left blank intentionally.)

1 Executive Summary

Given that the vast majority of Internet traffic relates to content access and delivery, recent research has pointed to a potential paradigm shift from the current host-centric Internet model to an information-centric one. In information-centric networks, named content is accessed directly, with the best content copy delivered to the requesting user given content caching within the network. Here we present COMET [1], an Internet-scale mediation approach for content access and delivery that supports content and network mediation. Content characteristics, server load and network distance are taken into account in order to locate the optimal content copy for maximizing the user quality of experience and optimize network utilization. The content mediation infrastructure is provided by Internet service providers (ISPs) in a cooperative fashion, with both decoupled/two-phase and coupled/one-phase modes of operation.

This deliverable reports the final specification of the *High-Level Architecture of the COMET System*. Previous versions of this document, namely D2.1 [2] and D2.2 [3] presented the initial and interim requirements and specifications of the system, including business models, functional block and entity specification for the system, as well as our use-cases. Furthermore, the detailed specification of the content resolution and delivery schemes, including our in-network caching functions, have been introduced in D3.2 [4] and D4.2 [6], while the evaluation of these schemes was presented in D5.2 [9].

Following the specifications in D2.2 [3], we summarise the specification of all the entities, as well as their functional operation in this deliverable. We update the basic functionality of our system in this document. In particular, as initially introduced in D2.1 [2] and further elaborated in D2.2 [3], the COMET system consists of two planes, the *Content Mediation Plane* (CMP) and the *Content Forwarding Plane* (CFP). The CMP is responsible for resolving names to contents, as well as for preparing the path for the content delivery. The CMP also incorporates network- and server-awareness functionality.

The CFP, the data plane of the COMET architecture, is responsible for delivering the content back to the content consumer. The CFP follows the path preparation and setup rules that the CMP has put in place. In this document, we introduce a new functional block for the CFP, which takes care of in-network caching functions. In particular, the Content-Aware Forwarding Entity (CAFE), introduced in D2.2 [3], is complemented with one extra functional block, the Content-Aware Caching Function CACF. The CACF is responsible for making caching decisions, that is, deciding which contents should be cached on the path to the content consumer and at which points along the delivery path should caching take place. The caching decisions can be influenced by the CMP. The caching algorithms used in COMET have been initially introduced in [14], [15] and [22].

The rest of this document is organised as follows. In Chapter 2, we introduce the concept of content mediation in *Information-Centric Networks* (ICNs). In Chapter 3, we describe the *High-Level COMET Architecture*, as this was introduced in D2.2. We also introduce the new *Content-Aware Caching Function* (CACF). As discussed in previous documents, in COMET, we have developed two approaches to content mediation, resolution and delivery. The first one, called the *decoupled approach*, follows the paradigm of the current Internet; it resolves content names similarly to the operation of today's DNS system and decouples the resolution from the delivery operation. The second approach, called the *coupled approach*, combines the content resolution with the path setup in a hop-by-hop fashion and prepares for the delivery of content from the resolution phase. We discuss in detail the two approaches in Chapters 4 and 5, respectively. In Chapter 6, we give an overview of the evaluation of our two approaches and conclude the document in Chapter 7.

Finally, in Appendix A, we provide answers to the reviewers' questions during the second year audit. Wherever applicable, these answers are also given in the description of the two approaches. In particular, Questions 1 to 3 are answered in Sections 4 and 5 for the decoupled and coupled approaches, respectively. Questions 4 to 6, which have common answers for both approaches are answered in Appendix A.

2 Introduction

The Internet has been enormously successful, with IP simplicity being a key factor that allowed it to reach an impressive scale. The original Internet model focused on interconnecting hosts for resource sharing purposes, but after significant evolution over the last two decades, the Internet is currently being used for a wide variety of applications and services. On the other hand, the vast majority of traffic relates to content access and delivery. This is evident from the proliferation of user-generated content, e.g., photos and videos, made available through social networking sites such as Facebook and Myspace, through video aggregators such as YouTube, etc., and also through overlay content distribution infrastructures, for example peer-to-peer (P2P) systems such as BitTorrent and eMule, and content delivery networks (CDNs) such as Akamai and Limelight.

A key aspect related to today's content access is *fragmentation*: users need to know the content location *a priori* in order to access it and content has to be searched through specific intermediaries, e.g., Youtube, BitTorrent, etc. As a result, a lot of content tends to be accessible only by particular user communities. Given the continuing exponential increase in content generation (both amateur and professional), a converged architecture for unified content access and delivery is necessary, providing *name-based content access*. In this context, recent research has pointed to a paradigm shift from the current host-centric Internet model to an information-centric one, with various architectural approaches proposed (refer to [21] for a detailed description of related approaches). The key aspect behind all these approaches is to address *named content* directly, with the best content copy delivered to the requesting user given that caching will take place within the network (see [14], [15], [22] for an elaborate discussion on in-network caching and recent advances in that field). In such an information-centric paradigm, content resolution and delivery functions will be natively realized by the network, enabling network operators to play a more active role in the future content-oriented Internet marketplace.

In this report, we present COMET, an Internet-scale mediation infrastructure for content access and delivery in information-centric networks. Our mediation approach is evolutionary, operating initially as a tightly coupled overlay over the current IP infrastructure, but it could eventually be supported natively within the network. Name-based content access is achieved through collaborative content resolution and delivery functions among Internet Service Providers (ISPs) who operate this content mediation infrastructure in a collaborative manner. Content providers and consumers publish or consume content through a set of *unified* content primitives, via which they interact with their local ISP. This is quite different from existing loosely-coupled overlay approaches, in which a content provider or consumer may have to interact with multiple content overlays in order to maximize accessibility of the published content, or in order to locate a specific piece of content.

The initial design of our architecture has been presented in D2.2 [3]. Here, we provide an update to the architecture and also present the final specification of our approach to content mediation. According to COMET, the players of the Internet ecosystem, which are affected by the host-to-host communication model are the following:

- *Content Creator*: the entity (individual or organization) that owns the rights to the content and wants to publish it on the Internet;
- *Content Provider* (which can also be the *Content Creator)*: the entity responsible for storing and making content available to the *Content Consumers* (usually a large organization, e.g., YouTube, Apple iTunes store);
- *Content Distributor* (which can also be the *Content Provider*): the entity that owns and maintains the infrastructure to distribute content to *Content Consumers* in the most effective way (e.g., CDNs, P2P networks);
- *Network Operator*: the entity that provides networking services, e.g., ISPs, or Internet Backbone Providers (IBPs);
- *Content Consumer*: the entity that consumes the content (usually the end-user). In case the end-user is both provider of content (e.g., user-generated content) and consumer of other contents, the term *content prosumer* is used.

Our content mediation infrastructure can be instantiated through two complementary approaches: the *decoupled* approach for the majority of Internet content, and the *coupled* approach for widely-accessed popular content which can benefit from in-network caching. In the decoupled approach, content resolution takes place first, followed by content access using the server and the path identified through the resolution process [18], [19]. In the coupled approach, content resolution and access are combined in a single phase, with content resolution following a gossip-like communication model, routing content consumption requests in a specific manner within the mediation plane in order to locate the targeted content source [13]. In both approaches, if multiple content copies are available at different servers, the one with good availability (e.g., with low or medium server load) in combination with the least network distance is selected [18]. In addition, monitored end-to-end path quality may be used together with the network distance. The coupled approach aims to optimize both network utilization and user quality of experience (QoE). We provide the operational details of the decoupled and the coupled approaches in Chapters 4 and 5, respectively.

The purpose of this document is to give an update of the *High-Level Architecture of the COMET System* as this was initially specified in D2.2 [3]. This architecture includes all the required properties needed in order to accommodate the corresponding content naming, resolution and delivery mechanisms as realized in COMET. The details of these mechanisms are included in the subsequent deliverables D3.2 [4] and D4.2 [6]. The evaluation of the resolution and in-network caching algorithms developed in COMET are reported in the related deliverables D5.2 [9], as well as the following papers [14], [15], [22]. As discussed later in this chapter, the main difference between the architecture presented in previous documents and the one presented here is the introduction of a *caching function* at the data-plane (i.e., the Content Forwarding Plane, CFP).

The objective of the COMET system as a whole is to develop a ***unified content naming, addressing and resolution architecture***, where the user's request points to the content or service itself, rather than to the machine that hosts the content. In addition, ***server, network and routing awareness will inherently improve QoS for*** content consumers, based on the content requirements, rather than on holistic bandwidth over-provisioning, as happens today.

2.1 COMET Content Mediation

The proposed information-centric ecosystem is based on the concepts of *content* and *network mediation*. As depicted in Figure 1, a mediation plane operates between the "content cloud" and the underlying network infrastructure, providing content access and delivery in a holistic manner. This mediation plane works as a tightly coupled overlay, being collaboratively provisioned and operated by ISPs. Current information-centric architectures are either native (e.g., [25]), requiring fundamental changes in the Internet fabric through content-aware network protocols (see [21] for related discussions), or evolutionary, operating as tightly coupled overlays [13], [20]. In both cases, there is intimate knowledge of the network characteristics and this is in contrast to current loosely-coupled network- agnostic overlays such as CDNs and P2P systems. In fact, realizing the relevant limitations, it has been recently proposed to pass network usage information to overlays for enhancing both overlay and network performance through optimized content source/peer selections, e.g., the IETF ALTO framework[1]. On the other hand, an ISP-operated tightly coupled overlay has intrinsic access to such information and can use it for selecting the best content source.

Our approach provides the following complementary mediation functions:

- *Content mediation* – the content mediation function gains awareness on both the content characteristics (e.g., quality requirements) and the content source conditions (e.g., server load). Based on this awareness, it is able to locate the best content copy in an intelligent manner.

[1] http://tools.ietf.org/wg/alto/

- *Network mediation* – the network mediation function gains necessary routing and network awareness for supporting content delivery through the best transport strategy in order to improve both user QoE and effective bandwidth utilization.

The control plane of the COMET architecture, the *content mediation plane*, or CMP, is realized through content resolution and mediation entities, which communicate with each other in order to provide inter-domain mediation (the two planes of COMET are shown in Figure 1 in the next section). Each domain or Autonomous System (AS) must operate at least one such entity, although it may operate more based on non-functional requirements such as availability, response time etc. In fact, these mediation entities are similar to today's Domain Name System (DNS) servers and every content publishing or consuming application should know its local mediation entity (through local configuration, in a similar fashion to today's local DNS server).

The key technical advantages that can be achieved thanks to this mediation are:

- Unified access to the content whatever its nature and location. Content delivery with guaranteed QoS.
- Point-to-multipoint content delivery capabilities, reducing bandwidth needs for live contents.
- Graceful handover of the content delivery path, providing more resilience and flexibility for multi-homed users.
- Advanced publication mechanisms, allowing Content Providers to update content servers on-the-fly, while switching among different ways of distribution.

As mentioned before and also reported in D2.2 [3], in COMET, we have developed two different approaches for the mediation plane to resolve and access content. In the first approach, the so-called *decoupled approach*, the resolution and mediation entities resolve the content name to a set of sources that hold that content, given that the content may be replicated [20]. This list may also include routers with content caching capability, if there is capability for content caching *within* the network (e.g., [14]). This resolution can be achieved through suitable organization of content records in the resolution entities, e.g., through a hierarchical Distributed Hash Table (DHT) approach or even through hierarchical content naming, in a similar fashion to the domain name system (DNS), although in the latter case it is difficult to cope with dynamic content caching in the routers. A list of content sources is returned through the resolution operation and the best possible source is then selected based on network distance, server load and other relevant information available in the mediation plane, for example, QoS parameters or average network load along the path. The content is finally requested from the selected source. We call this the *decoupled approach* as it decouples content resolution from content request and delivery, in a similar fashion to the resolution of host names to IP addresses through the DNS before establishing a session to a remote host in the current Internet. We provide a detailed description of this approach in Chapter 4.

In the second approach, information in the resolution and mediation entities about the "direction" in which a particular piece content can be found will guide the resolution message in a hop-by-hop fashion to the content source. This information is used together with information on network distance and server load in order to locate the best possible content source. We call this the *coupled approach* as it couples content resolution and access with content delivery: the content request message is routed through a chain of resolution and mediation entities across domains to the content source. The reasoning behind this approach is that it saves one round-trip as compared to the current DNS resolution system since it couples the name and content resolution into one phase. Another benefit of this approach is that it can support in-network caching and, given it emulates the function of native in-network approaches in the mediation plane it can constitute an interim migration step towards full native deployment of information-centricity. This is in line with native information- centric approaches such as NDN[2] and PURSUIT[3] and in contrast with the decoupled

[2] http://www.named-data.net/

[3] http://www.fp7-pursuit.eu/

approach, which separates resolution from content access and delivery. These approaches are also commonly called *two-phase* (the decoupled one) and *one-phase* (the coupled one) in information-centric architectures.

One key aspect in all information-centric architectures is content naming, as it dictates the content resolution process. In fact, in native information-centric architectures, packets are routed based on content names or IDs instead of host addresses. The same is the case in the mediation plane for the coupled approach described above. These names are not necessarily "human friendly" as they may be opaque; they may also be self-certified for security reasons, given that it is the content itself and not the content source that needs to be secured. Human users could find such a name or ID through search-engines based on the content object properties. In fact the mediation plane will also provide "hooks" to external search engines in order for them to index content based on meta-data.

The content mediation plane (CMP) can support information-centric operation over the current IP-based Internet, unifying content resolution, access and delivery for all types of content. In the decoupled approach, it can simply act as content-oriented enhanced "DNS" that could locate the best possible content source although network support may be optionally used for better-than-best-effort content delivery. In the coupled approach, network support is necessary, in the form of content-aware routers, which we call Content-Aware Forwarding Entities (CAFEs) and which have the capability to natively handle the delivery of content objects according to their IDs; CAFEs may also support in-network caching functions for achieving localized content access. Therefore, the CAFEs, initially introduced in D2.1 [2] and which included the Content-Aware Forwarding Function (CAFF) are now complemented by one new function to support in-network caching. This function is called Content-Aware Caching Function (CACF).

CAFEs may also cache popular content, thanks to the aforementioned Content-Aware Caching Function (CACF), that is traversing them guided by the local domain mediation entities. Caching in this case relates to content chunks (see Chapter 5 and [15]) and not to individual packets as in the NDN approach for instance. That is, while NDN proposes indiscriminate caching everywhere along the delivery path, our initial work on modeling and evaluating caching trees has shown that caching is more beneficial in specific network locations (D5.2 [9], D4.3 [7]). We discuss the exact operation of caching later on in Chapter 5. According to our approach in COMET, CAFEs are necessary in the domain edge (i.e., border routers *have* to be content-aware) but could also start being gradually deployed within the network for advanced information-centric network operation. In a target, native information-centric environment, the mediation plane for the coupled approach will be collapsed completely into the network, with ubiquitous native content-aware routers (e.g., CAFEs) routing packets based on names / IDs, as in recently proposed radical approaches.

The COMET architecture has been devised with a strong focus on addressing real-life (or almost real-life) use cases. Four of them were identified and described in depth in D2.2 [2], namely, a) *Adaptable and efficient content distribution,* b) *Handover of content delivery path in a multi-homing scenario,* c) *Webinar "All about CDNs"* and d) *P2P offloading*. During the project life only use cases a) and d) have been eventually implemented and demonstrated on the Federated Testbed set up in the context of WP6 (see deliverables D6.1 [10] and D6.2 [11]) since they were the most suitable for the sort of infrastructures available in the local testbeds federated in the COMET testbed. The tests carried out over the COMET federated testbed have proven the validity of the COMET Architecture for those two use cases, without precluding its extension/applicability to the remaining two other use cases.

3 High-level COMET Architecture

3.1 COMET Functional Architecture

In this Chapter, we present the functional view of the COMET system in terms of the contained functional blocks in the Content Mediation Plane (CMP) and the Content Forwarding Plane (CFP); this description is general, covering both the decoupled and coupled approaches. The overall functional architecture consists of two distinct planes as shown in Figure 1. The CMP is responsible for content resolution, i.e., for the optimal identification of the best content source according to the specific requirements of the content consumer, while the CFP deals with end-to-end content delivery.

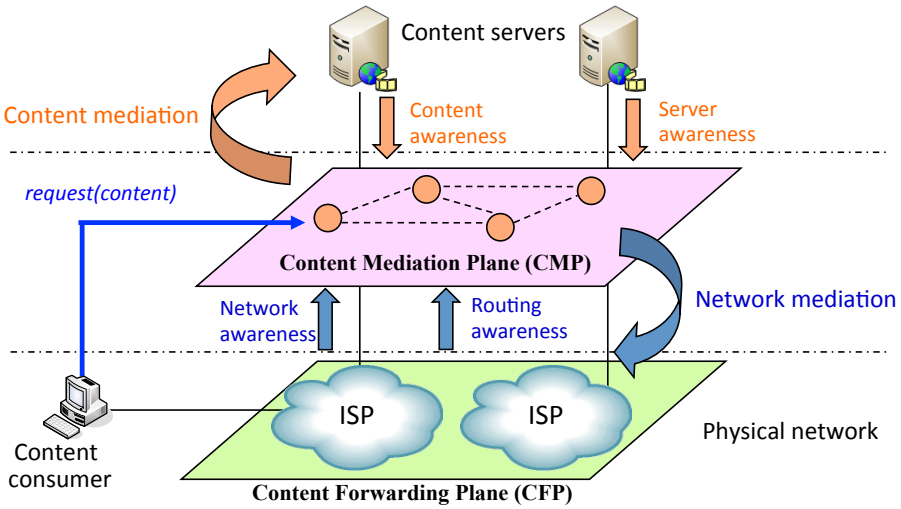


Figure 1. COMET Content Mediation and Forwarding Planes

The COMET system consists of:

- ***Functional blocks***, which accomplish the operations of the COMET system from content publication to content consumption. These functional blocks are realised via one or more entities, as we will explain later.
- ***Entities,*** which realises specific content manipulation functions. They interact with each other in the COMET system in order to i) *publish content* (i.e., register content to the COMET system), ii) *request for content* and iii) *provide content-server monitoring information*.

The functional architecture is depicted in Figure 2 and encompasses the following functional blocks, which are grouped in the CMP and the CFP, respectively:

- *The Content Resolution Functional block (CRF) of the CMP:* It is invoked during content publication and content consumption. Its main tasks are to maintain content records and to resolve content requests (i.e., resolve content names to content properties included in the content records and include content metadata). A content record provides the mapping information from the content name to the physical content sources in the Internet, and it may also include content metadata (e.g., alias, media type) to be used during the content resolution operation.
- *The Path Management Functional block (PMF) of the CMP:* It interacts with the underlying network to gather necessary network reachability information across domains through BGP, and also information concerning quality of service (QoS) capabilities and characteristics in QoS-capable networks, e.g., quality of service classes that are supported along the route. It is important to note that the information dealt with by the PMF is *long-*

term information.

- *The Server and Network Monitoring Functional block (SNMF) of the CMP:* The SNMF is responsible for gathering "just-in-time" information about both the server load conditions and also for potentially monitoring underlying path quality (e.g., bandwidth availability) at relatively short timescales for supporting optimized content resolution and delivery operations. The monitoring operation is typically time-driven, i.e., server and network conditions are periodically measured independently of the incoming content requests, although server load could also be event-driven.
- *The Content Mediation Functional block (CMF) of the CMP:* Being the core function as "decision maker" in the CMP, it gets necessary input from the CRF, SNMF and PMF, interacts with content clients during content consumption and configures the CAFEs in the CFP for setting up delivery paths during content consumption. Its main functionality is to make decisions regarding the selection of the best content copy based on information regarding server and network conditions received from SNMF and the information about the available paths from the PMF. Based on this information, it then determines the content delivery paths and performs necessary configurations.
- *The Content-Aware Forwarding Functional block (CAFF) of the CFP:* It is the main function in the CFP and is responsible for forwarding content through end-to-end paths determined by the CMF. It is actually concerned with data-plane aspects for handling content delivery, including appropriate forwarding behaviors.
- *The Content-Aware Caching Functional block (CACF) of the CFP:* This function is responsible for the in-network caching operations of the COMET system. The detailed operation of our in-network caching functions has been described in detail in [14], [15] and in [22].

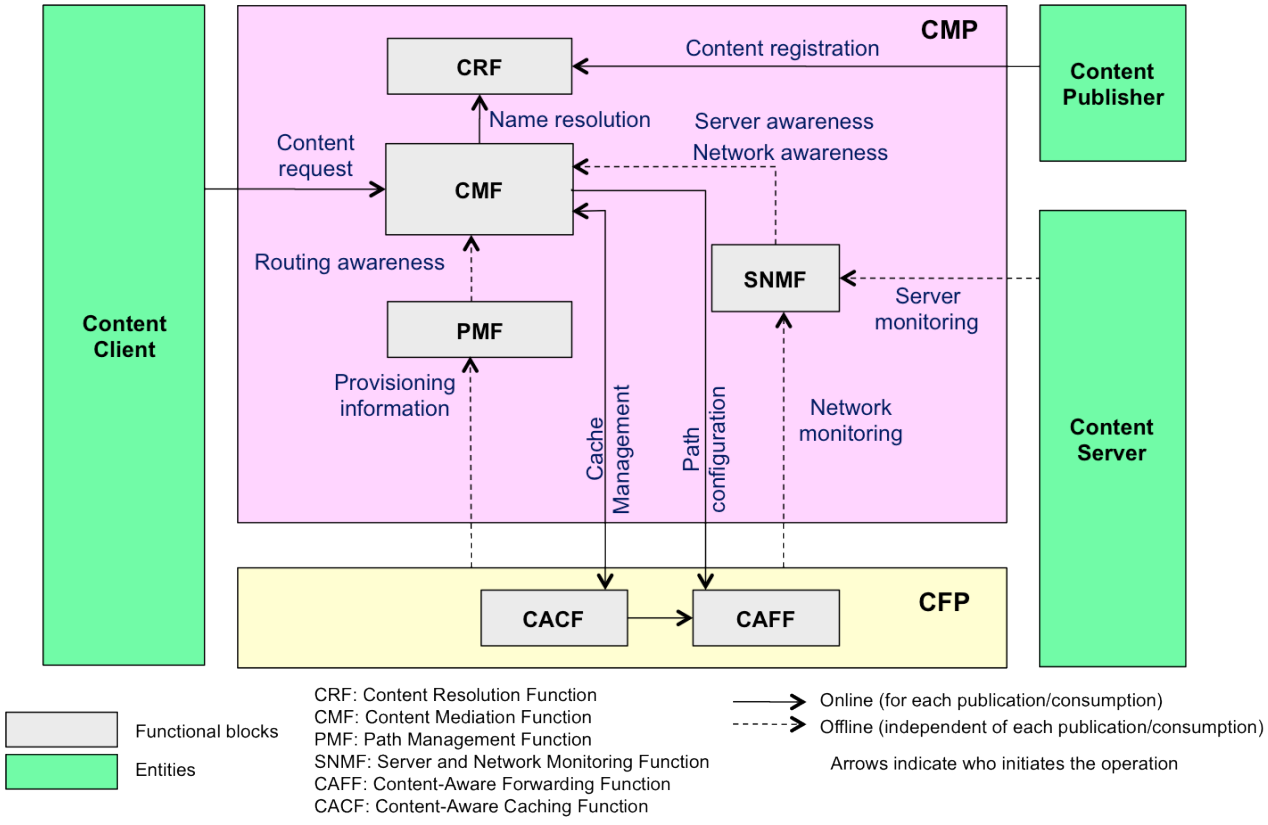


Figure 2. High-Level COMET Architecture

The overall content mediation system works as follows. Individual content publishers register their content to the CRF that is responsible for maintaining a global distributed content repository with records of all registered content items. When a content client requests a specific item, it uses a unified interface to contact the CMF for triggering the content resolution process. To start with, the CMF interacts with the CRF to identify the possible content locations according to the requested content name. Meanwhile, the CMF also receives both long-term and dynamic information as input from the PMF and SNMF respectively regarding route reachability and most recently monitored path conditions. An optimized decision based on the available information gathered is then made to identify the best server candidate together with the delivery path.

Thereafter, the CMF configures the underlying content delivery path (i.e., instructing the CAFF at the CFP) to be ready for delivering the content flow. At this stage configuration of the CAFFs also includes decisions on which contents should be cached in the network and at which place. This action involves interaction between the CMF and the CACF. In particular, the CMF is instructing the CACF on which incoming content chunks to cache and the CACF is informing the CMF of potentially cached contents upon content requests. As noted earlier, in-network content caching is mainly applied in the coupled approach. The detailed description of this operation (i.e., interaction between CMF and CACF) is given later in this document in Chapter 5, where we discuss the coupled approach to content resolution and delivery and in particular how individual chunks of a content object are requested.

As discussed in detail in [14], [15], [21], in COMET we have designed and developed two main in-network caching algorithms. The first one follows a resource-management approach to in-network caching [15], where the caching resources of the path are approximated and content chunks are cached probabilistically. The probabilistic nature of the algorithm introduced in [15] and further evaluated in [23] guarantees fair allocation of caching resources to all flows that share the same content delivery path. The second algorithm introduced in [14] and further evaluated in [22], follows a topology-driven approach. That is, the centrality of the participating domains is calculated and contents are cached in the most centrally located domains of the content delivery path [22]. The detailed support for content chunking by the high-level architecture presented in Figure 2 is discussed later on in Chapters 4 and 5 for the decoupled and the coupled approach respectively.

It is worth mentioning that the functional components of Figure 1 are all *logical* ones, and hence, can be realized through different physical instantiations in practice. In addition, not all functions are necessary in every instantiation. For example, in the decoupled approach, the CRF becomes a separate physical entity in a similar fashion to a DNS server while in the coupled approach both CMF and CRF can be realized in a common entity.

3.2 COMET Instantiations: From Functional Blocks to Entities

As mentioned before and further elaborated in D2.2 [3], based on the overall COMET architecture described in last section, two different approaches for mapping the functional blocks into entities have been designed and studied, namely decoupled approach and coupled approach, Both approaches differ not only in how functional blocks are mapped into entities, but also in the strategy to deploy these entities. Due to these differences, the two approaches may comprise different content resolution and delivery properties. We describe the mapping of the functional blocks into entities in the following two subsections and the detailed content resolution and delivery properties of each approach in the next two chapters, respectively.

3.2.1 Decoupled Content Mediation Approach

The decoupled approach is characterized by the following features:

- It follows the basic paradigm of the current Internet by allowing the physical signalling routes for content resolution and content delivery to be separated but coordinated.
- The content resolution sub-operation is based on a global directory system, which stores

content information, in a similar way as the current host-centric Internet uses global DNS directory system to support domain name resolution.
- There are specific and separate entities that hold the CRF functional block and act as the global directory system for contents. These entities are different from the ones holding the CMF functional block.
- The entity hosting the CRF functional block is not linked to any specific network domain. On the other hand, the entity hosting the CMF is associated to a specific network domain.

In this approach, the functional blocks described in Section 3.1 are mapped into the following entities:

- *Content Resolution Entity (CRE):* This is the entity that encompasses the CRF. A CRE keeps track of the Content Records and resolves Content Names into their assigned Content Records (i.e., the structures storing all the relevant info/metadata about a content see section 4.1 for a detailed explanation), whenever a Content Mediation Entity (CME) requests them. Replicating the typical DNS hierarchy, there are two types of CREs, the authoritative CREs, which store the Content Records for several domains, and the root CREs, which map domains with the authoritative CRE storing its Content Record.

- *Content Mediation Entity (CME):* This entity encompasses the CMF and is the one that the Content Client (CC) contacts first to consume contents. Apart from requesting the Content Record to the CREs, obtaining the paths linking the Content Servers (CSs) in the Content Record to the CC and the status of the CSs; CME's main function is the selection of an optimal pair (i.e., CS and delivery path) based on the content characteristics, the CC COMET Class Of Service (CoS, see section 4.1) and the awareness provided by other entities (Routing Awareness Entity or RAE, Server and Network Monitoring Entity or SNME, server's CMEs). Finally, this entity is also responsible of configuring the content delivery paths in the Content-Aware Forwarder Entities (CAFEs).

- *Routing Awareness Entity (RAE):* This is the entity that encompasses the Path Management Function (PMF). The RAE gathers Network Layer Reachability Information (NLRI) and provides it to the CME in a proactive offline manner to enable the CME to perform the path discovery process and make decisions. The path information supplied by the RAE is in the form of a list of Autonomous System IDs (AS IDs) linking the client's ISP with the server's ISP, each path being qualified with a COMET CoS and QoS Parameters (Bandwidth, Packet Loss, Packet Delay).

- *Server and Network Monitoring Entity (SNME):* This is the entity that encompasses the Server and Network Monitoring Function (SNMF). The SNME collects data about the status of content servers and the network conditions (basically the status of edge CAFEs, see section 4.1), which can be requested by the CMEs as needed. The network/server status data stored in the SNME is consolidated data provided by the server and network monitoring agents, in the form of predefined levels defined as HIGH/MEDIUM/LOW.

- *Content-Aware Forwarder Entity (CAFE):* This entity encompasses the Content-Aware Forwarding Function (CAFF), delivering the content through the paths defined in key format (see section 4.3 and 4.4) as instructed by the CME. In the decoupled approach the CAFEs are divided into two main types: edge CAFEs, located at the network edges and serving the CCs and CSs, and border CAFEs, those located at the ISP borders and managing the links connecting one ISP to its neighbours. Besides, CAFEs in the decoupled approach are stateless, therefore not storing any information about the flows crossing them, and CAFE configuration will be restricted to the edge CAFEs serving the CSs (more details on section 4.3). Edge CAFEs also report their status to the SNME, so this information can be taken into account in the CSs status evaluation (see section 4.3).

As it will be shown in Chapters 4 and 5, it must be noted that the RAE entity is the same in both approaches.

3.2.2 Coupled Content Mediation Approach

The coupled approach has the following characteristics:

- It follows a disruptive paradigm with respect to the current Internet, with the physical signalling routes for content resolution and the corresponding content delivery being coupled. More specifically, the *domain-level* content delivery path exactly follows the reverse direction of the original resolution path for each content consumption request.
- The content resolution sub-operation is performed on a hop-by-hop basis across intermediate domains.
- Content delivery paths in the CFP are maintained with content states installed during the resolution phase in the CMP.
- A unified entity holds both CRF and CMF blocks, and this entity is associated to a specific network domain.

The current Internet relies on the DNS system whereby a request is usually resolved first by querying the DNS system before the actual request is being sent to the resolved content server that hosts the requested content. The main strategy of this coupled approach is to simplify this two-phase process by combining the two round-trips into one single operation. Essentially, the resolution procedure is coupled with the content delivery procedure and thus saving one round-trip. This requires more radical changes to the intrinsic working of the current DNS-IP based Internet structure. These radical changes are realized in the coupled content resolution approach of the COMET system.

In this approach, these functions are grouped into three physical entities.

- *Content Resolution and Mediation Entity (CRME):* this entity, typically owned by individual ISPs, encompasses the CRF, CMF and SNMF functional blocks from the COMET architecture.
- *Routing Awareness Entity (RAE):* this entity encompasses on the PMF functional block. It is the same corresponding entity in the decoupled approach where the Network Reachability Information is obtained and fed to the CRME (more specifically, to the CMF block).
- *Content-Aware Forwarding Entity (CAFE):* similarly to the RAE, this entity is also common to both approaches and is the only entity of the CFP. As mentioned above, this entity is responsible for the delivery of content through the paths that have been decided by the CMF. Apart from the Content-Aware Forwarding Function (CAFF), the CAFE also contains the Content-Aware Caching Function (CACF), which is responsible for caching content chunks in edge-domain routers (i.e., CAFEs).

We note that according to this approach, each COMET-enabled domain has to have (at least) one CRME implemented or having an agreement with one of its neighbouring domains for the access of the neighbour CRME. The CRME interfaces with all internal and external COMET entities, in order to accomplish the publication, resolution and delivery operations. In addition, each CRME interfaces with other CRMEs, located in the same or other domains, resulting in a network of CRMEs, which constitute the CMP.

4 Decoupled Approach

4.1 Overview of the Decoupled Approach

The decoupled approach follows the basic paradigm of the current Internet by allowing the physical signalling routes for content resolution and content delivery to be separated but coordinated. By decoupling the content resolution from content delivery, it is possible to have different mechanisms for both tasks, thus implementing the most appropriate one for each purpose. Another important consequence is that the decoupled approach can be deployed as an overlay over the existing Internet, allowing coexistence with current systems and partial deployment during the interim period of adoption.

Central to the approach is the existence of a global directory system that stores *Content Records* identified by a unique *Content Name*. These records are data structures containing *Content Sources,* lists of *Content Servers (CSs)* grouped by common shared properties, which basically are the COMET Class of Service (CoS, see below), the Quality of Service (QoS) requirements for the distributed content and the protocol used for retrieval. Each CS is in turn characterised by its location, its IP address and content path, while Content Sources can be ordered by assigning priorities, imposing an explicit selection preference.

This global directory system resolves Content Names to its associated Content Records, using for that resolution a hierarchical architecture similar to the one used in DNS, with authoritative nodes for the Content Records and root nodes mapping the Content Names to the authoritative nodes. As such, the entities composing this global directory system are not directly mapped to ISPs, as it happens with the remaining decoupled entities.

Another central concept in the decoupled approach is the COMET CoS, which can adopt three predefined values: Premium (Pr), Better than Best Effort (BtBE) and Best Effort (BE), implying decreasing QoS levels/subscribed SLAs. The COMET CoS is used for characterising Content Clients (CCs), CSs and network paths, ensuring for example that CCs with Pr CoS are systematically assigned Pr CSs and Pr network paths by the decoupled entities, so that QoS requirements and subscribed SLA are warranted in the entire transmission chain from the CC to the CS.

The main entities that comprise the decoupled approach operation were given in section 3.2.1, while Figure 3 depicts the mapping of COMET functionalities to decoupled entities and the interactions between them. For simplicity, only the client's ISP, the server's ISP and the authoritative CRE have been included in the figure, but note that the entire process can involve a number of transit ISPs and their associated COMET entities.

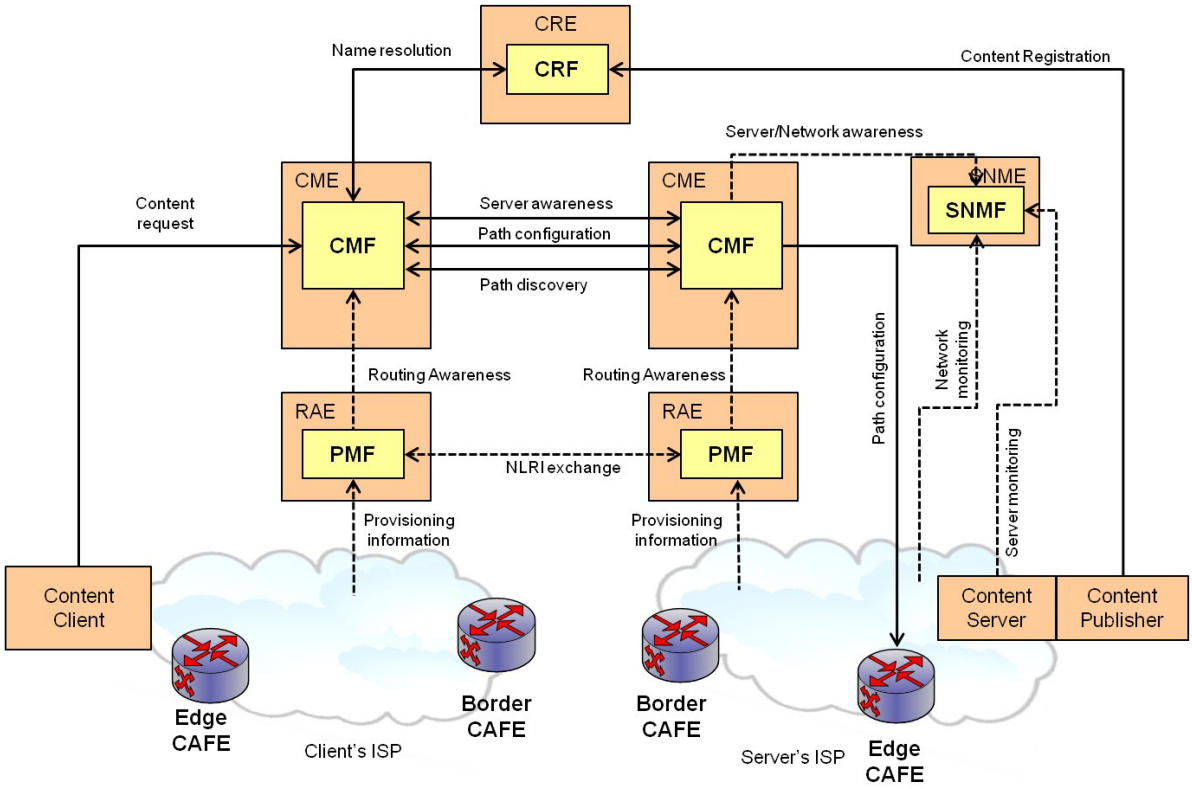


Figure 3. Decoupled approach – Architecture and Information Flow

Note that apart from the CRE hierarchy, all other entities are associated with and deployed in a specific ISP. The distribution is as follows:

- CREs are outside the ISPs, without a direct relationship with them. For instance, a single CRE can store Content Records from several domains while subdomains in a single domain can be distributed in different ISP.
- Each ISP manages an RAE, a CME and an SNME.
- An ISP typically operates several edge and border CAFEs, depending on the links to neighbouring ISPs and the managed access networks.

Fundamentally, there are two main operations: Content Publication and Content Consumption. The latter operation can be further divided into two sub-operations performed per content request: Content Resolution and Content Delivery. The main interactions between the different entities, as depicted in the previous figure, are as follows.

- *Content Registration*: The process used by a CP to create/update/delete a Content Record in the CRE.
- *Content Request*: How a CC requests a Content Name from the CME and receives the connection data of the selected CS.
- *Name Resolution*: How CME requests the CRE to resolve a Content Name into a Content Record.
- *NRLI Exchange*: How RAEs exchange path information so that any existing RAE learns how to get from its ISP to any other ISP.
- *Path Discovery*: How CMEs exchange information in order to find a path in AS format (as used by RAEs) if that piece of information is not available in one of them.
- *Path Configuration between CMEs:* How CMEs exchange information in order to translate a path in AS format (as used by RAEs) into a path in key format (as used by CAFEs).

- *Path Configuration between CME and CAFE:* How a CME configures an edge CAFE for content transmission with the path between the CS and the CC in key format.
- *Provisioning Information:* How a RAE obtains path information from the underlying network.
- *Routing Awareness:* How a RAE reports to a CME the information about the paths leading to other ISPs, in AS format.
- *Server Awareness between CMEs*: How a CME polls another CME about the status of server deployed in the ISP.
- *Server/Network Awareness*: How the CME retrieves status information from the SNME.
- *Server/Network Monitoring:* How the CS or the Network Elements report their status to the SNME.

4.2 Content Publication in the Decoupled Approach

Content Publication is the process of making content available to the COMET system and eventually, Content Consumers. It consists of 2 steps, Content Storage and Content Registration, both described below:

1. *Content Storage* is the operation through which a Content Provider or Owner stores its content in its Content Servers. However, this is an offline process and it is out of scope of COMET project.
2. *Content Registration* is the operation of registering the stored content to the CRE hierarchy. There are 2 types of CREs; a *root CRE*, which contains information about the authoritative CREs associated with specific naming authorities, and an *authoritative CRE*, which contains the content records associated with Content Provider's content.

   Initially, every Content Provider or Owner has to obtain a part of global COMET namespace through a registrar. This offline sub-operation is similar to acquiring a domain name from DNS and is only performed once. The Content Provider receives a globally unique identifier (naming authority), as well as a record in the root CRE(s), containing information about its naming authority and serving authoritative CRE(s).

   The Content Provider will then create content records for its respective contents. The content record contains an assigned content name, which is unique in Content Provider's namespace and other content-specific parameters, which were presented in D3.2 [4]. This process involves two COMET entities, namely, the assigned *authoritative CRE* for a specific part of its namespace, and the *Content Publisher*, which exposes both a web-based user interface and an API for creating a content record in the authoritative CRE. In the first case, a Content Provider administrator can securely access a Content Publisher's web interface by inserting his given credentials and create, update and delete content records, which are stored in the authoritative CRE. The exposed API also ensures the integration of Content Publisher software to existing applications and services, such as Youtube or Facebook, for the automatic content publication. Content record creation was described in D3.2 [4].

Once the content has been registered in the COMET system, the CMEs are responsible for making the content properties available throughout the Internet so this content can be reachable by all Content Clients of the different network domains.

4.3 Content Resolution in the Decoupled Approach

In the Decoupled Approach, content resolution is the sub-operation responsible for the discovery of the requested content, understood as the retrieval of the connection data to a CS hosting a requested content, from a Content Name supplied by the CC. This sub-operation is not restricted to this final outcome, but also involves the selection of an optimal CS and path linking the CS and CC,

according to the characteristics of both the CC (COMET CoS) and the requested Content, as well as the configuration of the selected path for retrieval in the edge CAFE serving the CS.

The following flow graph (Figure 4) illustrates the basic exchanges of data between the main entities involved in this sub-operation. A more complete explanation of this sub-operation and the internal processes in each entity can be found at D3.2 [4] and D4.2 [6]

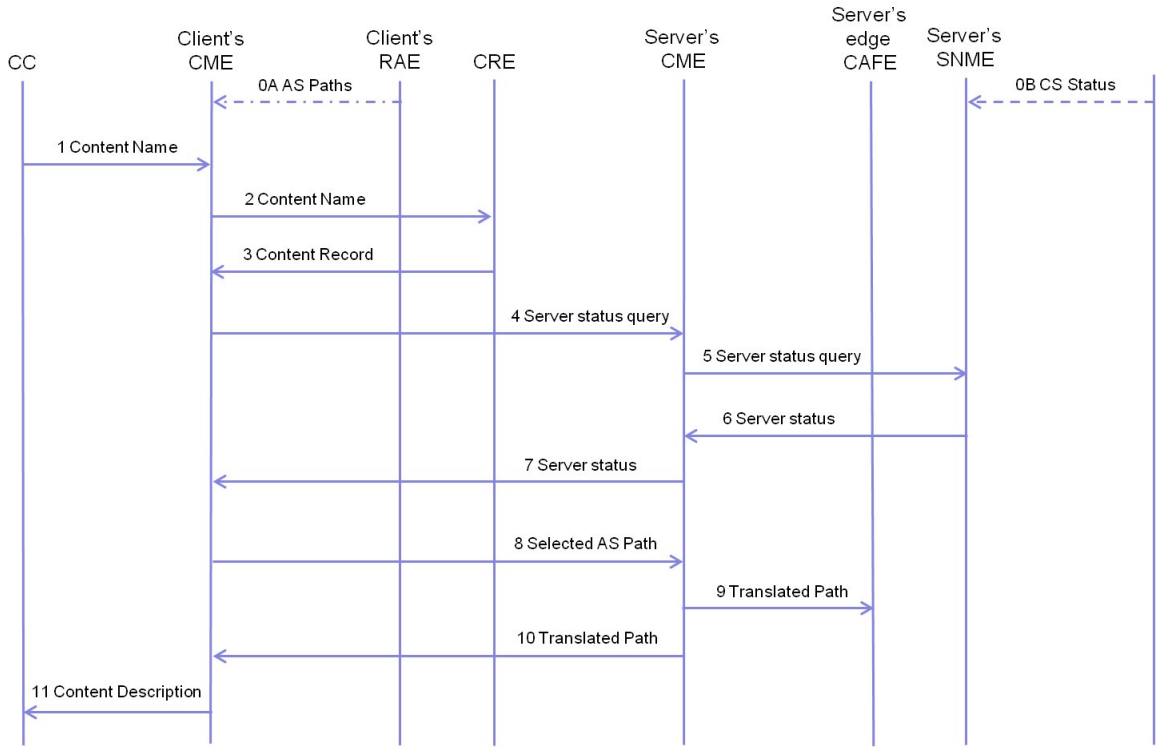


Figure 4. Content Resolution in the Decoupled approach

In general, before Content Resolution can start, several operations must have been internally performed in COMET.

Firstly, a Content Record for the Content Name needs to be published in the CRE, as explained in Section 4.2.

Secondly, RAEs need to have exchanged their path information (not illustrated in the figure), so that any RAE learns how to reach any given ISP. Paths in RAEs are expressed in AS format, as a list of AS IDs that have to be traversed to download contents from one ISP. Paths are also qualified with COMET CoS and QoS parameters (bandwidth, packet loss, packet delay), so they can be assigned to the right CC and the right contents.

RAEs can be understood as inter-domain routing entities similar to BGP speakers that perform the off-line routing awareness process in long time scale. Based on the inter-domain provisioning information from the inter-domain SLA agreements and the intra-domain provisioning information, each RAE exchanges Network Layer Reachability Information with other RAEs in peering domains to build inter-domain routing paths. Nevertheless, two main features make RAEs differ from BGP speakers:

- NLRI is exchanged in terms of COMET CoS that are globally known by COMET-enabled ISP networks. This feature enables RAE to assure end-to-end COMET CoS connectivity by mapping COMET CoSs into internal CoSs offered by particular domain;
- RAEs are able to propagate information about a number of alternative paths allowing for multi-path routing (instead of a single path routing performed by BGP) in order to offer differentiated content delivery across domains and improve availability and performance.

More details about RAEs and the routing awareness process are provided in D4.2 [6].

Thirdly, an RAE within an ISP has to report the discovered AS Paths to its CME (operation 0A in the figure), so the CME has them available locally and does not need to launch time-consuming requests to external elements.

Finally, the SNME has to collect the status of the CSs (operation 0B) deployed in its own ISP (and the edge CAFEs serving them, not illustrated in the figure). These statuses are periodically sent by the CSs and the edge CAFEs in a consolidated format, with three predefined values (LOW, MEDIUM, HIGH) which express increasing CS/CAFE loads.

Once these pieces of information are collected, Content Resolution can start. Some fail-safe mechanisms (not illustrated) have been created in case some of these pieces of data are not available.

- If an edge CAFE or CS does not send information to the SNME, the SNME assigns a predefined status of UNAVAILABLE to this element, so this CS/edge CAFE is not eligible for content distribution.
- If a path to a CS is not stored in the CME local path database, the CME queries the server's CME (as specified in the Content Record) to provide the path information.

In a normal situation, content resolution will proceed as follows.

1. The CC sends the Content Name (1) of the desired content to the CME deployed in its local ISP.
2. The CME queries the CRE (2) to retrieve the Content Record (3) for that Content Name. This operation involves two sub-operations:
    - A query to the root CRE to obtain the address of the authoritative CRE hosting the Content Name. As the CME keeps a cache mapping domains with authoritative CREs, this operation is only performed for a small percentage of the Content Resolution queries.
    - The actual query to the authoritative CRE to retrieve the Content Record for the Content Name. As explained before, a Content Record is a list of Content Sources with CSs sharing the same configuration parameters, namely, the COMET CoS, the QoS requirements of the Content, the protocol of retrieval and the priority of the Content Source (if there are several of them with the same CoS). Each CS in a Content Record is characterized by its IP address, the path for retrieving the content and the CME serving the ISP the CS is located.
3. The Client's CME then polls the servers' CME (4) for the status of the CSs listed in one of the Content Sources in the Content Record. The selected Content Source is the one matching the CoS of the Client and with the highest priority.
4. Each polled server's CME polls in turn its SNME (5) to obtain the CS statuses (6). As commented above, the SNME stores both CS and edge CAFE status. The CS status returned to the CME is refined by using the information of the edge CAFE serving it. For instance, if the edge CAFE is in HIGH status, all the CSs connected to the edge CAFE are labelled as HIGH, regardless of their individual statutes.
5. The server's CME returns the CS statuses to the client's CME (7).
6. At this moment, the CME has collected information about paths to the ISPs where the CSs in the selected Content Source are deployed and the status of those CSs. This data is fed into the decision process of the CME which returns an optimal solution expressed as the couple CS/path to reach it. The proposed decision algorithm belongs to the multi-criteria optimization approaches, which evaluate importance of decision variables based on two specific reference values, called aspiration and reservation levels. Finally, the candidate solutions are ranked based on objective function expressing operator policy. Details about decision algorithm are presented in D3.2 [4] and D5.2 [9]. If no solution is obtained the

next Content Source according to priority and CoS is assessed and all the operations from step 3 above are repeated.

7. Now, the edge CAFE serving the selected CS has to be configured to ensure the content transmission. However, the obtained path is in AS format and needs to be translated into key format. The key format is a list of the interface identifiers each CAFE in the selected path uses to forward a received packet and which belongs to the data flow from the CS to the CC. So, the client's CME sends to the server's CME (8) the path in AS format plus the list of required keys inside the client's ISP.
8. The servers' CME determines the keys inside its ISP and the keys leading to the next ISP in the list. If the client's ISP and the server's ISP are not directly connected, the servers' ISP will poll the CME of the next ISP in the AS list to provide the missing keys (operation not illustrated), until the AS path has been fully translated into a key path.
9. Once the translation is completed, the server's CME configures the edge CAFE serving the CME (9) with the path in key format (see in section 4.4 what this implies for Content Delivery).
10. If this operation is successful, the server's CME reports to the client's CME (10) that the content is ready for retrieval.
11. The client's CME then informs the requesting CC (11) about the connection data of the selected CS.

Note that in case of failure in any of these steps the CC is not notified. CC's default behaviour is wait for timer expiration and retry the query three times, before aborting.

4.4 Content Delivery in the Decoupled Approach

The content delivery sub-operation focuses on transferring the content from the Content Server to the Content Client. It involves three entities of the COMET architecture: the Content Client, the Content Server and the CAFEs. The CAFEs are network nodes that actually transfer content packets along the content delivery path selected during content resolution process.

In the decoupled approach, we design the stateless content delivery approach where CAFEs maintain only the neighbourhood (local) information, i.e., how to forward packet to the peering CAFEs. All information about content delivery path is stored in a COMET header attached to the original packet containing content payload. The COMET header is attached and removed by edge CAFEs located close to the content server and the client, respectively. The stateless content delivery follows the source routing principle at the domain level, which allows for flexible selection of content delivery path for each content request. Furthermore, the proposed approach is technology-agnostic in the sense that it allows each domain to use different packet forwarding technologies between peering CAFEs.

CAFEs are instructed during the content resolution on how to transport the content. These instructions include how to classify the content packets sent by the CS and how to forward the packets through a specific path of CAFEs. The detailed description on the content delivery operations corresponding to this part of the COMET architecture is described in D4.2 [6].

It must be noted that not all network routers are required to be CAFEs. We envision that the routers that will support the CAFF functionality are the ones that reside at the edges of the domain.

4.5 Extra Features of the Decoupled Approach

4.5.1 Support for system reliability

Robustness can be achieved by means external to COMET. Standard procedures (frontend/backend machines, entity redundancy, passive/active configurations, database

replication) already exist in current exploitation environments, and are sufficiently tested and tried to be trusted.

COMET needs to protect the following entities against failures/downfall:

CRE

- Optimal Approach: to have multiple CREs behind a frontend that exposes a single IP address to external entities. CREs are configured in active/passive mode (only one of them is active at any time) with database replication in order to avoid publishing operations getting lost.
- The CRE database should be updated to a version supporting replication and cluster mode, in case these capabilities are not provided.

SNME

- Optimal Approach: to have multiple SNMEs behind a frontend that exposes a single IP address to external entities. SNMEs are configured in active/passive mode with database replication in order to avoid server status reports from CSs getting lost.
- The SNME database should be updated to a version supporting replication and cluster mode, in case these capabilities are not provided.

RAE

- Optimal Approach: to have multiple RAEs behind a frontend that exposes a single IP address to external entities. RAEs are configured in active/passive mode with database replication in order to avoid path update operations getting lost.
- The RAE database model should be updated to a version that supports replication and cluster mode, in case these capabilities are not provided.

CME

- Optimal Approach: to have multiple CMEs behind a frontend that offers a single IP in the different interfaces, configured in active/passive mode with database replication in order to avoid configuration operations getting lost.
- The CME database model should be updated to a version that supports replication cluster mode, in case these capabilities are not provided.
- However, the CME implementation has to be modified to answer border/edge CAFE downfalls, because the CME is in charge of configuring the edge CAFEs with the list of keys to use for content download. Two different situations can arise:
- Edge CAFEs are down.
    - Since CME polls edge CAFEs for expired flows, this situation will be detected and this CAFE (and its attached servers) can be made not be eligible for new content downloads.
    - If a backup CAFE automatically replaces the one offline, the CME could reconfigure the new one with all the active flows previously assigned to the old one, ensuring that CCs will only experience small glitches in active streaming/download.
- Border CAFEs are down.
    - If an edge CAFE detects that a download path is broken, it can report the situation to the CME, so that this entity can suggest an alternate path linking the Content Client and the Content Server. Therefore, the CME implementation has to be modified to store all the possible path solutions, so a new one can be configured in the suitable edge CAFE.

- The old path will be marked as inaccessible in the CME database to avoid it being assigned to new queries.

CAFE

- Two scenarios have to be analysed: Edge CAFE vs. border CAFE failure
- Edge CAFE failure:
  - In this case, the CME will detect that the CAFE is out of service, because of the flow polling mechanism, and can therefore prevent new downloads from being placed at that CAFE.
  - If a backup CAFE is automatically awaken when the original one fails, the poll mechanism will enable the CME to detect the breakdown in communications, so the active flows could be reconfigured in the new CAFE, enabling current downloads to be resumed.
- Border CAFE failure:
  - The edge CAFE can be enhanced to check conditions for the flows being managing, enabling the detection of broken paths (e.g. because of a fallen border CAFE). The CME can be then informed so that it can provide an alternate path for content downloading.

4.5.2 Support for Content Chunking

Content chunking assumes that a given content is available in a number of fragments. Consequently, the user application may progressively download particular content chunks instead of downloading the whole content in one piece. Moreover, particular chunks may be downloaded from different servers in parallel. The main advantages of content chunking in content delivery are the following:

1. The content may be downloaded progressively following the content playback.
2. It supports scalable video coding by chunking particular video sub-streams.
3. Allows saving space in the content caches by placing content fragments instead of the whole content. Consequently, it may improve effectiveness of content caching.
4. The replicas of content chunks can be located on different content servers, allowing for:
   - adaptation to changing server and network conditions;
   - improving download rate by simultaneous download of chunks from different servers (this feature improves download rate from content servers connected by low rate links);
   - improving reliability in case on unstable content servers.

In the decupled approach content chunking may be supported in several ways that differ in the "system awareness" about available content chunks. Before we present the proposed approaches, we shortly discuss the COMET features that impact the gain achieved by making use of content chunking.

- The server load is controlled by the COMET system, so there is no need to change the server during content consumption.
- In Premium Class of Service (CoS), the network conditions are controlled by COMET in preventive way, so there is no need (and gain) to change paths during content consumption.
- In Better then Best Effort CoS, the network conditions are controlled by COMET in a reactive way, so there is limited motivation to change paths during content consumption.
- In Best Effort CoS, the network is uncontrolled by COMET. Therefore, it is reasonable to delegate decision about server and path selection to the user application.

For the decoupled approach, we identify the following options for content chunking.

4.5.2.1 Content chunking at the application layer (basic solution)

This approach assumes that content chunks are visible only at the application layer. The COMET system is not aware of content chunks. Therefore, the COMET system invokes the resolution process, selects the best server and content delivery path for the whole duration of content delivery. Then, the application sends a request to the content server and gets information about content chunks. In this case, it may decide to download particular chunks progressively following the playback process. Moreover, the application may take advantage of scalable video coding by downloading chunks related to sub-streams conforming terminal capabilities and conditions.

The proposed approach can be directly applied without any modifications in the name resolution, content resolution and content retrieval processes. Consequently, it has no negative impact on the system scalability. On the other hand, this approach does not allow the application to download chunks from different servers. Therefore, a particular content server must store all chunks related to a given content. This limitation comes from the fact that the content server is selected at the beginning of the content consumption.

We recommend this approach for content published in Premium or Better Then Best Effort services. These services engage pro-active and reactive congestion control mechanisms preventing the overload conditions. Consequently, there is no need to change server during content consumption.

4.5.2.2 Content chunking controlled by COMET (interim solution)

This approach enhances the basic solution by allowing the COMET system to control the list of content chunks provided to the user application. In this approach, we assume that the content owner publishes a list of all chunks, e.g. using media presentation description (MPD) of the MPEG DASH standard [12]. When a content request arrives, the COMET system answers to the application with a selected subset of chunks. In this way, the COMET system may decide which server and routing path should be used for content delivery. The final decision about downloading the appropriate chunk is delegated to the user application. In this approach, the COMET system may fully control the server and path selection by providing a subset of chunks located on a single server. On the other hand, the COMET system may fully delegate the decision process to the application by providing the complete list of chunks. The latter approach is recommended for content published in the Best Effort service, where content delivery is out of the COMET control.

In this approach, the content names refer to the content objects, while chunk identifiers are application specific. Consequently, the resolution and retrieval processes are not impacted. The required enhancements correspond to implementation of proxies, e.g. MPEG DASH proxy, which is responsible for the processing and creation of appropriate subsets of available chunks.

4.5.2.3 Content chunks registered with distinct IDs (extreme solution)

This approach assumes that content chunks are registered in the COMET system with distinct content ID (the chunk IDs should be linked with the content ID). The COMET system performs an independent content resolution procedure to retrieve a particular content chunk. This approach gains from the adaptation of chunks delivery based on the current conditions of the server and the network. The main drawback is the explosion of the number of registered content IDs, which leads to scalability problems. Consequently, this approach may be applied for limited number of contents with rather small number of chunks.

4.5.2.4 Content chunks cached in consumer domain

In this approach, we assume that content chunks are cached locally in the consumers' domains. They are available for local consumers, so the client CME located in the consumer domain must be aware of cached content chunks. When a content request arrives to the CME:

1. If no chunks of content are cached in the local domain, the CME follows usual content resolution procedure.
2. If some (or all) chunks of content are locally cached, the CME takes into account local chunks of content and returns the user application modified MPD DASH file with a list of chunks available on the local cache and other content servers. For Premium and Better Then Best Effort consumers, the CME initiates path setup between a) local cache server and client for locally cached chunks of content and b) chosen CS in different domains and CC (for chunks of content that are not locally cached). For Best Effort customers, the CME fully delegates the decision process to the application by providing the list of chunks enhanced by chunks available locally.

4.5.3 Support for Detecting and Adapting to Changing Network Conditions

Basically, the COMET system has been designed to adapt to changing network conditions on the content request level (request level adaptation). The COMET system determines the best tuple: the content server and network routing path for each content request during content resolution stage and takes decision for the whole duration of content consumption. This decision process exploits multi-criteria optimisation algorithm based on the following parameters: 1) provisioned network parameters such as COMET CoS, values of long-term quality of service parameters, 2) current network conditions, i.e. current load/quality on routing paths, 3) current server conditions, i.e., server load, and 4) content transfer requirements and consumer CoS. However, the COMET system may be extended to support finer granularity adaptation, i.e. the COMET system may be extended with a function that changes on the fly the content delivery path during the content consumption allowing for adaptation to current network conditions. Note however that such adaptation function may lead to route oscillations and network instability.

In the following sections, we present how the COMET system may adapt content delivery paths to changing network conditions. The discussed approaches depend on exploited COMET CoS.

4.5.3.1 Adaptation in Premium service

In the Premium service, the COMET system applies preventive congestion control, which strictly controls the traffic carried on network paths by applying Admission Control (AC) function. In the COMET prototype, we exploit the Declaration Based Admission Control (DBAC), which uses information about available resources from SLA provisioning and makes decisions based on traffic descriptors of both currently running flows and new flows. Optionally, we can also apply the Measurement Based Admission Control (MBAC), which additionally uses information about carried traffic measured by edge CAFE at the content server site. In this way, the COMET system may reuse resources lost due to over-declaration and adapt to actual traffic conditions.

As the load on premium paths is fully controlled by the COMET system, the only reasons to modify them during content consumption are network failures. In this case, we can follow the recovery mechanisms discussed above. Thanks to the COMET forwarding mechanisms, the premium path could be switched on-the-fly without breaking down the transport connection to the content server (see details in the next section).

4.5.3.2 Adaptation in BtBE service

In Better than Best Effort (BtBE) service, the COMET system controls the network conditions on the basis of long-term provisioning. The adaptation to current network condition is possible by applying the measurement system, which detects congestion and triggers switching content delivery to offload congested paths. The monitoring system consists of monitoring agents associated with the edge CAFEs, as depicted in Figure 5. These monitoring agents measure: 1) CAFE load, i.e. forwarded traffic load, CPU load, memory load (currently implemented in the prototype), and 2) quality COMET BtBE paths (optional extension of software). The information about edge CAFE load is propagated to SNME and incorporated into the metric of describing

content server load. The information about quality of BtBE paths is propagated to CME and stored in Path Storage component.

Upon receiving the content request, the CME at the consumer site retrieves: 1) the combined edge CAFE and the content server load, and 2) the path description (including current path quality). This information is used during the content resolution process assuring request level adaptation for BtBE traffic.

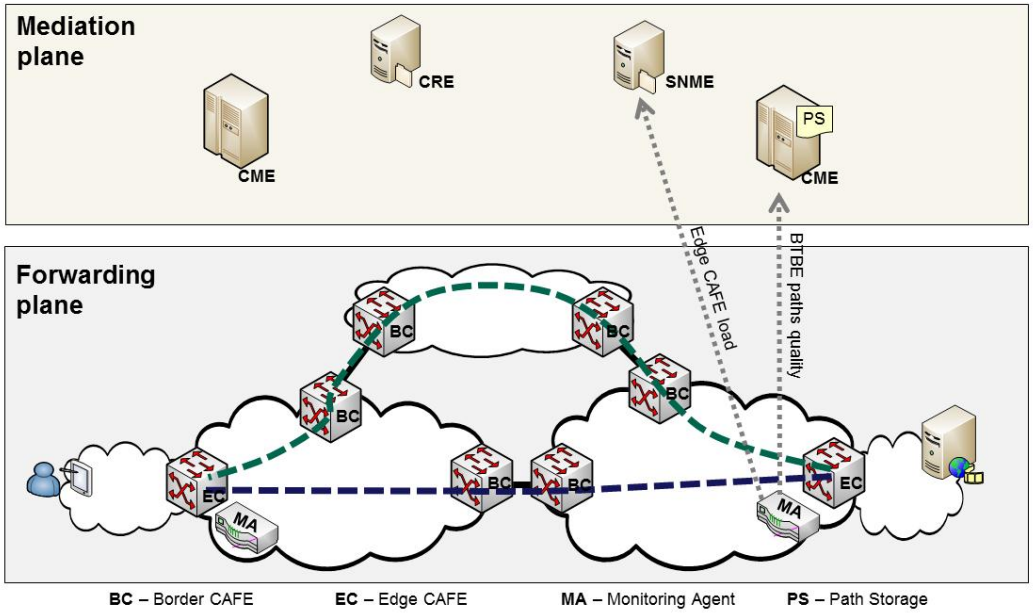


Figure 5: COMET monitoring system

The COMET system may be extended to support dynamic adaptation of BtBE paths to changing network conditions. When congestion or significant quality degradation on the BtBE path is detected, the CME at the server side may decide to switch some of the running flows to alternative paths. The path switching is performed on-the-fly under the control of the server side CME. This solution is feasible because the CME at the server side is:

- aware of all running flows. This information is kept by the edge CAFE, where a packet filter is configured for each running content consumption;
- aware of all alternative BtBE paths going towards consumer domain. This information is provided by the RAE and is stored in Path Storage component (jointly with current information about load conditions provided by measurement system);
- is able to run the decision algorithm and select the best alternative path;,
- is able to switch to the BtBE path for alternative path by reconfiguring the packet filter on the interception module of edge CAFE. A specific configuration command must be sent for each altered flow.

The BtBE path switching is performed transparently to both consumer and the content server without breaking down the TCP/UDP transport session. However, the on-the-fly path switching may cause temporary disturbances and packet reordering. In any case, this should not stand as a constraint since multimedia applications use playback buffer mechanisms and transport protocols that manage packet reordering, e.g. TCP, RTSP.

4.5.3.3 Adaptation in BE CoS

In Best Effort (BE) service, the content is delivered through the standard IP routing paths. The COMET system does not influence IP routing nor does it control the network resources. Therefore, adaptation to changing network conditions is delegated to the application.

4.6 Conclusions

The main advantage of the decoupled approach is that is provides an evolutionary solution for Content Mediation. The decoupled approach can be deployed over the existing Internet, without disrupting either current operations or services, as an enhanced network capability available for a reduced group of advanced users and later extensible to the general public.

This evolutionary quality has a strong repercussion in the following two issues, which facilitate future deployment and adoption.

1. As the COMET approach has been implemented, users do not need to modify their browsing habits, nor use a different set of tools for content retrieval and playing. The COMET Content Names are retrieved as specific COMET URLs and the COMET CC SW then simply registers the new comet protocol name (comet://) in the user's terminal, so that when a COMET URL is requested in the web browser, it is captured and resolved straightaway; in turn, when the CS connection data is returned to the CC, the application is configured by default in the terminal for the protocol/format specified in the connection data.
2. Likewise, CS operators do not need to modify their current CSs in operation, but just installing a COMET monitor which reports the CS status to SNME, as the only condition to join the COMET system, if deployed in their ISP.

More importantly, the decoupled approach entities are deployed in an ISP per ISP basis (except in the case of CRE). This means that the decoupled approach can organically grow from being deployed in a small number of interconnected ISPs, mediating between clients and contents in these ISPs, to reach the ever-growing expansion of the Internet, by adding new ISPs to the COMET club. The process to recruit a new ISP is as simple as deploying the COMET entities in the target ISP, configure the ISP's RAE to interact with neighbouring RAEs, so that paths to propagate across COMET realm, and register the ISP domains in the root and authoritative CREs, in order to allow content publication.

For the moment being, all the configuration operations in the COMET system are carried out by hand through web interfaces. Therefore, it remains an open issue to devise a procedure to feed this configuration information automatically into COMET from the ISP provisioning systems. Most importantly, this operation has be carried out in an transparent, flexible and backward compatible way, so that provision changes across the ISP are instantly reflected in COMET.

Concerning performance issues, our tests point at COMET being able to keep the Content Resolution Time (CRT) below the reference values, 2.5s, defined in D5.1 for the 95th percentile of the queries, if the load in the COMET entities is below 80 per cent. Performance can be improved by including in the CME design the use of caches, especially for the Content Records retrieved from the CRE and the CSs loads collected from the local SNME and/or remote CMEs.

A factor that impacts the resolution time is the number of remote CMEs that a CME needs to poll while resolving a content name. A possible improvement would be poll first those CMEs with a shorter round-trip time and stop polling when a critical mass of data allowing a CS/Path solution has been collected.

5 Coupled Approach

5.1 Overview of the Coupled Approach

In this Chapter, we discuss the coupled approach under the COMET architecture, as well as the *information flow* between the different functions and entities. This information flow is also shown in Figure 6, which includes all the required operations and functional component interactions from content publication to the COMET system, to content resolution and finally to path-setup for the content delivery.

In Figure 6, we illustrate a simple scenario where there are only two neighboring COMET-enabled domains with each having a single CRME respectively. In reality, additional backup CRMEs can be installed within each domain for reliability concerns. However, if this is not applicable, a reliable communication mechanism has also been introduced against potential failures of the single CRME (we give more details later on in this section). As described previously, the networking and server monitoring along with route awareness are done periodically in an out-of-band fashion. The publication and resolution processes, however, are triggered per content request.

We break the above figure in two, according to the COMET operations, namely the *content publication* (Figure 7) and *content consumption* (Figure 8). *Content consumption* is further divided into *content resolution* and *content delivery*. In Section 3.2, we discussed the responsibilities of each of the functions of the CMP and the CFP planes. Here, we discuss:

1. how the coupled content mediation approach implements the sequence of events,
2. how it handles the different functional blocks described before and
3. the interactions between different entities of the system.

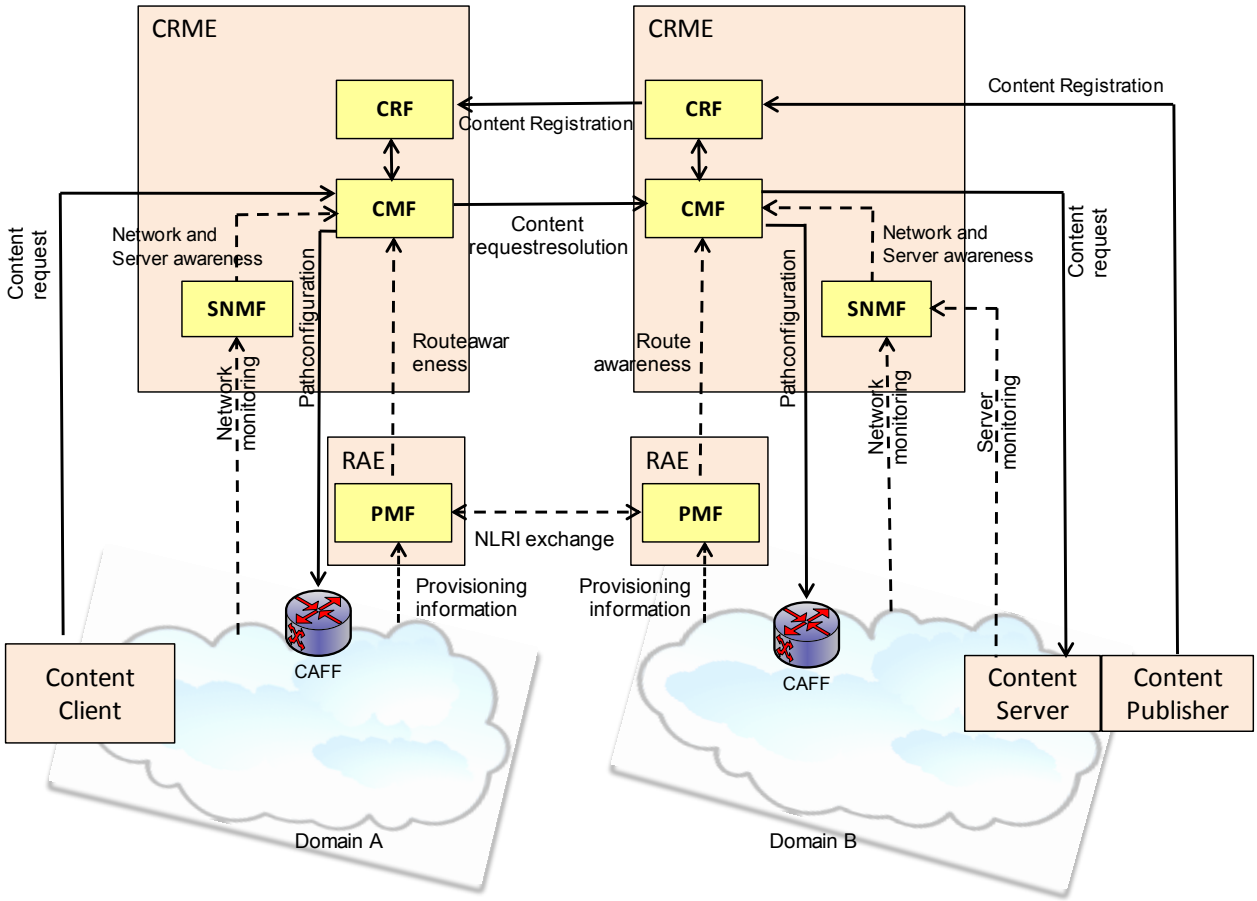


Figure 6. Coupled approach – Architecture and Information Flow

5.2 Content Publication in the Coupled Approach

According to the coupled approach, the preparation of the network entities for the content delivery starts during the resolution process. The general idea is to allow the business relationships (peer / customer / provider) amongst the domains (ISP networks) to dictate the publication path across domains. This is compatible with the current BGP routes, which are also configured according to inter-domain business relationships.

The first step of the content publication is done locally between the content publisher and the local Content Resolution Functional block (CRF) of the local CRME (step 1 in Figure 7). Basically, after uploading the content to the hosting content server, the content publisher issues a publish message to its local CRME (or immediate delegated CRME in the case where there is no local CRME). Upon reception of this message, the local CRME (specifically its CRF functional block) will create a new content record for this content, which includes the content identifier and the IP address of the local content server.

As a second step, the CRF block in the local CRME has to inform its counterpart in other CRMEs of the COMET system of the existence of this new content (step 2 in Figure 7). However, the propagation of this new content record is not done in a broadcast manner where all CRMEs in the entire Internet will know about the new content. This is due to scalability considerations. Rather, the approach defines specific publication rules that propagate the record to the ones that need to be informed, in order for all contents in the COMET system to be accessible from all content clients. Fundamentally, the publication is based on the communication between a chain of CRMEs located inside neighboring domains in a hop-by-hop manner according to the business relationship between them. The specifications of this process are described in detail in D3.2 [4].

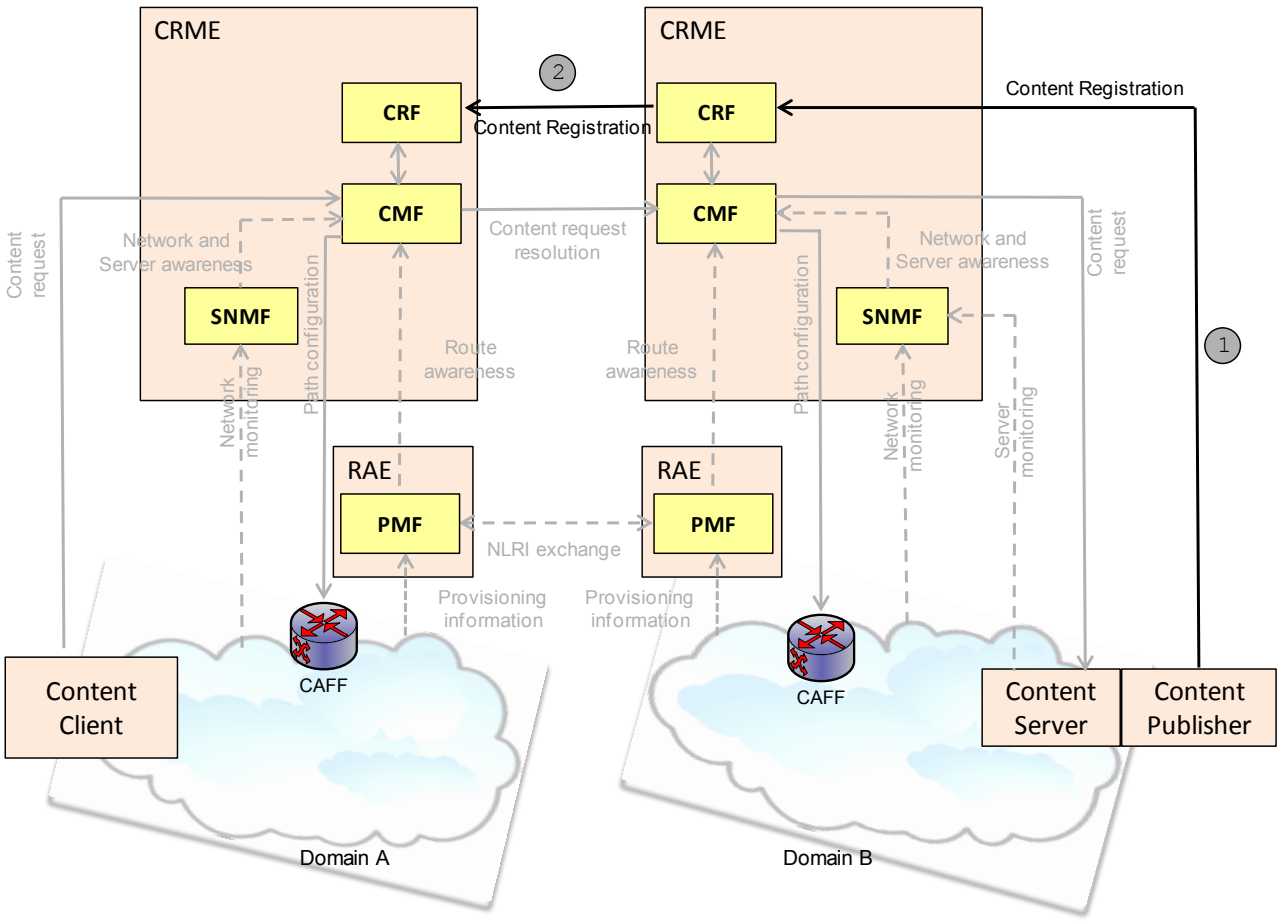


Figure 7. Coupled approach – Content Publication Information Flow

5.3 Content Resolution in the Coupled Approach

We further illustrate the sequence of events for the *content resolution* part of the *content consumption* in Figure 8, which consists of three distinct steps:

1) The content client sends its content request to the CMF block of the local CRME. During the *content resolution* process, the CMF consults the CRF (belonging to the same CRME) to resolve the received request. It will first check if the CRF block has the record of the requested content. Based on the result of this query, the CMF block will decide on the next resolution step.

2) The next step consists of two separate processes.

   i. The first process continues the resolution sub-operation. It is dependent on the answer obtained from the CRF. If a positive result is returned from the CRF, then either the implicit or the explicit location of the content is known. That is the local CRF block at the content client's domain knows either the exact location of the content, or the downstream path to follow in order to find the exact location. The CMF can then follow this information to find the targeted content server through hop-by-hop resolution operations across a chain of CRMEs towards the targeted source. On the other hand, if a negative result is returned from the CRF block, the CMF follows specific rules for the discovery of the content. Detailed procedures have been described in D3.2 [4].

   ii. The second process relates to the construction of the content delivery path in the local client's domain (i.e., Domain A in Figure 8). This is basically the preparatory stage that is required to couple the resolution and delivery processes. The CMF block gathers information about the available paths of the local domain (Domain A in Figure 8) from the PMF, and optionally the SNMF blocks, as discussed previously, and applies the required *Path Configuration* to the underlying paths (namely and ingress and the egress CAFEs within the local domain), or in other words, prepares the paths for the *content delivery in the client's domain*.

   This step is repeated until the request reaches the domain that hosts the content requested.

3) In this final step, the CRME of the domain where the content is actually located knows the explicit location of the content server (i.e., its IP address). Similar to the previous step, it has to *enforce the path*, but this time from the content server itself to the local egress CAFE. Then, it forwards the content request to the corresponding content server to initiate the transmission of the content. The content transfer itself takes place at the lower Content Forwarding Plane (CFP) through the chain of configured CAFEs. The specifics of these operations are detailed in D4.2 [6].

We note that in the above scenario (and in general in case the content client and the content server reside in different domains) the *Server and Network Monitoring Functional block* (SNMF) at the client's domain (Domain A, in Figure 8) provides *Network* information only, while at the server's domain (Domain B, in Figure 8) provides both *Network and Server* monitoring information.

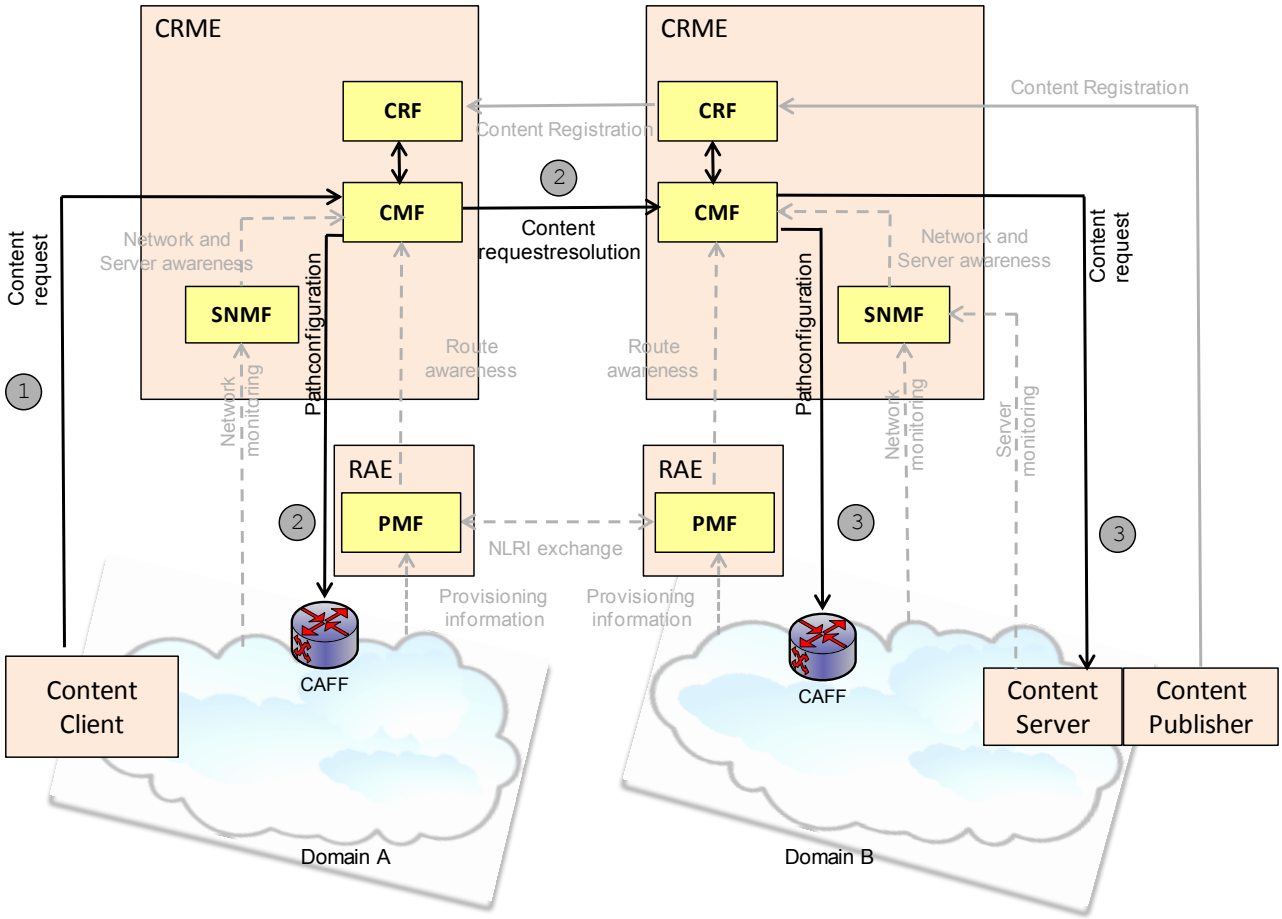


Figure 8. Coupled approach – Content Resolution Information Flow

5.4 Server-Awareness in the Coupled Approach

Content publication, server load updates and content resolution messages follow the provider route forwarding rule: each message is forwarded hierarchically to the provider domain(s) until the first tier-1 domain is reached. Content and server load records are shared among all tier-1 domains while lower tier domains only hold this information for themselves and their subordinate domains in the hierarchy. The rationale for this hierarchical structure is scale, given that higher-tier domains are more resourceful and able to support high-capacity CRMEs that will be able to cope with the full repository of Internet-wide content. Lower-tier domains are considered less resourceful and hence the amount of content and state information they keep relates to their position in the domain hierarchy. In strict consequence, lowest-tier leaf domains need only hold information pertaining to them. An additional reason for this organization in the resolution process is valley-free inter-domain routing [17]. Resolution messages reach tier-1 domains going upwards and then are forwarded downwards *only once*, based on information on server load and network distance that will identify the best possible source. This mode of operation is described in detail in [21].

In particular, server-load aware resolution is performed jointly by the CRF and CMF blocks of the functional architecture. Specifically, the CRF is responsible for identifying potential content server candidates, while the CMF is the actual decision maker to determine the actual target server by taking into account specific conditions such as server load information and network distance. In general, most server selection policies fall into one of the following two: (1) selection aims at achieving equal load distribution amongst servers and (2) selection strives to achieve equal average access delay at the group of servers. It has been shown in [16] that, in terms of average delay, both

approaches perform in increasingly similar fashion when the load increases. For further details regarding server-aware content resolution in the coupled approach, we refer the reader to [21].

5.5 Content Delivery in the Coupled Approach

The content delivery part of the content consumption involves the actual network (i.e., the data plane) and in particular the *Content-Aware Forwarding Functional block* (CAFF), as well as the newly introduced *Content-Aware Caching Functional* block (CACF) of the CFP. The CAFF and CACF blocks are implemented in the *Content-Aware Forwarding Entity* (CAFE). Basically, CAFEs are the physical entities that support the functionalities of the CFP.

The CAFE includes all the required mechanisms that guarantee smooth delivery of content back to the content client. The delivery path configuration mentioned in the previous section essentially means that the CMF installs specific content states on the relevant CAFEs within its own domain regarding the specific content request. The detailed description on the content delivery operations corresponding to this part of the COMET architecture is described in D4.2 [6]. Here, we note that we do not require all network routers to be CAFEs. We envision that the routers that are going to be enhanced with CAFF functionality are the ones that reside at the edges of the domain. As mentioned earlier these edge-domain CAFEs also implement the CACF and in co-operation with the CRME make decisions on which content should be cached in the network. The two approaches to in-network caching, namely the centrality-based approach and the probabilistic-based approach are described in more detail in [14], [15].

5.6 Extra Features of the Coupled Approach

5.6.1 Resilience Support

For the coupled approach, the delivery of content consumption requests can be based on reliable transport protocols between CRMEs. Regarding component failure such as CRME itself, one intuitive solution is to deploy backup CRMEs within each domain to provide shadow content publication/resolution services in case of the failure of the main CRME. While this is similar to the strategy of having multiple mirror DNS servers within each network, the key difference is that DNS information is very static, while content records in CRMEs are much more dynamic due to frequent content publications. As such, communication overhead for synchronization between the main CRME and its backups should be substantially higher than the DNS scenario.

Now we propose an alternative approach to overcome the limitation without necessarily deploying additional backup CRMEs. The basic idea is, when a CRME is not able to contact its next-hop counterpart across domains, it should be able to know the next-next-hop (NNH) CRME to continue the content publication/resolution process, so that the failed CRME in the middle can be seamlessly bypassed. To this end, during the bootstrap phase, each CRME should inform its neighbouring CRME about the location if its other neighbours, so that each CRME is aware of its counterpart in two hops away. On the other hand, we notice that there are different effects when CRME failures take place at specific stages such as content publication, resolution and delivery. Now we illustrate them separately using distinct examples.

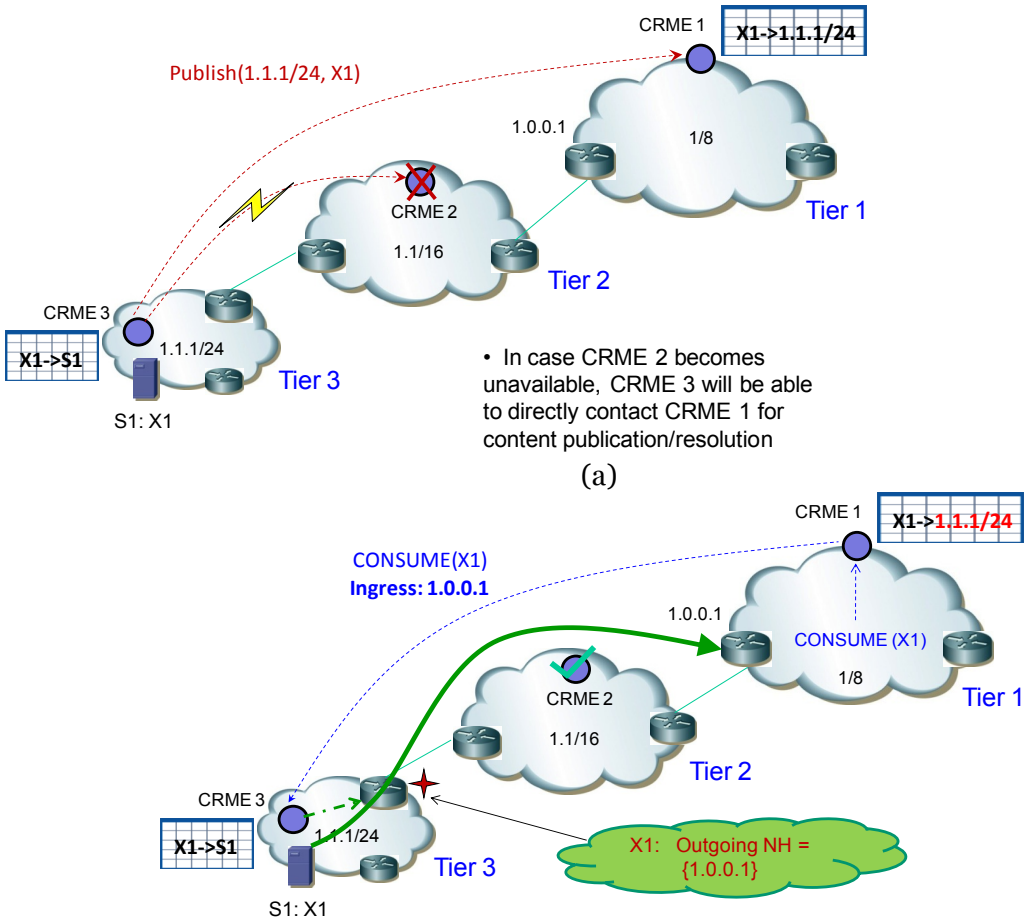
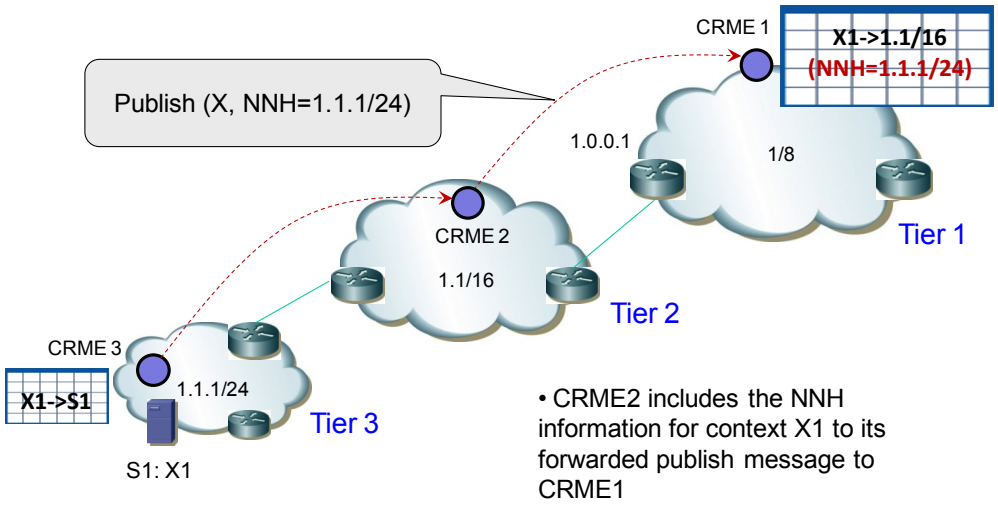


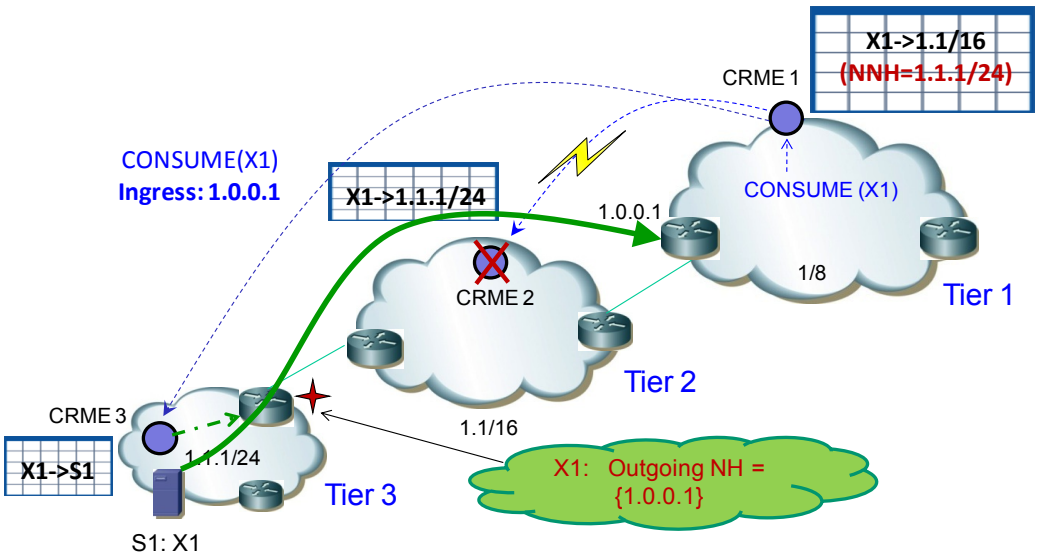
Figure 9. CRME failure protection during publication phase

Figure 9 shows the scenario where protection can be achieved during content publication phase. When CRME3 in the stub domain is not able to reach its provider-level CRME2, it will directly send the Publish message about content X1 to CRME1, which is two hops away. In this case CRME1 becomes aware that remote CRME3 has the content record for X1. Upon receiving a Consume request for X1, it will directly contact CRME3 for resolution while skipping CRME2 even if the latter has been recovered. In this special case the Consume message need to be revised to explicitly include the ingress CAFE address (in this example it is 1.0.0.1). This is because CRME3 needs to configure its local CAFE by installing the corresponding content state that needs to point to the remote domain through 1.0.0.1. This can be considered as a tunnel between the two remote domains that bypasses the middle domain whose CRME is not available. Effectively, no states are installed in the middle domain since the resolution has skipped CRME2.

Another scenario is the CRME failure during content resolution phase. This means the failure takes place after the content object has been successfully published. According to the basic coupled approach, if a CRME is not able to reach its default next-hop counterpart during resolution, the content consumption becomes unsuccessful. In order to make the system more robust, we extend the content publication protocol to carry NNH information, so that in case a CRME cannot reach its next hop due to its failure, it is able to continue the resolution by contacting its NNH.



(a)



(b)

Figure 10. CRME failure protection during resolution phase

In Figure 10, CRME2 makes a robust content publication by inserting the NNH information (1.1.1/24) for context X1 into the publish message that is sent to provide CRME1. As a result, CRME1 knows that, in case CRME2 in the middle is not available (or even is available), it can directly contact its NNH (CRME3) to finish the content resolution. In this case the Consume message should carry the ingress node 1.0.0.1 as the endpoint of the inter-domain tunnel to bypass the middle domain where CRME2 resides. As a result, when CRME3 receives the Consume message directly from CRME1, it configures its local egress CAFE to point to 1.0.0.1 for establishing the content delivery path.

5.6.2 Support for Content Chunking

To take advantage of in-network caching in the coupled approach, we introduce a slight modification to the requirements of Content IDs initially discussed in D3.2 [4]. That is, in order to avoid registering all content chunks in the COMET system, and therefore, raise scalability concerns, we extend the Content ID functionality discussed in D3.2 [4] as follows. Each unique Content ID is extended to include a sequence number for each of the content chunks it is fragmented to. For example, a content file with Content ID: *#some.file*, which is fragmented in 100

content chunks will comprise of chunks: *#some.file/c1* to *#some.file/c100*. The initial chunking operations on the entire content can normally take place at the original content server side.

These extensions to the main Content ID need to be universal, that is, all content providers have to adhere to the same semantics with regard to the specific label/ID extension, as well as the size of each chunk. We expect that as the field of Information-Centric Networks matures, the research community will form a set of guidelines for both the sequencing of content chunks and for their default size.

To support content chunking in the coupled content resolution approach, each of the Content Aware Forwarding Entities (CAFEs) makes use of the Content Aware Caching Function (CACF introduced earlier in this document) in order to perform the following functions in collaboration with its local CRME: i) check if the requested content chunk is stored in its local cache, ii) forward the requested chunk, in case it is found locally and inform the server not to transmit it, iii) forward the request onwards, in case the chunk is not found in the local cache. This procedure taking into account the related COMET entities is shown in Fig. 11.

In particular, the above functions are realised as follows. After the initial content resolution for the requested content through its Content ID (as specified in D3.2[4] and [13]), the end-user is required to periodically send subsequent requests for content-chunks of the same file. For scalability reasons, we avoid sending one request for each chunk of the file, but instead, we group content chunks into *windows* of fixed size. Therefore, the user sends one request for each subsequent window of chunks during his on-going content consumption session. Content state installation at the CAFEs follows the normal process described in D3.2[4] and [13] and is done during the initial content resolution, but it needs to be renewed upon subsequent requests. Upon each request for the next window of chunks all the CAFEs involved in this session are informed by the local CRMEs of the next few chunks that the user is requesting. Therefore, in case any of the CAFEs along the path has chunks within the incoming window range stored in its local cache, it notifies the local CRME and forwards the chunks back to the user (making use of the state installed in all other CAFEs down the delivery path – see also Fig. 11 below). In turn, to avoid redundant transmission of content, the local CRME of the CAFE, where the cache hit happened notifies the server to avoid transmission for the next window of chunks that are locally cached. We note that this message does not remove the state from the intermediate CAFEs. Instead, state is removed based on some timeout implementation. That is, if no subsequent requests have been received for a given number of window intervals, then the state is torn down.

It is worth noting that although content chunks do not need to be registered in the COMET system, the CRMEs can read chunk numbers. Content caching on the delivery path follows one of the algorithms proposed and evaluated in the context of COMET in [14] and [15]. However, the decision on whether to cache an incoming chunk or not is taken at the CRME, which then informs its local CAFEs accordingly. In this way, upon receiving an incoming request carrying a specific window ID, the CAFE is able to identify those locally cached chunks that belong to that window interval. In this case, the CRME is querying the local CAFE to see whether the content chunk that it has previously been instructed to cache is still held locally, or has been discarded. The CAFE replies with a positive or negative answer and the CRME constructs the corresponding message to be included in the related content request to be sent back to the server (Fig. 11).

The specific implementation of the timeout algorithm (to tear down the state from CAFEs) is a subject of further investigation and is outside the scope of the current study. Furthermore, the size of the *window* is also a factor that warranties further research. For example, a big window may have more hits for cached chunks, but those chunks may be far away from what a user wants at the time of request. As an example, imagine a YouTube video, where the user is requesting for the chunks of the 10th minute, but because of the big window, the CAFE is forwarding chunks for up to minute 25. It has been shown that users do not always watch the whole duration of the video, hence, transmission in this case results in waste of network resources. On the other, hand, single-chunk windows will increase both the frequency of requests from the user and the communication between CRMEs and CAFEs, something that brings up scalability concerns.

We note that content chunking and caching on the delivery path follows one of the algorithms proposed and evaluated in the context of COMET in [14] and [15].

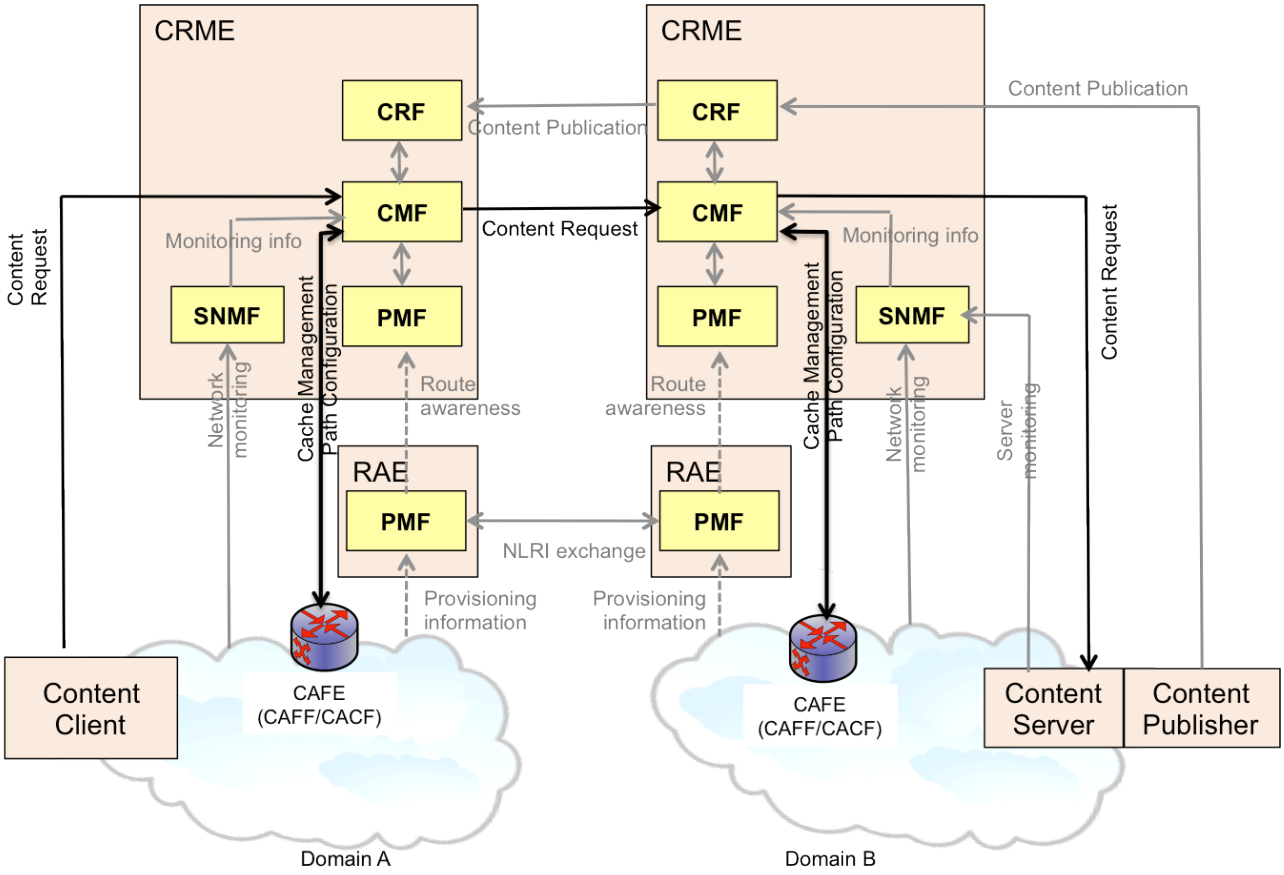


Figure 11. In-Network Content Caching for the Coupled Approach with Content Chunking

5.6.3 Support for Detecting and Adapting to Changing Network Conditions

For the coupled approach, when the SNMF detects congestion or significant degradation of the link quality, it will inform the CMF. The operation by CRMEs for adapting to this new network condition depends on whether the AS is multi-homed or not.

In the case where the AS is multi-homed, the affected CRME will initiate new content requests for affected content delivery sessions to establish new delivery paths with the alternative inter-domain link(s). We illustrate this with Figure 12.

1. Assume that an active content delivery session is established between A.A.A.A and A.A.B (see the solid red line).
2. At a certain point in time while this delivery session is still on-going, the inter-domain link between A.B and A.A.A degrades (e.g., suffers from congestion).
3. Since A.A.A is multi-homed, the CRME at A.A.A will initiate a new content request via it's alternative provider following the same provider route-forwarding rule defined in the coupled approach (see the dashed blue line). Note that if there is more than one alternative providers (e.g., if A.A.A is multi-homed to three or more domains), then the CRME can apply local policies to decide on which alternative inter-domain link(s) to forward the new content request.
4. Using this mechanism, the new resolution may result in finding a different source to serve the content request. For example, while the delivery is in progress, a new copy of the

content is stored/cached at A.A.C. Then, in this case, when A.A.A initiates a new request, it will resolve to A.A.C (see the dotted green line).

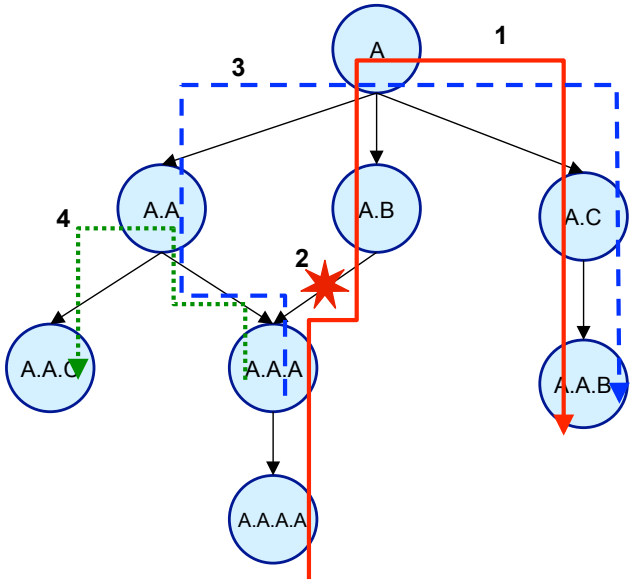


Figure 12: Multi-homed AS

In the case where the AS is single-homed, we propose a "crank back" mechanism. We illustrate this in Figure 13.

1. Assume that an active content delivery session is in progress from A.A.A.A via A (see the solid red line).
2. At a certain point in time while this delivery session is still on-going, the inter-domain link at A degrades (e.g., suffers from congestion).
3. Since A is not multi-homed, the CRME at A does not have an alternative provider to shift the content delivery session. It will then propagate a *notify* message to the previous hop. Since A.A is also single-homed, it repeats the same crank back operation by forwarding the *notify* message (see the dashed blue lines).
4. A.A.A is multi-homed. The CRME here will follow the procedure described above by initiating a new content request at the alternative provider link that is not affected (See the dotted green line).

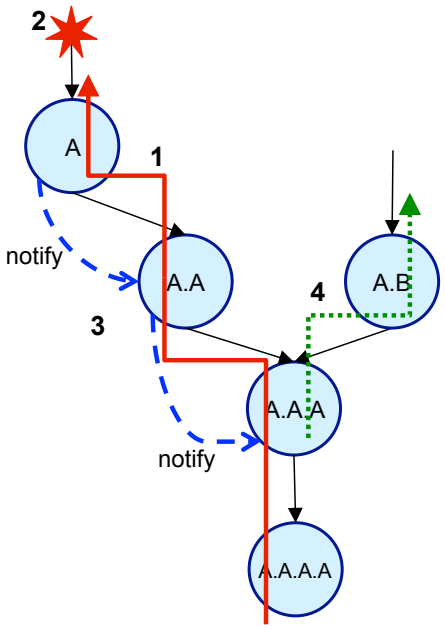


Figure 13: Crank back mechanism

The worst-case scenario is when the *notify* message reaches the client. In this case, the client will have to initiate a new content request to establish a completely new delivery path.

Note that in contrast with the decoupled approach, such scheme for the coupled approach does not require multi-path routing. While in the decoupled approach, the adaptation attempts to find an alternative path to the original content source, the scheme described here may identify an alternative content source whose associated path will not include the network component that suffers from local QoS deterioration.

5.7 Conclusions

In this section we make some key conclusions on the coupled approach. A number of items below are derived from our original design strategies/principles, while others are concluded based on our simulation-based experiments. More detailed elaborations on the simulation results for the coupled approach are included in D5.2 [9].

- The content resolution operation in the coupled approach is based on gossip-like communications hop-by-hop across a chain of domains. While such a feature allows intermediate domains to flexibly participate in content resolution according to own policies/conditions, intuitively this introduces longer content access time. Nevertheless, due to the nature of the power-low distribution of the Internet AS-level topology, content resolution messages are not expected to traverse a large number of domains before hitting the content source. According to our simulation-based experiment, the average number of AS hops is below 4 domains. On the other hand, the coupled nature of resolution-delivery/path saves one round-trip: path setup takes place during the resolution of the content, which means one round-trip less as compared to the common practice of DNS-based resolution operations.
- According to our evaluations based on the CAIDA topology dataset, both the random and the broadcast resolution modes produce similar results (in terms of hop count for the resolution path). Effectively this means the random resolution mode is more desired since it substantially reduces the communication overhead in terms of the number of content resolution requests across domains.

- The scalability of the coupled approach presented here is not designed to support the whole population of the content objects that are distributed in the Internet. The design principle can be regarded as a more sophisticated paradigm to enable Internet-scale point-to-multipoint content distribution. Such an approach is ideal for supporting extremely popular content objects where content consumers are distributed across a wide range of autonomous domains in the Internet. As such, the content records maintained on the CRME side and the content states installed at CAFEs become "cheaper" due to large number of consumers. From this point of view, the coupled approach is an ideal solution for supporting inter-domain multicast, overcoming both the technical and the business barriers that have limited its adoption to date.
- The coupled approach supports dissemination of server-load awareness across domains in a fully distributed manner across autonomous domains. In the conventional CDN-like environment, the operator of the CDN infrastructure is able to perform monitoring of the content servers in order to make sure content requests are optimally directed to specific locations. Such a function is not possible in the ICN environment, since there is no central entity that is responsible for such operation. In the coupled approach, server load context information is systematically disseminated in an out-of-band manner across CRMEs in order for them to make optimised resolution decisions. As far as our design is concerned, while it is not possible to achieve globally optimal content resolution in terms of balancing the server load due to scalability issues, our proposed dissemination mechanism certainly benefits in such a feature across local interconnected ISP networks.
- Upon the completion of initial content resolution operations, BGP based AS-path optimisation is possible since the content source has been identified. Intuitively, the best post-optimisation performance that can be achieved is the same as the result based on the routes advertised by BGP in the host-to-host model. However, it should be noted that the initial content resolution might lead to large number of AS hops between the identified source and the consumer in the first place. According to our simulation results, routing optimisation is able to achieve on average 40% performance gains in reducing the AS path length from the resolved content source back to the consumer.

6 COMET System Validation

6.1 Global Requirements

- ***"The content must be treated as a primitive itself.*** *The architecture must be oriented to deal with all aspects of content natively, facilitating the access and distribution of contents. Support for safe, based on trusted content publication, friendly and fast content retrieval for consumers through the COMET architecture and mediation functionalities is required."*
    - On the one hand, the entry point to COMET for consumers via the CMF is the Content Name, which effectively abstracts them from all the content characteristics, like location, protocols or connection means. Content characteristics as such are stored within the COMET System, and used across the entire system, in order to provide an optimal path and ensure QoS all the way long. Therefore, since COMET considers content characteristics as a central element of its working procedures, it is clear that content is treated as a primitive.
    - On the other hand, COMET cannot warrant the authenticity of the contents made accessible through it, as they remain stored in external servers and COMET only manages the content information/metadata. However, this content metadata provided by publishers is stored in specific COMET functions, the CRFs, which are protected against intruders/corruption, and which the other COMET functions can trust. Regarding ease of use, COMET maintains the set of tools (browser/players) with which a user normally interacts. A new protocol identifier, comet://, has been defined (see D5.1 [8]), so that the user only needs to write this new COMET URLs in the browser to retrieve COMET contents (after registering this protocol in its terminal). These COMET URLs are identified and intercepted by the COMET CC, and after resolution, played by the user's default tools. In terms of speed, the scalability studies in D5.2 [9] prove that COMET Mean Response Time for Content Resolution is below the reference values (2s) for 95% of the queries. So COMET will not worsen the current Internet experience of a normal user.
- ***"A global content naming and addressing scheme should be supported by an infrastructure capable of scalable content search and resolution.*** *The global content-aware mechanisms must be able to handle efficiently large amounts of content, being able to support significantly more objects than those handled by today's largest Content Distributors (YouTube, Flickr, Apple Store, for example). The protocols to be developed by the project will be capable of scaling to the order of billions (10^9) of content objects."*
    - As such, COMET does not store the actual contents, which remains in their current storage location, but the content metadata/information and the Content Names. This means that the information managed by COMET is much lighter than the actual contents, and comparable in the case of the Content Names to the URLs exchanged in the current Internet. The worst case would be when the actual content information/metadata stored in the Content Record is exchanged, but again these pieces of information are in the order of dozens of URLs, almost negligible if compared with the average size of the contents currently exchanged through the Internet. Basically a Content Record can be modelled as a set of download URLs, as many as CS, with QoS parameters attached to groups of URLs. So a typical Content Record is in the order of average URL length x number of CS in the Content Record.
    - In the decoupled approach, the hierarchical structure of the CRE enables the deployment of new authoritative CREs for new domains, so the system will be able to escalate as the amount of domains grows. Since the CREs store content metadata/information and not the actual contents, the test in D6.2 [11] have shown

that for a 1GB disk storage, at least 10^6 Content Names can be stored. Therefore, the target figure of 10^9 Content Names is well within the COMET manageable range.

- The coupled approach will apply only to popular content as it advertises subordinate domain content all the way up to tier-1 domains (unscoped content publication). Content popularity could be monitored, and as soon as content is deemed popular, it should revert from the decoupled to the coupled approach, with the relevant ID advertised all the way up to the tier 1 level. In addition, content providers expecting some content to be highly popular (e.g., Olympic games coverage, etc.) could choose to publish their content through the coupled approach in the first place. The content mediation plane needs to also keep globally track of which content has become popular, so that resolution and delivery for a requested content follow the appropriate approach. Content may also lose its popularity status, with access reverting back to the decoupled approach. At the CFP, content states at CAFEs follow the same usage as IP multicast addresses and thus, it will not incur more complexity.

- ***"The COMET system should be open for future evolution of the Internet.*** *This can be achieved by the modularity in the design of different components and with a flexible high-level architecture."*
  - All the COMET functions and entities are by design open for extension or enhancement in order to accommodate future evolution of the Internet. The design of all functions and components is generic and flexible enough to be easily extended in order to accommodate more functionalities.
  - In particular, it has been the aim of COMET to avoid make assumptions about how contents are distributed, transported and consumed through the Internet. Therefore, server managers do not need to modify their usual means of distribution nor consumers dispose off their usual means of retrieving/playing contents. This means that novel means of content distribution (i.e. protocols) can be inserted and managed by the COMET system without almost any changes.
  - For instance, in the decoupled approach, the concept of content source allows describing the CS characteristics in its entirety, so they can be used in configuring any required forwarding elements, like the CAFEs. Besides, the CC intercepts any COMET queries and hands over the connection data to the application defined by default in the user terminal, being, so to speak, agnostic to the involved protocols.

- ***"Support for gradual and economical embracement of the COMET system by ISPs.*** *The designed architecture for content mediation and the associated mechanisms for content discovery, resolution and access must be scalable to be deployed in the largest ISPs, consisting of the order of hundreds of point of presence (PoPs) and core routers. These mechanisms and protocols should be applicable for content distribution at Internet-scale, involving autonomous networks of the order of tens of thousands of ASs."*
  - The decoupled approach is specifically designed for an ISP per ISP deployment. In that respect, the name resolution system embodied in the CRE hierarchy is easily scalable by the straightforward expedient of deploying new authoritative CREs to serve new domains. Inside a single ISP, the SNME has been designed with the specific aim of enabling the growth of the CS population served by the ISP, just by allocating new Server Information Collector (SIC) SME sub-entities (as explained in D3.2 [4]) in individual dedicated servers. Last, the COMET architecture enables the deployment of dedicated CAFE entities in the ISP PoPs for their traffic management. Thus, the COMET system can grow as the ISP does, without losing pace. Note that CAFEs, although similar to routers, are not intended to substitute the ISP routers. They should be deployed as an overlay over the existing ISP infrastructure at the server domains and on the edges/borders of the ISP in order to allow COMET

system to optimise content delivery. Therefore, the number of CAFEs will be well below the number of routers managed by the ISP.

- The coupled approach, being the more disruptive approach, will require federated cluster of domains as the initial deployment. Then, gradual per ISP participation can be supported by linking to this initial cluster of participating domains.

- *"The content-aware mechanisms designed and developed for the network, when orchestrated by novel Content Mediation Plane (CMP) algorithms and protocols,* ***should facilitate the involvement of, potentially, all Internet users as Content Creators****. Thereby, creating the opportunity of a new, all-encompassing market where millions of small, medium and large Content Providers have access to efficient content distribution capabilities to reach billions of potential Content Consumers, taking advantage of a reduction of required resources, mainly bandwidth and processing capacity."*
  - In principle, COMET does not discriminate between users (e.g., individual users vs big corporations). As long as they have a registered account in a Content Publisher, they can publish their content metadata and obtain a Content Name. In turn, the Content Name automatically enables content retrieval to any consumer with COMET Content Client SW installed in its terminal. COMET will not provide for content storage space, though, which will have to be furnished externally by the publisher.
- *"The COMET system will support handover mechanisms which allow* ***a graceful switching of the content delivery path without impact on the application-layer****."*
  - Even though hand over mechanisms have not been included in the final COMET implementation nor selected for the set of use-cases tested in the scope of WP5, nothing precludes COMET from incorporating this functionality, which basically consists in issuing a new resolution request and obtaining a download server/retrieval path as already functioning in COMET.

6.2 Content Consumer and Client Requirements

- ***"Access to the contents must be independent from the content location.*** *The naming architecture should guarantee location-independence, which in turn would guarantee smooth transition from today's host-centric to a future content-centric Internet."*
  - Content in COMET is accessed using *Content Names*. The structure of both of these identifiers is not based on the content location. The actual location of the content is resolved within the COMET system and in particular, in the CRF, which makes content resolution *location-independent*.
- ***"The content identifier must be the same for different ways of distribution and nature of the content.*** *Also, different copies of content will be identified by the same Content-ID. It is, however, responsibility of the Content Providers to explicitly register the new copy of the content as such."*
  - All copies of the same content are indeed identified by one Content. The specifics of these operations are given in D3.2 [4]. All the parameters associated to a content, e.g., COMET CoS, QoS requirements, transmission protocols, server locations, etc., are provided by the content publisher at the time of the registration of the new content.
- ***"The Content Consumers must access the content in the same way as in current Internet*** *i.e. achieving user unawareness."*

- Both approaches in the COMET architecture allow "click-to-consume" user interfaces to access the content in the same way as in current Internet with hyperlinks.
- In the decoupled approach, the "click-to-consume" is embodied by the user writing a COMET URL in its browser, this COMET URL will be translated by the COMET entities (i.e., the CME) into a server connection data which will be automatically transferred back to the default application for consuming this content. The subsequent content request will then reach the CS and the content downloaded through the path set up by COMET. This approach follows a two round-trip approach, with two requests from the Content Client, the first one to get the content properties and the second one to perform the application-level request. Nevertheless, this whole process of two requests is transparent for the end user and perceived as a one-step procedure.
- The coupled approach, by design, requires only one round-trip to resolve a content request and will actually enhance the user experience and thus, will not adversely affect users' content consumption experience.

- *"The Content Client could optionally declare its capabilities during content resolution phase, but **it is up to the COMET system to decide how to deliver the content to the Content Consumer**."*
  - The decision on how to deliver the content is the job of the CMF. For the *Decoupled Approach*, the CME, based on the information that it gathers from the CRE (content properties), RAE (offline, long-term path condition) and SNME (online, near real-time network and server condition), it makes the decision of how to deliver the content to the client. The path(s) is then configured accordingly.
  - For the *Coupled Approach*, the delivery path is built while resolving the content request and thus, based on the business relationships between the intermediate ASes (possibly gathered from PMF). The path optimization function in the *Coupled Approach* may use the information from SNMF.
- ***"The Content Client will obtain all the parameters necessary to invoke the application level requests."***
  - In the decoupled approach, there is no need for modification of current applications in order to fit into the COMET architecture. l. The only requirement to properly interact with COMET is that the comet protocol must have been registered in the user terminal, by means of installing the COMET CC Client, which will intercept COMET URLs and return the connection data (server location, protocol, port, MIME types) to the player application configured by default in the user terminal.
  - In the coupled approach, the two round-trips are combined into one, thus forcing some kind of adaptation in the application or transport protocols to provide to the Content Server enough flexibility to deliver the content.

6.3 Content Provider and Server Requirements

- *"**There must be an interface that allows the Content Providers to update the content properties** (content location, server load, way of distribution, etc.)"*
  - For the decoupled as Explained in D3.2 [4] the Content Publisher not only allows basic edition operation (create, update, modify, delete) in the CREs, but will enable the publisher to characterise the contents by defining the QoS requirements the forwarding plane must honour and their means of distribution (location, protocols, ports). Server load is not stored in the CREs but in the SNME, since it is a transient system property which cannot be fixed in advance. The CME is the element in

COMET which will map the server information stored in the CREs with the load in the SNME.

- For the coupled approach, there are dedicated primitives for content providers to publish and update their content (including the options for filtering and scoping publication). In addition, CRMEs can also disseminate server load information in order to support optimised content resolution according to server load balancing objectives.

- *"**The Content Provider should be able to establish policies** to enforce the way to publish and deliver the contents to the Content Consumers."*
  - In the decoupled, when publishing a content, the publisher can assign different Comet CoS to different set of servers Content Sources, so some of the servers can be reserved for those user with higher CoS (in other words, better subscribed SLAs). The publisher can also impose specific QoS requirements per Content Source, so when a path for a content distributed from those servers is set up, the entities in the forwarding plane (the CAFEs) accept those requirements (e.g. by choosing a network path labelled as Pr in case of a Pr Content). Finally, the publisher can also define priorities for Content Sources with the same CoS, discriminating, for instance, among different means of content distribution.
  - The coupled approach includes dedicated scoping and filtering functions for publishing contents and thus, content providers can dictate / restrict the access of their published content.

6.4 Content Mediation Requirements (CMP)

- *"**There must exist a global content resolution architecture** for efficient and scalable name and content resolution."*
  - This is fully covered by the unified content access intelligence added to the COMET system by the two content resolution approaches. In particular, the CMF with the help of CRF can guarantee global content resolution. The details are given in D3.2 [4].
  - In the case of the decoupled, the CRE are organised in a DNS-like way, which allows easy growth as the amount of contents increases. The central content resolution element in the decoupled is the CME which are located on each ISP, allowing new ISP to join COMET by deploying their own CMEs, RAEs and SNMEs.
  - In the coupled scenario, content resolution is performed in a hop-by-hop manner across a chain of CRMEs based on the business relationship between neighbouring domains. According to our evaluations based on datasets with real domain-level Internet topologies, the average resolution hops is very small thanks to the power-law property of the Internet topology.

- *"**There should be an integrated traffic and resource management solution compatible with the content resolution architecture** to increase network efficiency and content delivery in order to reduce network congestion on the most highly loaded links."*
  - This requirement is covered by the cooperation of a couple of functions: PMF supplies the CMF with offline info about the availability of underlying paths. The SNMF, on the other hand, deals with near-real time, online information. Finally, the CMF makes the decision taking into account both sources of information and configures the corresponding delivery paths.
  - In the case of the decoupled, the multi-criteria decision algorithm was designed to optimize network efficiency and reduce congestion. The performance evaluation of

the proposed algorithm is presented in D5.2 [9]. Moreover, the concept of COMET CoS will help save specific links for traffic with stringent requirements, avoiding them being clogged by rogue traffic like P2P.

- In the case of the coupled, traffic and resource management functionalities are gracefully embedded in the hop-by-hop content resolution. A typical example is that a multi-homed domain may strategically forward content resolution requests to one of its provider networks in order to achieve content traffic load balancing on ingress CAFEs.

- ***"There should be an information gathering system in the CMP*** *for collection of various performance metrics on networks and servers. This is going to be implemented in the COMET Monitoring Module."*
  - The monitoring module that gathers online, near real-time information is realized in the SNMF, which gathers network and server monitoring information and feeds it to the CMF.
  - Specifically, in the case of the decoupled, the SNME supplies the CME with a consolidated measurement of a CS status (and the CAFEs serving the CS). This instructs the CMP to avoided those CS (and CAFEs) which are overloaded.
  - In the case of the coupled approach, the CRME within each domain may gather necessary server and network condition (e.g. traffic load on inter-domain links which are generally regarded as bottlenecks) based on which optimised decisions can be made for content resolution; this also affects decisions for the content delivery.

- ***"The protocol interfaces between the CMP and the Content Providers, publishers and end user devices must be efficient.*** *Specifically, the user terminals should be able to send their content consumption requests through well-defined common interfaces, and the Content Providers must announce their server condition and the information about the contents they publish using a common set of interfaces. To complete this requirement, some others have been extracted from the use cases:"*
  - Specific interfaces have been devised for those three tasks a) content request from user terminals, b) content publication and c) server load report.
  - In the decoupled, interfaces a) and c) have implemented by using IP datagrams, which provides a fast and compact method of sending small amounts of data (the Content Name and the CSs consolidated load) without wasting time in setting up a TCP connection, with the trade-off of tolerating packet loss, which can be recovered by simple retransmission. Interface b) has been implemented by means of a Web Interface, to ensure that no partial data concerning the content information is stored in the CRE because of network failures and to enable encapsulating it into secure protocol layers, like SSL.
  - For the coupled approach, both content publication (a) and consumption (b) requests are sent to the local CRME belonging to the same domain. On the resolution side, thanks to the "one-stage" content resolution operation, a content request from a consumer directly triggers the content consumption process (including finding content source and configuring delivery path), which is in contrast to the "two-stage" operations based DNS services. Regarding server load report (c), each CRME periodically gathers such information from local servers, and such information is efficiently aggregated for propagation across multiple domains in order to achieve inter-domain server-load aware content resolution.

- ***"The CMP must be able to dynamically modify the information related to the location of the servers in the content record."***

- Such information, once provided by the Content Publisher, is dealt with either by making use of specific "UPDATE" messages sent to the CRE, in the case of the decoupled, or by new "publish" commands in the case of the coupled as described in D3.2 [4].

- ***"The COMET system must offer to the Content Provider the possibility of registering different ways of distribution."***
  - This information has to be provided by the Content Publisher and is included in the content properties of the content. In the case of the decoupled approach, the Content Source allows grouping CSs that share the same distribution modes, and encapsulating these different sources in a single Content Record identified by a unique Content Name.
  - In the case of the coupled approach, both anycast and native multicast can be supported by the system. Broadcast mode of operation can also be supported, but we have not studied this distribution mode in detail.

- ***"The CMP in an ISP must be aware of network conditions*** *in order to take decisions oriented to reduce the latency in content retrieval that is due to network failures, network congestion or server load."*
  - There are two functionalities that gather network information and feed it into the CMF so that this element can choose the right delivery paths accordingly. These are the PMF, which gathers offline, long-term information, and the SNMF, which gathers near real-time information. Network failures, network congestion and server load information relate mainly to near real-time conditions that are provided by the SNMF block.
  - In both the decoupled and the coupled approaches, the SNME provides info about the CSs load, allowing the CME/CRME to avoid overloaded CSs and the status of the edge CAFEs serving them, so those overloaded CAFEs can be also avoided.

- ***"There should be specific interaction between the Content Mediation Plane and the Content Forwarding plane to enforce content delivery."***
  - The SNMF and the PMF provide the required information to the CMF, which in turn enforces the underlying paths accordingly and prepare the content delivery. The functionality of the CFP that receives these rules and enforces them in the network to prepare the content delivery is the CAFF.
  - For both the decoupled and the decoupled approaches, the CME/CRME will configure the edge CAFE serving a CS with the path in key format, which enables content retrieval from the requesting CC.

- ***"The CMP, upon the content request from a user device, should be able to request capabilities to enhance or facilitate the QoS and multicast in the network*** *for the delivery of that content to that user device."*
  - As previously stated, the Content Mediation Plane through the CMF function takes decisions about the path and prepares it for the content delivery. On the Content Forwarding Plane side, the CAFF, after gathering the required info from the CMF, regulates traffic according to the rules from the CMF.
  - The coupled approach mainly focuses on the enabling of content multicasting across multiple autonomous domains. The support of inter-domain multicast has been achieved through the joint operation in both the CMP and CFP. During the content resolution phase, CRMEs immediately install content states at the ingress and egress CAFEs such that they already know where to forward the content once it has been injected to the local network. Thanks to the states maintained at CAFEs at the

network edge, an inter-domain multicast tree can be constructed upon the resolution operations for multiple requests targeting the same content object.

6.5 Content Delivery Requirements (CFP)

- ***"There must be a content forwarding architecture able to perform content-based forwarding at speeds similar to the ones in IP-based forwarding."***
    - The content-based forwarding is performed by CAFE, which combines source routing with tag switching paradigms to assure flexible path selection and reduce routing tables. The CAFE prototype was implemented in Linux kernel to assure efficient content forwarding (details about implementation of CAFE are provided in D4.3 [7]). In order to evaluate CAFE performance, we followed methodology proposed in RFC 2544 and compare lossless throughput measured under the same conditions for both CAFE and software IP router. The results reported in D6.2 [11] confirmed that CAFE can forward content at the same speed as IP-router.
- *"The elements in the CFP should support* ***QoS-aware content delivery****."*
    - The QoS-aware content delivery in CFP is provided by CAFE nodes. The CAFEs exploits traffic control mechanisms to differentiate traffic handled in COMET. These mechanisms such as classifiers, policers, shapers, schedulers are supported in Linux kernel and can be directly used for COMET packets. The QoS-aware content delivery is controlled by CME, which enforces QoS-related rules in the edge CAFEs during path configuration process. Details on the operations carried out therein are given in D3.2 [4] and D4.2 [6]. Moreover, the results of functional tests are presented in D6.2 [11].
- *"The elements in the CFP should support* ***point-to-multipoint content delivery****."*
    - Point-to-multipoint content delivery is natively supported by the *Coupled Approach* described in D3.2 [4] and D4.2 [6]. For the *Decoupled Approach*, the point-to-multipoint feature is provided by the Content Streaming Relay (CSR). The specification of CSR, implementation details and reference scenarios are provided in D5.1 [8]. Moreover, the results of functional tests are presented in D6.2 [11]. As mentioned above, the coupled approach is able to support point-to-multipoint content delivery at the inter-domain level.
- ***"Content may be cached in the network*** *to optimise network resource usage."*
    - In-network caching is natively supported by the *Coupled Approach* described in D3.2 [4] and D4.2 [6]. The performance evaluation of in-network caching is presented in D5.2 [9].
    - In the decoupled approach, the content caching may be directly applied in the consumer domains as it is presented in section 4.5.2. For the coupled approach, content caching can be applied in all possible domains in the content delivery chain. This is a feature that nicely works together with multicasting in which any intermediate CAFE can cache content and provide local serve to multiple downstream content consumers.
- ***"There should be an interaction between the CFP and the CMP to provide information on network conditions and, optionally, routing information."***
    - The RAE is responsible for providing CMP information about available routing paths and their characteristics, e.g. supported CoS, QoS parameters, etc. It recalculates routing paths after any changes in network topology, e.g. prefix advertisement/withdrawal, link failure/repair, and updates this information to CME. These operations are performed in long time scale based on network provisioning information.

- On the other hand, the CAFEs monitor network conditions and provide CMP information about carried traffic and statistics of terminated flows. This information is collected in short time scale.

6.6 Validation Summary

In this chapter we summarise the results of performance tests performed by both WP5 and WP6.

For the decoupled approach we can conclude that:

The analysis of content resolution procedure in large scale scenario reported in D5.2 [9], confirmed that COMET is able to keep the Content Resolution Time below the reference values, 2.5s, for the 95 per cent of the queries, if the load in the client and server CMEs is below 80 per cent.

The tests of RAE reported in D6.2 confirmed that RAE correctly calculates routing paths in the federated testbed. Moreover, the routing Convergence Time (RCT) measured for basic stressing events, i.e., prefix advertisement/withdrawal, link failure/repair is at similar level as BGP-4 convergence times. The simulation results reported in D5.2 [9] confirmed that multi-criteria and muti-path feature provided by RAE significantly improves the number of satisfied destination comparing to standard BGP-4 protocol.

The tests of CAFE reported in D6.2 [11] confirmed that CAFE can forward content at the same speed as IP-router.

For the SNME module, our tests in D6.2 [11] have proved that a single SIC module could manage up to 500 CSs, however if we want to maintain time responses for the CME queries to the SNME in a reasonable range (target metric was about 0.2 response time for 95% percentile of queries from the CME, as explained in D5.1 [8]), the optimal amount of CS would be around 300 CS as explained in D6.2 [11].

The content publication tests (CRE tests) reported in D6.2 pointed out that average response time: was about 0.5 sec, with the maximum response time about 6 sec.

For the coupled approach we can conclude that:

First of all, we have conducted extensive simulation based experiments based on different datasets of the real domain-level Internet topologies, including the CAIDA dataset [26], [27]. As far as the content resolution operation is concerned, our validations indicate that, for both the broadcast mode and the random mode, 100% of valid content requests can be successfully resolved according to the business relationships between neighbouring domains. The average number of AS hops traversed by the content resolution requests in the broadcast mode is lower than the random mode, which is also consistent with our expectation. The performance gap between the two options is actually determined by the popularity of the multi-homed domains in the topology.

As far as content delivery is concerned, the initial content delivery paths are always in the reverse direction of the content resolution paths. Thanks to the BGP-based inter-domain routing optimisation functions, the average AS-hop count between content sources and destinations can be significantly reduced (up to 40%). The scenario with the most significant improvement is to use random content resolution followed by the routing optimisation. In this case both multi-homing and routes with peering inter-domain links will help to improve the performance. In case broadcast mode (which is more expensive) is used, then only routes with peering links are able to contribute.

We have also implemented a network emulator for the coupled approach based on the VLSP emulator platform. This network emulator is able to support over 100 inter-connected routers that can be clustered into autonomous domains. Business relationships are also assigned to these domains with validations. The current implementation of the emulator supports the following content functions: resolution, routing optimisation, inter-domain multicast and content caching. Since the emulator has been implemented as proof-of-concept at small scale, the aforementioned functions have been manually validated.

7 Summary and Conclusions

In this deliverable we have provided the final specification of the COMET architecture. This document comes to update D2.2, where we have presented our architecture for the first time. Although the main features of the architecture have remained intact throughout the lifetime of the project, we have added one function to the Content-Aware Forwarding Entity. That is, to support in-network content caching, we have added a Cache-Aware Caching Function in the CAFEs of the CFP. The update of the functional architecture has been discussed in Chapter 3. In Chapter 5, we provided more details regarding the exact operation of in-network content caching in the COMET architecture.

After the three-year period of design, evaluation and experimentation with the COMET architecture and the related entities and functional blocks, we conclude on the following points (some of which have already been included in individual chapters above):

- The combined mediation-data plane functionality presents many benefits when resolving and delivering content, guaranteeing optimal resolution and transmission characteristics.
- The mediation (control) plane can accommodate a lot of functionality and complement the data plane with monitoring, management, mediation and resolution tasks.
- The forwarding plane constitutes the main data plane; with the support and the additional functionality of the mediation plane, the data plane can achieve more than mere forwarding of data, such as for example, content caching.
- The decoupled approach of operation constitutes and evolutionary path towards the realisation of an Information-Centric Network. However, being evolutionary, it loses some of the performance benefits that the ICN paradigm can offer (e.g., support for in-network content caching).
- On the other hand, the hop-by-hop, coupled resolution approach provides a more disruptive way to the ICN paradigm, being also able to incorporate more performance enhancing features. However, the main properties of the coupled approach are not designed to support the whole population of Internet contents, and therefore, has to be applied selectively to the contents of limited content providers.
- Regarding deployment of the COMET system, we conclude that the decoupled approach has to be adopted first, to pave the way for the more disruptive decoupled approach to be deployed for selected, high-priority contents in the Internet.
- Both approaches require investment from the ISPs that adopt the system (i.e., implementation of related entities), but we have proved that the system performs below the reference values and provides a transparent and backward-compatible mode of operation that allows non-COMET ISPs to co-exist with COMET-compatible ISPs in the Internet ecosystem.
- Both approaches support system reliability and resilience to failures, smooth co-existence with CDNs and content providers, support for changing network conditions (in terms of link load) and robustness to security attacks.
- Our evaluation and validation tests have shown that the system operates below the reference values for the 95% percentile of the queries, when the server-load is below 70%. Furthermore, the coupled approach optimisation function, which is based on BGP inter-domain routing rules, can reduce the AS-hop count between sources and destinations up to 40%.

8 References

[1] The EU FP7 COMET Project, “Content Mediator Architecture of Content-Aware Networks”, http://www.comet-project.org/

[2] COMET Deliverable, “D2.1 Business Models and System Requirements for the COMET System”

[3] COMET Deliverable, “D2.2: “High-Level Architecture of the COMET System”, January 2011

[4] COMET Deliverable, “D3.2: Final Specification of Mechanisms, Protocols and Algorithms for the Content Mediation System”, November 2011.

[5] COMET Deliverable, “D3.3: Prototype Implementation and System Integration Interfaces for the Content Mediation System” January 2012.

[6] COMET Deliverable, “D4.2: Final Specification of Mechanisms, Protocols and Algorithms for Enhanced Network Platforms”, December 14th, 2011.

[7] COMET Deliverable, “D4.3: Prototype Implementation and System Integration Interfaces for Enhanced Network Platforms”, May 2012.

[8] COMET Deliverable, “D5.1: Integration of COMET Prototype and Adaptation of Applications”

[9] COMET Deliverable, “D5.2: Scalability of COMET System”

[10] COMET Deliverable, “D6.1: Demonstration Scenarios and Test Plan”

[11] COMET Deliverable, “D6.2: Final Report on Experiments and Evaluation of Results”

[12] ISO/IEC 23009-1:2012, Information technology -- Dynamic adaptive streaming over HTTP (DASH) -- Part 1: Media presentation description and segment formats, 2012

[13] W. K. Chai, N. Wang, I. Psaras, G. Pavlou, C. Wang, G. G. de Blas, F. J. Salguero, L. Liang, S. Spirou, A. Beben and E. Hadjioannou, "CURLING: Content-Ubiquitous Resolution and Delivery Infrastructure for Next Generation Services", IEEE Communications Magazine, Special Issue on Future Media Internet, pp. 112-120, March 2011.

[14] W. K. Chai, D. He, I. Psaras, G. Pavlou, "Cache ‘Less for More’ in Information-centric Networks", Proceedings of the 11th IFIP Networking, Prague, Czech Republic, 21-25 May 2012.

[15] I. Psaras, W. K. Chai, G. Pavlou, "Probabilistic In-Network Caching for Information-Centric Networks", In the Proc. of the 2nd ACM SIGCOMM Workshop on Information-Centric Networking (ICN'2012), Helsinki, Finland, August 2012.

[16] T. Wu and D. Starobinski, “A Comparative Analysis of Server Selection in Content Replication Networks,” *IEEE/ACM Transactions on Networking,* vol. 16, no. 6, pp. 1461-1474, Dec. 2008.

[17] Lixin Gao and Jennifer Rexford, "Stable Internet routing without global coordination," IEEE/ACM Transactions on Networking, December 2001, pp. 681-692.

[18] A. Bęben, J. Mongay Batalla, W. K. Chai and J. Śliwiński, "Multi-criteria Decision Algorithms for Efficient Content Delivery in Content Networks", Annals of Telecommunications, Special Issue on Networked Digital Media, Springer 2012, (DOI) 10.1007/s12243-012-0321-z.

[19] J. Mongay Batalla, A. Bęben, Y. Chen, "Optimization of the decision process in Network and Server-aware algorithms", IEEE Networks 2012, Rome, Italy, October 2012.

[20] G. Garcia, A. Beben, F. J. Ramon, A. Maeso, I. Psaras, G. Pavlou, N. Wang, J. Sliwinski, S. Spirou, S. Soursos, E. Hadjioannou "COMET: Content Mediator Architecture for Content-aware Networks", in Future Network and Mobile Summit 2011, Warsaw, Poland, 15-17 June 2011.

[21] G. Pavlou, N. Wang, W. K. Chai and I. Psaras, "Internet-scale Content Mediation in Information-centric Networks", Annals of Telecommunications, Special Issue on Networked Digital Media, Springer 2012, (DOI) 10.1007/s12243-012-0333-8.

[22] W. K. Chai, D. He, I. Psaras and G. Pavlou "Cache 'Less for More' in Information-Centric Networks (Extended Version)", Elsevier Computer Communications, Special Issue on Information-Centric Networking, 2013, (DOI) 10.1016/j.comcom.2013.01.007.

[23] I. Psaras, W. K. Chai, G. Pavlou "In-Network Cache Management and Resource Allocation for Information-Centric Networks", Technical Report, Submitted to IEEE TPDS.

[24] S. Kent et al, Secure Border Gateway Protocol, IEEE JSAC Vol. 18, Issue 4, 2000, pp. 582-592

[25] Teemu Koponen, Mohit Chawla, Byung-Gon Chun, Andrey Ermolinskiy, Kye Hyun Kim, Scott Shenker, and Ion Stoica. "A data-oriented (and beyond) network architecture. In*Proceedings* SIGCOMM '07".

[26] CAIDA dataset; http://www.caida.org/research/topology/#Datasets

[27] CAIDA, The CAIDA AS Relationships Dataset, 2010. Available from: <http://www.caida.org/data/active/as-relationships/>.

9 Abbreviations

AC	Admission Control
AS	Autonomous System
ATE	Automatic Test Equipment
BE	Best Effort
BGP	Border Gateway Protocol
BtBE	Better Than Best Effort
BW	Bandwidth
CACF	Content-Aware Caching Function
CAFE	Content-Aware Forwarding Entity
CAFF	Content-Aware Forwarding Function
CAIDA	Cooperative Association for Internet Data Analysis
CC	Content Client
CCN	Content Centric Networks
CDN	Content Distribution Network
CDNI	CDN Interconnection
CFP	Content Forwarding Plane
CMF	Content Mediation Function
CMP	Content Mediation Plane
CRF	Content Resolution Function
COMET	COntent Mediator architecture for content-aware nETworks
CoS	Class of Service
CP	Content Publisher
CS	Content Server
CSP	Content Server Providers
CSR	Content Streaming Relay
DBAC	Declaration Based Admission Control
DASH	Dynamic Adaptive Streaming over HTTP
DDoS	Distributed Denial of Service
DoS	Denial of Service
DNS	Domain Name Server
IAB	Internet Architecture Board
IP	Internet Protocol
ICN	Information Centric Networks
ISP	Internet Service Provider
MBAC	Measurement Based Admission Control
MPD	Media Presentation Description

MPEG	Moving Pictures Expert Group
NNH	next-next-hop
NRLI	Network Layer Reachability Information
P2P	Peer-to-Peer
PKI	Public Key Infrastructure
PMF	Path Management Function
PoP	Point of Presence
Pr	Premium Service/QoS
QoS	Quality of Service
RAE	Routing-Awareness Entity
RCT	Routing Convergence Time
RFC	Request For Call
RTSP	Real Time Streaming Protocol
S-BGP	Secure BGP
SIC	Server Information Collector
SLA	Service Level Agreement
SNME	Server and Network Monitoring Entity
SNMF	Server and Network Monitoring Function
STREP	Specific Targeted Research Project
TCP	Transport Class Protocol
UDP	User Datagram Protocol
VLSP	Virtual Link State Protocol

10 Acknowledgements

This deliverable was made possible due to the large and open help of the WP2 team of the COMET team within this STREP, which includes the deliverable authors as indicated in the document control. Many thanks to all of them.