Ibrahim Elkadi

Impact of Content Caching on Competitive Dynamics of Internet Content Delivery Ecosystem

School of Electrical Engineering

Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Engineering

Supervisor:
Prof. Heikki Hämmäinen

Instructors:
M.Sc. (Tech) Nan Zhang
M.Sc. (Tech) Tapio Levä

Aalto University
School of Electrical
Engineering

AALTO UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

ABSTRACT OF
MASTER'S THESIS

Author: Ibrahim Elkadi

Title: Impact of Content Caching on Competitive Dynamics of Internet Content Delivery Ecosystem

Date: 28.6.2012 Language: English Number of pages: 69

Department of Communications and Networking

Professorship: Networking Technology Code: S-38

Supervisor: Prof. Heikki Hämmäinen

Instructors: M.Sc. (Tech) Nan Zhang, M.Sc. (Tech) Tapio Levä

The current Internet architecture was never designed to deal with the type of traffic it is carrying today. This has led to the rise of add-on solutions such as web caching and CDNs as well as new revolutionary networking paradigms such as Information-Centric Networking (ICN). The common theme in all these solutions is the extensive reliance on content caching. Content caching has already been studied extensively from the technology perspective; however, few have analyzed the influence of caching on the market dynamics between the main stakeholders of the Internet content delivery ecosystem. In this work an analysis of these influences is conducted using a bottom-up approach. The analysis which is based on information collected during expert interviewing, first determines the main parameters affecting caching, then identifies the major shifts in control points, through which ICN caching influence the Internet content delivery ecosystem and finally the analysis is mapped to real stakeholder market movements. The results of the analysis highlight the importance of both access and content providers who retain two critical control points irrespective of the caching architecture in use. They also shed light on two shortcomings that ICN architectures suffer from. The first is the content provider contractual complexity and the second is the domination of most ICN caching control points by the Internet Service Providers.

Keywords: Caching, ICN, Content Delivery

Preface

This Master's Thesis has been written as a partial fulfillment for the Master of Science Degree at Aalto University, School of Electrical Engineering. The work was carried out in the Department of Communications and Networking as a part of the Socio-economics Task Group (T-A.4) in Work Package A (WP-A) of the EU-funded research project SAIL (Scalable & Adaptive Internet Solutions).

I wish to express my gratitude to the people who have supported me in this work. First of all, I wish to thank Professor Heikki Hämmäinen for providing the opportunity to work in his team and write this thesis under his guidance. I am especially grateful to Tapio Levä and Nan Zhang for their valuable advices and support throughout the research process. Furthermore, I wish to thank the Networking Business team for their comments and discussion as well as the interview participants for their valuable insights.

Last but not least, I wish to address my gratitude to my family for their support during the course of my studies. Special thanks to my fiancée for her continuous moral support during the last two years of my studies.

Espoo, 27th June 2012

Ibrahim Elkadi

Table of Contents

List of Figures

List of Tables

Acronyms

3GPP	3rd Generation Partnership Project
ASP	Active Server Pages
BSC	Base Station Controllers
CDN	Content Delivery Network
DNS	Domain Name System
DONA	Data Oriented Network Architecture
GGSN	Gateway GPRS Support Node
HTTP	Hypertext Transfer Protocol
IANA	Internet Assigned Numbers Authority
ICANN	Internet Corporation for Assigned Names and Numbers
ICD	Internet Content Delivery
ICN	Information-centric Networking
IETF	Internet Engineering Task Force
IO	Information Object
IP	Internet Protocol
ISP	Internet Service Provider
LFU	Least Frequently Used
LRU	Least Recently Used
LTE	Long Term Evolution
MGw	Media Gateway
NbR	Name-based Routing
NDN	Named-data Networking
NetInf	Network of Information
NR	NetInf Router
NRS	Name Resolution System
P2P	Peer-to-Peer
POP	Point of Presence
PSIRP	Publish/Subscribe Internet Routing Paradigm
QoE	Quality of Experience
QoS	Quality of Service VoD Video on Demand
RAN	Radio Access Network
RNC	Radio Network Controllers
RWI	Real World Identity
SLA	Service Level Agreement
SMTP	Simple Mail Transfer Protocol
UMTS	Universal Mobile Telecommunication System
URL	Uniform Resource Locator
VNI	Visual Networking Index
VoD	Video on Demand
WiFi	WirelessFidelity

1 Introduction

The Internet usage has changed drastically since its inception in the 1960s. It began as a purely research based network to interconnect research organizations but has grown to become an integral part of people's lives (Leinar, et al., 1997). Today's Internet delivers all types of traffic that was once delivered by other channels. From online shopping to Video-on-Demand (VoD), the thirst for higher and better content delivery is never quenched (Sandvine, 2012).

In recent years, Internet Protocol (IP) traffic has grown exponentially (Sandvine, 2012; Cisco 2011) due to the continuous increase in the Internet penetration levels and speeds. According to Cisco (2011), the number of networked devices globally in 2010 was equal to the global population, i.e. one networked device per capita, and is expected to double by 2015. Moreover, access speeds are increasing constantly and broadband Internet is experiencing a huge growth (Akamai, 2010).

Not only is the IP traffic growing but the nature of traffic traversing the Internet is also witnessing strong consolidation towards certain types of traffic. Figure 1 shows the change in Internet traffic mix from 2009 till 2011 as reported by Sandvine (2011) for Europe's fixed access networks. It can be easily seen how the share of Real-time Entertainment traffic such as movies, video clips and music is rising constantly from year to year.

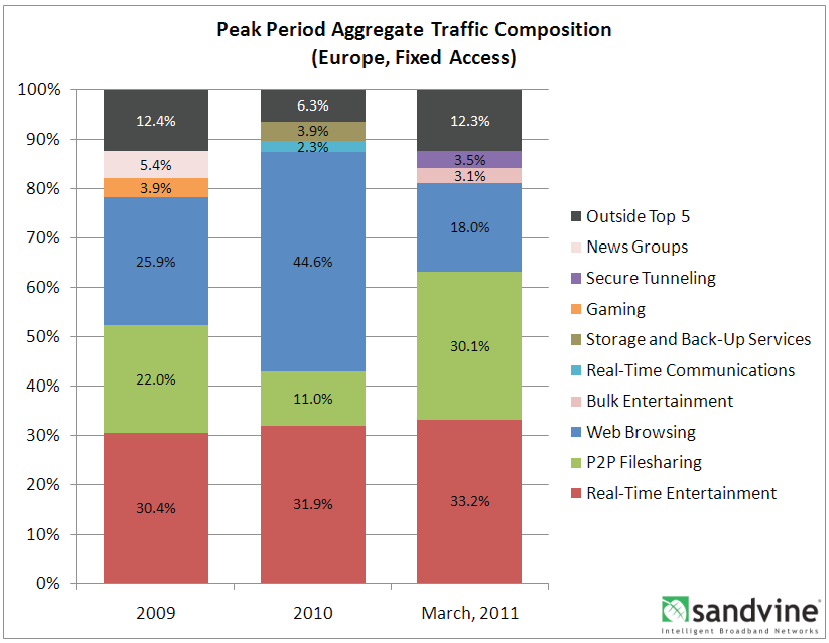

Figure 1: Europe Traffic Mix - 2009-2011 (Sandvine, 2011)

According to Cisco's Visual Networking Index (VNI) report (Cisco, 2011), the global Internet video traffic will account for over 50% of consumer IP traffic by 2012 and 61% by 2015. In 2010, video traffic surpassed peer-to-peer traffic to become the largest traffic type in the Internet. This growth in video traffic is driven by many factors. The main factor is the rise in the broadband penetration levels

allowing users to stream and download videos more conveniently than ever. Companies like Netflix (Netflix, 2012) and BBC (BBC, 2012) have built services that make use of the available high bandwidth, hence driving consumers more and more towards Internet video (Sandvine, 2012). Furthermore, websites offering short, user-generated videos like YouTube have also contributed heavily to this rise in video traffic (Sandvine, 2012).

Due to the surge in traffic volume fuelled by the increase in video traffic and migration of premium content to the Internet, the competitive dynamics between the different stakeholders in the Internet content delivery (ICD) ecosystem have changed significantly over the last few years. Network operators, fixed and mobile, are trying to avoid being turned into bit pipes as the exponential growth in video traffic is driving their networks to the limit, while more and more content providers are resorting to content delivery networks (CDNs) to deliver their content to end-users who are constantly expecting better quality.

Content caching is one of the key components that influence the competitive dynamics between the three main stakeholders in the ICD ecosystem: telecom operators, content providers and CDNs. This influence is translated through major control points "at which management can be applied and which can be rooted in business, regulatory, or technical regimes" (Riihijärvi, et al., 2009).

Moreover, with the rising interest in information-centric networking (ICN), which heavily depends on content caching, the control points are expected to shift. Such shift will lead to a change in the competitive scene between the three main stakeholders. Thus, the aim of this work is to analyze the influence of caching parameters as well as the impact of ICN on the competitive scene of the ICD ecosystem.

1.1 Research Question

Content caching is a platform through which different ICD architectures compete. This makes it a critical component to understand both on the technology level as well as on the ecosystem level, the latter being the focal point of this work. Hence, this work sets out to answer the following main question:

> How does content caching affect the competitive dynamics of the ICD ecosystem in the current host-centric networks and the future information-centric networks?

The answers for the main research question are reached by meeting the following objectives:

- Identifying the major caching architectures used in ICD
- Understanding how mobility and content naming influence caching
- Identifying all the parameters that influence the cacheability of content
- Understanding the role of caching in ICN in comparison with its role in host-centric networks
- Identifying the key control points of a caching system
- Identifying the shifts in caching control points due to the introduction of ICN
- Identifying the major structural movements in the current ICD market and its effect on the ICN adoption

1.2 Research Scope

This work takes a holistic view on the ICD market. On that account, any Internet based system which can be used to deliver content on a wide scale is considered as a content delivery system. This implies that content delivery networks (CDNs) are just one type of such systems and should not be mixed with the holistic meaning. Moreover, this work addresses ICN as one type of content delivery architecture rather than a technology which facilitates the comparison with other delivery architectures.

Since the main objective is to understand the influence of caching on the competitive scene, this work remains on the system level without diving deep into technical aspects such as caching algorithms and routing protocols. However, a basic technical understanding of Internet and caching architectures as well as routing protocols is needed in order to gain a complete understanding of the content of this work. Caching in this work refers to the act of storing of content in a place other than its original source with the aim of improving the delivery process.

Furthermore, because this work is conducted under the SAIL (SAIL, 2012) project, which adopts the NetInf (SAIL, 2011) reference architecture for ICN, any discussion regarding ICN in this work has its roots in the NetInf architecture. Aside from that reason, NetInf is a good platform for any discussion on ICN since it encompasses two different routing systems: Name Based Routing (NbR) and Name Resolution System (NRS) (D'Ambrosio and Dannewitz, 2011), while most of the other ICN architectures implement only one variation of the two.

1.3 Research Methods

In this work, two research methods are adopted. The first is an extensive literature review on topics covering the areas of caching algorithms, information centric networks, Internet architecture and Internet economics.

The second method is semi-structured interviews. The interviews are conducted with both technical experts on the Internet architecture as well as experts working on the business side of the industry. The main objective of these interviews is to build an understanding of the new ICN architectures by drawing insights on how the different stakeholders of the ICD ecosystem view ICN. This is in addition to drawing further insights from their predictions regarding the future shape of the market based on current competitive movements.

1.4 Structure of Thesis

The thesis will follow the structure shown in Figure 2:

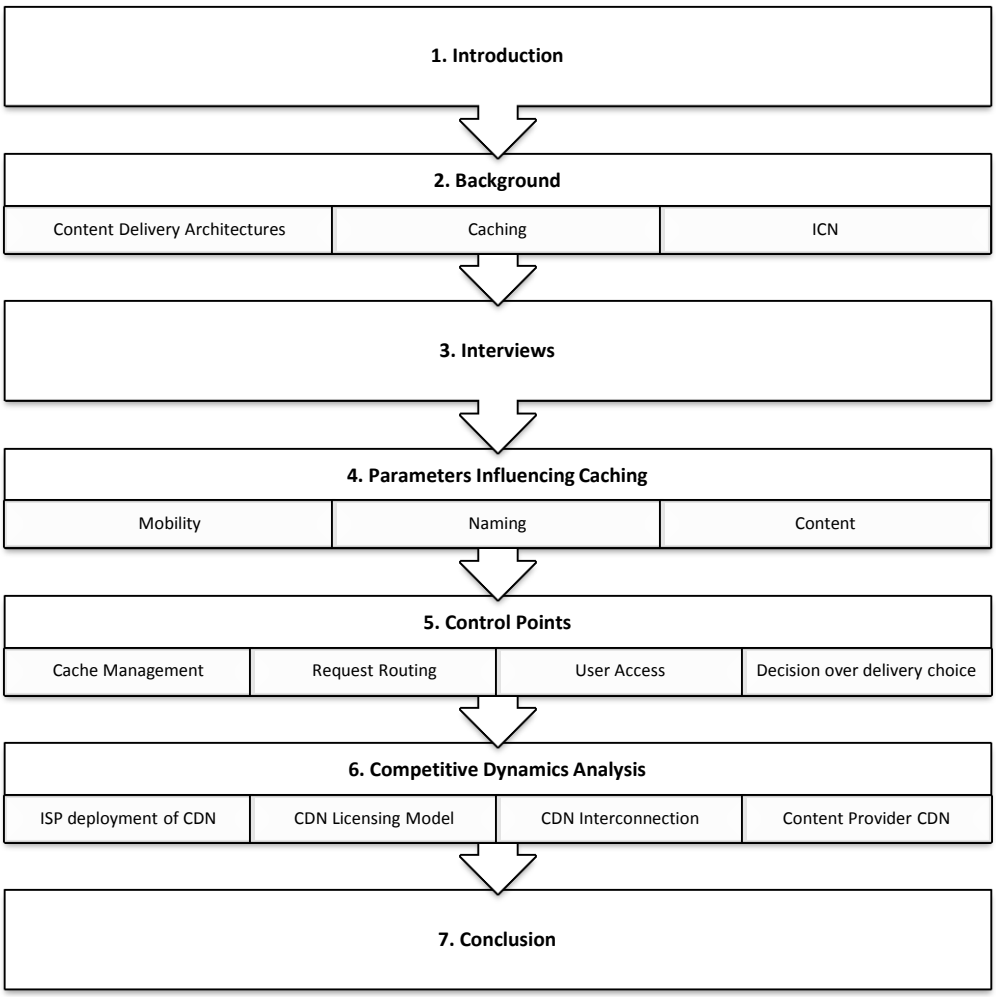

Figure 2: Thesis Structure

This chapter is followed by the background chapter that briefly discusses some of the key concepts, on which this work is built. Afterward a summary of the results of the interviews conducted during this work is presented in Chapter 3. In Chapter 4 an analysis of the major parameters influencing caching is presented. This is followed by the analysis of the caching control points and their shifts in Chapter 5 and analysis of the competitive dynamics of the ICD market in Chapter 6. Finally, the work is concluded by Chapter 7.

2 Background

This section gives an overview of the background topics on which this work is built upon. A survey of the major content delivery architectures is presented. This is followed by an overview on content caching and a presentation of the identified cacheability parameters. Finally, a brief introduction of Information Centric Networking concept and one of its implementation, the NetInf, is presented.

2.1 Current/Traditional ICD Architectures

A variety of content delivery architectures have been utilized to deliver content over the Internet. Strictly hierarchical systems adopting Client-Server architectures such as the World-Wide Web were the norm in the early days of content delivery. However, other systems emerged such as the highly controlled distributed Content Delivery Networks (CDN) (Vakali and Pallis, 2003) and the less controlled client-active Peer-to-Peer (P2P) distribution systems (Lua, et al., 2005).

2.1.1 Client-Server Architecture

The traditional model for ICD is based on the Client-Server architecture. In this model, a client running on an end-user's machine requests web objects from a web server. The request which is usually in the form of a Uniform Resource Locator (URL) address first passes by the end-user's access network. If needed, the access network translates the human-readable URL address into an IP address using a Domain Name System (DNS) and forwards the request to the web server holding the requested object through the core network as shown in Figure 3 (Funkhouser, 1996). Many of the Internet's main application protocols such as HTTP, SMTP, Telnet and DNS use this model (Reese, 2000).

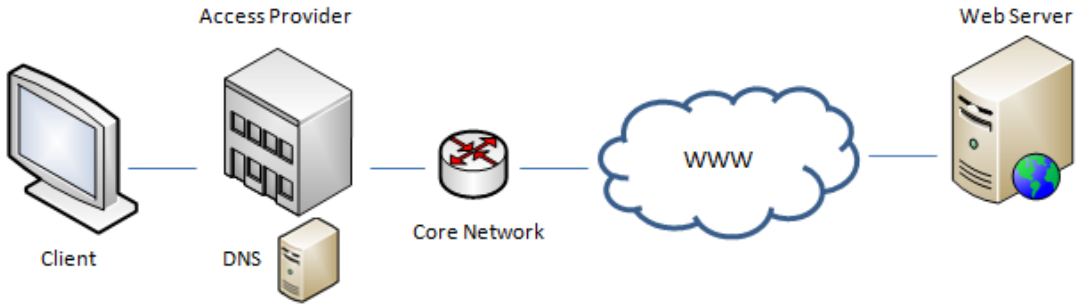

Figure 3: Client-Server Architecture

The Client-Server model offers several advantages such as: high security due to the tight control over the content, low capability requirements for clients since they do not need to store or distribute content and centralized management facilitating processes such as taking of backups, upgrading and recovery (Funkhouser, 1996). However, it also suffers from critical issues such as: risk of server overload due to high number of simultaneous client requests, network congestion since for each client request the server will reply with a copy of the same content filling the network pipes with redundant bits (Anand, et al., 2009) and, scalability problems due to the cost and management issues arising from the deployment of servers in multiple locations (Menascé, 2001).

2.1.2 Content Delivery Networks (CDNs)

CDNs are dedicated fully managed overlay networks. A CDN is comprised of a network of distributed surrogate servers strategically placed near clients' access networks offloading the content owners' origin servers by delivering the content on their behalf (Dilly, et al., 2002) as shown in Figure 4.

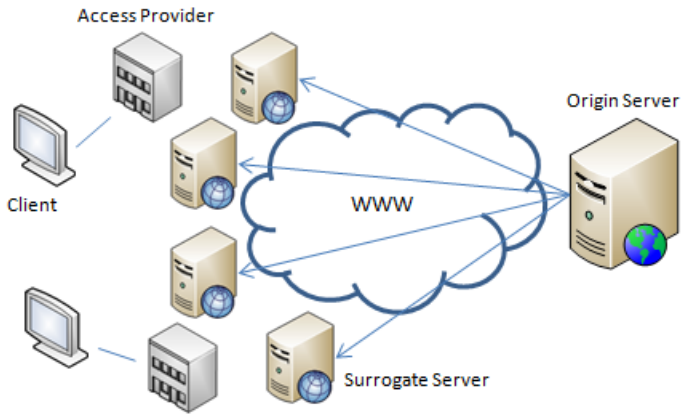

Figure 4: CDN Architecture

There are four basic steps in the CDN delivery process (Dilly, et al., 2002):

1. Surrogate servers cache the origin server's objects based on the service agreement between the CDN provider and the content provider.
2. The client requests the object in the same manner as in the client-server architecture.
3. The request is routed towards the most appropriate surrogate server using methods such as DNS redirection and URL rewriting (Vakali and Pallis, 2003).
4. The object is delivered from the closest and least crowded topological surrogate server to the client.

CDNs offer multiple benefits such as reduced origin server load, reduced latency for end-users, and increased throughput (Pallis and Vakali, 2006). Moreover, CDNs store only specific content based on agreements with content providers in contrast to web caches, which choose what to cache based on autonomous caching algorithms as explained later in this Chapter. Hence, CDNs can improve the delivery of typically uncacheable objects such as secured, streaming and dynamic contents (Vakali and Pallis, 2003).

A CDN can be deployed locally inside a single Internet Service Provider (ISP) or globally in many ISP points of presence (POPs) (Vakali and Pallis, 2003). The more globally distributed a CDN is, the higher in the network hierarchy its surrogate servers are located and vice versa. This comes from the fact that when the CDN is deployed locally (by an ISP for example), the ISP will not oppose having surrogate servers deep in their network closer to its clients since it retains full control over them. On the other hand, a global CDN provider will in most cases be denied access to deeper levels of other ISP networks since these ISPs will hesitate to relinquish some of their control over the content flow in their networks.

2.1.3 Peer-to-Peer Content Delivery Architecture

Peer-to-Peer (P2P) networks consist of self-organized peers forming a content delivery structure overlaid on top of the IP networks (Lua, et al., 2005). Unlike the

Client-Server and CDN architectures, the end-user devices in P2P networks are actively involved in the distribution process of the content as they act as servers besides their traditional role as clients (Figure 5).

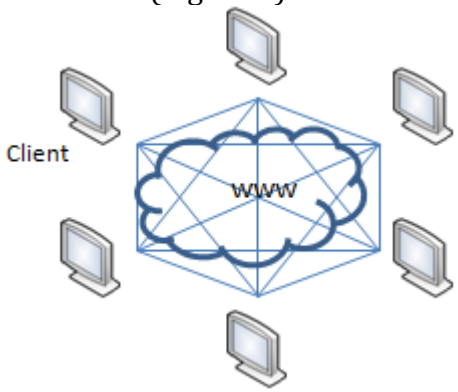

Figure 5: P2P Content Delivery Architecture

P2P networks introduce a mix of features, such as robust wide-area routing architecture, efficient research of data items, selection of nearby peers, redundant storage, permanence hierarchical naming, trust and authentication, anonymity, massive scalability, and fault tolerance (Lua, et al., 2005).

There are two main classes of P2P overlay networks: *structured* and *unstructured*. Structured P2P networks are tightly controlled as content is carefully placed in specified locations that optimize the querying and serving of subsequent requests. The once famous P2P file sharing network Napster is an example of a structured P2P network which adopted a centralized search system based on file lists provided by each peer. Such approach allowed it to scale well since the more peers joined the system the higher the aggregate download capability became without requiring much bandwidth for the centralized search. On the other hand, the system has a single point of failure due to the search centralization.

In contrast to the structured approach, unstructured P2P overlay networks like Gnutella adopt a completely decentralized system where both search and download capabilities are distributed among peers. Peers join the network following some loose rules. When a requesting peer is searching for an object it floods parts of the network with queries (Lv, et al., 2002). Upon receiving the flood query each network peer replies with a list of all matching objects. Although this system is robust against failures due to its fully distributed nature is suffers from two shortages. First, due to the flooding mechanism used in searching for items the unstructured P2P network cannot scale very well compared to the structured one (Lua, et al., 2005; Lv, et al., 2002). Second, while the system is reliable in finding highly replicated objects, it performs poorly when locating rare items (Lua, et al., 2005).

2.2 Caching

Caching is a technique utilized by different stakeholders in the Internet ecosystem by which they can limit the amount of traffic on certain links and speed up the process of retrieving content on the network. The cache normally lies in an intermediate position between the content consumer and content provider and it works on replicating and storing the content of the content provider allowing a quicker and more efficient access to subsequent consumer requests leading to better quality of experience for the end-user and cost savings for the service provider.

Caching mechanisms can be implemented in different parts of the network. Generally, they are deployed near the content consumer, near the content provider or at strategic points in the network (Barish and Obraczka, 2000). When deployed near the content consumer, web caches can reduce the traffic flowing between different ISPs and hence lowering the high costs usually associated with such transit traffic with an enhanced delivery to the end user. Another option is to deploy the cache directly near the content provider improving scalability and availability as different instances can retrieve the same content quicker from the cache rather than original server. The last option is to deploy the cache somewhere in the middle between the consumer and the content provider which can give some of the advantages of the previous two deployment configurations. In reality, all three types of deployment co-exist providing different benefits to the different stakeholders in the ecosystem.

2.2.1 Client-side Caching

One of the earliest and most basic caching mechanisms to be deployed was the web browser caching. Web browsers can do caching on a per-user basis using the local file system (Barish and Obraczka, 2000), storing the content the first time a user accesses it and serving the content from the local file system for subsequent requests. This method of caching was very efficient when the internet was still dominated by static content, however, with the rise of rich internet applications that reconstruct web objects on the fly, browser caching was deemed inefficient.

Another client-side caching mechanism is the P2P caching. The rise in the storage and processing capabilities of client terminals has allowed using these terminals to store and forward content directly among themselves without the need for a central server unit. P2P caching is the basis of P2P content delivery architectures (2.1.3).

P2P's popularity as a caching architecture has grown a lot in recent years. With the exponential increase in internet traffic due to the explosion in the usage of video traffic, many have suggested P2P caching as a solution for reducing expensive transit traffic by exploiting the locality of interest for internet content. P2P client caches temporarily store frequently-requested content and subsequent requests are served from the client cache rather than from the original server. Client-side caching releases the pressure off the original server who only needs to transfer the content to few clients who will then share it among the rest of the P2P network. From the network perspective, P2P client caching can reduce transit traffic only if the P2P caching mechanism attempts to fetch and deliver content from peers on the same network (Zhu and Hu, 2003).

2.2.2 Server-side Caching

Caches can also be deployed at the server end of the network; such architecture is often known as "Reverse Proxy Caching" (Barish and Obraczka, 2000). Reverse proxy caching can play a decisive role in reducing the load on the server holding the content, especially those servers expecting a high number of requests and want to maintain a high level of service and availability (Barish and Obraczka, 2000). Similar to most types of caching, reverse proxy caching, exploits the Zipf-like popularity distribution that most content exhibits to cache popular content (Hefeeda and Saleh, 2008). Different content types can be cached in this architecture. For static content, caching is an easy and straightforward process as

the content does not change quite often. On the other hand, dynamic content remains a challenge for server-side caching.

Different studies have been made to investigate caching techniques for dynamic content, among those efforts, Datta (2001) introduced a server-side caching engine that caches dynamic page fragments in "order to reduce dynamic page generation processing delays on a Web site" and Zeng and Veeravalli (2008) proposed a novel server-side web caching for multimedia applications. Other efforts introduce server-side caching enhancements targeting specific services like web map services (Zhang, Li and Zhu, 2008) and online auctions (Menascé and Akula, 2007). Moreover, companies like Microsoft, through its ASP.NET web application framework (Microsoft, 2012) and IBM (IBM, 2012) are offering proprietary server-side caching solutions, while there are also free and open source solutions like "Memcached" (Memcached, 2012).

2.2.3 In-Network Caching

The last category of caching architectures is the in-network caching. In-network caching is currently the dominating caching architecture because it combines the advantages of both client-side and server-side caching. Deploying caching servers in the ISPs' networks, for example, would reduce the transit traffic flowing outside the ISP's network hence benefiting both the customer, who gets a quicker local response for his/her requests from the cache, and the ISP itself, who will save costs of the expensive transit traffic that it would otherwise has to pay for. Moreover, CDNs which can be considered as an in-network caching architecture, provide scalability and better user Quality of Experience (QoE) for content providers (Pathan and Buyya, 2006).

There are mainly two broad categories of in-network caching: agreement based caching and transparent caching (Salo, 2011). Agreement based caching is based on a business relationship between the content provider, who needs his content to be cached and the cache server owner, who charges the content provider for caching its content. This category is represented mainly by CDNs like Akamai and Limelight. However, this category also includes content providers who have their own CDNs, like Google and Facebook. On the other hand, transparent caching is agnostic to the underlying content it is caching; the main aim of transparent caching is to provide an overall benefit from caching rather than raising the efficiency of certain content. Web caching (Barish and Obraczka, 2000) and P2P network caching (Zhu and Hu, 2003) are two examples of transparent caching.

In-network caching architectures can also be classified as either cooperative or non-cooperative (Hosanagar and Tan, 2006). Non-cooperative caching architectures are stand-alone caches that serve two purposes only: deliver the content requested if it was found in the local storage or forward the request to the remote server in case of a miss. Stand-alone proxy web caches belong to this category. They can be deployed in a transparent or a non-transparent manner. In non-transparent architectures, the client is forced to forward any requests to the cache server which acts like a gateway, while in transparent architectures; a router intercepts the traffic and forwards it to the cache (Barish and Obraczka, 2000). On the other hand, cooperative caching allows "an array of distributed caches to cooperate and serve each other's' web requests" (Yu and Macnair, 1998).

Similar to a non-cooperative cache, when a cooperative cache receives a request, it first looks it up in its local cache. If the requested information object was

not found, the cooperative cache resorts to a secondary level of caching by interrogating other caches in its network searching for the information object before it requests it from the original remote server. There are many caching mechanisms that exploit such architecture; adaptive web caching "consists of multiple distributed caches which dynamically join and leave cache groups based on content demand" (Barish and Obraczka, 2000); push caching keeps cached data near the requesting clients by dynamically mirroring the data to close by caches (Barish and Obraczka, 2000).

2.3 Information-centric Networking (case: NetInf)

Information-centric networking (ICN) is a new networking paradigm in which the routing is based on the content rather than the location of the content. Such content location independence makes the task of distributing and exchanging information more efficient compared to today's host centric approach (Ahlgren, et al., 2011) as it reduces the redundancy of the traffic and the distance between the user and the content. In order to achieve this independence, ICN is built around two major themes. The first is the ubiquitous in-network caching of information objects (IO) and the second is a namespace to name and identify information in a manner which is independent of the storage location (Dannewitz, 2009).

In ICN, IO refers to the group of all representations of a certain piece of information. Hence it can span multiple different representations, for example, a video file with different encoding formats, as well as multiple copies, for example, a copy in a web cache or in a user's device. This enables the network to identify any piece of information irrespective of its location and irrespective of its representation (SAIL, 2011).

In order to achieve independence of location, the security of the data is not anymore an add-on to the network architecture where it is relies on trusting the sources of information; rather, the security is integrated with the IO itself (SAIL, 2011). Such security integration is implemented in the naming scheme of the IOs. Hence, a name does not only identify the IO but it is the means to authenticate it as well.

There are many proposed architectures for ICN (Ghodsi, et. al, 2011). Some of them adopt a clean slate approach such as the Data-Oriented Network Architecture (DONA) (Koponen, et. al, 2007) and the Publish/Subscribe architecture (Eugster, et al., 2003) proposed in PSIRP/PURSUIT (Trossen, et al., 2011). While others proposed architectures that can be deployed over the current Internet such as the CCN (Jacobson, et al., 2012) architecture proposed in NDN (Zhang et al., 2010) and the NetInf architecture (Dannewitz, 2009) proposed in 4WARD (Aranda, et al., 2010) and SAIL (SAIL, 2012) projects.

Figure 6 shows a high level presentation of the NetInf architecture.

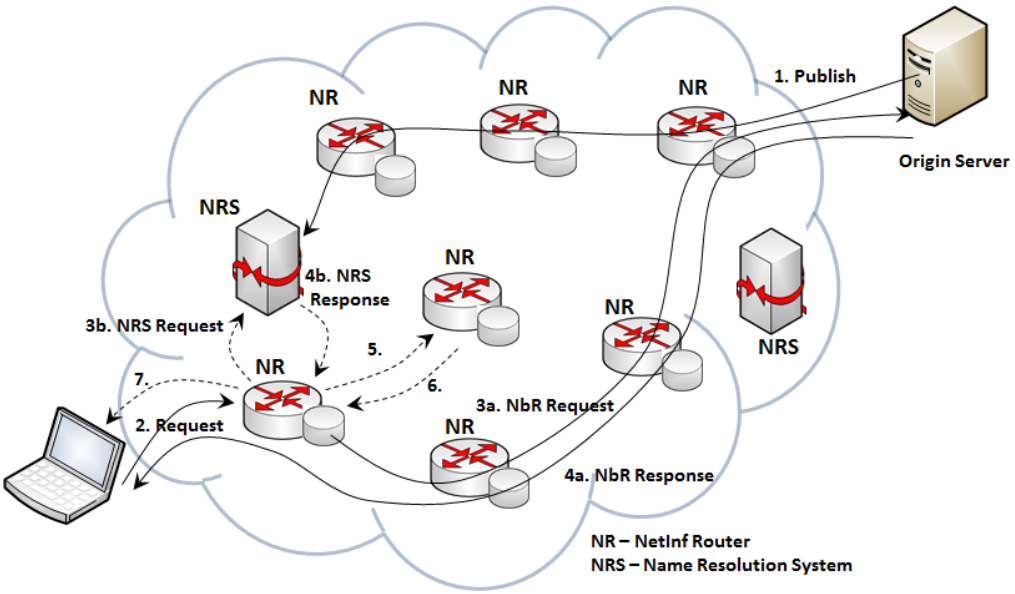

Figure 6: High level Description of NetInf

The NetInf reference architecture adopts two routing mechanisms. One of them is the Name-based Routing (NbR), which is built around the idea of caching content on the same path on which the response for a request goes through. The other one is the Name Resolution System (NRS) routing, which relies on an extra server (the NRS) that keeps track of the caching locations of the IOs in a similar fashion as today's DNS servers. Figure 6 illustrates both routing mechanisms: first (1) the publisher publishes the IO by adding it to the NRS registry. When a (2) request comes for the IO the network can (3a) route the request towards the original server with the possibility of finding the IO cached in any of the NetInf router on-path, which is an example of NbR. Then (4a) the response will be cached in all NRs on the way back to the requester (same path as travelled by the request). Another possibility is that the NR receiving the request (3b) forwards the request to the NRS which will (4b) reply with a set of locators of nearby NRs which have the requested IO already cached in its storage. The receiver NR then (5) requests the IO directly from the nearest NR, (6) gets the reply from the caching NR and finally (7) serves it to the requester.

3 Interviews

Informational interviews were conducted with experts representing four stakeholders in the ICD market: network operators, equipment vendors, research institutions and a commercial CDN operator/provider. The aim of the interviews was to obtain a better understanding of the current content delivery market, a better understanding of ICN (focusing on NetInf) and information about caching and cacheability of content.

3.1 Interview Process

There are mainly three methods to use when conducting informational interviews: structured interviewing, semi-structured interviewing and unstructured interviewing (Robson, 2002). In structured interviewing the same set of questions in the same order and words is used when interviewing. The interviewers does not add or remove any question which means the flow of the interview from beginning to end is fully known beforehand. Likewise, semi-structured interviewing method depends on a pre-defined set of questions. However, the set of questions act more or less as guidelines that loosely control the flow of the interview. Based on the interviewee's answers to the questions, the flow of each interview can be different. The third method of interviewing is the unstructured method. In this method the interview is quite informal and the conversation can develop freely within the interest area.

This work employs a semi-structured interviewing method based on the list of questions in Appendix A. The interviews were conducted in January 2012. The interview questions are divided into four topic areas: content caching, CDNs, NetInf and naming in ICN and competitive dynamics questions. In total, nine interviews were conducted each lasting from 45 minutes to 1 hour. Two interviews were conducted in the headquarters of a global network operator (NO1-NO2), two with persons from research institutions working on ICN architectures (RI1-RI2) and four with persons from equipment vendor companies involved in ICN projects (EV1-EV4). And finally one interview with a previous manager in a global commercial CDN operator/provider (CC1).

Only three interviewees had before been involved with CDNs either by being an employee in one of them, collaborated with them in partnerships in the process of building their own CDN. The interviewees are listed with their position and referencing in Table 1.

Table 1: Table of Interviewees.

Field of Expertise	Position in Company	Referencing
Network Operator	Content Delivery Program Manager	NO1
Network Operator	Research Unit Team Manager	NO2
Equipment Vendor	Senior Research Engineer	EV1
Equipment Vendor	Senior Researcher	EV2
Equipment Vendor	Researcher	EV3
Equipment Vendor	Senior Specialist	EV4
Research Institution	Research Scientist	RI1
Research Institution	Research Scientist	RI2
Content Delivery Network Provider	Ex-Senior Service Line Manager	CC1

The chosen Network Operator is a global one and offer access services as well as backbone interconnectivity services. Moreover, this operator is currently building its own CDN based on its own proprietary solutions.

The interviewees (EV1-EV4) work for three global network equipment vendors who sell their network equipment for both fixed and mobile networks. Their main focus is on the access network equipment rather than backbone equipment, but at least one of them is working actively to enter the backbone equipment market. While, RI1 and RI2 are researchers conducting extensive research in future Internet architectures.

CC1 is the only interviewee who has not worked in the field of ICN before. He is currently working for a new multi-platform management consultancy firm and hence deals quite closely with content providers. Furthermore, he was a senior manager in a global CDN company and hence knew their strategic thinking and their stance regarding ICN.

Most of the interviewees were at one point or another involved with ICN projects and hence had deep knowledge of its architecture and market potential. However, since most of the interviewees worked or were working in the SAIL project they were biased towards exaggerating such potential.

3.2 Results

This section presents a summary of the interviews. The summary is divided based on the four topics discussed during the interviews: caching, NetInf and naming, pure play CDNs and competitive dynamics. CDNs are highlighted in these interviews due to the important role they have in the value network of the ICD ecosystem as they have business relationships with both content providers and ISPs.

3.2.1 Caching

There were many reasons cited by the interviewees for the failure of web caching (especially cooperative caching) as a solution for efficient content delivery. One of

the main reasons was a large trace-based study done by Wolman et. al (1999) which showed that there is little point in deploying cooperative caches beyond the level of medium-sized cities since the benefits do not scale with costs. Other reasons cited were the drop in bandwidth prices, the change in traffic nature from static content to dynamic content, the lack of persistent naming and the failure of cache interconnection standardization efforts.

Although most interviewees agreed that web caching in general did not take off as expected, large network operators (e.g. NO1) do deploy them on a local scale to deal with highly popular content as well as to shape the traffic of P2P file sharing networks which generate a lot of transit traffic and hence increase the costs. Some interviewees, such as CC1 and EV1, did not even agree with the "failure of caching" statement and hinted that commercial CDNs are in the end one type of caching and it is highly successful and hence the problem was finding the right business model rather than the right technical solution. All interviewees also agreed that the traffic profile is drastically changing (e.g., static video traffic dominates the internet) and hence it is worth revisiting the earlier caching studies.

When asked about the current CDN interconnection efforts and why the outcome would differ from the previously failed cache interconnection efforts, EV1 and NO1 stated that the scope is quite different. CDN interconnection efforts seek to connect networks rather than connecting just caches. Also, the fact that CDNs work mostly on proactive caching means they know beforehand where they will place certain content making collaboration an easier task than in the case of on-path caching. Since, on path caching depends on caching algorithms which are difficult to co-exist with other algorithms from other caches.

As part of the caching questions, the interviewees were presented with a list of the types of traffic and asked to rank each content type by its level of cacheability. As expected, and supported by recent studies, most content was ranked as highly cacheable except for clearly non-cacheable content such as Real-time communications. One significant finding is that all interviewees ranked Live Video Traffic as medium to highly cacheable type of traffic. Such finding supports heavily the efforts done by ICN projects such as NetInf for enhancing the delivery in flash crowd scenarios by depending on caching.

On a technical level when asked about where in the proposed NetInf network will caching occurs, EV1-EV2 and RI1-RI2, all agreed that although the NetInf architecture is still a work in progress however, the goal is to make any node in the network, this includes end-user devices, capable of caching and distributing content. Hence, the ubiquitous caching of NetInf might not only replace the functionalities of web caching and CDNs but also extend to cover delivery schemes such as P2P content delivery.

3.2.2 NetInf and Naming

One general observation from the answers received for questions in this category is the sometimes conflicting replies to the same questions. This is certainly due to the fact that not all NetInf architectural components and functions are well defined yet since NetInf development is still in a very early phase. However, answers were consistent for high level questions about ICN.

One critical aspect of NetInf is the unique naming of information objects. IR1 and EV1 emphasized the fact that in NetInf architecture there will be no need for a centralized service that ensures the uniqueness of names similar to today's host

centric architectures. Instead, hash functions generating the names would provide acceptably statistically unique names. This means that entities such the ICANN and IANA who store all the addresses on the Internet will not be needed and their functions will be distributed among the different NRSs. However, there was no consensus regarding which mechanism would be utilized to ensure that all names are globally unique. Experts divided into two groups: one group adopting the opinion of using hash tables to produce statistically unique names while the other group's viewpoint was that the network will indeed require a global name resolution system which ensures such uniqueness.

Another point which was discussed during the interviews relates to the translation between non-human readable self-certified names and human readable names. Although some interviewees (IR2 and NO2) agreed that such translation service, although not yet implemented or planned, should exist in order to provide the same addressing and easiness to end-users like in today's host-centric networks. Yet, the experts on this topic (IR1 and EV1) stated that such translation service would not be needed since already today most end-users search for links on search engines instead of typing full URLs. Moreover, today's links are already complicated and in many cases not human friendly. Also, implementing such translation service means that the certifying process will need to be built-in the service making it more complex.

As for the effect of the naming on current search engines, all interviewees agreed that the introduction of named objects will not affect search engines much since their crawlers already move from webpage to another through in-page links so whether these links are regular URLs or self-certified names will not make much of a difference. On the other hand, the introduction of named objects might lead to new ways of searching. For example it could be possible to perform searches that identify the relationships between objects and leverage them to get better results. Also, search results might be ranked by the number of replicated copies of the object in the caches of a certain geographical area allowing for locally optimized search results.

In response to questions regarding the misuse of the NRS by NetInf providers to direct requests towards certain caches over others, most of the answers challenged such an assumption. The majority of interviewees (especially NO1 and NO2) stated that such misuse can is possible also in today's network since the network is the one that chooses where to route requests. Also, they emphasized that the interest of the network and the end-users are usually aligned and hence the optimum route (or cache) for the ISP to serve the content is usually the optimum from the end-user's perspective as well.

All interviewees agreed that the management of caches in an ICN is much simpler and cheaper than in today's web caches or CDNs. This goes down mainly to the fact that unique naming will provide an easy way to identify different instances of the same information object and hence caching decision-making will be simpler.

One final comment about the NetInf was that although caching is the most visible function of it, it does not mean that it is the main motivation for the adoption of NetInf. According to RI1, the benefits of NetInf go far beyond caching. Inherent multicasting capabilities and ad-hoc network formation in flash crowd and disaster scenarios are just examples of other benefits of NetInf deployment.

3.2.3 Pure Play CDNs

The second set of interview questions was related to the CDN market. Not all interviewees had experience with CDNs and much of the answers were at best personal opinions and guesses. Generally, the pure play CDNs stance on ICN is a bit vague. Most of the interviewees did not know whether CDNs do internal research regarding ICN architectures or not, but it is obvious that they are not partnering in big research projects on this topic although it is quite related to their own business. CC1 stated that ICN is not really a new concept for CDNs and that the name based routing and caching techniques proposed are to a large extent implemented already inside CDN networks albeit on current Internet protocols rather than on new proposed ones. CDNs in general put much more emphasis on proprietary research rather than collaborative standardized research.

Regarding CDNs relationships with ISPs, CC1 and EV1 stated that the type of relationship depends heavily on the size of the ISP. For smaller small ISPs, peering or hosting a CDN can add a lot of value to their networks and hence most small ISPs allow CDNs to be located deep in their network. On the other hand, most large ISPs are reluctant to do so and usually CDNs are hosted only at the top levels (backbone). This was confirmed by NO1 who's company does not allow CDNs to deploy inside the network but rather can only peer with the network on the backbone exchange levels. Hence, any content distribution system that can go deeper towards the edge might be able to provide higher value and benefit from a competitive advantage versus the current global pure play CDN players.

On the other hand, the business relationship between CDNs and content providers is quite direct. Content providers pay CDNs to deliver their content and get charged either by the average speed in Gbps/Mbps (usually on 95/5 rule) or by the GB/TB delivered. However, according to CC1, the market is getting more fractured and commoditized, which gives rise to brokerage services that cut the direct relationship between the content providers and CDNs.

When asked about the recent trend of ISPs deploying their own CDNs and its effect on the CDN market, almost all interviewees agreed that this is the biggest threat on the current pure play CDN players. Moreover, CC1 and NO1 highlighted the licensing model that many big pure play CDN players such as Akamai and Edgecast are currently adopting as a sign of the eminent threat of such rise in ISP owned CDNs. According to NO1 and NO2 the benefits that an ISP might get from licensing a CDN rather than building one from scratch are: 1) faster deployment with a proven technology, 2) access to the licensed CDN customers, and 3) facilitation of interconnectivity with other ISP CDNs that license the same technology.

Finally, one important trend is that of large content providers contracting with multiple CDNs to deliver their content. Sometimes all CDNs are delivering content in parallel but in most cases one or two are delivering and the rest act as offload CDNs only used when load increase or the primary CDNs fail to deliver for any reason. This gave rise to a new genre of players who might play a decisive role in success or failure of any CDN. These are consultants for multiplatform content providers who provide live rating of CDN performance and control the way the content of their customers (i.e., the content providers) is distributed among the different CDNs. CC1 is currently working for one such company and suggested that they are important for the content delivery ecosystem.

3.2.4 Competitive Dynamics

The questions in the competitive dynamic category were meant to clarify the position of the different stakeholders in the content delivery market as well as the future direction of each. Also, where ICN projects fit in the plans of the different stakeholders was an important question that needed to be answered in this work.

Currently, large ISPs are focused heavily on building their own CDNs, which is causing a lot of changes in the whole ecosystem. According to most interviewees, ISPs are well positioned to regain the CDN market that they have lost to commercial pure play CDNs since the early 2000s.

On the other hand, NO1 stated that network operators are not thinking so much of competing with current pure play CDNs as much as they are aiming at optimizing the traffic in their own networks. ISPs are focusing more on the revenue streams coming from end-users than on revenue streams coming from content providers. According to NO1, both ISP owned CDNs and pure play CDNs can exist together and actually complement each other since pure play CDNs can optimize content delivery on the backbone level while locally deployed ISP CDNs optimize content delivery on the access level.

Although this might be the case for NO1 Company, the goals of building a CDN seem to differ from one network operator to another. This is evident by the pure play CDN players' movement towards licensing their technology to ISPs. According to EV1, pure play CDNs have started adopting this licensing model as a market repositioning strategy after sensing the threat that ISP CDNs poses to their business.

Getting back to ICN, EV1 stated that his company, a major equipment vendor in the access network market, is looking to add ICN capabilities as extra features to their products rather than as a standalone line of products. According to him, this will lead to slow viral adoption of ICN and is more realistic than offering a totally new, dedicated ICN system.

When asked about how ICN can compete against traditional CDNs and the role of pure play CDNs in an ICN, most interviewees agreed that ICN is an evolutionary step from CDN and that it will slowly take over. This means that ICN concepts can be adopted by traditional pure play CDNs to improve their own networks as well. However, all interviewees also agreed that in an ICN era the role of a pure play CDN may diminish to a brokering role between content providers and the different ISP NetInfs since content providers will not accept to contract with thousands of NetInf providers to deliver their content but will need one interface.

Finally, regarding the adoption scenarios of ICN, there was no clear consensus on a certain scenario. However, most interviewees agreed that ICN adoption will be driven by ISPs who are constantly trying to optimize the traffic flow inside their networks, and ICN can be a mean to achieve just that. Some interviewees, such as RI2 and EV1, suggested that new, popular applications using the additional functionalities of ICN may drive the adoption as well.

4 Study of Domains Influencing Caching

There are many domains that influence the way content is cached inside networks. Analyzing and understanding these domains and the way they influence caching is of upmost importance when conducting holistic studies related to caching. Three main domains are identified in this work as the ones that heavily influence caching:

1. Mobility: The more mobile a user becomes, the more dispersed the points in the network from which he request content and the more challenging it is to optimize the caching inside the network.

2. Naming: Different naming schemes can be used to identify objects on the Internet, and they affect the way content is handled inside caches. For example, a loose naming scheme such as the one used in P2P networks can lead to unnecessary redundancy in caching whereas a controlled unique naming scheme might lead to better cache management.

3. Content: Regardless of the access and addressing of content, the nature of content itself has a huge effect on the way it is cached. Some content are cacheable by nature (e.g., they are popular) while other are not (e.g., personalized content).

4.1 Mobility

The continuous growth in processor and memory capabilities met by the constant shrinkage in their sizes has led to the emergence of a variety of powerful handheld devices (e.g., smartphones and tablets) capable of performing tasks which were one day limited to PCs. These advancements in device capabilities are paralleled by advancements in mobile access technologies (e.g., Long Term Evolution (LTE)(Sesia, et al., 2009)) providing end-users more bandwidth and a quality of experience level matching the one they receive on fixed broadband networks.

Caching has not been discussed in the context of mobile networks with the same vigor as it has been in fixed networks. This rises from the fact that content delivery striving heavily on caching is quite new to mobile networks that traditionally have been focused on voice services and simple text messaging. It also owes to the fact that mobile network architectures are standardized by few organizations such as 3GPP[1] who do not include much about caching in their specifications in contrast to the more open standardization approach of IETF[2] which includes numerous caching specifications on different network layers.

4.1.1 Caching Challenges in a Mobile Environment

There are obvious challenges when implementing caching in mobile context. These challenges arise from issues relating to user mobility such as the continuous change in user population per access point and the constant entry and exit (churn) of users during transfers; device resource constraints such as the limited capabilities of mobile devices and mobile access constraints such as the physical

[1] http://www.3gpp.org/
[2] http://www.ietf.org/

limitations of the wireless channels and the simultaneous use of multiple access technologies.

User Mobility Challenges

Users in the same geographical area tend to access the same content within relatively small time durations, a phenomenon referred to as temporal locality, which is highly exploited by caching. In a mobile context, where the users are always in motion, the level of temporal locality diminishes making caching more complex and less efficient.

In many wireless environments the churn is quite high making it difficult to establish an efficient caching system based on active caching in end-user devices. At the same time, this high churn means that many of the content inside caches will only be accessed partially since a user might leave the network while in the middle of a request session leading to a non-optimal utilization of storage space inside caches.

Most of the CDNs today use DNS redirection to route request to the nearest cache server holding the requested content. This method relies on the assumption that the end-user is always near the DNS and hence the network routes the request to the closest cache server to the DNS. Since, the user in a mobile network might be on the move, sometimes moving quite fast (e.g., on trains) and hence becomes nearer to another DNS, the cache server serving the content might not be optimal anymore leading to degradation of performance (Dong, Ge and Lee, 2011).

Mobile Access Challenges

The radio access network part of the mobile networks represents a huge bottleneck in the delivery of content. This puts more constraints on the locations in which caching can take place, since in addition to a cache miss delay the user will face the latency of the radio access network. If both of these delays are large, it can negatively affect the end-user quality of experience.

Many mobile devices switch automatically and seamlessly between different wireless access technologies (Joseph, Manoj and Murthy, 2004) (e.g. from 3G to WiFi and vice versa) while fetching content. Although there is only one interface, that of the device, that the user is dealing with, there are multiple networks being utilized to fetch the content. This means that to implement an efficient caching system, the system should be managed coherently across the borders of different networks giving rise to many challenges.

Device Resource Constraints

Contrary to fixed networks where user devices are usually PCs and laptops that enjoy ample storage space, processing power and constant power sources, mobile devices have much lower capabilities. Hence, caching in the end-user device is quite challenging in a mobile context since there are high constraints regarding the caching space as well as battery power constraints if the device would act as a caching node serving other users in the network (e.g., P2P caching system)(Heikkinen, 2012).

4.1.2 Caching in Mobile Networks

There are many possible methods to implement caching in mobile networks; some of them are already implemented in the market while others are still being discussed in the research community. Generally, caching can take place in one of three locations in the mobile network (Figure 7): the core, the radio access network (RAN) or the mobile device.

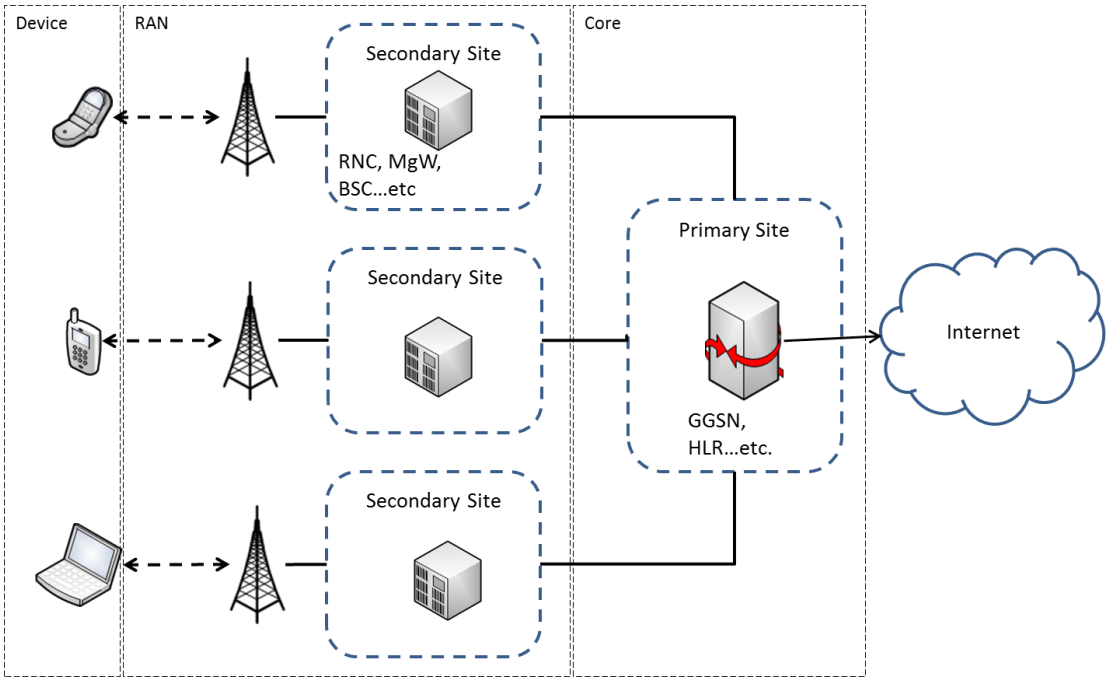

Figure 7: Mobile Network (Kaaranen, et al., 2001)

Primary Site Caching

In a mobile network (e.g. Universal Mobile Telecommunication System, UMTS) a data request is routed from the RAN up till the Gateway GPRS Support Node (GGSN) which is responsible for the interworking between the mobile network and the external packet switched networks such as the Internet. Hence, it represents a logical place to perform caching since all the aggregated data requests pass by it giving it the advantage of accurate estimation of the most popular data objects. A study conducted by Caterin et. al (2011) showed that a caching system implemented in primary sites, such as the ones hosting the GGSN, can already be efficient if only 5.1% of traffic is suitable for caching. However, since mobile networks usually cover vast geographical areas (e.g., whole countries) with only few GGSNs inside the network, the caching is done quite far from where the requests originate, which leads to high latencies. Another problem in caching so deep in the core is that a large storage space is required to permit the cache server to handle and cache massive amounts of aggregated requests leading to higher costs. Finally, having a few number of caching servers in primary sites increases the risk of failure.

Secondary Site Caching

The second category of caching methods in mobile networks is performing caching in secondary sites of the network. These secondary sites serve as concentration and distribution sites hosting network equipment such as media gateways (MGw), radio network controllers (RNC) and base station controllers (BSC) (Catrein, et al.,

2011). There are more secondary sites distributed in the network than primary sites and they are usually smaller and closer to the request origin. Catrein et al. (2011) showed in their study that although caching in secondary sites might not be as efficient as in primary sites from the perspective of cache hits and profitability, it still represent an attractive location for caching as it provides better end-user experience. One advantage is of course the closeness of cache to end-users, which reduces the latencies experienced by the end-users. Also, since the cache will be serving a smaller population, it can have a relatively smaller storage and hence be more cost efficient. However this might not always be the case since the users under the coverage of a secondary sites will change frequently (since they are mobile) changing the request pattern in turn and reducing the temporal locality of requests which means the cache may still require a large storage space to achieve a high hit rate.

Access Point Caching

A recently discussed caching method in the mobile network context is performing caching in the wireless access points of the network. These access points can be outdoor tower antennas (e.g. BTS or Node-B) or smaller indoor antennas (e.g. WiFi access point or Femtocells). Some studies proposed (Lungaro, Segall and Zander, 2010a;b) that the deployment of caches at the base stations can improve the end-user experience in back-haul limited scenarios as well as save transmission resources when multiple users access the same data object in the same geographical area. Other studies (Mishra, Shin and Arbaush, 2004) proposed deploying caches in the WiFi access points which offload some of the data traffic of the mobile networks. Such caching implementation would depend on traditional caching techniques that are already used in IP networks, making it much easier to deploy and manage. On the other hand, there are also studies (Golrezaei, et al., 2011) that suggest attaching a caching storage space to the Femto-cells that mobile network operators are increasingly deploying. Such configuration will depend on a weaker and cheaper backhaul connection which is compensated by the local serving of requests from a large storage space.

Device Caching

The most ambitious caching techniques are the ones which utilize the user devices storage space for caching content. These techniques can be divided into two categories: active user involvement in content delivery (Chow, Leong and Chan, 2007) and passive local caching of content (Lungaro, Segall and Zander, 2010). In the first category, the content is cached on the end-user device which itself act as a caching node for other devices in the same geographical area. This is achieved by having a broker which manages the caching and the searching of content in the end-user's devices. Although, such method has proved successful in fixed networks, it introduces a lot of problems in the mobile context due to the limitations of the storage space and the limitations on the battery life. On the other hand, there have been many proposals (Lungaro, Segall and Zander, 2010) to push content towards end-user devices for local consumption. Such methods require accurate prediction models to predict the content the user will request in the future by utilizing mobility and contextual information gathered by the network or by the device. Of course, for such methods to introduce savings to the network the

content is pushed during non-peak periods and a priority mechanism to serve some critical requests before others is put in place.

Although mobility introduces a lot of caching challenges in a host-centric networking environment as can be seen from the previous examples, it does not heavily affect the caching in an information-centric networking (ICN) environment (Giannaki, et al., 2011). In NetInf, there is no need for additional mechanisms to implement caching in a mobile environment since the inherent properties of the network locates the best copy of a request data object and serve it. Also, caching in the end-user's devices can be done seamlessly, since devices can be NetInf enabled transforming them into another caching node in the network that can serve requests (SAIL, 2011).

Even with all the challenges that caching brings in a mobile network, caching remains under the control of just one stakeholder that is the network operator, allowing for efficient and conflict free caching management. This is in contrast to fixed networks, in which the request might be served from caches that belong to content providers inside the network, or by CDNs hosted in the network which in many times lead to conflict of interests and complicate the cache management.

4.2 Naming

Objects in networks are given names which act as unique identifiers that are utilized by the network to find the requested objects. Different naming mechanisms exist, each providing its own level of security, scalability and flexibility (Ghodsi, Koponen and Rajahalme, 2011).

A naming mechanism consists of three main components as shown in Figure 8, these are (Ghodsi, Koponen and Rajahalme, 2011):

1. *Name*: this is the name which is used to identify and fetch an object within a network
2. *Real World Identity (RWI)*: This is the entity which the name belongs too, or has been created by. It can refer to a person or organization.
3. *Public Key*: Each RWI is associated with a public-private key which is a cryptographic key that verify that a name does indeed belong to a certain RWI.

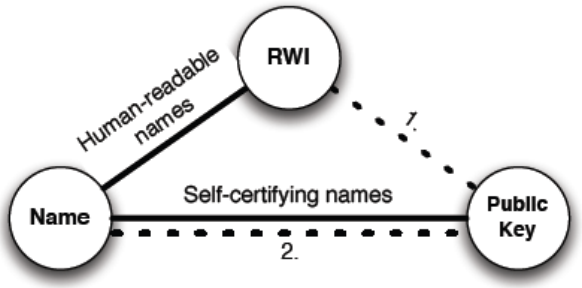

Figure 8: Three components of naming (Ghodsi, Koponen and Rajahalme, 2011)

In order to ensure a secured naming architecture, there must be bindings between the three components: name, RWI and public-key. The three bindings are essential to the provenance of content. The binding between the name and the RWI allows users to identify that a certain name belongs to a certain entity. The

bindings between a public-key and a corresponding RWI ensure that the RWI that claims that a certain name belongs to it has the means to prove it by providing a key which can validate this claim. Finally, the bindings between the name and public-key are needed to allow the user to identify which public-key to use to validate the provenance of an object. Due to the transitive relation that combines the three bindings, if two bindings are supplied the third is directly implied. One binding should be intrinsic to the naming system while the other is supplied by an external authority (Jacobson, et al. 2012).

Generally, naming mechanisms adopt one of two architectures. Names can be chosen based on a hierarchical architecture in which each additional prefix in the name narrows down the possible locations where the object that belong to this name resides. Or they can be based on a flat architecture in which each name is unique and interpreted as a single, whole label without any internal structure (Ghodsi, et al., 2011). Each of these two naming architectures enjoys some advantages as well as suffers from some drawbacks.

4.2.1 Hierarchical Human-readable Naming

In a hierarchical naming architecture, the name consists of prefixes, which are read sequentially (Ghodsi, et al., 2011). When the network routes the first prefix it is guaranteed that it will be routed towards the entries for the next prefix which in turn guarantees the routing towards the next prefix and so on. For example, when a network routes a hierarchical name such as *com.google.mail* it first follows the routing entries of com which guarantees the finding of the routing entries for *google*. When routing *com.google*, the network will eventually find the sought after object which is the mail.

Although each entry in the hierarchical name is not globally unique, for example, different content providers can use the *mail* entry (e.g. *mail.yahoo.com*) while a company like Google can utilize the *google* entry in different addresses (e.g. *scholar.google.com*), yet the concatenated prefixes are always unique. Hence, under *com* there exist only one entry for *google* and under *google* there exist only one entry for *mail*. Such configuration allows for the re-use of popular names, making it easier to form highly human understandable and readable names that consist of sequences of readable strings.

Other advantages of hierarchical naming architecture are the scalability (D'Ambrosio and Dannewitz, 2011) it offers which comes from the use of hierarchical naming servers as well as the flexibility of design which comes from the fact that the data path makes no assumption about the naming model. No entity needs to store all the mappings for each global address (D'Ambrosio and Dannewitz, 2011), rather it is enough to store the high level prefixes in global name servers (e.g. DNS) which can utilize local name servers (e.g. local DNS) to translate the lower prefixes, making the whole architecture very scalable and flexible.

The main drawback of hierarchical naming though is the week intrinsic binding between the name and the RWI (Ghodsi, Koponen and Rajahalme, 2011) since it relies on the user understanding of the name and his interpretation of which entity this name refers to. For example, a user who wants to visit Aalto University website might search for the word *Aalto* and finds two possible URLs: *aalto.fi* and *aalto.com*, there is nothing in the URLs that can ensure that one of these URLs is the right one except the personal interpretation of the end-user.

Another drawback of hierarchical naming is the difficulty it introduces when performing object replication inside the network. Since each entry in a hierarchical name points towards a certain location, the location and the name are strongly bonded together and in order to do any off-path replication the network needs to assign extra names to the copies of the objects. These names in turn need to be mapped to the original name which the user provides when requesting content requiring the use of extra techniques such as DNS redirection and URL rewriting, which introduce more complexity to the system.

4.2.2 Flat Self-Certifying Naming

Unlike hierarchical naming architectures, flat names are routed directly without resorting to the sequential interpretation of entries in the name (Dannewitz, et al., 2010). Since flat names do not map to certain locations the verification of the name is done directly on the object (self-certifying) rather than by the host of the object. Such self-certifying names enable the verification of data integrity as well as the authentication and identification of the content owner.

Flat self-certifying names are very well suited for Information-centric networks where caching is an inherent part of the architecture and where there is a need for a naming scheme which facilitate such caching to take place. Since, the verification is done on the object rather than the host (Ghodsi, Koponen and Rajahalme, 2011); objects which belong to a self-certifying naming system can be replicated in multiple locations around the network without compromising the security associated with them. In self-certifying names, the name and the public-key (Figure 8) are intrinsically bonded together either by attaching the hash of a public-key in the object metadata, or making it a part of the name.

Although flat self-certifying names provide a better mechanism for content replication and integrate many security properties, they do suffer from some drawbacks such as usability and migration challenges. Since self-certifying names are machine generated names they are either totally or partly non-readable by humans and cannot be understood or remembered. This makes it challenging for users to directly address a certain object and make the system heavily reliant on metadata accompanying the name to be used in identifying and fetching the object. Also, moving towards such naming mechanism would require challenging reworking of existing hierarchical names which might make it unfeasible to achieve.

Another important issue which arises from the use of flat self-certifying naming is the management of the name resolution service. Unlike in hierarchical naming where the name resolution takes place in a rigid hierarchical DNS system, flat self-certifying names require a flat name resolution system which is capable of identifying a list of locations in the network where copies of the requested object are located. Since, one of the main objectives of flat self-certifying naming is the elimination of centrality; the name resolution service should be fully distributed making it more challenging to scale and manage the system as the number of names increase (Ghodsi, Koponen and Rajahalme, 2011). Moreover, ensuring that all the names are globally unique can be quite challenging in such a naming mechanism.

4.2.3 Naming in ICN – Case: NetInf

This work is interested in the naming schemes proposed in the context of information centric networking and particularly in the NetInf use case. Naming and caching in NetInf are two faces of the same coin. Both are being heavily relied on in order to build a truly content-based replication-inherent networking system. Moreover, the naming scheme influences heavily the choice of the caching mechanisms that are to be utilized in the network. NetInf proposes the use of both hierarchical and flat names (Dannewitz, 2009). The main motivation behind the support of both naming schemes is to handle both new and old names without the need for expensive migration between both.

Moreover, in NetInf, the name directly identifies the information object without relating to any network node or file structure. Since these information objects are independent of any location, the burden of securing the content lies on the content itself rather than on the hosts of the content. One approach, which is adopted in NetInf, is to integrate the security aspects into the naming concept. Hence, the name in NetInf provides security features such as: confidentiality, data integrity, accountability, availability and controlled access (SAIL, 2011).

There are however many aspects which are not very clear in the naming scheme of NetInf. Three questions in particular were investigated in this work, these are:

1. What are the mechanisms that ensure that the flat names which are created and published in NetInf are globally unique? And is there an entity which keeps all the global names?
2. How will flat self-certifying names which are human non-readable affect end-users interactions?
3. How will flat self-certifying names affect the search engines?

Hash tables can produce names with acceptable uniqueness. Even though such mechanisms would not hundred percent ensure names are unique it brings many advantages. Using such a mechanism would mean there is no need to have a global entity which is knowledgeable of all the names, but rather the responsibility will be divided among the local NRSs of each network. Not only will this improve the scalability of the whole architecture but at the same time it does not hand such an important control point to any entity making the architecture a truly distributed one in both technical and control point of views. On the other hand, looking from a routing point of view, not having a global name resolution system which keeps track of all the objects in the network would give rise to the possibility of having separate NetInf island networks which do not know about each other's objects nor have a mechanism which can be consulted to get such information (e.g. like the global DNS does in host-centric networks).

Many in the ICN research community have come to a conclusion that end-users no longer need to remember addresses or full names of content in order to retrieve it. One proof they give, is that most users currently use search engines to find their content instead of writing URLs. Although, this might be true to a certain extent, it seems that such conclusion has been reached not through personal conviction but rather through necessity. Such necessity is born from the fact that for self-certifying names the name partly consists of a non-human readable product of a hash function which acts as a security measure that ensures the integrity of the

object. If there is to be another mechanism which translates these self-certifying human non-readable names to human-readable names, the security might be compromised since the point of control for securing the content will shift from the name of the object to the translation mechanism itself.

Finally from a search engine point of view, using self-certifying names instead of URLs would not introduce much change to its operations. Search engines work by crawling webpages and moving from one page to another through the URLs which link them. In a NetInf context, those URLs will simply be replaced by NetInf names and the search procedure will be the same. However, one point to note here is that the search engine will present results (information objects) in the form of NetInf names and hence each user when requesting an IO using the NetInf name returned by the search engine will receive this IO from the nearest cache to him.

4.3 Content

Having analyzed the caching from the user access (Chapter 5.1) and the network addressing (Chapter 5.2) perspectives, in this section an analysis of the content itself and how it influences caching is done. Although theoretically caching is a very efficient mean to optimize content delivery, in reality not all content that traverses the internet is cacheable. Cacheability is defined in this work as: *"*the technical ability and economic feasibility of caching a type of Internet traffic without degradation in the user quality of experience or the violation of any copyright law*"*.

In this work a list of parameters which influences the cacheability of content has been constructed. The parameters were identified based on extensive literature review of caching studies and were verified by experts during the interviews (Section 3). A total of 13 cacheability parameters belonging to three domains are identified as shown in Table 2. The technical domain parameters relate to technical and physical aspects of the content such as size, latencies and rates of change. The economic domain parameters deal with the economic aspects of the content such as costs, values and monetization. The contractual parameters deal with issues of control over the content and its distribution. The parameters are not mutually exclusive and strong correlations exist between them.

Table 2: Cacheability Parameters

Domain	Parameter	Description
Technical	Rate of Information Change	The rate at which an IO changes from one state to another.
	Size of Information Object	The size that an individual IO takes on the cache disk storage.
	Level of Personalization	A measure of the extent of the personalization of an IO to a certain user or set of users.
	Time between production and consumption	The elapsed time between the production of an IO (in the original server) and the consumption of this IO by the end-user.
	Delay and Jitter Tolerance	A measure of how tolerant to delays and jitters a user is when consuming an IO.
	Concentration of Requests	A measure of how spatially concentrated the requests for an IO are.
Economic	Rate of Value Erosion	The rate at which an IO lose its value (either in popularity or in price).
	Popularity of Content	A measure of how popular a certain IO is in terms of number of requests.
	Cost of Caching	The cost related to caching an information object. This cost can be the total cost of storing and processing a content in a data centre for example or the price paid by the content provider to any caching entity to cache its content.
	Content Monetization	A parameter that shows whether an Information object is monetized by its owner or not.
Contractual	SLA Existence	A parameter that shows whether an SLA exist in the process of delivering the content or not.
	Distributional Control	Measures the extent of control needed to be exercised over the IO by the content provider to ensure the IO is only available for the persons with permissions or in permitted geographical areas.
	Request Statistics Control	Measures the extent of control needed to be exercised over the IO by the content provider to ensure the proper statistical information about usage and requests of the IO is available.

4.3.1 Rate of Information Change

One of the most obvious parameters when looking at caching is the rate at which content change. Since caching in its essence is a time-shift activity in which the content is served in a time later than its original delivery, there is always the risk that the content served from the cache is stale. When the rate of change of a certain piece of content is high, it is more challenging to cache this content and in many cases might not even be cacheable at all.

Several caching studies refer to content as either static or dynamic. By static they mean that the content does not change frequently once it is created and hence can be cached and served very easily. Video files, photos and icons on web pages are some examples of such content. Although some of the static content might be generated by dynamic scripts which generate content only at the time the user request a page, they are still considered static in this work, since from a caching perspective the content itself is not altered and can be served from other locations than the original one by either caching the results of the dynamic scripts or by

implementing the scripts on the caching servers themselves (i.e., caching the script).

Dynamic content on the other hand constantly changes in short time periods. A weather update or a stock quote can be considered dynamic types of content, since they tend to change very quickly making it inefficient to cache them since the cache will need to refresh the content constantly. This makes the whole caching process unnecessary and may actually slow the delivery process of critical data, e.g., receiving a stock quote late.

There is a fine line between content that are considered static and content that are considered dynamic. Since, the rate of change of content is dealing with time, each caching system depending on its speed and closeness to the original data might differ in their assessment of the cacheability of a certain piece of content. For example, if a server-side caching is implemented at the original server site, a piece of content which changes quickly might still be regarded as static enough to cache, since the refreshing of the stale content may be quick and does not consume much bandwidth. In order to clearly draw the line between static and dynamic content, a rule of thumb is proposed in this work (Figure 9): "A piece of content can be considered static to a caching system only if the rate of serving a subsequent request is higher than the rate at which content becomes stale."

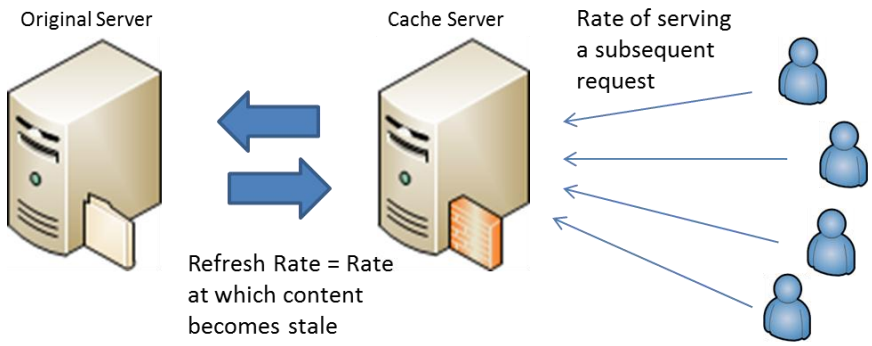

Figure 9: Rate of Content Change

4.3.2 Size of Information Object

The main constraint which every caching system is trying to optimize is the storage space. Caching systems aim at reaching the highest byte hit rate using the limited space they have. In order to do so, there are two variables which they need to optimize, these are the number of requests for a particular cached object and the size of this cached object.

The smaller the size of cached objects become the higher the number of objects that can be cached which in turn increase the probability that a requested object is found in the cache. On the other hand, the larger the size of the cached objects become, the smaller number of objects can be cached, which in turn lowers the probability that a requested object is found in the cache. However, large objects need lower number of requests to provide the same level of byte hit rate as small objects and hence there is an optimum size which can maximize the byte hit rate.

One challenge when trying to optimize caching is the high variation in sizes of objects traversing the Internet. Not only does sizes vary but also the popularity in terms of the number of requests fluctuate heavily from one object to another complicating the optimization problem even more.

4.3.3 Level of Personalization

The level at which a piece of content is personalized to a certain group or persons plays a big role in the cacheability of such content. An object which is personalized is by nature less cacheable than an object which is meant to be accessed by anyone since the number of users who will request a certain personalized object will be limited to those who the object was personalized.

4.3.4 Time Between Production and Consumption

The time between the production of an object and its consumption can draw the line between a cacheable and a non-cacheable object. Theoretically, content consumption can be done in parallel, where the consumption begins as soon as the production begins, in series, where consumption starts only after the production process has finished, or in delayed parallel, where consumption begins in a slightly later time after the production has begun.

Since caching time-shifts the consumption of content from the original production time, it can only be utilized in the latter two consumption cases. For serial consumption it is pretty straightforward, the cache stores the object produced and when the user requests the object, it is served from the cache instead of the original server. On the other hand, caching in a delayed parallel consumption context is not so straightforward. Depending on length of time permitted before consumption commence, caching might be possible or not. Assuming it takes time T1 to transfer an object from the original server to the cache, and it takes T2 to serve the object to the user as soon as the cache has stored it, then if T is the maximum delay which is permitted before the content consumption starts, T1 + T2 should be equal or less than T if caching is to take place (Figure 10).

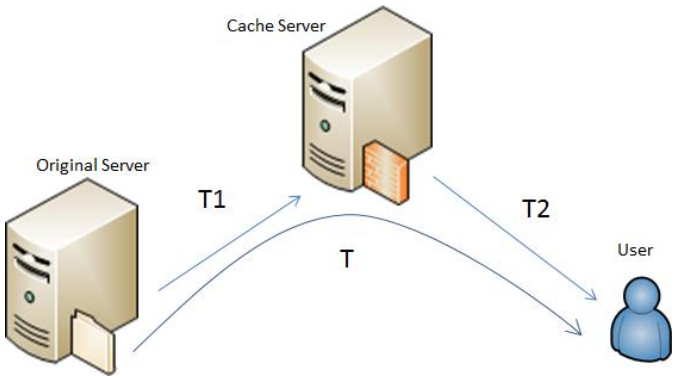

Figure 10: Time between Production and Consumption

P2P file sharing is one use case of serial consumption. Users who utilize P2P file sharing programs such as BitTorrent and Gnutella start consuming the content after it has totally been downloaded at their machines and hence caching is easily implemented in this use case. On the other extreme, when broadcasting premium live events, the consumption time is quite close to the production time and hence latencies introduced by caching can have a detrimental effect on the delivery process.

4.3.5 Delay and Jitter Tolerance

Users' delay and jitter tolerance towards an object is closely related to the previous parameter Time between Production and Consumption. Although some content

might be requested as soon as they are produced, i.e., time between production and consumption is short, such as for live video content, the end-user tolerance for delays and jitters for this content might differ from one case to another. Taking live sports matches as an example, some users subscribe to content providers in order to receive the live broadcast of matches with the same quality as they receive it on TV (e.g. ESPNplayer (ESPN Network, 2012)), while others resort to free live streaming methods such as P2P broadcasting channels (e.g. SopCast (Sopcast, 2012)). The first group is more sensitive to delays and jitters than the second group and hence the caching constraints in the premium delivery case are much higher than in the free delivery one.

4.3.6 Concentration of Requests

Although an object might be popular in terms of the total subsequent requests, yet it is the distribution of these requests which is critical to the performance of a cache. If an object is requested 100 times from 100 different networks, then it should not be cached in any one of them, since each network is experiencing just one subsequent request making the caching process not economically feasible.

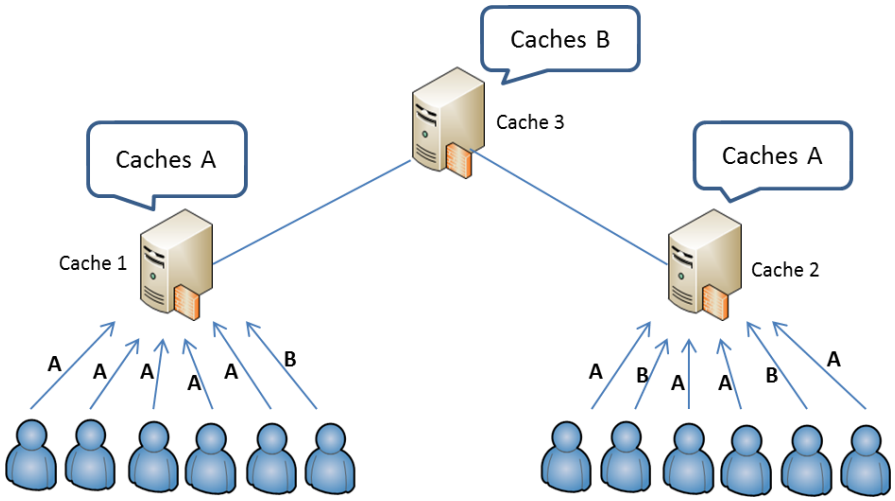

Figure 11: Concentration of Requests

The importance of this parameter is especially obvious in networks deploying hierarchical caching systems. In hierarchical caching systems the caches on the edge of the network usually carry small storage space compared to caches located deeper in the network. Hence, to optimize the caching, networks try to reduce unnecessary redundancy and cache only highly localized content in the edge caches whereas less localized content are cached higher in the hierarchy. As seen in Figure 11, object A seems to be quite popular in the vicinity of both edge caches (Cache 1 and Cache 2) and hence it is cached in both. On the other hand, although object B is to a certain extent popular in the network, it is not very popular from a single edge cache perspective. Hence it is better for the network to cache this object deeper in the hierarchy in order to benefit from the aggregation of requests coming for object B to achieve high hit rates.

4.3.7 Rate of Value Erosion

Most of the caching studies which have been conducted in recent years have focused on optimizing caches based on the popularity of content measured by the number of subsequent requests which the content generates (Breslau, et al., 1999;

Cheng and Kambayashi, 2000; Davidson, 2001). However, very few have studied how the erosion of popularity influences caching. An object such as a data chunk of a live match stream may become popular very quickly and for a short time after which its popularity vanishes. These types of objects are usually treated in a non-optimal way by most caching algorithms.

Today's algorithms, such as LRU, LFU and MRU (Aggarwal, Wolf and Yu, 1999), assign values to the cached content based on the number of requests and the time passed since the content were last requested. These algorithms perform well under the assumption that the popularity of content will drop gradually moving the content step by step towards the end of the queue until it finally get evicted. This is not the case for content which experience sudden drop in popularity, since they are initially placed at the front of the queue due to their sudden popularity but then they remain for an unnecessary long period of time inside the cache even when there is no or very few requests generated.

Hence, the rate at which content loses its value in terms of the number of requests it generates can have a detrimental effect on the caching performance and should be taken into consideration in the future caching algorithms. However, the challenge in incorporating such a parameter in the decision making of a cache remains in coming up with accurate prediction models of the rates at which the values of different content erodes. This might be easy for certain content such as news videos and live matches but can be challenging for other content such as music and movie files.

4.3.8 Popularity of Content

The entire caching ecosystem is built around the fact that much of the content traversing the Internet are popular enough to generate subsequent requests after their initial transfer and hence can be stored and reserved from a point nearer to those requests than the original serving point. Hence, all caching algorithms aim at identifying and storing the most popular objects traversing the network in order to achieve high cache hit rates.

Most of today's caches implement a variation of one of two caching algorithms (Aggarwal, Wolf and Yu, 1999): Least Recently Used (LRU) or Least Frequently Used (LFU). In LRU the cache keeps track of the time since each object was last requested and when a new object needs to be admitted to the cache it replaces the object which was least recently requested. In LFU, the cache assign hit rate counters for each object, and when a new object needs to be admitted it replaces the object in cache which has the lowest value in its counter. These two caching algorithms attempt to keep the most popular objects in their storage for the longest time possible.

4.3.9 Cost of Caching

One of the main goals of caching is saving transfer costs of objects. In order to calculate these savings the cost of content delivery without caching must be compared to the costs of delivery using caching which can be quite challenging. The motivation behind caching comes from the fact that memory prices are dropping faster than bandwidth prices (Malik, 2011) and hence the more the content the network stores and serve locally the less bandwidth it consumes and the more savings it achieves. However, since both prices are dropping the importance of other costs such as energy are getting higher as presented. Hence, in

order to decide whether to cache an object, it is crucial to know in real-time the costs associated with storing this object and based on its expected popularity calculate if there are any savings to be made by caching this object or not.

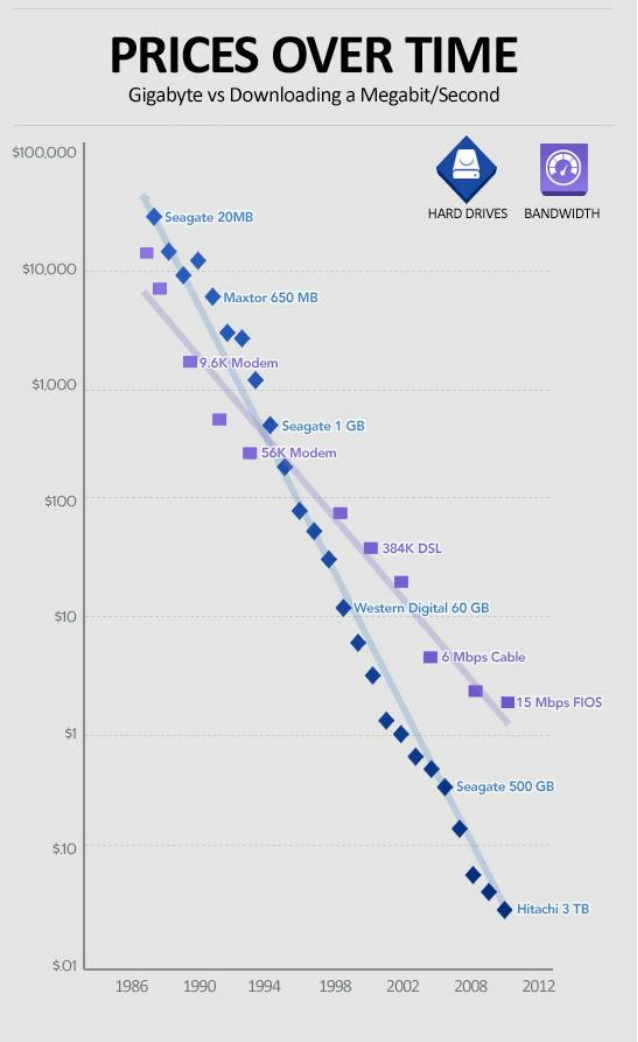

Figure 12: Memory and Bandwidth Price Evolution (Malik, 2011)

4.3.10 Content Monetization

Much of the content traversing the Internet is monetized by their owners one way or another. Some content providers charge the end-users directly while others adopt ad based revenue models and sell advertisement space and usage patterns while providing the content freely to the end-users. Such monetization is dependent on aspects such as the quality the end-user receives and the accuracy of usage statistics.

Caching cuts the direct relation between content providers and end-users threatening in the process the profitability of the content providers' businesses. Therefore, many content providers tag their content as non-cacheable to avoid this problem. In order to solve this predicament, it is essential that a caching system is transparent to both end-users in terms of the quality of delivery and to content providers in terms of accurate usage statistics reporting.

4.3.11 SLA Existence

A Service Level Agreement (SLA) is a contract in which minimum levels of performance are guaranteed by the provider of the service to its customers. Most of the content delivery providers such as ISPs and CDNs go into such agreements with their customers and hence need to ensure that their service level does not drop below the requirements stated in the agreement. One way to satisfy these requirements is to cache content in order to retain an acceptable throughput, latency and jitters.

Hence, objects which are guaranteed to be delivered with a certain level of quality may need to be treated in a privileged way in the caching system compared to other objects. For example, a commercial CDN caches its customer's objects irrespective of the number of requests (popularity) these objects will generate. Therefore, having a SLA in the delivery process can alter the way caching is done in networks and should be considered when analyzing the parameters that influence caching.

4.3.12 Distribution Control

One parameter that has become especially important in recent years is the distribution control over the content. Since lots of the services which deal with premium content such as subscription TV channels and video rentals are migrating to the online world, more restriction are in place regarding the geographical locations in which the content is allowed to be served online. For example, NetFlix (Netflix, 2012), the biggest online video rental service in the US, does not serve its content outside the borders of the US since the royalties it had paid for the content producers restrict any wider distribution. Hence, a lot of the content traversing the Internet is not allowed to be stored except in certain locations introducing restrictions on caching. This is especially visible in some countries were regulations are put in place to stop caching of sensitive information about the population outside the border of those countries. Hence, if a network spans more than one country it needs to have a mechanism in place which ensures that caching is done only in the caching servers located in data centers inside the respective country.

4.3.13 Request Statistics Control

How tightly a content provider wants to keep track of the statistics of its content usage heavily influence the caching choice for this content. Many content providers tag their content as non-cacheable in order to have complete visibility over their content even if the content itself is highly cacheable from all other perspectives. This parameter is tightly linked to the content monetization parameter (section 5.3.11) since the need for tight analytics control in most cases is motivated by content monetization. However analytics control is not limited to monetized content only, but can be required for other content, especially content belonging to content providers adopting advertising based revenue model who charge advertisers on per view or per click basis. Hence, this parameter is identified as a separate parameter.

5 Caching Control Points Shifts

In this chapter an analysis of the critical control points of the different content delivery architectures is presented. This work adopts the definition of control points used by the "MIT Communications Futures Program" (MIT, 2012) which state that:

"A control point is a point at which management can be applied. Control points can be rooted in business, regulatory, or technical regimes"

The main goal of this chapter is to identify and understand the major control points of caching architectures. This understanding facilitates the studying of the effects of deploying ICN, which relies extensively on such control points, and hence its influence on the competitive dynamics of the whole ICD ecosystem.

5.1 Caching System Control points

Before going into the analysis of the shifts of control points expected to happen in the different caching systems if ICN was globally adopted, it is essential to first identify and understand the main control points of a generic caching system as well as the main functional entities of it. Moreover, it is equally important to analyze the nature of control point shifts. Figure 13 presents a generic caching system, which consists of five main functional entities and four control points.

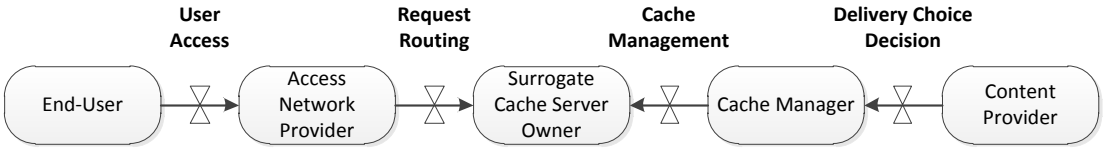

Figure 13: Generic Caching System

The functional entities are:

1. Content provider: provides the content to be delivered through the caching system. It can be the owner of the content or act on behalf of the owner.
2. Cache Manager: oversees the distribution of content over the surrogate cache servers.
3. Surrogate Cache Server owner: performs caching and can differ from one caching system to the other.
4. Access Network Provider: provides connectivity to end-users allowing them to request content.
5. End-User: requests the content and consumes it.

And the control points are:

1. *Delivery Choice Decision* is traditionally owned by the content provider who has the power to choose which delivery method to utilize in order to deliver its content.

2. *Cache Management* gives its owner the power to choose which caches the content will be stored in and the power to manage the admittance and eviction of content.
3. *Request Routing* gives its bearer the power to choose which caches the request should be routed to in order to fulfill it.
4. *User Access* gives its bearer control over the end-users' last mile access network and hence acts as a very important gatekeeper for the whole delivery system.

5.1.1 ICN Caching Architecture

ICN is a new concept which regards information as a first class citizen in its architecture rather than the hosts of information as in the traditional host-centric networks. There are two proposed ICN routing architectures: the name based routing (NbR) (Figure 14), in which requests are satisfied by any on-path cache, or the name resolution system (NRS) based routing (Figure 15) in which requests are routed by a name resolution system to an off-path cache holding a copy of the requested content.

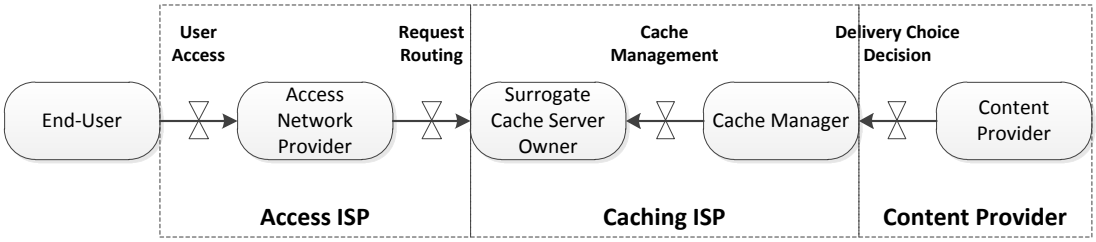

Figure 14: Name Based Routing ICN Caching System

Analyzing the ICN name based routing architecture on a high level, three major roles are identified. The first is the content provider which provides the content to be consumed by end-users. This content provider holds the first important control point that is of the decision regarding which delivery system to be utilized in order to deliver the content to the end-users.

The second role is that of the caching ISP. The caching ISP is the ISP which manages and owns the cache servers that holds and serves the requested content if it lies in the path between the requester and the content provider and hence it has authority over the cache management control point. Although the caching ISP is identified separately from the access ISP, in some scenarios, the access ISP plays both roles.

The final role is that of the access ISP. The access ISP in ICN NbR holds two critical control points. The first is the user access, as it acts as the gatekeeper through which all user requests pass, and the second is the request routing, since in ICN NbR, the request is routed normally towards the content provider, a task traditionally controlled by the access ISP.

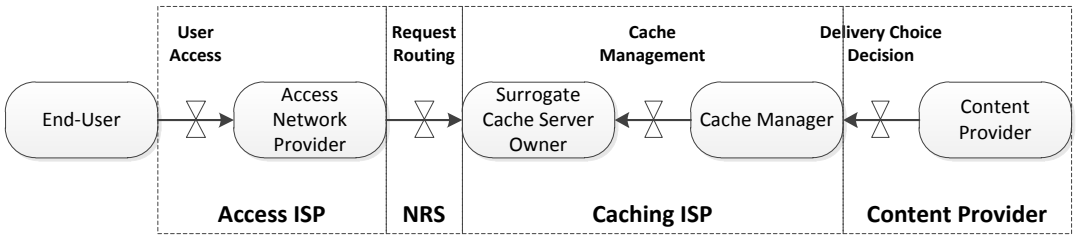

Figure 15: Name Resolution Based ICN Caching System

The ICN NRS has quite similar roles as the ICN NbR, however, the main difference is the presence of a fourth role that of the NRS. The NRS acts as the mediator between the access ISP and the caching ISP since it receives the requests coming from the first and lookup for caches in its vicinity and route the requests towards the ones it deem fittest to serve the requested content. Hence, the request routing control point is under the direct authority of the NRS. Similarly, the NRS and Caching ISP roles can be performed by the Access ISP in certain scenarios (e.g. NRS can be deployed locally and content can be cached locally).

5.2 Control Point Shifts

Due to technological and market changes, control points can shift from the authority of one or more entities to be settled under the authority of another entity or entities. Although there are numerous technical and economic studies in which control points are identified and analyzed in a variety of use cases (Trossen, Sarela and Sollins, 2010) there has been no systematic way that was put forth to study the nature of shifts that control points might experience during technological and market disruptions. In this work, in order to facilitate the study of control point shifts which are expected to occur if ICN is globally adopted, a classification of the generic type of shifts is presented.

Control point shifts can occur on two levels. They can occur between whole industry sectors or between businesses (i.e. single companies). The first shift is usually due to disruptive innovations that give an existing or newly formed industry sector power over control points, which traditionally belonged to other sectors. For example, the advent of VoIP applications such as Skype has led to a shift in control over voice services from the traditional telecom operator sector to the software provider sector. On the other hand, shifts in control points between businesses are quite common and occur due to the continuous change in the market dynamics. An end-user changing from one access provider to another in the market is an example of such shift since the control point over the user access moves from a single business to another.

Three generic control point shifts are identified in this work. These shifts are not mutually exclusive meaning that a change in one control point can be attributed to more than one of these generic shifts. Since, shifts can happen both on the sector as well as the business level; each generic control point shift is analyzed in both cases and hence a total of 6 shifts are discussed in this section.

5.2.1 Authority Shift

Authority shift is a one-to-one shift by which a control point moves from the vicinity of one entity to the vicinity of another. An authority shift can happen between sectors as shown in Figure 16. These types of shift tend to occur when new revolutionary technical architectures or technologies belonging to a different domain than the current architectures are introduced to the market. Therefore,

these shifts usually cause huge market architecture changes as they alter heavily the power balance between the different market sectors. The movement of an important control point such as cache management from the vicinity of ISPs to the vicinity of CDNs is one relevant example of such a shift which brought to the world a new business sector that of the CDNs and relegated the role of ISPs in content delivery to mere last mile bit pipes.

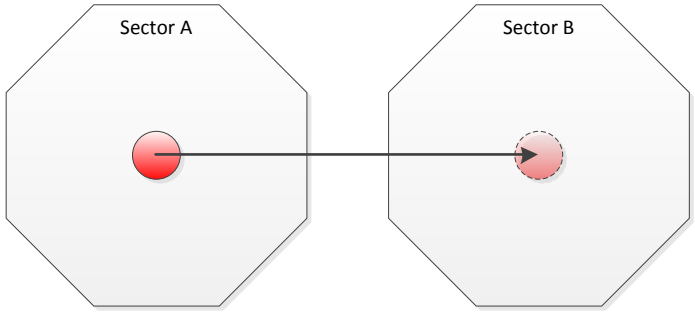

Figure 16: Sector Authority Shift

An authority shift can also occur between businesses as shown in Figure 17. These shifts occur more often than sector authority shifts since they result from the traditional and continuous competitive behavior between businesses. A user changing his ISP to join another one in the market is an example of such shift. In this example, the control point over the user access shifts from one business to another.

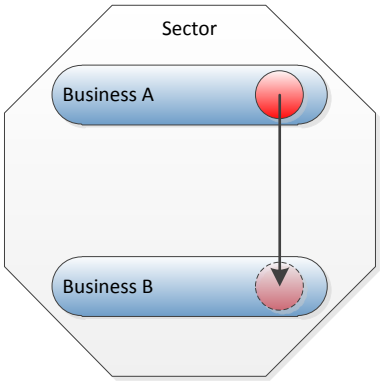

Figure 17: Business Authority Shift

5.2.2 Distributional Shift

Distributional shift is a one-to-many shift in which a control point that was under the control of one entity is divided among multiple entities each holding a fraction of the original control point. Distributional shifts may occur between sectors in which case the control point which was exclusively under the authority of one particular sector is now shared among multiple sectors as shown in Figure 18. This shift leads to the democratization of the markets in which services dependent on such control point are offered, however, it can also increase the contractual and architectural complexities in such markets. With the advent of broadband 3G mobile data services and cable Internet, the control point over the Internet user

access has spread from a single sector (that of the ISPs) to include the mobile operator and cable providers sectors. Although this gave more flexibility and power to the end-user, it also meant end-users needed to sign contracts for each access separately making it more complex to manage.

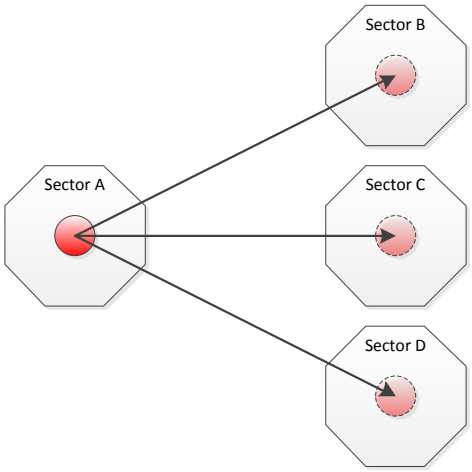

Figure 18: Sector Distributional Shift

Distributional shifts can also occur among single businesses. In this case, the control point which was once under the exclusive authority of one business moves to be under the authority of multiple businesses. Such shift leads to fiercer competition among businesses and provide a healthier competitive atmosphere. One example of such shift is the movement of content providers from contracting with single CDNs to contracting with multiple ones in order to deliver their content, hence increasing the pressure on the CDNs who are threatened to become a mere offload channels for other CDNs.

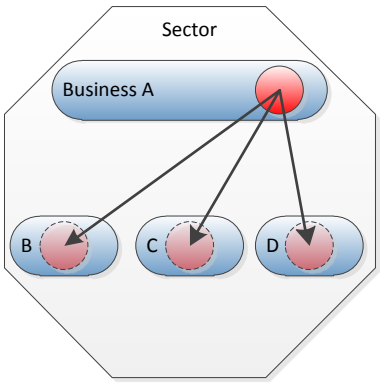

Figure 19: Business Distributional Shift

5.2.3 Consolidation Shift

Consolidation shift is a many-to-one shift in which multiple control points, which were under the authority of multiple entities, consolidate and move to be under the authority of a single entity. Like the previous two shifts, consolidation shifts may occur between different sectors, moving multiple control points from the realms of multiple sectors to the realm of a single sector as shown in Figure 20. This shift leads to the concentration of power in the hands of a single sector, which although weakening the position of other sectors might still introduce much efficiency to the services concerned. The shift from content providers contracting

with multiple ISPs, cable providers and other delivery networks to deliver their content online towards contracting with a single CDN is an example of such consolidation shift which has put the control over the distribution of content under the authority of the CDN sector instead of the other sectors.

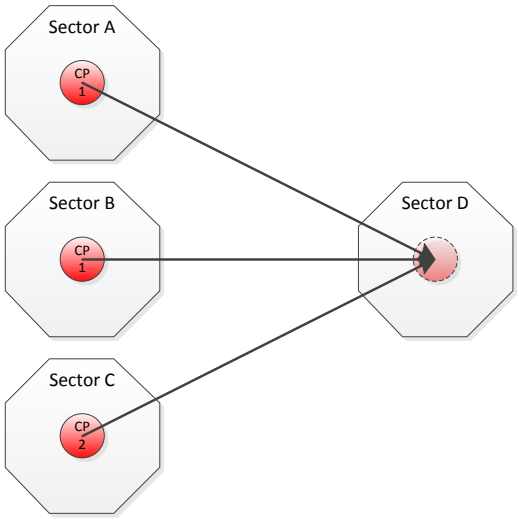

Figure 20: Sector Consolidation Shift

Consolidation shift can also occur among businesses (Figure 21). Such shift is a characteristic of monopolistic behavior in which one business dominates multiple control points which one day belonged to other businesses. Google's continuous expansion into other services such as videos and maps search is an example of a company that is trying to collect all control points of search (e.g., search for webpages, search for maps, and search for videos) under its vicinity and is in the process monopolizing the web search market.

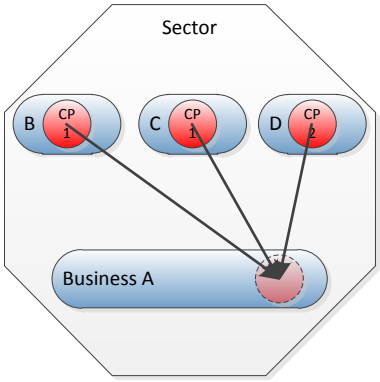

Figure 21: Business Consolidation Shift

5.3 Control Point Shifts from ISP Web Caching to ICN Caching

Since both ISP web caching (Figure 22) and ICN adopt a transparent caching architecture in which content is cached based solely on the popularity of content rather than on agreements with the content provider, the shift in control points is minimal when moving from an ISP web caching delivery architecture to ICN. This shift is studied in two scenarios: the shift from ISP web caching to ICN NbR caching and the shift from ISP web caching to ICN NRS caching.

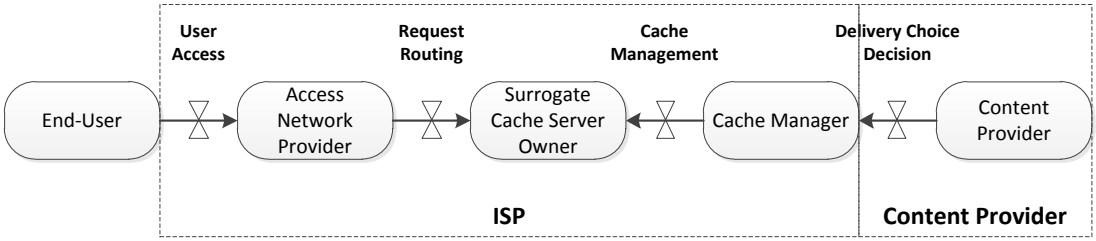

Figure 22: ISP Web Caching System

5.3.1 1a: Delivery Choice Decision

The delivery choice control point does not experience any significant shifts when moving from an ISP web caching system to any of the two ICN caching architectures. It remains under the authority of the content provider which has the final decision regarding which delivery method to choose to deliver its content. Like ISP web caching, ICN caching will not provide content providers with the means to control the end-to-end quality nor the distribution of their content, since caching in both systems is done transparently and no privileges are given to content over other except based on the popularity.

5.3.2 1b: Cache Management

Moving from a web caching architecture to an ICN NbR caching architecture does not cause any significant shifts in the cache management control point. Since, an ICN NbR caching architecture is based on on-path caching in routers which is similar to the web caching concept. With each hop the request travels, the router which receives the request checks whether the requested object is already cached in its internal storage (or a default nearby cache). If it is, it serves the request; if not, it forwards the request to the next hop. In such a system the request is routed towards the original server irrespective of whether there is a nearby off-path cache holding the content requested or not. Hence, from a single user request perspective, there is only one place from which content can be served, either a router belonging to one of the ISPs on path to the original server or the original server itself, making the control point of cache management from a single transfer perspective under the authority of a single entity.

On the other hand, moving from a web caching architecture to an ICN NRS caching architecture causes a business distributional shift in the cache management control point. This happens because the cache management moves from being under the authority of a single ISP to be under the authority of multiple caching ISPs who depending on how close they are to the requester have an equal chance to be the serving entity for the requested content.

5.3.3 1c: Request Routing

The request routing control point does not experience any type of shifts when moving from an ISP web caching architecture to an ICN NbR caching one since the access ISP remains the entity which routes the request towards the content's original server. Hence, the route from the requester to the original server holding the desired content is chosen based on traditional routing protocols similar to those utilized in today's host-centric networks and is heavily influenced by the access ISP and its peering and transit agreements.

The story is quite different when moving towards ICN NRS though. In a truly global NRS based ICN, NRSs will be deployed around the globe and whenever a caching entity cache a new object it registers it with the NRSs it belongs too. The access network in this architecture consults the NRS for possible locations in which the requested content is cached, and hence the NRS is the entity which is responsible for routing the request (if not physically then at least by giving guidance) and therefore holds this control point. The assumption here is that the NRS is a separate unbiased entity which is not influenced neither by the access or the caching ISP which in reality might not be the case.

5.3.4 1d: User Access

The user access control point remains under the authority of the access ISP when moving from an ISP web caching architecture to any of the two ICN caching architectures. However, due to the inherent support for ad-hoc networking in ICN, the user access is not as strongly bounded to the access network provider as in the ISP web caching architecture.

5.4 Control Point Shifts from Pure Play CDN Caching to ICN Caching

Pure play CDNs are potentially the sector which might get affected by the introduction of ICN architectures the most. One of the primary goals of ICN is the facilitation of ubiquitous caching on a global scale, in other words, moving content as close as possible to the end-user. This puts ICN in direct competition with pure play CDNs who have been providing such service for over a decade and causes a serious shift in the four control points of caching (Figure 23).

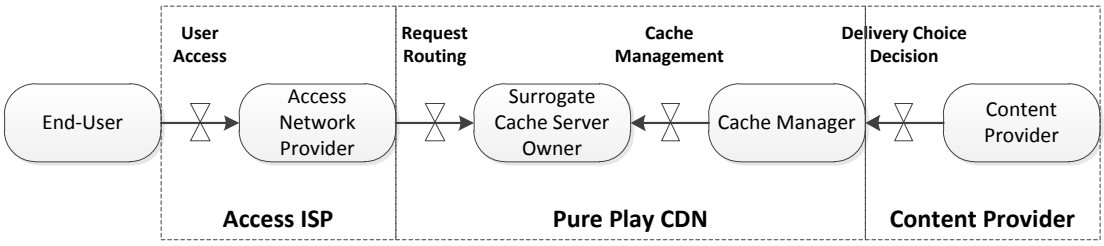

Figure 23: Pure Play CDN Caching system

5.4.1 2a: Delivery Choice Decision

The delivery choice decision control point once again remains under the authority of the content provider when shifting from a Pure Play CDN caching architecture to any of the two ICN caching architectures. However, the choice might be heavily influenced by the fact that in order to deliver its content on a wide geographic area, content providers will have to contract with several geographically limited ICN providers instead of contracting with one or few delivery providers as in the case of CDN caching. Such contractual complexity undermines the benefits which ICN brings to content providers such as closer caching to end-users.

5.4.2 2b: Cache Management

The cache management control point experiences significant shifts when moving from a CDN caching to ICN caching. When moving to an ICN NbR caching architecture, a sector authority shift will take place as the control point moves from under the realm of a single CDN to be under the realm of a single ISP in a similar fashion as explained in section 5.3.2. On the other hand, when moving to an

ICN NRS caching architecture, two shifts will occur, one is the sector authority shift similar to ICN NbR and the other shift is a business distributional shift. The business distributional shift occurs because the cache management control point moves from under the realm of one business that is the CDN to be under the authority of multiple caching ISPs in the ICN NRS architecture. Such shift leads to the democratization of the content delivery since there are fewer restrictions on the locations where content can be cached and hence the probability that content gets cached near the location where the request originated is higher than in CDN caching.

5.4.3 2c: Request Routing

When moving to an ICN NbR caching system, the request routing control point will experience a sector authority shift by which it moves from under the authority of the CDN to become under the authority of the access ISP. This is not the case though when moving to an ICN NRS caching architecture. In this case, the control point experiences a sector authority as well as a business distributional shift not towards the access ISP, but rather towards the NRS entity. Theoretically, the NRS entity could be a standalone unbiased entity and there could exist multiple NRSs from which the end-user could choose from in order to locate close by cached content. Shifting the request routing control point from CDN single ownership to an unbiased NRS multiple ownership is one of the main attractiveness of ICN NRS caching architecture compared to other caching architectures.

5.4.4 2d: User Access

The user access control point will not be affected by adopting any of the two ICN caching architectures. The access ISP remains in full control over this control point, however, given the inherent support for ad-hoc interconnection in ICN architectures, the end-user has slightly more freedom than in the CDN caching architecture.

5.5 Control Point Shifts from Content Provider Caching to ICN Caching

The consolidation of content and the rise of hyper giants like Google, Amazon and Facebook have led to the increase in the phenomena of content providers deploying their own caches either inside ISP's point of presence or in dedicated data centers having direct peering connections with ISPs. This caching architecture performs a similar task to pure play CDNs, the difference being in the elimination of a broker between the content provider and the access network. This gives full control for content providers over the three control points: delivery choice decision, cache management and request routing as shown in Figure 24.

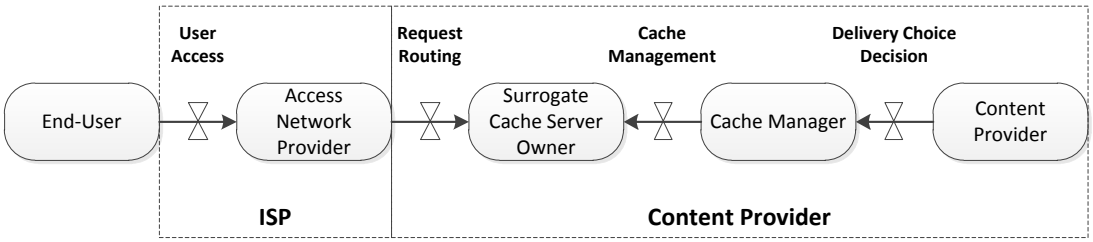

Figure 24: Content Provider Caching

5.5.1 3a: Delivery Choice Decision

Like in the previous two shifts the delivery choice decision will not be affected by moving from a content provider caching architecture to any of the two ICN caching architectures. However, such movement will impact the two ends connected by this control point. In the content provider caching architecture the content provider does not interface with any external entity when taking the delivery decision since it is usually large enough to benefit from the economies of scale and tight control over content when deploying its own caches globally. Hence, the delivery choice decision control point in content provider caching architecture is less prone to external influences by other entities providing different delivery methods and hence it is one of the biggest challenges for ICN to overcome.

5.5.2 3b: Cache Management

The cache management control point will experience a sector authority shift when moving from a content provider caching architecture to an ICN NbR caching architecture. The authority moves from single ownership by the content provider to single ownership by one of the ISPs on the path between the user and the origin server. The same shift will occur when moving to the ICN NRS caching system in addition to a business distributional shift as the control point moves from under the authority of a single content provider to be under the authority of multiple caching ISPs. Those ISPs will have equal opportunities of being chosen by the NRS to deliver the content to the end-user.

5.5.3 3c: Request Routing

The request routing control point will experience similar shifts as those experienced when moving from a CDN caching system to the two ICN caching system. Again, the control point will experience sector authority shifts when moving to both ICN architectures, in the NbR architecture the point moves to the access ISP while in the NRS architecture it moves to the NRS entity. Moreover, when moving to the NRS caching architecture, the control point will encounter a business distributional shift since the end-user will have more than one NRS to choose from to route its request.

5.5.4 3d: User Access

The user access like the previous two movements will not experience any shifts and will remain under the authority of the access ISP. However, as in the CDN shift case, the end-user will enjoy slightly more access liberty due to ICN inherent support for ad-hoc interconnection.

5.6 Summary of Findings

The study on control point shifts highlights many aspects concerning the dynamics of the content delivery architectures and the effects on the major caching control points when moving from current architectures to the two proposed ICN architectures. These insights must be taken into consideration by the people overseeing the development of the different ICN architectures in order to ensure the attractiveness of ICN to the different stakeholders in the ecosystem.

One obvious conclusion from this analysis is that the content delivery choice decision and user access control points remain under the control of the two same entities, the content provider and the access ISP respectively, in all the delivery architectures, and are thus minimally affected by the adoption of ICN. This sheds

light on the importance of these two key entities which the success of ICN depends upon.

However, in reality only ISPs and their equipment vendors back ICN till now whereas the content providers are to a certain degree outside the picture. The main assumption by ICN developers is that providing content providers with an economic and technically reliable solution is enough to attract them in the future. Such an assumption might turn out to be invalid since one of the most important concerns for any content provider, especially with the increase in premium online content delivery, is the ability to control their content and have a say on who has the right to cache it, a concept that conflicts with one of the major themes of ICN – content location independence.

One way to overcome such conflict is to create a mechanism by which content providers can retain control over their content by having strict caching rules which are deeply ingrained in the system to ensure that ICN caches will stick to them. For example, a content provider can include in the metadata file attached to its information object the locations in which this information object is allowed to be cached and for how long. The cacheability parameters presented in section 4.3 are ideally suited for building such intelligent caching mechanism.

Another important observation is the lask of any consolidation shifts in the movement from current caching architectures to any of the ICN caching architectures. This confirms the claims made by many involved in the designing of ICN architectures, which state that ICN will democratize the ICD ecosystem and eliminate centralization. Although such ideal Internet architecture theoretically sounds attractive, it is important to understand whether hyper giants generating more than 50% of the Internet traffic (Labovitz, 2011) find it so as well. From historical records, it seems that large content providers have opted for delivery systems that are centrally controlled by one or few players, such as CDN caching and content provider caching architectures.

From the above analysis it can also be seen that from a pure control points perspective, the ICN NbR caching architecture is quite similar to the ISP web caching architecture. The latter has failed for many reasons including the inability to perform cache interconnections between different ISPs; the existence of a maximum number of caches after which cache hit rates stop increasing; and the complexity that faces global content providers which arise from contracting with multiple ISPs to deliver their content globally. It remains to be seen whether the proposed ICN NbR can fare better than its predecessor and whether it can overcome the mentioned issues.

Furthermore, the analysis clearly shows that CDNs are the entities most threatened by the introduction of ICN since in both ICN caching architectures, the cache management and request routing control points, which are the bases of the CDN competitive advantage, will shift away from CDN to other entities. CDNs are in a relatively weak situation since they do not own the other two critical control points, the user access and the delivery choice decision, and hence can only survive by building upon their close relationship with content providers who themselves own one of the critical control points – the delivery choice decision. One advantage which CDNs can count on is the fact that they provide a single interface that enables global content delivery for content providers easing the contractual burden on them. This is not the case, at least until now, in the ICN, since for a

content provider to ensure a preferential treatment of its content in the ICN caches it needs to contract with each local ICN provider (ISP) independently.

One other observation regarding the movement from a content provider caching model to ICN caching is that such movement means content providers are giving up authority over two important control points: cache management and request routing. Hence, ICN caching should compensate such loss in order to be attractive enough for large content providers. This can be achieved, for example, by introducing significant cost savings without impacting quality of delivery, or by creating a control mechanism by which content providers can maintain control over cache location and access rights. This observation is particularly critical for the success of ICN since more than 50% of the Internet traffic is generated by a few hundred hyper giants (Labovitz, 2011), many of whom have their own deployed global network of caches and would therefore require much convincing in order to give up on their investments and backup the ICN caching model.

Finally, although they were not explicitly mentioned in the above control point shift analysis, naming and mobility impact the shifts significantly. Naming obviously impacts the request routing control point as the choice of the naming system, either hierarchical, flat or some other system would give advantages and enable the usage of one request routing method over the other. For example, a flat naming system perfectly fits with the ICN NRS architecture while hierarchical naming fits more with the ICN NbR architecture. Moreover, naming affects one control point which does not appear in the above analysis but that is quite relevant, that is the control over the global naming of content. Currently, the Internet Corporation for Assigned Names and Numbers (ICANN) is the entity responsible for the coordination and uniqueness of the addresses (e.g. IPv4 and IPv6) and names (e.g. URL). When moving, for example, to a flat naming system, the authority over this important control point might be distributed among more than one entity (e.g., NRS or the content provider).

Mobility on the other hand impacts the user access control point. With the increase in the number of wireless hotspots and the rise of mobile device based Internet traffic consumption, end-users these days have a broad choice of access networks to choose from when deciding to request content. For example, a user might choose to use the cellular Internet service provided by his mobile operator, the Wi-Fi Internet connection available in his area or the fixed Internet on his PC provided by his cable provider. Such distributional shifts undermine the power of access networks and give content providers, who are the only entity enjoying full and uncontested control over a critical control point, a big say in the direction of the future ICD ecosystem.

6 ICN Adoption and the Competitive Dynamics of the ICD Market

The previous chapters have illustrated the increasing importance and impact of caching on the ICD ecosystem. These impacts were analyzed independently and from them a set of conclusions and scenarios were built. In this chapter, a look on the actual competitive moves by the important stakeholders that are taking place in the market is done. The goal is to first link the theoretical analysis of the previous chapters to the actual market movements in order to understand the motives behind such movements. Secondly, an analysis of the effects which the current competitive movements will have on ICN adoption is done.

Competitive dynamics is the "analysis of competition at the action and response level to predict how a firm will act or react against opponents" (Chen and Leong, 2005). Although traditionally used to analyze firm level intra-market competition, competitive dynamics can also be used to analyze the competitive movements on an ecosystem inter-industry level. In this chapter, the competitive dynamics between the three most important actors in the ICD ecosystem, the ISPs, CDNs and content providers, are studied.

Content caching can be regarded as a platform on which the three main actors in the ecosystem compete and by which the recent competitive dynamics are enabled. Content caching and cloud storage in general have grown in importance with the increase of cacheable traffic traversing the Internet and the constant drop in storage prices. Hence, today's competition revolves more around efficient in-network storage rather than ownership of high-speed pipes.

Although ICN is a futuristic technical architecture, which does not significantly impact the current competitive dynamics of the ICD ecosystem, such competitive dynamics will certainly shape the way ICN develops. Depending on how the ecosystem will be shaped as a result of the strong competitive movements taking place right now, the ICN might end up being attractive to many actors and hence be quickly adopted or might remain in the realm of research without attracting any real market interest.

6.1 Current state of the ICD market

The ICD market is characterized by a set of competitive and collaborative behaviors between the access ISPs, CDNs and content providers. The relationships among those three stakeholders steers the content delivery market towards certain directions. Hence, to understand the current state of the ICD market, it is necessary to analyze the dynamics of each of the three relationships that combine those three major stakeholders.

6.1.1 Access ISPs – CDNs Relationship

ISPs compete with CDNs over content providers' premium delivery contracts; yet, they enter into mutually beneficial relationship with them. The CDNs need to locate inside or directly peer with ISP POPs in order to have a global footprint which is the base of its offering to content providers. On the other hand, ISPs benefit from hosting the CDNs close or inside their networks in two ways, first they save transit costs on requests coming for content stored in CDNs as they are either delivered locally if the CDN's caches are hosted in the ISPs POPs or delivered through peering links which do not generate extra charges for the ISP. Secondly, the ISPs gain a competitive advantage in their local markets since they can serve

popular content with a superior quality due to the content being served locally inside their network and closer to their subscribers.

With the rise in video traffic and premium content, CDNs have been generating a lot of revenues and profits from content providers while at the same time getting a free ride on the access ISPs networks who are the ones paying the bill of hosting the CDNs servers in their networks. In contrast, the ISPs are suffering from the exponential increase in video traffic which is burdening their networks and are barely able to monetize it as they are forced to keep their flat-rate pricing in order to compete in markets dominated by flat-rate prices. Due to this as well as losing control over traffic patterns inside their networks due to the hosting of black boxes (i.e. CDNs), accesses ISPs are reevaluating the relationship with CDNs and many of them have already started expelling CDNs from their networks.

6.1.2 Access ISPs – Content Providers Relationship

Depending on the size of the content provider and whether or not it owns a distributing network, the relationship between it and the access ISPs can differ (Norton). Traditional content providers who are not operating their own delivery networks (e.g. eBay, Hertz and General Electric) view ISPs as transit vendors. Such relationship is a strictly customer-vendor relationship in which ISPs charge content providers based on the volume of the traffic transferred through their networks. The upper hand in this relationship is usually in the hands of the access ISPs who owns the end-user access control point. These types of relationships are not very common these days, as most traditional content providers resort to few global CDNs to deliver their content instead of having to contract with each access ISP independently.

On the other hand, large scale network savvy content providers (Norton) who own and operate their own delivery networks (e.g. Google, Microsoft and Sony Online) enjoy a more privileged relationship with access ISPs. Access ISPs usually deal with these content providers as peers in a similar fashion as they do with ISPs they enter into peering agreements with. The motives behind peering with large scale network savvy content providers (e.g. as Netflix does) or hosting their servers inside the access ISP's own POP (e.g. as Google Global Cache does) are: 1) the reduction of the volume of transit traffic that results from requests generated by these content providers and 2) provisioning of a better QoE to the access ISPs subscribers as the content is served from a nearby location.

In recent years the number of content providers who deploy their own delivery networks have risen which meant a rise in the number of servers the access ISPs need to accommodate in their networks or directly peer with. Also, the huge rise in traffic of these content providers have led to an imbalance in the relationship with the access ISPs in favor of the first who are leveraging such growth and influencing the traffic patterns inside the access ISPs networks. Hence, many access ISPs have started to shy away from these types of relationships in order to avoid being relegated into mere bit pipes for the content providers.

6.1.3 CDNs – Content Providers Relationship

The relationship that brings together CDNs and content provider is a purely customer-vendor business relationship. Content providers use the services of CDNs for at least one of four reasons: 1) Increase the performance to their end-users; 2) ensure the availability of their websites; 3) achieve scalability without the need to build huge server farms; and 4) increase security and mitigate distributed

denial of service (DDOS) attacks. The choice of the CDN to contract with is usually based on the type of functionality that the CDN can provide, the geographical regions where it has servers, cost, level of support and of course the brand of the CDN and its track record. From the CDN side, CDNs usually charge content providers based on the volume of traffic and provide SLAs which defines the minimum QoE in terms of latencies and availability.

Traditionally, most content providers have had direct relationships with CDNs. However, the market is currently in flux as more offerings are being made which are leading to the appearance of brokers between CDNs and content providers. Moreover, with many content providers, especially the large ones, contracting with multiple CDNs, a new breed of companies have emerged who potentially can be major influencers of the market. These companies are platform managers (e.g. Cedexis) who monitor the performance of the CDNs which their customers, the content providers, are using. Based on the performance, these platform managers orchestrate the flow of traffic among the different CDNs.

With many large content providers (e.g. Netflix) building their own CDN networks or contracting directly with access ISPs, the CDN providers are left to compete for smaller content providers in the market. Moreover, the CDNs are also facing a tough competition with ISPs who are developing their own CDN services which potentially can provide superior QoS for content providers particularly for video content.

6.1.4 Major Challenges of the ICD Market

The current ICD market is facing three major challenges that have destabilized the market and triggered the recent competitive moves that are analyzed later on in this Chapter. The three challenges that need to be overcome are:

1) CDNs cannot locate deeper inside ISP networks and hence can only guarantee the delivery of their customers' content up till the points where their cache servers are located. Such deficiency was not an issue when the content delivered were mostly webpages that do not have high delay and jitter requirements. However, with the increase in video traffic, especially high definition and 3D video, the requirements have become higher and content providers are increasingly demanding higher levels of QoS which traditional CDNs simply cannot guarantee.

2) Unlike CDNs who enjoy a wide geographical footprint due to the global distribution of their cache servers, local access ISPs' footprints are limited to the markets where they operate. Hence, local access ISPs can only offer content providers local delivery services. This is not an issue for localized content such as TV-series and movies in local languages. However, with the rise of global content providers and the internationalization of demand, more content providers are demanding global delivery with QoS guarantees that only ISPs can provide. The only option for content providers to achieve this today is to contract independently with multiple access ISPs around the world, which leads to high transaction costs and increased complexity.
3) With the migration of premium content to the online world, the demand by content providers for tighter control over the delivery of their premium content has been rising heavily in recent years. Hence, there have been

many conflicts between ISPs and content providers regarding caching and the illegal dissemination of copyrighted material. Such need for tighter control is also one of the reasons why many content providers have resorted to CDNs who provide a single contractual interface and can guarantee a higher level of control over the distribution of the premium content. Moreover, large content providers have gone all the way to building their own global network of caches in order to maintain such tight control.

6.2 Analysis of competitive dynamics in the current ICD Market

The three major challenges the ICD market is facing have led to a number of drastic movements by the three major stakeholders in the market. Figure 25 illustrates these movements and the motivations and triggers behind them. The causal relationships that join the components in the figure can either be supportive (denoted by a + sign on the arrow) or hindering (denoted by a – sign on the arrow). A total of 18 interesting trends and movements happening in the market are presented in the figure, among them four movements in particular, highlighted in yellow in Figure 25, are promising to change the structure of the market significantly.

1) The ISPs' deployment of CDNs.
2) The adoption of licensing model by pure play CDN providers.
3) The increase in CDN interconnection efforts.
4) The content providers' deployment of cache networks.

6.2.1 Rise of Telco CDNs

The major and most exciting disruption currently taking place in the ICD market is the movement by many ISPs (access and Tier-1) towards building their own CDN networks, called Telco CDNs (Idate, 2010). There are many motives behind such move towards a service segment already dominated by strong pure play CDN providers such as Akamai[3] and EdgeCast[4]. ISPs, especially Tier-1 ISPs, are trying to avoid being turned into delivery bit pipes that just enable new services (e.g. over the top (OTT)) without being able to generate revenues and profits from them. Hence, these ISPs are looking to get a share of the content providers' revenue going mostly to pure play CDNs whom they have been giving reduced infrastructure costs with no revenue sharing agreements. Another development driving this movement is the huge rise in video traffic which is overwhelming the ISPs' networks, especially the access ISPs, which cannot scale quickly enough to the increase in volume of traffic being generated. This trend is expected to continue as more cable and satellite TV content migrate to the Internet. Finally, access ISPs are looking to provide their subscribers a high quality and delay free service and in the process take the driving seat in the "single pipe for all services" race.

Access ISPs have a natural competitive advantage over pure play CDNs. The full control over the last mile connection will enable them to deploy caches much deeper than pure play CDNs can ever do. This will give them the ability to guarantee the QoS of content delivered through their CDNs by minimizing the distance the content needs to travel. Moreover, Telco CDNs will have a lower cost

[3] http://www.akamai.com/
[4] http://www.edgecast.com/

structure than pure play CDNs who need to factor in the bandwidth leasing expenses into their cost structure. Finally, access ISPs can bundle premium content with their other services to provide subscribers with attractive entertainment packages that generate more revenues both for the access ISPs and the content providers.

Nevertheless, Telco CDNs still suffer from a few downsides. These CDNs will have limited geographical footprint, especially in the case of access ISPs who are able to deploy them only in their local networks. This can make them unattractive for global content providers who are looking to deliver their content on a wide scale. Moreover, ISPs, both access and Tier1, are coming very late to the content delivery market and hence they are in a very early stage in the learning curve compared to the well-established pure play CDNs. This can lead to two situations: 1) the ISPs try to gain experience by building their own CDNs from scratch, which requires a lot of investments and time and might mean missing short term opportunities, or 2) access ISPs license the pure play CDNs' mature technology and benefit from the quick deployment but risk being at the mercy of their competitors.

In conclusion, Telco CDNs are in a very early stage and different ISPs are taking different routes in developing their CDNs. There are the ones who are adopting a do-it-yourself model such as Orange while others like AT&T have resorted to licensed technology. But judging by the level of excitement and concrete steps by ISPs towards the CDN market, we can expect a huge structural change in the ICD market coming our way.

6.2.2 Pure-play CDNs Adopting Licensing Model

The movement by many ISPs towards the CDN market has led to the adoption of licensing models by many pure-play CDNs. This movement mainly comes as a defensive reaction from the pure play CDNs who are forced to search for a new position in the value chain other than being the delivery providers of premium content or else they will be driven out of business. This is especially highlighted by the recent expulsion of many pure play CDNs from the ISP networks, on which they have been enjoying a free ride for many years. Nevertheless, not all pure play CDNs offering licensed technology have done so out of fear from losing their positions. Many pure-play CDNs (e.g. Jetstream[5] and Edgecast) have seen this as an opportunity to establish themselves in a new area of the market and avoid head on competition with the traditional big CDN players.

Adopting a licensed CDN technology rather than building one from scratch has its advantages and disadvantages. From ISPs' perspective, given the long experience and track record of pure play CDNs, they can benefit from using a proven technology which can get their CDNs up and running faster and with fewer hassles than developing one of their own. They can also exploit the strong relationship existing between pure play CDNs and content providers to bring more value to their networks.

By providing such licensed services, pure-play CDN providers can capitalize on the already established relationships with ISPs who have been hosting or directly peering with their servers for long. In addition to that, the pure-play CDNs allow for the interconnections of Telco CDNs through their licensed technology and hence give ISPs a two way benefit from such licensing agreement while obtaining a critical position in the value chain.

[5] http://www.jet-stream.com/

On the downside, most of the pure-play CDNs have been working independently from the ISPs in which they were deployed. This creates two challenges for the licensing offering. Firstly, the pure-play CDNs lack experience in solutions which are deeply embedded in the ISPs networks and face tough competition from traditional network equipment vendors like Ericsson[6] and Cisco[7]. Secondly, despite their established relationship, there is a lack of trust from the ISPs towards the CDNs as those CDNs have always been looked at as unmanageable black boxes residing in their networks.

It remains to be seen whether the licensing model will prosper or not, especially after the ISPs accumulate experience in managing CDN networks. But at least on the short term, the success of the licensing model is tied to the ISPs compromise between a quick CDN solution with some short-term gains and little few long-term competitive advantages, or a slow CDN development process which is costly but can provide long-term benefits.

[6] http://www.ericsson.com/
[7] http://www.cisco.com/

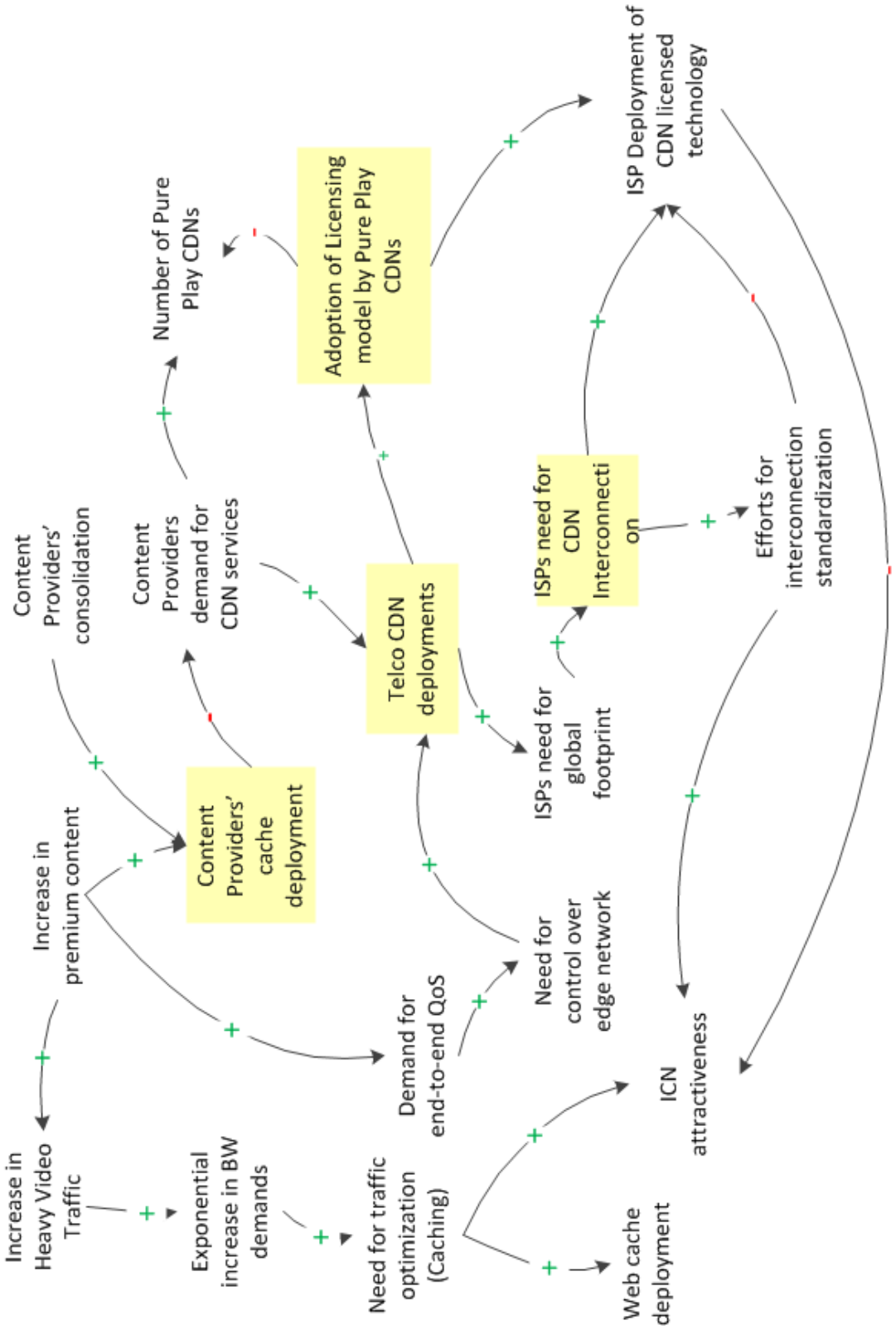

Figure 25: Major Competitive Dynamics of the ICD Market

6.2.3 Move Towards CDN Interconnection

The deployment of CDNs by ISPs who operate on a regional rather than a global level gives rise to a few problems which need to be addressed in order for such delivery model to be successful. The main issue is the ISPs' inability to provide content providers a truly global access to all their customers; forcing content providers to either go into independent agreements with each ISP or resort to a global best effort service provided by pure play CDNs. Moreover, many network operators who offer bundled content subscription along with their traditional services seek to deliver content to their subscribers even when they are roaming in networks other than theirs, which is referred to as off-net access, something which the separate island Telco CDN model cannot enable. These problems have driven ISPs to look for an interconnection solution, sometimes referred to as federated CDN (Robinson, 2012) through which they can benefit from providing cross CDN offerings and services to the content providers and to their subscribers.

The idea behind CDN interconnection is to find technical as well as contractual solutions for interconnecting disparate CDN systems. This will give the Telco CDNs the possibility to provide a single interface to content providers through which they can achieve global distribution for their content as well as allow for off-net delivery of content across multiple operators in a similar fashion to roaming services offered by mobile operators these days. Moreover, having such interconnection gives the network operators the possibility to offload some of their traffic to other CDNs during the peak hours, and hence provide a better QoS as well as reduce capital costs.

Although CDN interconnection promises to bring a lot of efficiency and possibilities for new business models for the ISPs, also many hurdles exist. Technically it is challenging to interconnect disparate systems; especially that many ISPs have resorted to licensed proprietary solutions from pure play CDNs and do not have the expertise to take on such challenge without the pure play CDN involvement. The licensing CDNs themselves might not be so eager to solve interconnection issues if they cannot profitably capitalize on them. Hence, many ISPs are forced to use the interconnection capabilities provided by their licensing CDN which typically limits the interconnection to networks deploying the same licensed technology.

On the other side, network equipment vendors such as Cisco have been driving the efforts for standardizing the CDN interconnections which can be a viable solution for the future of their customers, namely the ISPs, especially if no pure play CDN was able to provide a de facto interconnection standard. The IETF Content Delivery Networks Interconnection (cdni) working group (IETF, 2012) and the Open ContEnt Aware Networks (OCEAN) projects (OCEAN, 2012) are prominent examples of such standardization efforts. However, standardization takes a lot of time and many ISPs are looking for quick solutions in order to deal with the rapid increase in video traffic on their networks.

The second hurdle facing CDN interconnection is a business related one. There are no solutions yet to ensure a reliable cascading of payments and SLAs between the interconnecting CDNs, so conflicts may arise due to the different interests of different telecom operators. There are generally two ways of approaching the agreements between interconnecting CDNs, one is that each operator go into bilateral agreements with the operators they want to interconnect with, the other

way is to have an interconnection hub which can manage the interconnections and the settlements fees based on the volume of traffic exchanged (Skytide, 2011).

6.2.4 Deployment of Caches by Content Providers

With the consolidation of traffic in the hands of few hyper giants, many of them have deployed their own network of caches globally. Such movement is especially visible in video content providers who are seeking to deliver their content as close to their customers as possible, hence, the likes of Google's YouTube[8] and Netflix[9] are currently delivering either all or part of their traffic though their own distributed network.

There are numerous obvious benefits to the content providers from adopting such delivery model. It gives them full control over their content and a better visibility of request patterns which allows them to react faster to any changes occurring in requests coming their way – something that is increasingly important as more premium content migrate online. Furthermore, given the huge volume of traffic they generate, the economies of scale make the deployment of their own infrastructure more cost-effective than paying a brokerage fee to middlemen (pure play CDN) to deliver their content.

For many years the telecom operators have entered in mutually beneficial relationships with content providers by freely hosting their cache servers and in return saving on transit and peering traffic costs (Dolce, 2011). However, this has changed recently, especially with the significant growth in those content providers traffic who went on to dominate much of the traffic inside their hosting telecom operators' networks. This has caused disruptions in telecom operators' control over traffic patterns inside their networks and relegating their role to mere bit pipes. Moreover, the increase in this trend means that telecom operators are expected to host racks for each content provider in their sites which is obviously a sub-optimal solution compared to having just one CDN hosting the entire content providers' content.

Unlike the relationship between pure play CDNs and telecom operators in which the latter can survive without the earlier, the relationship between content providers and telecom operators is more sensitive, since the former needs to reach their customers through the latter, while the latter needs the content of the former to provide attractive offerings for their subscribers. Such balance will determine how telecom operators will deal with those content providers insisting on having their own caches rather than utilizing the telecom operator CDN. Solving this issue could be one of the main driving forces for the success of the Telco CDN model.

6.3 ICN Adoption

The structural changes occurring in the ICD market as a result of the four major dynamics taking place will significantly influence the adoption of ICN. ICN is considered both a technology and an architecture. It can be implemented on a small scale starting with individual interconnecting devices up to local deployment by ISPs, or on a large scale to replace the current host-centric global Internet architecture. This flexibility in the scale of adoption means that ICN can have a place in most of the scenarios unfolding in the future although this place will vary in size and importance from one scenario to another.

[8] http://www.youtube.com/
[9] http://www.netflix.com/

By far, the biggest disruption that the content delivery market is currently witnessing, as has been discussed earlier in this chapter, is the rise of Telco CDNs. This disruption will surely influence the adoption of ICN since the CDN architectures have many similarities to the ICN architectures. There are generally two directions which the Telco CDN market may take. The first is the proprietary solutions direction, which will be heavily dominated by the experienced and well placed pure-play CDNs and their licensing offerings. While the second is the standardization direction which will depends on the success of CDN interconnection standardization efforts and will, to a certain extent, be dominated by the ISPs and the network equipment vendors.

6.3.1 ICN Adoption in a Proprietary Telco CDN Market

If the trend of ISPs licensing CDN technology from pure-play CDNs continues, then the ICD market will head to a future heavily dominated by pure-play CDN proprietary technologies. In such future, even though, the actual CDN services will be offered by ISPs, the technology development will be totally in the hands of pure-play CDNs in a similar fashion as in the mobile market where the equipment vendors are the ones directing the technology development while operators are buying their products. The pure-play CDNs will not only act as vendors for CDN technology but will also form the backbone of the new content network to interconnect the different Telco CDNs.

Such scenario might slow down the ICN adoption. ICN concepts are built around the idea of unrestricted ubiquitous caching and automatic interconnectivity between the different networks. These concepts clearly conflicts with those of the pure play CDNs who will be looking to further promote their proprietary technologies to generate more revenues from ISPs as well as ensuring interconnection happens through their platforms to maintain the revenue coming from interconnection charges.

Nevertheless, the ICN as a technology can still have a place in such a future which is dominated by pure-play CDNs. Pure-play CDNs, even if they are currently not so much involved in ICN research as ISPs and network equipment vendors are, will definitely be interested in differentiating their offerings from each other as the competition heats up. And one of such differentiators will definitely be the adoption of ICN technology as a superior way to perform in-network caching and request routing.

6.3.2 ICN Adoption in a Standardized Telco CDN Market

Many ISPs are developing their own CDNs using their internal expertise or assisted by network equipment vendors. Such efforts follow the pattern which ISPs have traditionally adopted regarding new infrastructure deployments in their networks which is limited to cooperation with network equipment vendors. If this pattern continues then it is expected that strong standardization efforts will take place in parallel. This is to ensure interconnection and interoperability between the different ISPs and network equipment vendor technologies in a similar way as in today's highly standardized network infrastructures.

Such highly standardized Telco CDN market can be considered the optimum from the point of view of ISPs as it removes the pure play CDNs from the value network totally and increases the ISPs' share of the revenues coming from content providers in the ICD market. Such a scenario will be similar to today's mobile networks, where the mobile operators own and operate their networks locally, but

can serve their roaming off-net subscribers through other mobile operators who they are in direct roaming agreements with.

As from ICN architectural adoption perspective, such standardized ICD market can be seen as a positive step towards global adoption. With the limited influence of pure-play CDNs on the ICD market, equipment vendors and ISPs who are already championing the ICN technology and are the ones involved in most of its research activities will dominate the technology development. Hence, ISPs will have more freedom in experimenting with ICN technology both on the local scale as well as on the global scale by building upon well-defined standards for interconnection.

7 Conclusion

This study set out to determine the effect of caching on the competitive dynamics of the Internet content delivery market. The work especially focused on the change which ICN caching brings to the market and its effect on the three main stakeholders, the content providers, the access networks and the content delivery networks (CDNs).

7.1 Key Findings

The key findings that emerged from this study are the identification of the domains influencing content caching, the identification of key control point shifts stemming from ICN caching, the singling out of two critical shortcomings of the proposed ICN architectures and the identification of major market disruptions.

This work took a bottom-up approach in studying the effects of caching on the competitive dynamics of the content delivery market. The analysis began by identifying the domains which influence caching in order to gain a clear understanding of the whereabouts and dynamics of caching. Three main domains were found to be major influencers on caching: 1) the mobility domain, which introduces new types of constraints due to the high churn rate of users and the physical limitation of the wireless channels; 2) the naming of content, which directly influence the feasibility of some caching architectures; and finally 3) the nature and usage context of content, which directly influence its cacheability. These identified domains are in line with many previous studies; however, the main contribution in this work lies in the holistic approach which was adopted when studying these domains especially when identifying the cacheability parameters in the third domain. This work was interested in identifying not only the traditional known technical parameters, but expanding the analysis to encompass economic and contractual parameters as well. A total of 13 cacheability parameters, technical, economic and contractual were identified and built upon in this work.

The second contribution is the identification of the key control point shifts that are due to occur if the content delivery market adopts ICN. Control points of the ICD ecosystem were heavily studied before; however, this work was interested more in identifying the consequences of control point shifts rather than identifying the control points themselves. Hence, in an effort to understand their nature, the control point shifts were classified into three categories: authority shift, distributional shift and consolidation shift. This classification facilitated the analysis of the four main caching control points highlighted in this work: delivery choice decision, cache management, request routing and user access.

It was found that the request routing and cache management control points are the ones which will witness significant shifts if today's caching architectures were replaced by an ICN caching architecture. In contrast, the delivery choice decision and user access control points are the least effected by the shift to ICN. Also, the analysis supported the claim that ICN will democratize the market as no consolidation shifts occur when adopting it. Moreover, it was found that the shifts in control points will heavily affect the position of the pure play CDNs who lack control over any of the two critical non shifting control points: the delivery choice decision and the user access.

The analysis has further highlighted two major shortcomings that the proposed ICN architectures suffer from. The first being the contractual complexity which premium content providers will face in case ICN was fully adopted and pure play CDNs retreated from the market. The second is the handing of most of the control points to one player, which is the ISP in this case, and hence making the whole ICN preposition unattractive to other important players in the market namely content providers and pure play CDNs.

Finally, the work has identified four major developments that the market is witnessing as part of the competitive dynamics analysis. The most important of these developments is the rise of Telco CDN which promises to significantly change the ICD market structure and is responsible for two of the other developments, the adoption of licensing model by pure-play CDNs and the increase in CDN interconnection efforts. These developments will drive the market in one of two directions, a proprietary or a standardized Telco CDN future with each scenario influencing ICN adoption in its own way.

7.2 Exploitation of Results

There are many exploitation paths for the results presented in this work. The holistic study which resulted in an exhaustive list of cacheability parameters in particular can be beneficial in designing intelligent caching mechanism. Such mechanisms can take into consideration not only technical aspects such as the number of requests and size of objects but can also be designed to exploit economic and contractual information about the content in order to make more informed decisions. These intelligent caching systems can especially be utilized in ICN networks, which are inherently knowledgeable of the type of content they are delivering, making the exploitation of such knowledge for improving caching decisions very attractive.

On the other hand, the comprehensive analysis of the shift in critical control points resulting from moving from current content delivery architectures to the newly proposed ICN architectures can be utilized in the socio-economic studies which relate to future Internet projects. The generic shifts identified in this work form the basis of a methodology which can be utilized to study any architectural changes which include shifts in control points from certain stakeholders to other ones.

Finally, this work encourages ICN research projects such as SAIL to take into consideration the two shortcomings, which were highlighted in the analysis. The main challenge lies in the fact that ICN research is being pushed by telecom operators who are motivated by the idea of increasing the efficiency of content delivery inside their networks. However, they also need to consider the shift in control, which such architectural change will bring and ensure that there is enough value for all stakeholders in the new ICN ecosystem in order for it to prosper.

7.3 Future Research

This work took a holistic approach in analyzing caching influence on the ICD market and the important role it plays in ICN. It can be considered as an exploratory kind of study on which more concrete analysis can be constructed in the future.

One limitation of this study was the lack of diversity in the experts who the interviews were conducted with. Only one telecom operator and one CDN were represented during the interviews, and hence the results were expected to be

slightly biased towards the interests of these two stakeholders Hence, more diverse interviews with the three main stakeholders of the ICD market are needed to support the conclusions reached during this study.

Moreover, since this study was conducted under the umbrella of the SAIL project; it focused to a large degree on the NetInf reference architecture without taking into account other ICN architectures. Therefore, it will be valuable in future studies to analyze all the ICN architecture variations in order to be able to generalize the results reached in this work.

Finally, a more detailed analysis of the technology development inside pure play CDNs would be a very interesting future research topic. Pure play CDNs have traditionally been viewed as black boxes and very little information is known about their proprietary technologies or development tracks. Nevertheless, with more pure play CDNs adopting licensing technologies, their solutions are expected to be more visible to the research community. This will especially be valuable for the research efforts in ICN technologies.

References

Abedini, N. and Shakkottai, S., 2011. Content Caching and Scheduling in Wireless Broadcast Networks with Elastic and Inelastic Traffic. IEEE International Symposium of Modelling and Optimization of Mobile, Ad Hoc, and Wireless Networks, pp. 125-132

Ager, B., Schneider, F., Juhoon, K. and Feldmann, A., 2010. Revisiting Cacheability in Times of User Generated Content. INFOCOMM IEEE Conference on Computer Communications Workshops, pp. 1-6

Aggarwal, C., Wolf, J.L. and Yu, P.S, 1999. Caching on the World Wide Web. IEEE Transactions on Knowledge and Data Engineering. pp. 94-107

Ahlgren, B., Dannewitz, C., Imbrenda, C., Kutscher, D. and Ohlman, B., 2011. A survey of information-centric networking.

Akamai Technologies Inc., 2010. The State of the Internet, 4th Quarter, 2010 Report, Volume 3, Number 4. [online] Available at: < http://www.akamai.com/stateoftheinternet/> [Accessed 15 January 2012].

Akamai Technologies, Inc., 2012. About Akamai. [online] Available at: <http://www.akamai.com/html/about/index.html> [Accessed 18 April 2012].

Alimi, R., Rahman, A., and Yang, Y., 2011. A Survey of In-network Storage Systems. IETF, DECADE Internet-Draft.

Anand, A., Muthukrishnan, C., Akella, A. and Ramjee, R., 2009. Redundancy in network traffic: findings and implications. ACM, pp. 37-48

Aranda, P.A et al., 2010. Final Architectural Framework. [online] Available at:< http://www.4ward-project.eu/> [Accessed 23 June 2012]

AT&T Enterprise, 2012. Content Distribution. [online] Available at: <http://www.business.att.com/enterprise/Family/content-delivery/distribution/> [Accessed 18 April 2012].

Barish, G. and Obraczka, K., 2000. World Wide Web Caching: Trends and Techniques. IEEE Communications Magazine, 38(5), pp. 178-184.

BBC Worldwide Ltd., 2012. BBC iPlayer. [online] Available at: < http://www.bbc.co.uk/iplayer/tv> [Accessed 17 January 2012].

Breslau, L., Cao, P., Fan, L., Philips, G. and Shenker, S., 1999. Web caching and Zipf-like distributions: evidence and implications. Computer and Communications Societies. Proceedings. IEEE.

Catrein, D., Löhrer, B., Meyer, C., Rembarz, R. and Weidenfeller, T., 2011. An analysis of Web Caching in Current Mobile Broadband Scenarios. 4th IFIP International Conference on New Technologies, Mobility and Security (NTMS), pp. 1-5

Chen, M. and Leong, J., 2005. Competitive Dynamics as Action-Response. Darden School of Business.

Cheng, K. and Kambayashi, Y., 2000. LRU-SP: a size-adjusted and popularity-aware LRU replacement algorithm for web caching. IEEE Computer Software and Applications Conference.

Chow, C.Y., Leong, H.V. and Chan, A.T.S., 2007. GroCoca: Group-based peer-to-peer cooperative caching in mobile environment. IEEE Journal on Selected Areas in Communications, 25(1), pp. 179-191.

Cisco Systems Inc., 2011. Cisco Visual Networking Index (VNI), Entering the Zettabyte Era. [online] Available at: <http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/VNI_Hyperconnectivity_WP.pdf > [Accessed 15 January 2012].

D'Ambrosio, M. and Dannewitz, C., 2011. MDHT: a hierarchical name resolution service for information-centric networks. ICN '11 Proceedings of the ACM SIGCOMM workshop on Information-centric networking, pp. 7-12

Dannewitz, C., 2009. NetInf: An Information-Centric Design for the Future Internet. Proc. 3rd GI/ITG KuVS Workshop on The Future Internet.

Dannewitz, C., Golic, J., Ohlman, B. and Ahlgren, B., 2010. Secure naming for a network of information. INFOCOM IEEE Conference on Computer Communications Workshops, pp. 1-6.

Datta, A., 2001. A comparative study of alternative middle tier caching solutions to support dynamic web content acceleration. 27th VLDB Conference. Roma.

Davidson, B.D, 2001. A web caching primer. Internet Computing, IEEE. pp. 38-45

Dilley, J., Maggs, B., Parikh, J., Prokop, H., Sitaraman, R. and Weihl, B., 2002. Globally distributed content delivery. Internet Computing, IEEE, 6(5), pp. 50-58

Dolce, J. 2011. Global Caches Everywhere?. [online] Available at: http://www.verivue.com/blog/index.php/uncategorized/global-caches-everywhere/ [Accessed 26 June 2012]

Dong, W., Ge, Z. and Lee, S., 2011. 3G Meets the Internet: Understanding the Performance of Hierarchical Routing in 3G Networks. Proceeding of ITC 2011, pp. 15-22

Edgecast, 2009. Deutsche Telekom ICSS and Edgecast networks partner to launch innovative content delivery solution (CDS). [online] Available at: <http://www.edgecast.com/pr_deutsche_telekom_edgecast.htm> [Accessed 18 April 2012].

Eugster, P.T., Felber, P.A., Guerraoui, R. and Kermarrec, A.M., 2003. The many faces of publish/subscribe. ACM Computing Surveys (CSUR), 35(2), pp. 114-131

ESPN Network, 2012. ESPN Player. [online] Available at: <http://www.espnplayer.com/espnplayer/console> [Accessed 18 May 2012]

Fotiou, N., Nikander, P., Trossen, D. and Polyzos, G.C., 2010. Developing information networking further: From PSIRP to PURSUIT. Proc. 7th International ICST Conference on Broadband Communications, Networks, and Systems.

Funkhouser, T.A, et al., 1996. Network topologies for scalable multi-user virtual environments. Proceedings of the 1996 Virtual Reality Annual International Symposium.

Ghodsi, A., Koponen, T. and Rajahalme, J., 2011. Naming in Content-Oriented Architectures. SIGCOMM ICN '11.

Ghodsi, A., Shenker, S., Koponen, T., Singla, A., Raghavan, B. and Wilcox, J., 2011. Information-centric networking: seeing the forest for the trees. Proceedings of the 10th ACM Workshop on Hot Topics in Networks.

Giannaki, V., Vasilakos, X., Stais, C., Polyzos, G. and Xylomenos, G., 2011. Supporting Mobility in Publish Subscribe Internetwork Architecture. Proceedings of the IEEE ISCC.

Golrezaei, N., Shanmugam, K., Dimakis, A.G., Molisch, A.F. and Caire, G., 2011. FemtoCaching: Wireless Video Content Delivery through Distributed Caching Helpers.

Google, 2012. Google Global Cache. [online] Available at: <http://ggcadmin.google.com/ggc> [Accessed 18 April 2012].

Hefeeda, M. and Saleh, O., 2008. Traffic modeling and proportional partial caching for peer-to-peer systems. IEEE/ACM Transactions on Networking, 16(6), pp. 1447-1460

Heikkinen, M., 2012. Techno-Economic Analysis of Mobile Peer-to-Peer Systems and Services. Doctoral Dissertation, Aalto University publication series.

Hosanagar, K. and Tan, Y., 2006. Cooperative Cashing? An Economic Analysis of Document Duplication in Cooperative Web Caching. Information Systems Research.

IBM, 2012. IBM Server-side request caching. [online] Available at: <http://publib.boulder.ibm.com/tividd/td/ITAME/SC32-135900/en_US/HTML/am51_webseal_guide67.htm> [Accessed 15 January 2012].

ICANN, 2012. Internet Corporation for Assigned Names and Numbers. [online] Available at:<http://www.icann.org/> [Accessed 2 June 2012]
Idate Consulting and Research, 2010. Evolution of the CDN market. Presentation given during the CDN World Summit 2010.

IETF, 2012. Content Delivery Networks Interconnection (cdni) - Working Group Charter. [online] Available at: http://datatracker.ietf.org/wg/cdni/charter/ [Accessed 26 June 2012]

Jacobson, V., Smetters, D., Thornton, J., Plass, M., Briggs, N. and Braynard, R., 2012. Networking named content. Communications of the ACM, 55(1), pp.117-124.

Joseph, D.A., Manoj, BS and Murthy, C., 2004. Interoperability of Wi-Fi hotspots and cellular networks. Proceedings of the 2nd ACM international workshop on Wireless mobile applications and services on WLAN hotspots, pp. 127-136

Kaaranen, H., Ahtiainen, A., Laitinen, L., Naghian, S. and Niemi, V., 2001. UMTS Networks. Wiley Online Library.

Koponen, T., Chawla, M., Chun, B., Ermolinskiy, A., Kim, K. H., Shenker, S. and Stoica, I., 2007. A data-oriented (and beyond) network architecture. Proceedings of SIGCOMM'0, 37(4), pp. 181-192

Labovitz, C., 2011. Internet Traffic Evolution 2007-2011. Global Peering Forum

Leiner, B.M., Cerf, V.G., Clark, D.D., Kahn, R.E., Kleinrock, L., Lynch, D.C., Postel, J., Roberts, L.G. and Wolff, S.S., 1997. The past and future history of the Internet. Communications of the ACM, 40(2), pp. 102-108

Lua, E.K., Crowcroft, J., Pias, M., Sharma, R. and Lim, S., 2005. A Survey and Comparison of Peer-to-Peer Overlay Network Schemes. IEEE Communications Surveys, 7(2), pp. 72-93.

Lungaro, P., Segall, Z. and Zander, J., 2010. Context-aware RRM for Opportunistic Content Delivery in Cellular Networks. 2010 Third International Conference on Communications Theory, Reliability, and Quality of Service, pp. 175-180

Lungaro, P., Segall, Z. and Zander, J., 2010. ContextShift: a Model for Efficient Delivery of Content in Mobile Networks. IEEE Communications Society.

Lv, Q., Cao, P., Cohen, E., Li, K. and Shenker, S., 2002. Search and replication in unstructured peer-to-peer networks. ICS '02 Proceedings of the 16th international conference on Supercomputing, pp. 84-95.

Malik, O., 2011. The Storage vs bandwidth debate. Gigaom, [online] 24 June. Available at: <http://gigaom.com/broadband/the-storage-vs-bandwidth-debate/> [Accessed 30 May 2012].

Memcached, 2012. Memcached - a distributed memory object caching system. [online] Available at: < http://memcached.org/> [Accessed 15 January 2012].

Menascé, D. and Akula, V., 2007. Improving the Performance of Online Auctions Through Server-side Activity-based Caching. World Wide Web, 10(2), pp. 181-204.

Menascé, D. and Almeida, V., 2001. Capacity Planning for Web Services: metrics, models, and methods.

Microsoft, 2012. ASP.NET Caching. [online] Available at: <http://www.asp.net/moving-to-aspnet-20/tutorials/caching> [Accessed 15 January 2012].

Mishra, A., Shin, M. and Arbaush, WA, 2004. Context caching using neighbor graphs for fast handoffs in a wireless network. INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies, 1.

MIT, 2012. MIT Communications Futures Program. [online] Available at: < http://cfp.mit.edu/> [Accessed 23 June 2012]

Netflix Inc., 2012. How Netflix Works. [online] Available at: <https://signup.netflix.com/MediaCenter/HowNetflixWorks> [Accessed 15 January 2012].

Norton, B. The Global Internet Peering Ecosystem. [online] Available at: < http://drpeering.net/AskDrPeering/blog/articles/The_Internet_Peering_Ecosystem__The_Content_Provider.html> [Accessed 24 June 2012]

OCEAN, 2012. Open ContEnt Aware Networks (OCEAN). EU FP7 project. [online] Available at: http://www.ict-ocean.eu/ [Accessed 26 June 2012]

Pallis, G. and Vakali, A., 2006. Insight and perspectives for content delivery networks. Communications of the ACM, 49(1), pp. 101-106

Pathan, A. and Buyya, R., 2006. A taxonomy and survey of content delivery networks. Fifth International Joint Conference on INC, IMS and IDC, IEEE. pp. 44-51.

Reese, G., 2000. Database Programming with JDBC and Java, Second Edition. O'Reilly Media. Ch.7.

Riihijärvi, J., Trossen, D., Marias, G., Burbridge, T., Zahemszky, A., Ylitalo, J., et al., 2009. Deliverable D4.2, First report on quantitaive and qualitative architecture validation. PSIRP.

Robinson, D., 2012. CDN Federation: A Badly-Defined Solution in Search for a Real Problem?. [online] Available at: < http://www.streamingmedia.com/Articles/Editorial/Featured-Articles/CDN-Federation-A-Badly-Defined-Solution-in-Search-of-a-Real-Problem-80757.aspx> [Accessed 26 June 2012]

Robson, C., 2002. Real World Research. 2nd Edition. Blackwell Publishing. p. 270.

SAIL, 2011. D-3.1 (D-B.1) The Network of Information: Architecture and Applications. [online] Available at: < http://www.sail-project.eu/wp-content/uploads/2011/08/SAIL_DB1_v1_0_final-Public.pdf> [Accessed 25 April 2012]

SAIL, 2012. Scalable and Adaptive Internet Solutions (SAIL). [online] Available at: <www.sail-project.eu> [Accessed 18 April 2012].

Salo, J., et al., 2011. New Business Models and Business Dynamics of the Future Networks. SAIL, D-2.7.

Sandvine Inc., 2011. Global Internet Phenomena Spotlight, Europe, Fixed Access. [online] Available at: < http://www.sandvine.com/news/global_broadband_trends.asp> [Accessed 15 January 2012].

Sandvine Inc., 2012. Global Internet Phenomena Report: 1H 2012. [online] Available at: < http://www.sandvine.com/news/global_broadband_trends.asp> [Accessed 10 June 2012].

Saroiu, S., Gummadi, K.P., Dunn, R.J., Gribble, S.D. and Levy, H.M., 2002. An analysis of ICD systems. ACM SIGOPS Operating Systems Review, 36(SI), pp. 315-327.

Sesia, S., Toufik, I. and Baker, M., 2009. LTE - The UMTS Long Term Evolution. From Theory to Practice, 66.

Skytide, 2011. 7 Online Video Trends to Watch in 2012. White paper. [online] Available at: http://www.skytide.com/blog/7-online-video-trends-to-watch-in-2012.html [Accessed 26 June 2012]

Sopcast, 2012. Sopcast Technology. [online] Available at: <http://www.sopcast.org/info/sop.html> [Accessed 30 May 2012].

Telefonica, 2012. Telefónica Content Delivery Network. [online] Available at: <http://www.telefonica.com/cdn/en/solutions/solutions_overview.shtml> [Accessed 18 April 2012].

Trossen, D., et al., 2011. Conceptual Architecture: Principles, Patterns and Sub-components Description. [online] Available at: < http://www.fp7-pursuit.eu/PursuitWeb/> [Accessed 23 June 2012]

Trossen, D., Sarela, M. and Sollins, K., 2010. Arguments for an information-centric internetworking architecture. ACM SIGCOMM Computer Communication Review, 40(2), pp. 26-33

Vakali, A. and Pallis, G., 2003. Content Delivery Networks: Status and Trends. IEEE Internet Computing, 7(6), pp. 68-74.

Wolman, T., Voelker, M., Sharma, N., Cardwell, N., Karlin, A. and Levy, H.M., 1999. On the scale and performance of cooperative Web proxy caching. ACM SIGOPS Operating Systems Review, 33(5), pp. 16-31

Yu, P. and Macnair, E., 1998. Performance study of a collaborative method for hierarchical caching in proxy servers. Computer Networks and ISDN Systems, pp. 215-224.

Zeng, Z., & Veeravalli, B., 2008. Hk/T: A Novel Server-Side Web Caching Strategy for Multimedia Applications. 2008 IEEE International Conference on Communications, pp. 1782-1786.

Zhang, L., Estrin, D., Burke, J., Jacobson, V., et al., 2010. Named data networking (NDN) project. PARC Tech Report 2010-003.

Zhang, Y., Li, D., & Zhu, Z., 2008. A Server Side Caching System for Efficient Web Map Services. International Conference on Embedded Software and Systems Symposia, pp. 32-37.

Zhu, Y. and Hu, Y., 2003. Exploiting client caches: An approach to building large web caches. Proceedings of International Conference on Parallel Processing, pp. 419-426.

Appendix A

Content Caching Questions

- Why did content caching fail to fly inside ISPs, is it because:
    - Most of the contents were not cacheable?
    - There were agreement problems with content providers?
    - The return on investment did not justify deployment of caches?
    - Other reasons?
- When an ISP decides to implement transparent caching:
    - Does it specify certain content providers to cache their content? or,
    - Are the caching algorithms transparent to the source of the content?
- From this list can you tell me which content do you think is 1) highly cacheable (H), 2) moderately cacheable (M) and, 3) lowly cacheable (L)?
    - Live Video
    - Video-on-Demand
    - P2P shared files
    - Web Pages
    - Real-time communication
    - Bulk entertainment
    - Online Gaming
    - Secured traffic
    - Software updates
    - Social networking
    - Storage and backup services
    - Emails
- In reality which of these content is usually cached? And where does the caching take place (ISP, CDNs, client-side, other)?
- What is the difference between the current efforts of CDN interconnection and the failed cooperative caching initiatives?

CDN Questions

- What are the channels that CDNs use to sell their services to content providers?
    - Direct, through brokers, both
    - Is it the same in pure play CDNs as ISP owned CDNs
- Are there any exclusivity terms in the contracts between CDNs and CPs?
- What is the stance of pure CDN players from the different ICN projects?
- Are there any agreements between CDNs to mutually increase their geographical footprints?
- Do ISPs who offer CDN services allow external CDNs (e.g. Akamai) to locate in their POPs?
    - If yes, what is the motivation behind that?
- Do CDNs mostly use proprietary or standard solutions in their networks?

NetInf and Naming Questions

- External trust mechanisms are needed in both hierarchical naming and self-certifying names. In the NetInf context which entity will play this role?
    - Is it the NRS?
    - What about in the NbR?
- Will self-certifying names require another service to translate between human-friendly and self-certifying names?
    - If yes, which entity will be responsible for that?
- Will the IANA (ICANN) manage the unique names like it manages IP addresses? What are the benefits of having the same entity controlling IO names and underlying IPs?
- What will stop ISPs from misusing NRSs to direct their users to the least costly IO rather than the most efficient?
- What is the main advantage for ISPs for deploying NetInf nodes over traditional caches or proprietary CDNs?
    - Akamai offering licensed CDNs
    - What are the costs of deploying NetInf nodes in the network compared to deploying regular caches?
    - Are there technical constraints on the caches that will be deployed in the NetInf network or can the already deployed proxy caches and CDNs be used?
- How will naming of objects impact the web search engines like Google? Especially that searching will be inherent in the IO?
- How will naming of objects affect the end-user?
    - Will he use search engines to reach IOs?
    - Can he look-up information objects directly?
    - If yes, then how?
- If a publisher publishes pirated copyrighted IO what is the mechanism that will counter act such behavior in NetInf?
- If a publisher wants to delete an IO is there a mechanism to achieve this so that all cached instants of this IO are permanently deleted?
- Does NetInf target end-user devices as potential caching nodes?

Competitive Dynamics Questions

- Increasingly CDNs and server farms are bypassing the public backbone altogether and connecting to their caches through private peering links, in the process transforming CDNs into a fundamentally different network architecture. How will this affect the deployment of NetInf? And isn't such an architecture (using redirection for close by caches) similar in concept to the NetInf?
- Some CDNs such as Akamai are offering licensed CDN services to operators, how far can this be considered direct competition against the adoption of NetInf?
- CDNs are present in different POPs of major ISPs around the world and hence present an attractive distribution solution for CPs. If NetInf is deployed by each ISP

    - Will content providers need to contract with each ISP separately?
    - Does this mean that there's a need for a broker layer between NetInf ISPs and CP?
    - If so, who will play this role?
- What is the stance of leading IP router vendors such as Cisco, Huawaui, Alcatel and Ericsson from the different ICN projects?
- How does vertical integration (Content-Transport) impact the NetInf attractiveness to ISPs?
- How will the NetInf adoption affect the entry barriers for new players to the content delivery market?
- What are the first mover advantages for deploying NetInf nodes?
- Who will be most harmed by the introduction of NetInf?
- Who is the player that can truly drive the adoption of NetInf?
    - ISPs, CDNs, Vendors, CP, Standardization Organizations