New Review of Hypermedia and Multimedia

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/tham20>

Digital forgetting in information-centric networks—the CONVERGENCE perspective

Fernando Almeida^a, Helder Castro^a, Maria T. Andrade^a, Giuseppe Tropea^b, Nicola Blefari Melazzi^b, Salvatore Signorello^b, Aziz Mousas^c, Angelos Anadiotis^c, Dimitra Kaklamani^c, Iakovos Venieris^c, Sam Minelli^d & Angelo Difino^e

^a Faculty of Engineering of Oporto University, INESC TEC, Porto, Portugal

^b UdR Roma Tor Vergata, CNIT, Rome, Italy

^c School of Electrical and Computer Engineering, NTUA, Athens, Greece

^d R&D Unit, Alinari 24 ORE, Florence, Italy

^e CEDEO.net, Via Borgionera, Villar Dora (TO), Italy

Published online: 30 Jun 2014.

To cite this article: Fernando Almeida, Helder Castro, Maria T. Andrade, Giuseppe Tropea, Nicola Blefari Melazzi, Salvatore Signorello, Aziz Mousas, Angelos Anadiotis, Dimitra Kaklamani, Iakovos Venieris, Sam Minelli & Angelo Difino (2014) Digital forgetting in information-centric networks—the CONVERGENCE perspective, *New Review of Hypermedia and Multimedia*, 20:2, 169-187, DOI: [10.1080/13614568.2013.877088](https://doi.org/10.1080/13614568.2013.877088)

To link to this article: <http://dx.doi.org/10.1080/13614568.2013.877088>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims,

proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

Digital forgetting in information-centric networks—the CONVERGENCE perspective

FERNANDO ALMEIDA^{†*}, HELDER CASTRO[†],
MARIA T. ANDRADE[†], GIUSEPPE TROPEA[‡],
NICOLA BLEFARI MELAZZI[‡], SALVATORE SIGNORELLO[‡],
AZIZ MOUSAS[§], ANGELOS ANADIOTIS[§], DIMITRA KAKLAMANI[§],
IAKOVOS VENIERIS[§], SAM MINELLI^{††} and ANGELO DIFINO^{‡‡}

[†]Faculty of Engineering of Oporto University, INESC TEC, Porto, Portugal

[‡]UdR Roma Tor Vergata, CNIT, Rome, Italy

[§]School of Electrical and Computer Engineering, NTUA, Athens, Greece

^{††}R&D Unit, Alinari 24 ORE, Florence, Italy

^{‡‡}CEDEO.net, Via Borgionera, Villar Dora (TO), Italy

(Received 27 May 2013; final version received 16 December 2013)

The Web is rapidly becoming the prime medium for human socialization. As it evolves towards an information-centric operation, it records everything and forgets nothing, assuming that every online resource disclosed by people (photos, posts, multimedia files, etc.) is permanently valid and is to be stored forever. However, throughout their lives, people tend to change, both in their habits as well as in their views and opinions. In many situations, as the years go by, information released loses relevance or people may decide they no longer want others to access information they have previously published. The work presented in this paper strives for a new information persistence paradigm, whereby the enforcement of “digital forgetting” is implemented over an information-centric model for the Internet. The defined solution enables the definitive elimination of digital objects, either on-demand or on a pre-scheduled basis, and, hence, their “forgetting.” The solution, conceived within the framework of the European project CONVERGENCE, is based on the employment of metadata descriptions about resources, which unambiguously identify their rightful owners. This additional data is efficiently bound to the resource through the use of an extended version of the MPEG-21 Digital Item specification, and its prescriptions are enforced by CONVERGENCE’s distributed provisions.

Keywords: MPEG-21; Information-centric networking; Digital forgetting; Metadata; Right management; Orphan works

*Corresponding author. Email: almd@fe.up.pt

1. Introduction

In today's digital world, it is extremely easy for users to publish information via websites, blogs, video-sharing sites, social sites etc. Once that information has been published, it becomes instantly available to a potentially global audience. Search engines and web archiving tools ensure that copies of the information are rapidly disseminated beyond their original publication venues. Private users may also make copies of and republish the information, usually without the knowledge of the original author. As a result, it is practically impossible to locate and eliminate all copies of any specific information item, even when the author is able to remove the information from the site where it was originally published.

There are, however, various circumstances in which this complete elimination is precisely what authors need. The most commonly cited reasons, for this necessity, are personal (Bannon 2006), such as, for instance, the wish to eliminate references to opinions the user no longer holds, or the wish to eliminate personal information that has become embarrassing for the user or for the user's relatives. According to Rosen (2010), the fact that the Internet never seems to forget is viewed as threatening, at an almost existential level, in regard to our ability to control our identities, to reserve the option of reinventing ourselves and starting anew, or simply to overcome our checkered pasts.

In other circumstances, there may be business reasons to withdraw previously published information (e.g. a company's desire to update product information that has been shown to be inaccurate or information on promotional offers that are no longer valid such as content licenses). Additionally, the EU legal framework, published at the beginning of 2012 by the European Commission, establishes new rights and rules associated with Internet-based data collection systems. EU leaders have described the new provisions as "a right to be forgotten" (Lund 2012).

There is also an interesting debate in Europe about licensing and digital right management policies (AMPAL 2012). For one side, it is important to protect the right of digital providers and their investments in the production of new contents; however, for other side, it is fundamental to guarantee that end-users have greater clarity on legitimate and non-legitimate uses of protected material and authorization to use these media contents for non-commercial purposes (e.g. educational fields, scientific research domains, etc.).

In summary, there is a mismatch between the Internet's tendency to preserve information forever and the, legally acknowledged, needs of providers, consumers, and end-users. In this paper, we present the design and implementation of a digital forgetting mechanism that may be applied to Internet's emerging information-centric paradigm, which preserves the rights of users to delete their "digital footprint." The adoption of information-centric networking brings a shift from the traditional host-to-host communication to a content-to-user paradigm which focuses on the delivery of the desired content to the intended users. This content-oriented access control provides access to specific information items as a function of time, place (e.g. country), or profile of user requesting the item, which enables the use of digital forgetting, to ensure that content generated at one period in a user's life does not come back to haunt him later on. Furthermore, it also

enables the performing of garbage collection, deleting from the network expired information.

The proposed solution, developed within the framework of the European-funded project CONVERGENCE (2013), is based on the use of the MPEG-21 Digital Item (DI) concept (MPEG 21 2012), which allows the binding of descriptive data to every resource published on the Internet, as a unique logical package. This metadata includes the unambiguous identification of the rightful owner of the resource, as well as indications concerning the validity or expiration date of the resource, among other information. The use of MPEG-21, together with the adoption of an information-centric networking paradigm and a middleware layer comprising engines capable of efficiently processing the data packages in a distributed fashion, provides the solution to the challenge of implementing digital forgetting in the Internet.

This paper is organized as follows: Section 2 presents an overview about the concept of information-centric networks and of the digital representation of objects, highlighting the role of MPEG-21. It also introduces the CONVERGENCE project and its relevant associated concepts, such as the Versatile Digital Item; Section 3 presents the defined digital forgetting mechanism, with its context of application, i.e. the CONVERGENCE system; Section 4 details the implementation of the digital forgetting mechanism and of its associated workflows; Section 5 exemplifies how the mechanism may be exploited; and Section 6 draws the final conclusions and presents some venues for future work.

2. State of the art

2.1. The need of information-centric networks

The Information Centric Networking (ICN) paradigm is one of the more prominent research topics in the context of the future Internet. The communication paradigm within ICN is different from what it is with Internet Protocol (IP). Current IP architectures revolve around a host-based conversation model (i.e. a communication link is established between two hosts before any content is transferred), and the delivery of data, in the network, follows a predominantly source-driven approach (i.e. the path is set up from the sender to the receiver) (Mathieu *et al.* 2012), especially with regard to video content. Furthermore, said delivery is predominantly performed through a strict client-server interaction model, where the client side's characteristics are volatile, but the server side is static and centralized.

The current architecture presents several issues when we have billions of connected devices, increased bandwidth needs, or critical applications (e.g. video and audio communications) interacting over the Internet. Furthermore, if we consider a wireless scenario, and take into account the massive growth of mobile devices, the bottleneck becomes even more apparent. Moreover, the system does not account for the number, owner, or location of digital copies. Druschel *et al.* (2011) considers that in such an open system, it is not generally possible for a person to locate all personal data items (exact or derived) stored about them, and

it is difficult to determine whether a person has the right to request removal of a particular data item.

Several approaches of digital forgetting over the current Internet paradigm have been proposed. Meyer-Schönberger proposed to tag sensitive data with an expiration date and to require all servers handling such data to obey that date. Geambasu *et al.* (2009) propose the adoption of share keys in a distributed hash table, a data structure that underlies many P2P networks. Finally, another potential solution to prevent data duplication is to adopt techniques from Digital Rights Management (DRM). However, DRM has faced various problems at the technical level, which can be bypassed with moderate effort (Druschel *et al.* 2011). All existing approaches to ensure the right to be forgotten in open systems are vulnerable to unauthorized copying, while the data is publicly accessible and a re-dissemination of such unauthorized copies once the data has expired.

On the contrary, within the ICN operational paradigm, users request content without knowledge of the host that can provide it, communication follows a receiver-driven principle (i.e. the path is set up from the receiver to the provider), and data follows the reverse path. Within this paradigm, it is the network that is responsible for doing the mapping between the requested content and the location [i.e. host(s)], where it can be obtained from, as explained in Figure 1. It is, thus, the matching of content offered by hosts to the requested content that dictates the establishment of a communication in ICN. According to Ahlgren *et al.* (2012), the main components of the information-centric approach are the Named Data Object (NDO), Naming and Security (names are uniquely used for identifying objects), Application Program Interface (API), Routing and Forwarding (two general approaches to routing are possible: routing of NDO requests; and routing of NDO back to the requester).

ICN offers a lot of advantages in terms of routing efficiency and simplicity (Detti *et al.* 2012, Haris 2012, Perino *et al.* 2012, Ruidong and Harai 2012), security (Detti *et al.* 2012, Mohaisen *et al.* 2012), and naming persistence (Haris 2012, Sollins 2012). Additionally, Detti *et al.* (2012) states that, by providing different performance in terms of both transmission and caching, the ICN

Figure 1. Communication paradigm in ICN (Mathieu *et al.* 2012).

paradigm offers a per-content quality-of-service differentiation. ICN would, thus, enable network operators to provide different levels quality-of-service to different types of content, without complex procedures, making it possible to improve the overall quality of experience in a relatively simple manner. Additionally, the potentialities offered by ICN make it easier to preserve the integrity of an object and its provenance (Sollins 2008).

Superior performance in areas such as routing and content delivery may also be achieved within a host-based interaction model, if technologies such as Content Delivery Networks (CDNs) and Multipath TCP are adequately used. However, even if Multipath Transmission Control Protocol (TCP) within a host-based interaction model may increase the data throughput between two communicating entities, it is still inscribed within a context where content is statically bound to specific locations (hosts), and not dynamically diffused throughout a vast network tissue in a temporally evolving manner, as it is in the context of ICN. Furthermore, the advantages offered by Multipath TCP are not exclusive to a host-centric interaction context, as a simultaneous multipath exchange of data and content may also be done within a system operating according to an ICN paradigm.

The employment of CDNs, within a host-centric interaction context, does enable the achievement of data diffusion performances similar to those attained within an information-centric interaction context, as it makes client-server distribution more scalable and dynamic. However, this is done in a manner that is not universal and transparent, because CDNs are overlay structures that serve only a specific subset of all content, whose delivery implies the CDN's conscious and explicitly employment. Furthermore, they require a constant central administration. Comparatively, within an information-centric interaction context, all content and data storage and diffusion is distributedly, automatically, and transparently handled by the networking layer without the need for any explicit employment of an auxiliary structure (the CDN). Furthermore, the elimination of the added complexity implied by the interaction of a CDN with the networking layer and the exploitation of the synergies that become possible when content diffusion is handled directly at the networking layer, grant information-centric based data and content diffusion an overall greater efficiency and, especially, scalability.

In light of the above, it becomes apparent that even when aided by technologies such as Multipath TCP and CDNs, content and data exchange within a host-centric interaction context is still at a disadvantage when compared to the same activity within an information-centric one, in such aspects delivery efficiency, transparency, universality and scalability.

2.2. Digital objects representation with MPEG-21

A digital object can be seen as a compound artifact that wraps digital material in terms of four elements: its content, its metadata, its relationships with other objects, and its behavior. From a digital object encoding perspective, various XML formats, including Metadata Encoding and Transmission Standard (METS), MPEG-21 Digital Item Declaration Language (DIDL), Fedora Object XML (FOXML), and Resource Description Framework (RDF), help encode arbitrary

Figure 2. Abstract models & MPEG-21 DIDL (Bekaert and Kooning 2006).

digital object variations through the use of common XML schema (Saidis and Delis 2007). The adoption of XML brings relevant advantages, because it is text-based and position-independent, which brings a higher level of application-independence. However, JavaScript Object Notation (JSON) could also be adopted and represents a novel approach, bypassing the need for language processing and is better suited to data-interchange (Downes *et al.* 2010).

MPEG-21 is an open and comprehensive standard framework for multimedia delivery, and consumption, designed by the ISO/IEC Moving Picture Experts Group that can be used to combine video, audio, text, and graphics (Pereira *et al.* 2005). As described by Timmerer and Hellwagner (2010), the standard provides normative methods for content identification, rights management, and seamless adaptation. With MPEG-21, MPEG created a new interoperable unit for multimedia delivery, and transaction, called DI, which is essentially a virtual container for metadata and content.

The MPEG-21 DI is, thus, the basic unit of transaction within the MPEG 21 framework. It contains two distinct sections: the MPEG-21 DID Abstract Model and the MPEG-21 DIDL. The former is an abstract data model for declaring DIs; the latter is a serialization of the Abstract Model in XML. The structure of these sections is shown in Figure 2.

A DI is a combination of media resources (video, audio tracks, images, etc.), and metadata pertaining to such resources. The latter may describe the resources' creation context, (creator, date of creation, etc.), semantic context (what the context is about), usage rights context (intellectual property, licensing and encryption information), usage event monitoring context, etc.

2.3. The CONVERGENCE framework

The CONVERGENCE system adopts a publish-subscribe content-centric architecture that focuses, simultaneously, on providing access to content (documents, media, files, etc.) and to interactive services (i.e. HTTP sessions, Telnet, voice calls, etc.) (Gkonis *et al.* 2011). CONVERGENCE deals with the delivery of both static content and dynamic services, by considering that both content and services (service-access points) are identified by a name and referred to by such name, instead of by their locations.

The CONVERGENCE system's architecture is structured into three different functional levels: the network level; the content level; and the application level, which is, further, divided into the tool layer and the user application layer. Each of these levels aggregates a distinct set of conceptually similar functions, which in turn are implemented by one or more functional blocks. The following main blocks, or block groups, have been defined:

- CONVERGENCE Network (CoNet)—operating at the Network level, this block provides access to named-resources on a public or private network infrastructure. A named-resource is any resource that can be identified by means of a name. Such resources may be either data or service-access-points. Examples of named-resources include: a Versatile Digital Item (VDI); an electronic document; an image; a source of information with a consistent purpose; the point of access to a service; or a collection of other resources.
- CONVERGENCE Middleware (CoMid)—CoMid is the set of all Content level blocks. Whereas the CoNet block handles named-resources only in terms of their names, content level functionality (provided by CoMid's constituting blocks) delivers the means to handle resources on the basis of "what" they contain and offer a distributed publish/subscribe matching mechanisms employing various CoMid engines.
- Community Dictionary Service (CDS)—this functional block belongs to the Content level of CONVERGENCE. It provides the CoMid collective with all the concepts that users' subscriptions, searches, requests, and publications may pertain to. The CDS parses and interprets ontologies created by CONVERGENCE users, thus enabling the understanding of user subscriptions, search requests, or publications.
- CONVERGENCE Applications and Tools—these blocks are part of CONVERGENCE's application level. Collectively they provide the services that handle the interfacing of users with the overall system. Tools provide useful APIs for common CONVERGENCE operations, like publish or subscribe of VDIs, hiding the complexity of CoMid engines from the application layer. Applications, in their turn, employ CONVERGENCE tools on behalf of the user and interact with functional blocks of other CONVERGENCE functional levels.

CONVERGENCE relies on MPEG-M for its middleware protocols. [Figure 3](#) shows an abstract architectural scheme of CONVERGENCE. In this scheme, the middleware consists of Protocol Engines, Technology Engines, and an Orchestrator. A Protocol Engine is responsible for processing the protocol of the corresponding Elementary Service, each time it is called. A Technology Engine is a component of the CoMid providing technological functionalities via a standard API. An Orchestrator Engine is a component of the CoMid capable of organizing chains of Technology Engines to provide aggregated functionalities.

In defining the protocols for the protocol and technology engines, CONVERGENCE has followed a unified approach. Each engine is split into a schema handler (mandatory) and a technology handler (optional). The schema handler is required for performing essential operations on the schema. For example, the CreateContent Protocol Engine relies on the schema specified in MPEG-M Part 4 for the

Figure 3. CoMid abstract architecture.

CreateContent Elementary Service. The technology handler provides the API for the operations that a given engine can perform. Thus, a Protocol Engine technology handler provides the API for making a call (local or remote) to the corresponding Elementary Service. For example, the technology handler for the Security technology engine provides functionality for signing, encrypting, decrypting, etc.

2.4. The VDI

A fundamental concept within the CONVERGENCE architecture is that of the VDI. This is a common container for all types of digital content that is derived from the MPEG-21 DI (Hang *et al.* 2012). VDIs serve the new requirements of the “Internet of Things,” by providing a homogeneous way of searching and handling structured information, and by incorporating privacy mechanisms to support the needs of information providers and information consumers.

Every VDI is uniquely identified by a given VDI identifier, assigned to the specific VDI when it is first created, and finalized as a self-contained piece of information, to be injected into the CONVERGENCE system, and traded between communication peers. Additionally, each VDI may also include a license which identifies its owner, its authors and defines its access rights context. According to Future Internet Architectures (FIA) (2011), the act of modifying a VDI is in fact the creation of a new VDI, semantically related with the original, but with a new digital signature recalculated on the whole updated package of bits, as well as with a new identifier. This new VDI can thus be trusted, based on the author of the change, and guarantees that the original VDI has not been tampered with after the publication. Furthermore, a VDI is composed of a dynamic part which can be

manipulated as a shared resource, where actors can add or replace new pieces of information, and a semantic part that contains the metadata and a stable identification block which link back, through a level of indirection, to the specific VDI regardless of its successive manipulations. These VDI updating operations are designed in a way that implies no alterations to already published VDIs, but allows for removal of obsolete VDIs (enabling digital forgetting) and for retrieval of all old versions of a VDI, if requested by the user.

2.4.1. The Resource VDI. A Resource VDI carries or references an actual information asset of service. It is thus the actual container, or referrer, of consumable resources in the CONVERGENCE system.

2.4.2. The Publish VDI. The purpose of a Publish VDI is to publicize the existence of a specific information resource, (static information asset or dynamic information service), represented by a Resource VDI, throughout CONVERGENCE's middleware, by way of asynchronous gossiping.

2.4.3. The Subscribe VDI. The purpose of a Subscribe VDI is to publicize the interest of some specific user on some specific subject (or set of them) or on some individual publication, throughout CONVERGENCE's middleware, by way of asynchronous gossiping. At specific points in the middleware, matches between Subscribe and Publish VDIs will be identified (in terms of the subjects subscribed to by the earlier and involved by the latter), and the subscribing user will be notified of the potential relevance of the matching Publish VDI (and its associated Resource VDI).

2.4.4. The Unpublish VDI. The purpose of the Unpublish VDI is to enable the asynchronous gossiping, through the Overlay Technology Engine (TE) collective, of an unpublication request of a specific VDI. The Unpublish VDI must thus carry information for the following purposes:

- (1) Inform all receiving parts (Overlay TEs) that a specific VDI is to be removed;
- (2) Specify (in a valid form) the identity of the user which is calling for the revocation;
- (3) Supply the rights information pertaining to the permission of revocation of the VDI.

In order to satisfy the first two points, the Unpublish VDI should carry the original Revoke Content Request. To satisfy point 3, it should carry the license (obtained from the corresponding Publish VDI), that expresses the permission (given by the content owner) of the calling user to revoke the VDI in scope.

Figure 4 depicts the overall structure of an Unpublish VDI.

It thus has a root element sequence composed by elements *did:DIDL*, *did:Container*, and *did:Item*. The *did:Item* element represents the actual digital object to which the VDI pertains. That object is the corresponding Publish VDI. That is, the Unpublish VDI declares the Publish VDI as its "remote resource," and binds information to it. That information is not merely descriptive, but instead specifies

Figure 4. Overall structure of the Unpublish VDI.

an action to be performed upon the Publish VDI, and provides proof (in terms of property rights) of the adequacy of that action.

The mentioned information is contained within two *did:Descriptor* child elements of *did:Item*:

- *did:Descriptor* (A) (Figure 3)—carries the original Revoke Content Request which refers to the Publish VDI and
- *did:Descriptor* (B) (Figure 3)—carries the license (obtained from the corresponding Publish VDI) that expresses rights panorama regarding the revocation of the Publish VDI.

The final child element of the *did:Item* is a *did:Component*, which carries a *did:Resource* element that references the Publish VDI.

3. The digital forgetting mechanism in CONVERGENCE

One of CONVERGENCE's key goals is to support the “digital forgetting” of content published onto it, allowing users to completely and irreversibly remove

(or update, i.e. remove and re-insert a new VDI) information they have previously published.

The CONVERGENCE offers an open source middleware that exposes network functionality to information providers and consumers, allowing them to search for, retrieve, modify, and revoke VDIs. The CONVERGENCE architecture makes it possible to maintain access to VDIs when they move from one host to another, protect user security and privacy, and provide support for “digital forgetting,” ensuring that VDIs are deleted when they pass a user-defined expiry date.

3.1. Functional requirements

One of CONVERGENCE’s key goals is to support the “digital forgetting” of content published onto it, allowing users to completely and irreversibly remove (or update, i.e. remove and re-insert a new VDI) information they have previously published.

The digital forgetting operation can be formalized in the following four requirements:

- (1) A VDI shall carry a specification of its expiry date (which may be some specific future point in time or an “unlimited” date);
- (2) All VDIs shall have a default expiry date;
- (3) Users are allowed to, at any moment, unpublish all copies of the VDIs that they have published; and
- (4) The CONVERGENCE system shall include a standard tool providing automated garbage collection of VDIs, residing on the system, that have passed their expiry date. This implies that users will no longer be able to retrieve such VDIs.

To implement these requirements effectively, it will also be necessary to meet a number of secondary requirements, namely in terms of authorization, security, and performance.

Firstly, metadata and resources in expired or unpublished VDIs shall be deleted from the CoNet. The creator of a VDI shall be able to define who can change the expiry date for a VDI and who can “unpublish” the VDI. Only authorized users shall be able to change the expiry date on a VDI or unpublish it. Furthermore, derived VDIs shall, by default, inherit expiry dates, rights to change expiry dates, and rights to unpublish from original VDI, if the publisher does not provide new ones.

Secondly, VDI search and subscribe services shall not return references to expired or unpublished VDIs. However, users should be able to specify that a VDI has no expiry date (i.e. that it is intended to be “eternal”).

Thirdly, VDIs stored off the CoNet shall expose expiry data to third party applications, allowing users to verify if they have expired (this supports regulation to make use of expired VDIs illegal).

3.2. Digital forgetting mechanisms

In order to satisfy the requirements defined above, CONVERGENCE supports two forms of digital forgetting: pre-planned digital forgetting and on-demand digital forgetting.

In the first scenario, the CONVERGENCE system defines a default expiry date for all the Publish VDIs published to the CONVERGENCE system. The publisher of a VDI has the ability to override the default value at publication time. The user can also define who (if anyone) is authorized to change the expiry date. This information is, by default, inherited by derived VDIs. Publish VDIs that have passed their expiry date are no longer retrievable by CONVERGENCE users and are no longer to be referenced in results from CONVERGENCE search and subscribe services. The associated Resource VDI will also be eliminated from CoNet.

For the second scenario, when publishing a Publish VDI, the publisher defines who (if anyone) has the right to unpublish it. The CONVERGENCE system shall enable, authorized users, at any time after publication, to request that a Publish VDI, (and associated Resource VDI), be removed from the system. The system will then proceed to asynchronously eliminate all stored versions of the Publish VDIs (and their associated Resource VDIs), and eliminate all semantically qualifying metadata about it, scattered throughout the system. Once this operation has been completed, the Publish VDI (and its associated Resource VDI) will no longer be retrievable and will no longer appear in search or subscription results. The removal of the Publish VDI will not prevent access to, retrieval, and consumption of other Publish VDIs in the same VDI sequence (set of subsequently updating VDIs). Other Publish VDIs that declare relationships with the unpublished VDI will remain accessible. Still, the mentioned relationship will then point at an irretrievable object.

4. Implementation

The implementation process of the digital forgetting mechanism is slightly different for the pre-planning and on-demand digital forgetting implementations. Therefore, we detail both processes and highlight their differences.

4.1. Pre-planned digital forgetting

The pre-planned digital forgetting includes two mandatory steps. The former is the publication of the VDI in the network; the latter is the elimination of the VDI from the CoNet.

At Publish VDI publication time, the publisher specifies the date and time at which the Publish VDI will expire. This date should be anterior to the expiry date of the targeted Resource VDI. At the CoMid, the *Publish Content Aggregated Service (PCAS)* uses the *Overlay TE* to gossip the new Publish VDI throughout the system. Then, *PCAS* submits the Publish VDI to *CoNet* for storage and distribution. Finally, the *CoNet* stores and disseminates the submitted digital objects as sets of Named-data CoNet Information Units (CIUs) with a field specifying their expiry date.

When the Publish VDI gets expired, more precisely, from the expiry time onward, *Match TE* which handles the matching between Publish and Subscribe VDIs performs the following operations:

- (1) Police the expiration dates of the Publish/Subscribe VDIs they are responsible for;
- (2) Detects that the corresponding entry in the table is no longer valid; and
- (3) Removes the entry.

At the *CoNet* side, independently from the *CoMid* operation, when the *CoNet* deems it appropriate, nodes mark as “stale” the named-data CIUs in question (those that contain the Publish and Resource VDIs), and no longer cache or distribute them.

From the Publish VDI’s expiry time onward, there will no longer be a trace of the expired Publish VDI’s existence at the *CoMid* level and the Publish VDI will no longer be retrievable from the *CoNet*. Looking to the Resource VDI’s expiry time, which is independent from that of any Publish VDI, it will also no longer be retrievable from *CoNet*.

4.2. On-demand digital forgetting

Within this mode of digital forgetting, an authorized user specifies his/her desire for the removal (unpublishing) of a Publish VDI (published sometime in the past), for which he/she has “unpublish rights.”

Given that CONVERGENCE incorporates a dynamic set of peers coming and going in and out of the system in an arbitrary fashion, it is not possible to have a complete and synchronous removal of a Publish VDI from *CoMid*. Therefore, an asynchronous approach is followed through the employment of an Unpublish VDI which is injected into the system notifying the anticipated caducity of a specific Publish VDI.

The on-demand digital forgetting includes two mandatory steps. The former is the creation and distribution of an Unpublish VDI; the latter is the elimination of the VDI from the *CoNet*.

At some time after the publication, but prior to the originally specified expiry date, an authorized user issues a request for the removal of a Publish VDI. At that time, the *CoMid* employs the *PCAS*, to notify the relevant peers that all the metadata that was contained in the Publish VDI under revocation, which is stored at the match tables, must be eliminated. This is done in the following way:

- (1) *PCAS* creates an Unpublish VDI which references a specific Publish VDI. The Unpublish VDI declares that its targeted Publish VDI is no longer valid and should be eliminated from the match tables;
- (2) The Unpublish VDI is gossiped throughout the peer collective (employing the *Overlay TE* collective) and will end up in the same peers as the corresponding Publish VDI; and
- (3) Any peer, upon receiving the Unpublish VDI, proceeds to the appropriate “forgetting” activities.

After these operations, the *Unpublish Content Aggregated Service* issues a revocation command to *CoNet* requesting the deletion of the Publish VDI in question. Then the *CoNet*, upon reception of the revocation command from

CoMid, proceeds to the appropriate activities for marking as “stale” on the local node the named-data units correspond to the unpublished VDIs.

The forgetting of the Unpublish VDI will also be dealt with by the system. At Unpublish VDI production time, *PCAS* uses the *Overlay TE* to gossip the new Unpublish VDI throughout the system. Given that the Unpublish VDI has the same expiry date as that of its associated Publish VDI, for as long as there is some possibility that some metadata of the Publish VDI is still circulating in the system (which may occur up to the time of the Publish VDI’s originally specified expiry date), the Unpublish VDI (more accurately, its metadata specifying the caducity of the associated Publish VDI) will also be circulating in the system notifying any latecomer peers of the anticipated caducity of the target Publish VDI.

At Unpublish VDI expiry time, the *Match TE*, when policing the expiration dates of the match table’s entries, detects that the corresponding entry in the table is no longer valid and removes it.

5. Exploitation and scenarios

The relevance of the previously defined mechanism may be demonstrated by showing the added value of its employment/exploitation in different situations of consumption of multimedia resources that are flourishing on the Web.

In a first instance, we will look to the main opportunities offered by the market by the implementation of a VDI that supports digital forgetting mechanisms. After that, we will present a scenario that could exploit the relevance of pre-planned digital forgetting. Finally, a slightly different adaptation of the previous scenario will be given to demonstrate the usefulness of on-demand digital forgetting.

5.1. Opportunities offered by digital forgetting

In the context of creative contents made available online, one of the most increasing market segments is the visual production (let it be video or photography).

The most interesting application of the VDI concept to this domain is the so-called orphan work domain: the orphan works are creative content productions where the rights holders cannot be identified. Most of the visual content created nowadays can be considered “orphan,” and most of it is an objective of market interests. In the case of a large usage of the VDI modeling and digital forgetting/remembering, any content would have matched its right holders.

A second application of VDIs can be considered. During an average photographic campaign, a professional photographer on a dynamic scene generates one to four thousand pictures and up to 50 short videos per day. The professional workflows are more and more managed by batch and automated processes to select, improve, annotate, and distribute (at low prices) large amount of content.

Larger volumes of photographic and video content are flooding on the digital media and consumed in less than one day. Few creations, compared to the total created content, will last for more than one day.

One opportunity for the authors (after such huge work of shooting, organizing, post-processing) is to sell time-limited licenses and, at the same time, be sure that their clients do not resell to third parties nor reuse their collection after some time. A VDI could satisfy this need. The general practice of customers is to pay time-limited licenses and then use as much as they like the creative content.

5.2. Pre-planned digital forgetting

An illustrative such example is this hypothetical “Augmented e-Learning” scenario:

Professor Watson publishes an InfoObject (IO), containing a work group activity for her students, as an Activity IO in a distributed Publish-Subscribe based e-Learning Platform. This Activity IO is made available for subscription to all students. Professor Watson also wants to associate an expiry date to this activity, because the deadline for the submission acceptance for the students was already previously established at the classroom. After the defined expiry date, the Activity IO must no longer be available to the students.

First of all, the CONVERGENCE system creates a new Publish VDI with the contents of the work group activity. An example of such a VDI is presented in Figure 5.

The Publish VDI defined in the Figure 4 carries an Item element, (representing a DI), which contains the following elements:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<mpegm-didl:IDIL>
  <mpegm-didl:Item id="A0">
    <mpegm-didl:Descriptor>
      <mpegm-didl:Statement>
        <dii:Identifier>urn:eu:convergence:server1/eida9c08-ef7fad62927f</dii:Identifier>
      </mpegm-didl:Statement>
    </mpegm-didl:Descriptor>
    <mpegm-didl:Descriptor>
      <mpegm-didl:Statement>
        <didl_msx:Metadata>
          <didl_msx:StructuredData ref="urn:conv:metadata:schema:2011">
            <ns21:ConvergenceMetadata>
              <ns21:SequenceIdentifier>urn:444e7cb4-2dee81b0f13f</ns21:SequenceIdentifier>
              <ns21:VDIkind>FVDI</ns21:VDIkind>
              <ns21:ExpiryDate>2013-03-15T14:30:15.117+02:00</ns21:ExpiryDate>
              <ns21:PublicationMetadata>
                <ns21:Dimension>
                  <ns21:FractalAlgebra>
                    <ns21:Fractal>podcast</ns21:Fractal>
                  </ns21:FractalAlgebra>
                </ns21:Dimension>
              <ns21:RVDIid>urn:eu:convergence:server1/98a69572-3af7-4706-b571-7031a4ee38a8</ns21:RVDIid>
              <ns21:Keyword>MMP</ns21:Keyword>
              <ns21:Keyword>Test</ns21:Keyword>
            </ns21:PublicationMetadata>
          </ns21:ConvergenceMetadata>
        </didl_msx:StructuredData>
      </didl_msx:Metadata>
    </mpegm-didl:Statement>
  </mpegm-didl:Descriptor>
</mpegm-didl:Item>
</mpegm-didl:IDIL>
```

Figure 5. Example of the Publish VDI for the pre-planned scenario.

- a Descriptor which carries the VDI's unique identifier;
- a Descriptor containing CONVERGENCE-specific metadata which carries:
 - the definition of the VDI type (i.e. Publish VDI (PVDI));
 - the expiry date associated with the Publish VDI;
 - the identification of the fractal-like semantic space to which the contents of the associated Resource VDI pertain.

After this step, the Publish VDI will be injected in the system and, therefore, the *CoMid* and *CoNet* will receive the notification about the existence of a new VDI. At the *CoMid*, the *Overlay TE* offers a method *overlayPropagationMessageConsumer* that is responsible for the proper propagation of messages in the overlay network among the considered semantic fractals. The *Overlay TE* is also responsible for maintaining the overlay and ensuring that all peers are in communication with at least one other peer. On the other side, *CoNet* offers the method *storeAndAdvertise* that is responsible for storing a resource in the *CoNet* with a given Network Identifier (NID). The NID uniquely identifies a named-resource at *CoNet* level and has the following form:

$$NID = \langle namespace\ ID, name \rangle \quad (1)$$

The *NID namespace ID* determines the structure of the rest of the NID (i.e. of the name field). Thus, the *name* field is a namespace-specific string. In *CoNet*, objects are split into different parts or chunks. Each chunk is packaged in a *named-data CIU*, whose control information contains the network-identifier, the chunk sequence number, temporal-data, and the security-data (Jacobson *et al.* 2009).

The data formats of the CIUs are defined by XML schemas and encoded with explicitly identified field boundaries. This design permits field values of arbitrary length. The use of XML structures does not imply that field values are text strings nor does it require that messages be encoded as human-readable text. Most fields, including those that identify content, are defined to contain arbitrary binary values. Furthermore, the format of carrier-packet complies with a bit-level description, as in classical network protocols, like IP, TCP, User Datagram Protocol (UDP), etc. This approach facilitates high-speed forwarding of carrier-packets, eliminating the processing effort required to parse XML.

When we reach the established expiry date defined in the Publish VDI, the *PCAS* invokes the *Match TE* at the *CoMid*. This invocation employs two interfaces:

- *Publication Manager*—this interface offers the *revokePublication* method, which is responsible for the removal of Publish VDIs.
- *Subscription Manager*—this interface offers the *revokeSunscription* method, which is responsible for the removal of Subscribe VDIs.

The information regarding the expiration time must also be sent to *CoNet*. For that, *CoNet* offers the *revoke* method that can revoke (flags as unreachable or “stale”) a specific resource, identified by its NID, in the cache of the node, after a *FreshnessSeconds* timer has expired. The resource is also flagged as unavailable

from the local repository of the local *CoNet* node, if that node explicitly serves as repository and if resource is stored thereby.

5.3. On-demand digital forgetting

The hypothetical “Augmented e-Learning” scenario considered for pre-planned digital forgetting will be slightly modified so as to comprise an on-demand digital forgetting situation. We consider the following situation:

Professor Watson has forgotten to associate an expiry date to the Activity IO at the time of its creation. After some time, she would like to remove this activity from the system, making it no longer available to her students.

As a first step, *PCAS* creates an Unpublish VDI which references the specific Publish VDI declared in Figure 4. An example of the Unpublish VDI, in scope, is given in Figure 6. In Figures 5 and 6, all namespaces declarations to Intellectual Property Management and Protection (IPMP), Rights Expression Language (REL), media streaming, and content identification are omitted.

The Unpublish VDI presented in Figure 5 is composed by the following elements:

- a Descriptor containing the original revocation request where a user requests the revocation of the specific Publish VDI in question;
- a Descriptor containing a license that grants the caller user the right to revoke the Publish VDI in question.

```
<?xml version="1.0" encoding="UTF-8"?>
<DIDL xmlns="urn:mpeg:mpeg21:2002:01-DIDL-NS">
  <Container> <Item> <Descriptor>
    <!--Descriptor containing the original revocation request -->
    <Statement mimeType="text/xml">
      <RevokeContentRequest>
        <mpegmb:TransactionIdentifier>12345</mpegmb:TransactionIdentifier>
        <saml:Assertion xmlns:saml="urn:oasis:names:tc:SAML:2.0:assertion" ID="Watson" Version="***"
          IssueInstant="2010-03-17T09:30:47.0Z"/>
        <RevokedEntity xsi:type="mpegmb:ContentEntityType">
          <dii:Identifier>urn:gu:convergence:server1/elda9c08-ef7fad62927f</dii:Identifier>
        </RevokedEntity>
      </Statement> </Descriptor> <Descriptor>
    <!--Descriptor containing a License that enables the caller user to revoke the Publication VDI -->
    <Statement mimeType="text/xml">
      <license> <inventory>
        <!-- Specification of the Governed Resource -->
        <digitalResource licensePartId="governedResource">
          <nonSecureIndirect URI="urn:gu:convergence:server1/elda9c08-ef7fad62927f"/>
        </digitalResource> </inventory> <grant>
        <!-- Specification of the Principal who is given a Right over a Resource. -->
        <keyHolder licensePartId="Watson"> <info>
          <dsig:KeyValue> <dsig:RSAKeyValue>
            <dsig:Modulus>KtdToQyyZA==</dsig:Modulus>
            <dsig:Exponent>AQABAA==</dsig:Exponent> </dsig:RSAKeyValue>
          </dsig:KeyValue> </info>
        </keyHolder>
        <digitalResource licensePartIdRef="governedResource"/> </grant> </license>
      </Statement> </Descriptor>
    </Component>
    <Resource mimeType="text/xml" ref="urn:gu:convergence:server1/elda9c08-ef7fad62927f"/>
  </Component> </Item>
</Container>
</DIDL>
```

Figure 6. Example of the Unpublish VDI for the on-demand scenario.

After the injection of the Unpublish VDI in the system, *PCAS* calls the method *handleRevocationVDI* that is offered by *Overlay TE*. This method is responsible for identifying the corresponding peers where the Publish VDI is available and for forwarding to them the Unpublish VDI. Any peer, upon reception of an Unpublish VDI, uses the *Match TE* to detect and remove the corresponding entry in table.

Furthermore, *PCAS* is also responsible for delivering a revocation command to *CoNet* requesting the marking of the Publish VDI in question as unavailable. For that, *CoNet* uses the same *revoke* method adopted in the first scenario, invoked on the local node.

6. Conclusions

The content-centric network paradigm offers a new paradigm centered around content distribution rather than host-to-host connectivity. This change from host-centric to content-centric brings some great benefits in terms of network load reduction, content search efficiency, and expansibility. Another great improvement that Content-centric networking (CCN) offers is the potentiality to implement an efficient and reliable digital forgetting mechanism.

Forgetting is currently a very relevant issue when we look to the World Wide Web. Facebook is enshrined in cyberspace for future employers to see. Google remembers everything we have searched for and when. The digital realm remembers what is sometimes better forgotten, and this has profound implications for us all. Therefore, a digital forgetting tool would be particularly welcome in the context of a DI life.

In CONVERGENCE, the digital forgetting allows users to completely and irreversibly remove a VDI from the network. This implies that the hosted VDI will be removed from all peers. For that, CONVERGENCE offers two digital forgetting mechanisms: the pre-planned digital forgetting and on-demand digital forgetting. The former is responsible to eliminate a VDI based in a default expiry date; the latter deals with online unpublishing requests from a publisher that intends to eliminate a particular Publish VDI.

As future work, it would be interesting to expand the digital mechanism to make it possible to eliminate associated and derived resources. Another topic of interest would be to conduct a quantitative research in order to test the performance and scalability issues of the digital forgetting at *CoNet* and *CoMid* using Ofelia or PlanetLab.

References

- B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher and B. Ohlman, "A survey of information-centric networking", *Communications Magazine IEEE*, 50(7), pp. 26–36, 2012.
- AMPAL, 2012. European commission issues paper on copyright in the digital economy. Available online at: <http://www.ampal.com.au/news-and-events/2012/12/6/european-commission-debate-on-copyright-in-the-digital-economy> (accessed 18 September 2013).
- L. Bannon, "Forgetting as a feature, not a bug: The duality of memory and implications for ubiquitous computing", *CoDesign*, 2(1), pp. 3–15, 2006.
- J. Bekaert and E. Knooning "Representing digital assets using MPEG-21 digital item declaration", *International Journal on Digital Libraries*, 6(2), pp. 159–173, 2006.

- CONVERGENCE, 2013. The CONVERGENCE project. Available online at: <http://www.ict-convergence.eu/> (accessed 29 April 2013).
- A. Detti, M. Pomposini, N. Blefari-Melazzi and S. Salsano, "Supporting the web with an information centric network that routes by name", *Computer Networks*, 56(17), pp. 3705–3722, 2012.
- S. Downes, L. Belliveau, S. Samet, A. Rahman and R. Savoie, "Managing digital rights using JSON", in *Proceedings of the 7th IEEE Conference on Consumer Communications and Networking*, Piscataway, NJ: IEEE Press, pp. 1065–1074, 2010.
- P. Druschel, M. Backes and R. Tirtea, 2011. The right to be forgotten – between expectations and practice. European Network and Information Security Agency (ENISA). Available online at: http://www.enisa.europa.eu/activities/identity-and-trust/library/deliverables/the-right-to-be-forgotten/at_download/fullReport (accessed 18 September 2013).
- R. Geambasu, T. Kohno, A. Levy and H. Levy, "Vanish: Increasing data privacy with self-destructing data", in *Proceedings the 18th Conference on USENIX Security Symposium*, Berkeley, CA: USENIX Association, pp. 299–316, 2009.
- P.K. Gkonis, C. Patrikakis, A.C. Anadiotis, D. Kaklamani, M.T. Andrade, A. Detti, G. Tropea and N. Blefari-Melazzi, "A content-centric, publish-subscribe architecture delivering mobile context-aware health services", in *Proceedings of Future Network & Mobile Summit*, 15–17 June, Warsaw, Poland, New York, NY: ACM, pp. 1–9, 2011.
- A. Hang, F. Almeida, H. Castro, M.T. Andrade, L. Chiariglione, N. Blefari-Melazzi and H. Hussmann, "Converging podcasts: A proposal for a content-centric approach for social learning environments", in *Proceedings of the Internal Conference on Information Society (i-Society)*, 25–28 June, London: IEEE Press, pp. 33–38, 2012.
- A. Haris, 2012. Different faces of information centric networking. Available online at: <http://netlab.pkusz.edu.cn/wordpress/wp-content/uploads/2012/05/Different-Faces-of-Information-Centric-Networking.pdf> (accessed 7 March 2013).
- V. Jacobson, D. Smetters, J. Thomson, M. Plass, N. Briggs and R. Braynard, "Networking named content", in *Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies (CoNEXT 09)*, 1–4 December, Rome, Italy, New York, NY: ACM, pp. 1–12, 2009.
- J. Lund, "The end of forgetting and administrative rights to our online personas", *IP Theory*, 2(2), pp. 83–96, 2012.
- B. Mathieu, P. Truong, Y. Wei and J. Peltier, "Information-centric networking: A natural design for social network applications", *Communications Magazine IEEE*, 50(7), pp. 44–51, 2012.
- A. Mohaisen, X. Zhang, M. Schuchard, H. Xie and K. Yongdae, "Protecting access privacy of cached contents in information centric networks", in *Proceedings of the 19th ACM Conference on Computer and Communications Security (CCS 12)*, 16–18 October, Raleigh, New York, NY: ACM, pp. 37–40, 2012.
- MPEG 21, 2012. MPEG 21 – ISO/IEC 21000 multimedia framework. Available online at: <http://mpeg.chiariglione.org/standards/mpeg-21> (accessed 12 March 2013).
- F. Pereira, J. Smith and A. Vetro, "Introduction to the special section on MPEG-21", *IEEE Transactions on Multimedia*, 7(3), pp. 397–399, 2005.
- D. Perino, M. Varvello and K. Puttaswamy, "ICN-RE: Redundancy elimination for information-centric networking", in *Proceedings of the 2nd Edition of the ICN Workshop on Information Centric Networking (ICN 12)*, 13–17 August, Helsinki, New York, NY: ACM, pp. 91–96, 2012.
- J. Rosen, 2010. The Web means the end of forgetting. Available online at: http://www.nytimes.com/2010/07/25/magazine/25privacy-t2.html?pagewanted=all&_r=0 (accessed 21 February 2013).
- L. Ruidong and H. Harai, "ANBR: An aggregatable named based routing for information-centric network", in *Proceedings of the 9th Asia-Pacific Symposium on Information and Telecommunication Technologies (APSITT)*, 5–9 November, Koganei, Piscataway, NJ: ACM, pp. 1–6, 2012.
- K. Saidis and A. Delis, "Type-consistent digital objects", *D-Lib Magazine*, 13(5/6), pp. 1–8, 2007.
- K. Sollins, 2008. Identity in an information-centric internet. CFP privacy and security working group. Available online at: http://cfp.mit.edu/publications/CFP_Papers/Identity%20Whitepaper%20v13_clean-3.pdf (accessed 18 September 2013).
- K. Sollins, "Pervasive persistent identification for information centric networking", in *Proceedings of the 2nd Edition of the ICN Workshop on Information Centric Networking (ICN 12)*, 13–17 August, Helsinki, New York, NY: ACM, pp. 1–6, 2012.
- C. Timmerer and H. Hellwagner, "MPEG-21 digital items in research and practice", *Proceedings of the 1st Digital Preservation Interoperability Framework Symposium*, 21–23 April, Dresden, New York, NY: ACM, pp. 1–8, 2010.