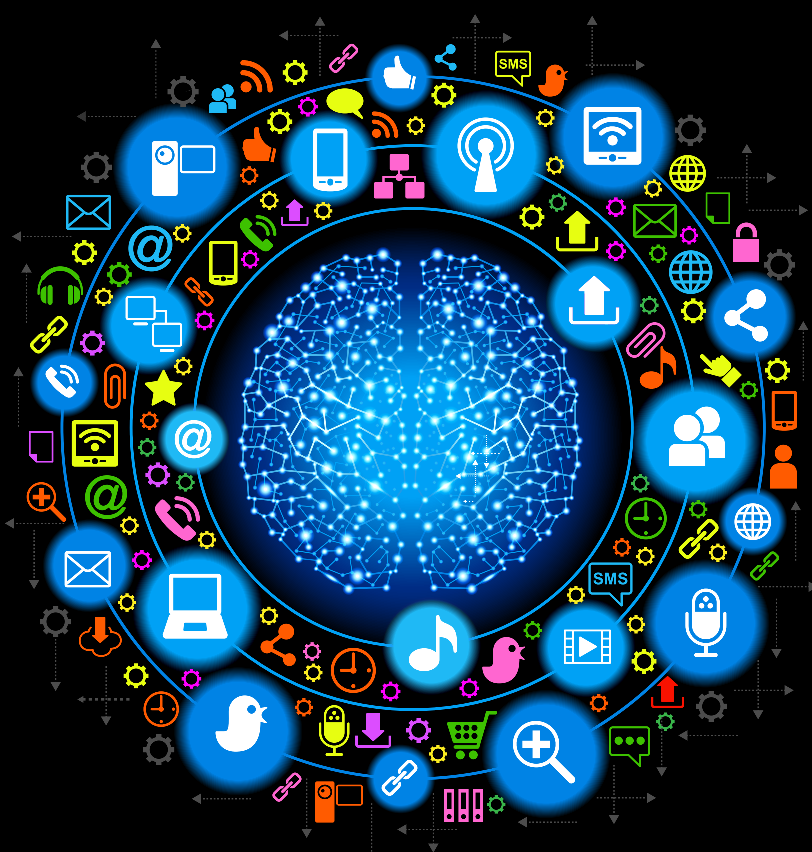
March 2018

Evolution to Content Optimized Networks, Issue 2

Value Assessment Report: A Comparative Study of ICN Versus Conventional Approaches



Abstract

eCON Value Assessment Report – A Comparative Study of ICN versus Conventional Approaches provides the reader with an in-depth analysis of ICN-based solutions (compared to IP-centric architectures) from the perspective of a network architect or network planner. The value assessment is intended to convey an understanding of the design choices and their impacts on network performance and application enablement, as they relate to market-impacting use cases that are on the horizon for most network operators. While ICN is still in a developmental and experimental phase, early results at both a quantitative and qualitative level are presented in this report. These results are then applied to develop a set of early findings that can assist in making informed network evolution choices.

Foreword

As a leading technology and solutions development organization, the Alliance for Telecommunications Industry Solutions (ATIS) brings together the top global ICT companies to advance the industry's business priorities. ATIS' 150 member companies are currently working to address 5G, cybersecurity, robocall mitigation, IoT, artificial intelligence-enabled networks, the all-IP transition, network functions virtualization, smart cities, emergency services, network evolution, quality of service, billing support, operations, and much more. These priorities follow a fast-track development lifecycle – from design and innovation through standards, specifications, requirements, business use cases, software toolkits, open source solutions, and interoperability testing.

ATIS is accredited by the American National Standards Institute (ANSI). ATIS is the North American Organizational Partner for the 3rd Generation Partnership Project (3GPP), a founding Partner of the oneM2M global initiative, a member of the International Telecommunication Union (ITU), and a member of the Inter-American Telecommunication Commission (CITEL). For more information, visit www.atis.org.

Notice of Disclaimer and Limitation of Liability

The information provided in this document is directed solely to professionals who have the appropriate degree of experience to understand and interpret its contents in accordance with generally accepted engineering or other professional standards and applicable regulations. No recommendation as to products or vendors is made or should be implied.

NO REPRESENTATION OR WARRANTY IS MADE THAT THE INFORMATION IS TECHNICALLY ACCURATE OR SUFFICIENT OR CONFORMS TO ANY STATUTE, GOVERNMENTAL RULE OR REGULATION, AND FURTHER, NO REPRESENTATION OR WARRANTY IS MADE OFMERCHANTABILITY OR FITNESS FOR ANY PARTICULAR PURPOSE OR AGAINST INFRINGEMENT OF INTELLECTUAL PROPERTY RIGHTS. ATIS SHALL NOT BE LIABLE, BEYOND THE AMOUNT OF ANY SUM RECEIVED IN PAYMENT BY ATIS FOR THIS DOCUMENT, AND IN NO EVENT SHALL ATIS BE LIABLE FOR LOST PROFITS OR OTHER INCIDENTAL OR CONSEQUENTIAL DAMAGES. ATIS EXPRESSLY ADVISES THAT ANY AND ALL USE OF OR RELIANCE UPON THE INFORMATION PROVIDED IN THIS DOCUMENT IS AT THE RISK OF THE USER.

NOTE - The user’s attention is called to the possibility that compliance with this standard may require use of an invention covered by patent rights. By publication of this standard, no position is taken with respect to whether use of an invention covered by patent rights will be required, and if any such use is required no position is taken regarding the validity of this claim or any patent rights in connection therewith. Please refer to [http://www.atis.org/legal/patentinfo.asp] to determine if any statement has been filed by a patent holder indicating a willingness to grant a license either without compensation or on reasonable and non-discriminatory terms and conditions to applicants desiring to obtain a license.

Copyright Information

Alliance for Telecommunications Industry Solutions
1200 G Street, NW, Suite 500
Washington, DC 20005

Contents

1. Introduction

ATIS' evolution to Content Optimized Networks (eCON) project is assessing the network implications of transitioning IP-based architectures to solutions that are based on named data (or content). From a practical perspective, the solutions outlined in this report do not assume a greenfield (or end-to-end replacement) approach to deploying networks based on Information-Centric Networking (ICN) principles, but rather an evolutionary path that may encompass existing or planned IP-based networks, including 5G. It is assumed that IP-centric architectures and ICN-based approaches could co-exist for a significant period of time, and that new capabilities such as network slicing could allow optimization of ICN in specific applications, while IP-based networking could meet the needs of many existing applications.

In February 2017, ATIS issued the *Evolution to Content Optimized Networks Report*[1], providing a broad assessment of ICN-based architectures, including market drivers, architectural approaches, technology choices, and standards development plans. In the *eCON Phase I Report*, it was recognized that a critical next step would be a focused deployment analysis that could assist network planners in making decisions regarding the value and viability of named content solutions. This *eCON Value Assessment Report* serves as a compendium document to the previous report and provides an in-depth analysis of key use cases, including a comparative analysis of IP and ICN-based solutions. Use cases have been selected that represent significant marketing opportunities in the near future. Evaluation criteria have been applied to each use case to more fully understand the network-based tradeoffs and potential benefits of ICN-based solutions with respect to existing best-in-class IP-based solutions.

The outcome of these assessments is presented in the form of Key Findings, as well as recommended Conclusions and Next Steps. It is hoped that network planners and architects can use the information in this report to make well-informed, tactical planning decisions to meet the growing demands for content optimized networks, today and well into the future. Further, this report may serve as a valuable implementation-level guide to related standards development activities addressing ICN requirements, protocols, and architectures.

---

[1] ATIS-I-0000055, *Evolution to Content Optimized Networks,* February 2017.

2. Reference Architecture and Evaluation Criteria for Use Cases

2.1 State-of-the-art Baseline Architecture

IP-based architectures are always evolving as both network infrastructure and application layer improvements are developed and introduced into the market. A prime example is how NFV/SDN-driven flexibility has unleashed sweeping implementation and network architecture transformations. Therefore, it is important to compare ICN-based solutions in context with a forward-looking, state-of-the-art, IP-based, baseline network reference architecture.

Figure 1 illustrates such a transformed network architecture, which flexibly centralizes access (e.g., cloud RAN) functions and distributes core functions, consolidating them into distributed edge clouds to deliver 5G-era latency-sensitive and bandwidth-intensive services. This proximity of the "IP edge" to end users is necessary to meet stringent 5G URLLC performance targets. It also enables more efficient and effective localized content distribution via local caches (with agreement or participation from content owners).

This functional distribution is important to consider when applying new content networking mechanisms such as ICN to the network and evaluating their benefits relative to the baseline. Indeed, the functional distribution and most of the functional elements shown in Figure 1 will be common to both a baseline IP network and one employing ICN mechanisms. Thus, ICN implementations can be defined relative to the baseline as adding, subtracting, and substituting particular functions.

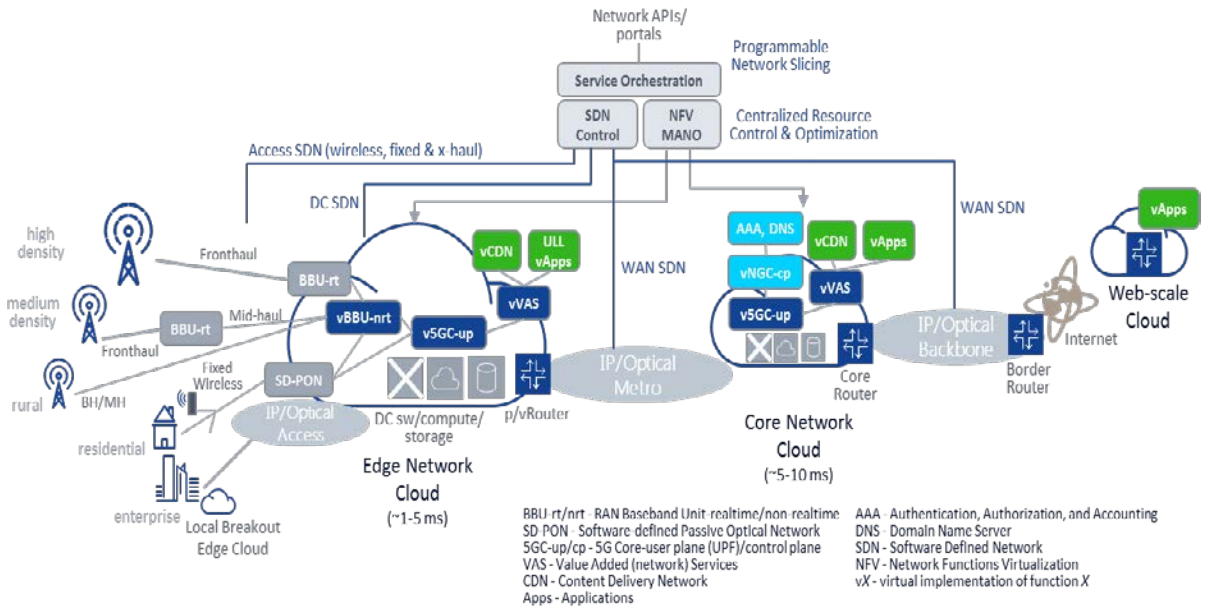


Figure 1: High-Level IP-Based Baseline Reference Architecture

The following list details further network assumptions for the IP-based baseline:

- NG/5G network, based on SDN/NFV.
- Core clouds are located at a major metro level.
- Edge clouds appear at multiple locations within a major metro/regional area.
- Cloud RAN widely adopted for new wireless deployments.
  - Mix of centralized, hubbed (split), and distributed configurations.
- Traffic classes tailored via network slicing.
  - Varied levels of user plane GW/anchor (core, edge, and selectively at far edge/enterprise).
  - Degree of mobility, control plane vs. user plane resources, etc.
- Content distribution pushed to the core and edge clouds (hosted and operator's).
- Low-latency applications placed at the edge cloud (hosted and operator's).
- IP/optical infrastructure centrally resource-controlled and optimized.
  - Multi-layer PCEs, segment routing.
- Routing at edge clouds is combination of Virtual Network Function (VNF) and Physical Network Function (PNF) (depending on scale).
- Potential for future enhancements such as interest-aware DNS, application of object security to content distribution.

The baseline defined in this section provides detail beyond what is utilized in the individual use case comparisons but serves as a reference for further network-level use case cost-benefit assessments, which are beyond the scope of this document.

2.2 ICN-based Comparison Architectures

In the future, an ICN-based architecture cannot be singularly viewed as a replacement strategy but will likely co-exist and, in many cases complement, IP-centric networks. The introduction of network slicing capabilities will support a range of ICN implementations, including ICN enabled at the end points and in edge or core routers. In this context, network slicing might create multiple logical networks on a common physical infrastructure. Each of these network segments (or slices) can support specific services with pre-designed performance characteristics. Network slicing can enable ICN packets to be delivered between ICN-enabled end points across a core IP network. Consequently, network slicing can significantly contribute to the co-existence of ICN- and IP-enabled network services as part of an evolutionary stage.

One impediment to ICN adoption has been that, until recently, most ICN proposals either suggested a native ICN approach, in which IP is completely replaced by ICN, or an overlay approach *ICN Over IP* (IoIP), where ICN becomes a sub-layer running over an existing transport layer protocol (e.g., over UDP). Both approaches have drawbacks in the sense that they require major changes to existing networks.

In *ICN within IP* (ICN-IP), routable ICN content names are mapped transparently into the IPv6[2] packet header, facilitating a more incremental and graceful adoption, and are more efficient than overlay solutions. One advantage of ICN-IP implementations compared to native ICN implementations is that ICN awareness and forwarding capabilities can be deployed only at key locations in the network and utilized for services where they do the most good (i.e., not in every forwarder), without affecting the forwarding layer. In this scenario, ICN packets can be routed and forwarded between locations using a traditional IP network.

Figure 2 illustrates the similarities and differences (denoted by color) between IP networking, ICN-IP, and native ICN.

---

[2] ICN-IP can be supported on either IPv6 or IPv4, but IPv6 is the preferred implementation.

IP Content Networking	ICN-IP	ICN
Names into IPv6 addresses	Names into IPv6 addresses	Variable length routable names
L4-7 request routing based on names (e.g. with SR)	L3 Name-based routing and hop2hop dynamic forwarding	L3 Name-based routing and hop2hop dynamic forwarding
Connection-based sender-driven transport	Partially symmetric routing	Symmetric routing
Tunnel-based security	Connectionless receiver-driven multipath transport	Connectionless receiver-driven multipath transport
Anchor-based mobility	Object-based security	Object-based security
Application layer caching	Anchorless mobility	Anchorless mobility
	In-path reactive caching	In-path reactive caching

Figure 2: Comparison of IP, ICN-IP, and ICN

One of the key terms highlighted in this report is *ICN multicasting*. While unicast and multicast techniques are well understood with respect to current network architectures, it is important to define the role of multicasting as it applies to ICN-based solutions. In a general data communications model, multicasting refers to a one-to-many or many-to-many forwarding of packets to a set of receivers identified by a group address. Today, in an IP infrastructure, one-to-many multicasting refers to applications such as video streaming, multipoint collaboration, and IPTV networks.

In an ICN context, multicasting is accomplished via two mechanisms. The first is a consequence of caching of data packets for some configured period of time in the ICN forwarder's content store (or packet buffer). This approach allows subsequent requests for the same content object to be served directly, without any communication with the original content server. The second mechanism, interest aggregation, is the cataloging of duplicate pending interests in the ICN forwarder's pending interest table (PIT) so that only one interest packet for a particular content object is forwarded toward the content server. When the requested data packet is received in response to that interest, it is returned to all the requestors represented by the interests in the PIT. These two mechanisms can greatly reduce the traffic between the downstream ICN forwarder and the content server, also making it possible to serve multiple users via a single transaction with the content server.

2.3 Comparative Criteria

All use cases were evaluated across the following set of core criteria:

- Latency (start-up time, steady-state transmission delay, handover duration).
- Efficient utilization of network resources (radio, backhaul, mid-haul, backbone, etc.).
- Complexity (architectural and operational).
- Impact on applications and user devices.

Additionally, the following list of criteria and functions was considered and applied where appropriate, in the context of each use case and application:

- Per endpoint throughput.
- BW efficiency in RAN/X-haul.
- BW efficiency in core transport.
- QoS (SLAs).
- QoE (optimization).
- Reliability.
- Content security.
- User privacy.
- Network security.
- Mobility.
- Load-balancing.
- Access aggregation.
- Multi-path support.
- Dynamic path discovery.
- Multi-server support.
- Dynamic server discovery.
- Device/service scale.
- Device cost/power.
- Interoperability.
- Network operational complexity.
- Network evolution complexity.
- Ecosystem operational complexity.
- Ecosystem evolution complexity.

Opportunities and challenges related to several of the criteria above, while not assessed in the individual use cases, are addressed generally in the *eCON Phase 1 Report*, *Evolution to Content Optimized Networks*. This includes areas such as resolution/alignment of naming structures, privacy/trust models, content ecosystem evolution to object security, HTTP-based applications, and standards development. Read both reports for a more complete view.

For both ICN and IP state-of-the-art comparison architectures, the 3GPP R15 architecture was adopted as the skeleton upon which an ICN and state-of-the-art non-ICN solution would be implemented and compared. Figure 3 illustrates this architecture.

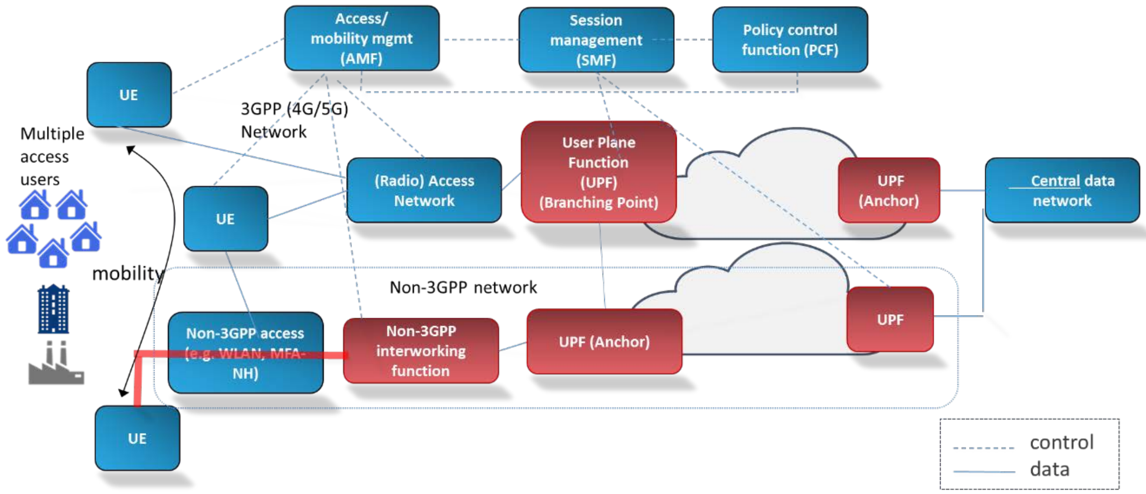


Figure 3: Functional View of Multi-Access Network Based on 3GPP Release 15 Next Gen Core (NGC) Architecture

3. Use Case Assessments

The first step in assessing the value of ICN-based solutions with respect to state-of-the-art IP-centric solutions is to target the use cases that have high market-relevance and will have profound impact on network design and performance. On this basis, an assessment of the business and technology landscape identified three significant use case classes that acknowledge the importance of video distribution, augmented reality, and real-time IoT applications such as Smart Mobility.

An in-depth assessment of each use case is presented in the following sections.

3.1 Linear Video Delivery

3.1.1 Introduction

In recent years, video consumption has shifted away from watching TV over multichannel subscription-based services and toward video delivery over broadband IP connections (wireline, wireless, fixed, mobile). As Figure 4 shows, video already consumes the largest share of network bandwidth today, and that share is expected to continue to grow.

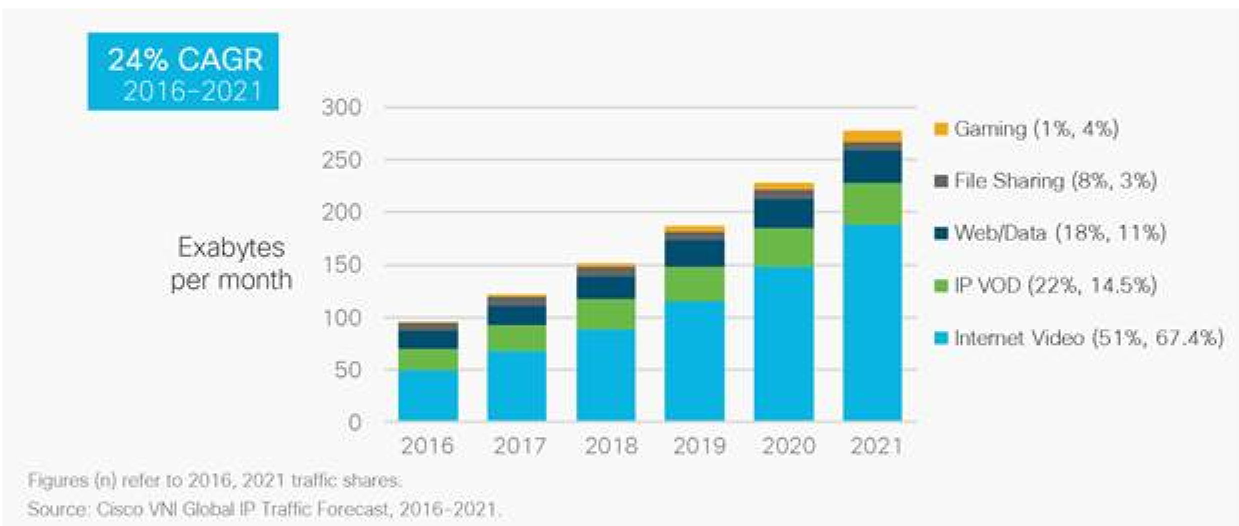


Figure 4: Past and Forecast Network Traffic by Traffic Type

Five factors are driving video traffic growth:

- An explosion in the amount of available content.
- Ever increasing video resolution (going from 720p to 8K over the span of a few years).
- The need to support multiple encoding bit rates simultaneously.
- Continued growth in the number of users and devices.
- A fundamental shift in content production and consumption behavior.

In the traditional video distribution model, operators provided a fixed number of channels, whose content could be pre-provisioned (or driven by a live feed tied to a particular channel) and pushed simultaneously to multiple subscriber end-points at a scheduled time. This model is being replaced by one where the majority of linear (including live) video content is more dynamic, created on an ad-hoc basis, and accessed via pull by end users.

In the past, consumers of video services were essentially end-points (e.g., a household set-top box). Their location was fixed, and the path to the end-point was pre-determined. That model is rapidly being replaced by one where the consumer is an end user, mobile, and accessing video content via many different devices, locations, and access mechanisms.

These changes have driven an evolution in architectures and mechanisms to better handle the new challenges presented. For networks, as described in Section 2, the direction of evolution is toward greater and more dynamic distribution of (virtualized) network functions and compute/caching capabilities closer to the end user. These architectural changes proceed in parallel with higher-capacity and lower-cost transport.

At the same time, new content delivery network (CDN) mechanisms have evolved to address the problems associated with linear video delivery over existing and proposed networks. These mechanisms support the new consumption models while minimizing cost and maximizing the user experience given the constraints imposed by the networks.

The following sections describe the issues surrounding linear video distribution, the unique problems faced by CDNs, and how these problems are being addressed. The report explores how incorporating ICN principles in the transport layer of these networks could work together with evolving networks and CDNs to improve performance and user experience while lowering costs.

3.1.2 Description

This section discusses linear video delivery over multi-access networks, including 4G, 5G, wireline, Wi-Fi, and other access technologies.

The desired features of a next-generation video distribution system will:

- Support seamless mobility among multiple access mechanisms (e.g., wireline and wireless, 4G/5G, and Wi-Fi) to:
    - Minimize handover latency between technologies.
    - Support minimally interrupted content delivery to the end user.
    - Provide the highest quality video to the end user through potentially rapid mobility events.
- Take advantage of all access paths to the user simultaneously, ideally balancing the load among them dynamically (determined by policy) to deliver the highest available bandwidth to the end user while using network resources most efficiently.
- Correct transmission errors/losses locally as much as possible, thereby reducing overall traffic and server loading and improving the end-user experience.
- Make the most efficient use of transport within the network, dynamically balancing loading over available routing/forwarding paths.
- Deliver the highest possible video quality to the end user, within the constraints imposed by network bandwidth and policy.
- Minimize start-up latency.
- Minimize signaling and control overhead.
- Manage potentially large, dynamic, and non-stationary loading scenarios.
- Require the lowest cost to deploy and operate.
- Scale gracefully to accommodate projected future growth in video traffic.
- Provide security even when content (and the path from the user to the content) is located in different points within the network and may change dynamically.

3.1.3 CDNs in the Multi-Access Network

Networks, protocols, and CDN overlays have evolved to cope with the challenge of ever-increasing traffic and ever-changing consumption models. This section examines some of the challenges and solutions that have been proposed/implemented beginning by presenting a schematic, hierarchical CDN (Figure 5), and using it to outline some of the areas for optimization.

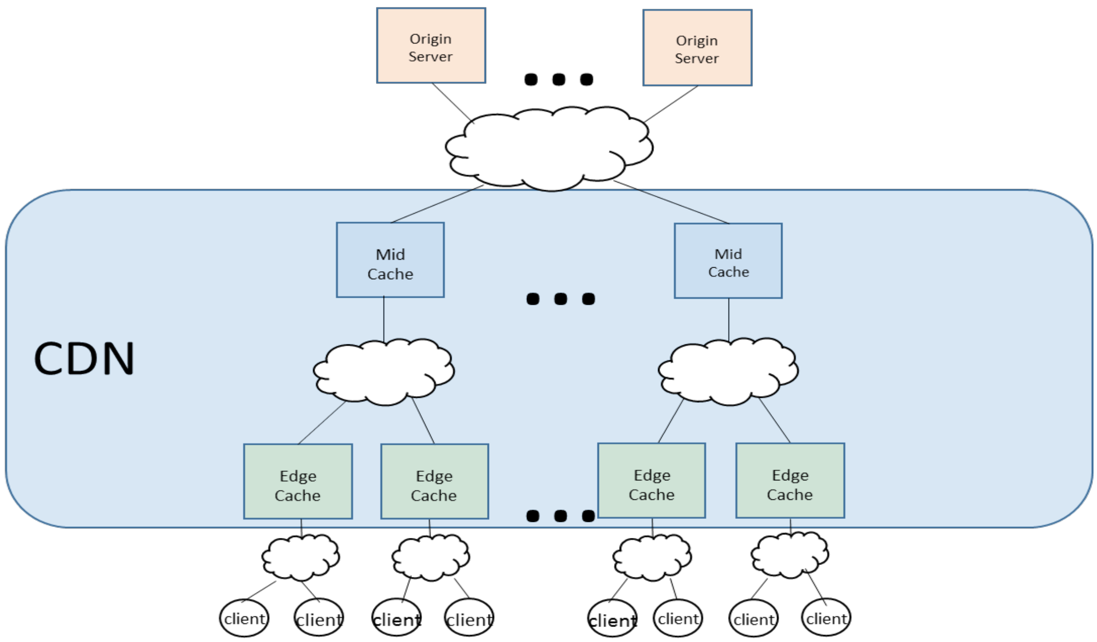


Figure 5: Schematic Representation of a Hierarchical CDN

The tiered CDN approach reduces the load on the origin server and reduces latency and transport costs by migrating content closer to the client endpoint. For linear video distribution of multi-bit rate content, the origin server will encode and store content (potentially live material) in multiple bit rate formats. To access the content, the client will make an HTTP request for a video chunk, which might comprise one or more IP packets containing video encoded at a particular bit rate. The request is routed from the client to the serving edge cache according to request assignment rules adopted in the particular implementation.

If the requested content chunk is present in the edge cache, it is returned to the client. If the cache does not contain the chunk (i.e., a cache miss occurs), the request is forwarded to the next higher tier of caches in the CDN. If there is a cache miss at the highest tier, the request is forwarded to the origin server. Every cache in the CDN serves as an HTTP proxy to handle the HTTP requests from clients and to communicate with the next higher tier.

The CDN will employ a set of rules to determine how to populate the various caches in each tier. The cache may or may not store each new chunk received as the result of a cache miss (i.e., the caching strategy may be proactive, reactive, or some combination of the two). The rules typically operate and are communicated out of band in a signaling/control layer.

Figure 5 shows a two tier CDN, but in theory, any number of tiers could be supported, with edge caches located arbitrarily close to the end user and mid-caches distributed at various locations in the network. As the number of layers and cache instances increases, performance may be enhanced, but cost is also increased.

Figure 6 shows a schematic representation of a dual-technology multi-access network, based on the 3GPP Release 15 Next-Generation Core (NGC) architecture. This diagram assumes control-user plane separation (CUPS), such as when the serving data gateway function is separated from the control gateway function and may be distributed to the optimal location in the network. The diagram also assumes distributed compute/caching capability at the edge of the network, as in multi-access edge computing (MEC). The Branching UPF and an instance of a CDN caching function could be located within an operator's edge cloud, as shown in the reference diagram of Section 2.

A user device may be connected to either the Wi-Fi/non-3GPP access, the 4G/5G/3GPP access, or both simultaneously. The user device may also move freely between the two networks while consuming the same live/linear content. An added complication is that each wireless link's characteristics are highly non-stationary, making support for video rate adaptation (i.e., multi-rate) a necessity to provide a high-quality user experience and to manage load. Also, because of this non-stationary of link characteristics, the "best path" to the user, as well as the best edge cache, may vary dynamically.

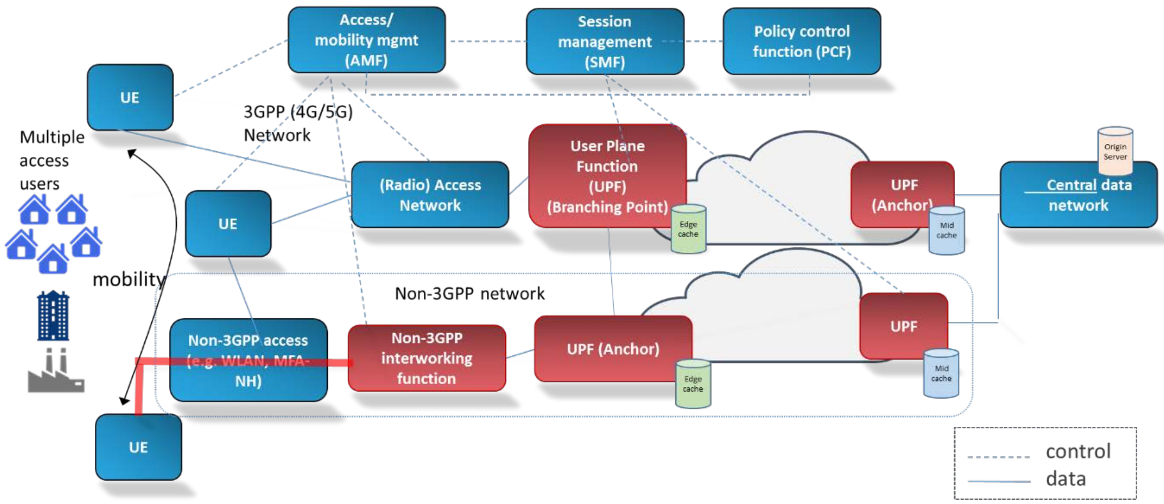


Figure 6: Functional Representation of a Multi-Access Network Comprising a 4G/5G/3GPP Access Network and a Wi-Fi/Non-3GPP Access Network, with a Multi-Tier CDN Overlay, but Without ICN

In this example, the network is controlled via 3GPP Release 15 NGC mechanisms, with the control exercised by the 3GPP network. It is important to note that this is not (and will not be) the only case. In many cases, the 3GPP and non-3GPP networks will be operated and controlled by different service providers with completely separate and distinct control planes.

3.1.4 CDN Incorporating ICN

Before beginning a discussion of the challenges of linear/live video distribution, it is useful to first address how the multi-access network shown in Figure 6 could be enhanced by incorporating ICN in the forwarding layer and potentially in the CDN elements themselves.

As discussed in Section 2.2, *ICN within IP* (ICN-IP) transparently maps routable ICN content names into the IPv6[3] packet header. This approach facilitates more incremental and graceful adoption. One advantage of ICN-IP implementations, compared to overlay or native ICN implementations, is that ICN awareness and forwarding capabilities can be deployed at key locations in the network where they do the most good (i.e., not in every

[3] ICN-IP can be supported on either IPv6 or IPv4, but IPv6 is the preferred implementation.

forwarder), and ICN packets can be routed and forwarded between these locations using a traditional IP network.

Figure 7 shows the network of Figure 6, enhanced with ICN-IP capabilities at possible strategic locations.

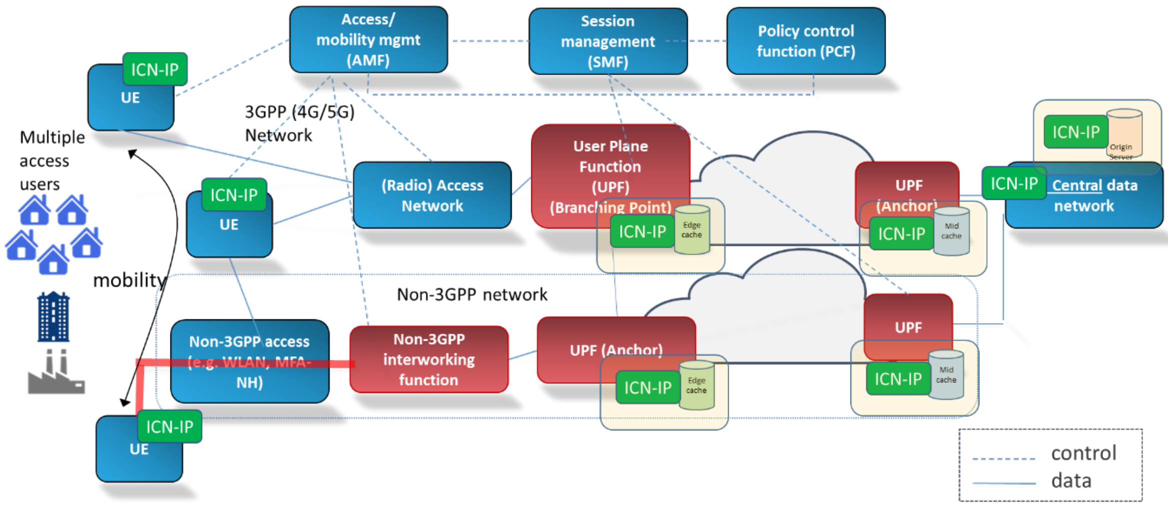


Figure 7: Functional Diagram of Multi-Access Network with ICN-IP at Potential Key Locations, and a CDN Overlay

In Figure 7, ICN capabilities are shown at several potential locations:

- **Client devices:** An ICN-IP stack is assumed at the client device. This enables the device to support ICN capabilities without requiring any native ICN support on the part of particular applications. It also enables multi-access load balancing, access aggregation, and seamless inter-access mobility. An ICN-IP stack can be implemented in the device user space, without requiring recompiling of the OS kernel for Android or iOS devices and can be downloaded to existing devices.
- **Edge Node:** An ICN-IP forwarder can be located at the very edge of the network, potentially in the operator's edge cloud network or at the access node. This enables very rapid local packet loss detection and retransmission. In the diagram, the edge node in the 4G/5G network is shown as also serving as the branching point between 3GPP and non-3GPP networks. This enables the most rapid inter-access mobility, access aggregation, and load balancing.
- **Core Router and/or Border Router:** ICN-IP forwarding capabilities within, and at the borders of, the operator's network enable dynamic load balancing, local loss recovery within the PDN, and inter-network mobility.

- **Origin Server, CDN Caches:** Ideally, the origin server and the CDN caches natively support ICN-IP and understand the naming conventions for the video application. This support is not mandatory. For example, in the case where the origin server for the application does not support ICN-IP, the CDN cache at the PDN border could also serve as a Transmission Control Protocol (TCP)/ICN-IP proxy.

  In the diagram, a CDN cache instance is shown at every location where an ICN-IP forwarder would also be placed. This may not be desirable in all cases. For each case, it will be important to evaluate the trade-off between engineering a larger content store in the ICN-IP forwarder at a particular location versus offloading content storage to a dedicated CDN cache. In practice, the ICN-IP forwarder would incorporate at least as much buffering capability as would be needed to enable ad-hoc, asynchronous multicast, and local error recovery.

3.1.5 CDN Challenges and Solutions

Some broad categories of challenges presented by linear/live video distribution, in terms of the objectives outlined in Section 3.1.2, are:

- Reducing latency, reducing network traffic, and server loading. These capabilities avoid race conditions.
- Increasing reliability (loss recovery), particularly over wireless and highly non-stationary paths.
- Using all access paths optimally; load balancing, access aggregation, seamless mobility (including inter-access), and multisource. These capabilities make the best use of multiple networks and servers simultaneously in a dynamic, mobile, heterogeneous environment.

The ultimate yardsticks for measuring any solution that addresses these challenges are the resulting end-user video experience and the costs associated with deploying and operating the network. The following discussion will present today’s state-of-the-art responses to these challenges and will offer a qualitative comparison of them versus a CDN architecture that incorporates ICN. A more quantitative comparison on individual performance metrics, based on laboratory experiments conducted by Cisco, appears in Section 3.1.6. References to other works are also included below.

3.1.5.1 Reducing Latency, Overall Traffic and Server Loading

The tiered CDN architecture was initially developed to help deal with the issues of latency, reducing overall traffic, and reducing server loading by migrating content closer to the network edge. The content in a cache may be proactively provisioned for pre-programmed video, but for live video content, this is not generally possible if the content is to be delivered without significant latency. Consequently, a reactive cache management strategy is employed.

As described in Section 3.1.3, a user is assigned to an edge cache based on operator/implementation rules. A request for a content chunk is made via HTTP to the edge cache. If the edge cache does not include the content, the request is forwarded up to the next tier. If there are cache misses at every tier in the CDN, the request is forwarded to the origin server. For a very popular live video (i.e., one being consumed by many nearly simultaneous users), many requests for the same new content chunk may be received by the proxy/cache at close to the same time, resulting in many cache-misses. This results in the thundering herd problem, in which the origin server, or higher tier CDN caches, are swamped with requests for the same content chunk at nearly the same time. This can have a severe impact on traffic, on server loading, and on the end-user experience (because of increased latency).

Modern live video delivery systems[4,5] use a request-coalescing mechanism to mitigate the thundering herd problem, whereby multiple requests for the same content (not currently present in the cache) are held by the cache proxy, and only one request is forwarded up to the next tier. While these solutions mitigate the problem in terms of server loading and controlling the volume of upstream traffic, the overhead of keeping the session state of many HTTP sessions at the proxy grows larger with increasing numbers of consumers.

ICN provides a similar mechanism, whereby pending interests for the same object (not currently present in the content store or forwarder buffer memory) are aggregated in the forwarder's Pending Interest Table (PIT), and only the first interest received is forwarded up to the next tier. The advantage of the ICN mechanism is that because there is no

[4] Facebook live streaming: http://highscalability.com/blog/2016/6/27/how-facebook-live-streams-to-800000-simultaneous-viewers.html

[5] PBS proxy cache: https://www.nginx.com/blog/mitigating-thundering-herd-problem-pbs-nginx/

concept of a “session”, there is no need to maintain session state for each user. Where HTTP/TCP-based approaches scale as a function of the number of users (sessions), ICN approaches scale as a function of the number of content objects that are relevant at a given time. For applications with large numbers of users accessing very popular content, the ICN approach has decided advantages over an HTTP/TCP approach. Moreover, because the ICN interest aggregation mechanism operates at the transport layer, it could be argued that it results in lower latency than HTTP-level request coalescence given the same computational power.

Existing mechanisms to reduce latency typically make use of RTMP[6]. RTMP works well in the absence of losses, but because it is based on TCP, it introduces delays in the presence of packet loss. Other strategies for mitigating perceived latency are based on cache-priming or server push, in which content is requested by (or pushed to) the lower tier cache before it receives a user request, potentially in response to an out-of-band control mechanism. These push strategies are not unique to HTTP-based CDNs and could be implemented in ICN, as well. However, they (regardless of the protocol) are not compatible with client-driven rate adaptation, which is a required capability in highly non-stationary channels.

3.1.5.2 Increasing Reliability (Loss Recovery)

Packet losses in a network such as the one shown in Figure 7 typically occur for two reasons:

1. The wireless channel is unreliable and non-stationary. Packets can be lost because of RF conditions (e.g., shadowing, fading), as well as by interference.
2. There are mobility implications, particularly in a small cell/multi-access network environment. In this case, a user may request (or may be expecting) some content object but may have moved before it could be delivered and can no longer be reached via the old path.

The previously discussed approaches based on TCP/HTTP (e.g., RTMP) can reduce latency, but they do so at the expense of reliability: They perform poorly in the case of losses and highly dynamic channels. Other approaches have been proposed to make

---

[6] http://wwwimages.adobe.com/content/dam/Adobe/en/devnet/rtmp/pdf/rtmp_specification_1.0.pdf

HTTP sessions more robust. An example is ZiXi[7], which achieves reliability by using forward error correction (FEC). The FEC makes packet loss much less likely to occur, but it does so at the cost of about 30-percent increase in packet size, thereby increasing traffic volume. When a packet loss does occur, it falls back on HTTP/TCP mechanisms for recovery, incurring the associated delays. This reliance on TCP for loss handling is a drawback of any HTTP/TCP-based mechanism.

ICN forwarders deal with losses by serving missed packets from copies retained in the buffer of the nearest upstream forwarder, such as at the network edge if the loss occurs over the wireless link. Mechanisms have been developed to provide sub-RTT error/loss recovery for ICN-IP in mobile and wireless networks. Examples are *Wireless Loss Detection and Recovery (WLDR)* and *Mobility Loss Detection and Recovery (MLDR)*[8]. These mechanisms can be supported on an ICN-IP implementation. Together, they provide very rapid loss recovery in mobile networks operating at L3-L4.

3.1.5.3 Support for Access Aggregation, Multipath, Dynamic Load Balancing, Multi-Server, and User Mobility.

At any given time, many user devices can access more than one network (e.g., Wi-Fi and 4G/5G). However, the link quality on each visible network can vary greatly (and rapidly) over time as a result of loading, interference, mobility, etc. It is desirable to take advantage of all the paths (access links as well as within/between PDNs) for delivering content to the end user, resulting in the best experience and the most efficient utilization of network resources.

Besides the case of a single server having multiple paths to a user, it is often the case that multiple servers (or intermediate content repositories) could be utilized to serve content to that user if a means existed for doing so. This is particularly relevant if the end-user device is connected via two different networks, where the quality of the connection on each may vary over time.

At any point in time in a mobile environment, the optimum path(s) from a server to a client endpoint, or the best server based on the client's current location and network

[7] http://www.zixi.com/PDFs/Adaptive-Bit-Rate-Streaming-and-Final.aspx

[8] G.Carofiglio, L. Muscariello, M.Papalini, N. Rozhnova, X. Zeng, "Leveraging ICN In-network Control for Loss Detection and Recovery in Wireless Mobile networks," *in Proc. of ACM Sigcomm ICN 2016.*

conditions, may not always be known, *a priori*. Dynamic discovery of paths and servers is an important capability for live video delivery in the multi-access mobile network of Figure 7.

Multipath TCP (MPTCP) is a mechanism that can be used to take advantage of multiple known paths between a single server and a client device. This mechanism provides rapid fail-over in the event that an active link fails or to aggregate bandwidth among two or more known paths. When used with HTTP for video delivery (e.g., DASH), MPTCP will attempt to select the best link for transmitting each segment via a decision based on an estimate of instantaneous path quality. It has been shown that although this is usually beneficial, the coarse granularity of a segment-based decision approach can actually degrade performance relative to Single-Path TCP (SPTCP) when applied to highly dynamic, lossy channels[9]. Because MPTCP does not provide any intrinsic security, it is usually run over HTTPS, which adds some additional processor load.

Multipath Quick UDP Internet Connections (MPQUIC) is another protocol for bandwidth aggregation and mobility support over multiple known paths between a single server and a client device[10]. Rather than extending the behavior of TCP, QUIC (and its multipath extension, MPQUIC) runs over UDP and provides its own mechanisms to ensure end-to-end delivery that avoids some of TCP's shortcomings in terms of latency. MPQUIC has been shown to achieve higher file transfer rates than TCP[11]. QUIC intrinsically provides security by encrypting all content and headers, specifically to prevent observation or interference by middle boxes; hence no additional security is required. However, MPQUIC may degrade the performance of HTTP video delivery (i.e., DASH).[12]

ICN and ICN-IP support bandwidth aggregation and load balancing both from an end-to-end point of view and from hop to hop. These capabilities make instantaneous routing decisions for interest packets on a per-packet basis based on estimates of

---

[9] C. James, E. Halepovic, M.Wang, R. Jana, and N. K. Shankaranarayanan. 2016. "Is Multipath TCP (MPTCP) Beneficial for Video Streaming over DASH?" In 2016 IEEE 24th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS). 331–336. https://doi.org/10.1109/MASCOTS.2016.75

[10] Quentin De Conink, Olivier Bonavanture, "Multipath QUIC: Design and Evaluation", https://multipath-quic.org/conext17-deconinck.pdf

[11] Ibid.

[12] Divyashri Bhat, Amr Rizk, Michael Zink, "Not So QUIC: A Performance Study of DASH over QUIC," Proceedings of the 27th Workshop on Network and Operating Systems Support for Digital Audio and Video, pages 13-18

instantaneous link quality. Experiments presented in Annex A demonstrate that a higher aggregated rate (and thereby higher user-perceived video quality) is achieved by making per-packet path decisions relative to the per-segment path decisions made by MPTCP. ICN-IP performs better than both MPTCP and MPQUIC in highly non-stationary channels.

Unlike MPTCP and MPQUIC, paths do not have to be known by ICN *a priori* because ICN supports dynamic path discovery. Also, unlike MPTCP and MPQUIC, ICN supports multi-server delivery and aggregation even when the multiple servers are on different networks. Because ICN secures the individual content packets themselves, rather than the path, there is no need to establish a secure path (e.g., HTTPS, QUIC).

In terms of supporting user mobility (inter- and intra-access), ICN has the advantage over both MPTCP and MPQUIC by virtue of packet-by-packet path selection, sub-RTT error and loss recovery, and dynamic path discovery. Experimental results in Annex A demonstrate that ICN-IP supports high-quality video delivery even for very high-frequency, inter-access mobility events.

3.1.6 Linear Video Delivery Summary and Discussion

The emergence of ICN within IP solutions has made it unnecessary for a network operator to choose between ICN-based and IP-based solutions and has greatly mitigated the sudden migration costs that have been a barrier to ICN adoption in IP networks. The questions that remain are whether adopting ICN-IP in an evolutionary way would improve performance relative to non-ICN solutions at the same cost, and whether performance improvement is necessary or desirable. The latter question can be answered only by prospective providers of such a service, but the former question was addressed in previous sections of this document.

Exploration has occurred on desired characteristics of a linear video distribution service that operates over a distributed network and that delivers the highest quality video experience to users who are mobile and who access the network via multiple access paths and technologies, sometimes simultaneously.

The previous sections have distinguished between two classes of ICN benefits. The first class of benefits accrues from the increased distribution of caching capabilities and tailored cache management strategies. This class of benefits includes reduced traffic volume, reduced server loading, and the ability to avoid CDN race conditions such as the thundering herd problem. The second class of benefits is not dependent on caching in the network, but is intrinsic to ICN, and distinct from other state-of-the-art mechanisms for delivering content in a multi-access, dynamic network. This class of benefits includes simultaneous multi-path support, multi-access mobility support, local error recovery, and support for multiple simultaneous servers.

In designing a CDN for linear video distribution, the architect must trade off the costs and performance impacts of providing additional caching capabilities in the forwarding layer (i.e., additional memory) at a particular forwarding node against the placement of a L4/L7 application-layer cache at a strategic (perhaps the same) location. This decision will generally be driven by the amount of content that needs to be stored in a location, and the optimal overall solution will comprise a combination of distributed caches/proxies and a delivery mechanism that is suited to supporting the objectives stated in Section 3.1.2. Annex A provides experimental results that can be used to infer engineering rules for the amount of ICN-IP in-forwarder buffering required by an application to obtain optimal performance. For live video, for example, a small amount of buffer memory in an ICN-IP forwarder can be used to provide all the benefits listed in Section 3.1.2.

Setting aside the benefits of in-forwarder caching and assuming only a small amount of in-forwarder buffering capability, ICN outperforms other state-of-the-art delivery mechanisms with respect to reliability (loss recovery) over wireless and highly non-stationary paths. It also outperforms other methods with respect to using all access paths optimally: load balancing, access aggregation, seamless mobility (including inter-access), dynamic path discovery, and support for multiple sources (even when the sources are on different networks).

Deployment of a hybrid solution combining distributed L4/L7 caches and ICN-IP can be approached incrementally. Support for ICN-IP at the client device is mandatory but can be delivered as a download because it does not require any modification to the OS kernel. ICN-IP forwarding can be implemented in software within existing routers and also can be implemented in a MEC node in the operator’s edge cloud. With respect to cache/proxies, the designer can choose to natively support ICN-IP at the cache or can make use of a separate proxy to mediate between the cache/server and the ICN-IP forwarding layer.

3.2 AR/VR Live Event

3.2.1 Introduction

Augmented Reality (AR) and Virtual Reality (VR) are becoming increasingly common. Platforms such as Facebook and YouTube have deployed support for some immersive videos, including 360-degree versions. Facebook, Google, and Microsoft are among the many companies offering devices to view VR, ranging from simple mechanical additions to smartphones to full-fledged, dedicated devices. While the first commercial AR deployments were inauspicious, this new wave of products promises an enhanced user experience.

Figure 8 and 9 quantify the growth in AR and VR traffic, which is driven by increasing end-user adoption, as well as by the higher rate of traffic generated by such applications. A 360-video stream may require as much as six times the bandwidth of a typical video stream at the same resolution.

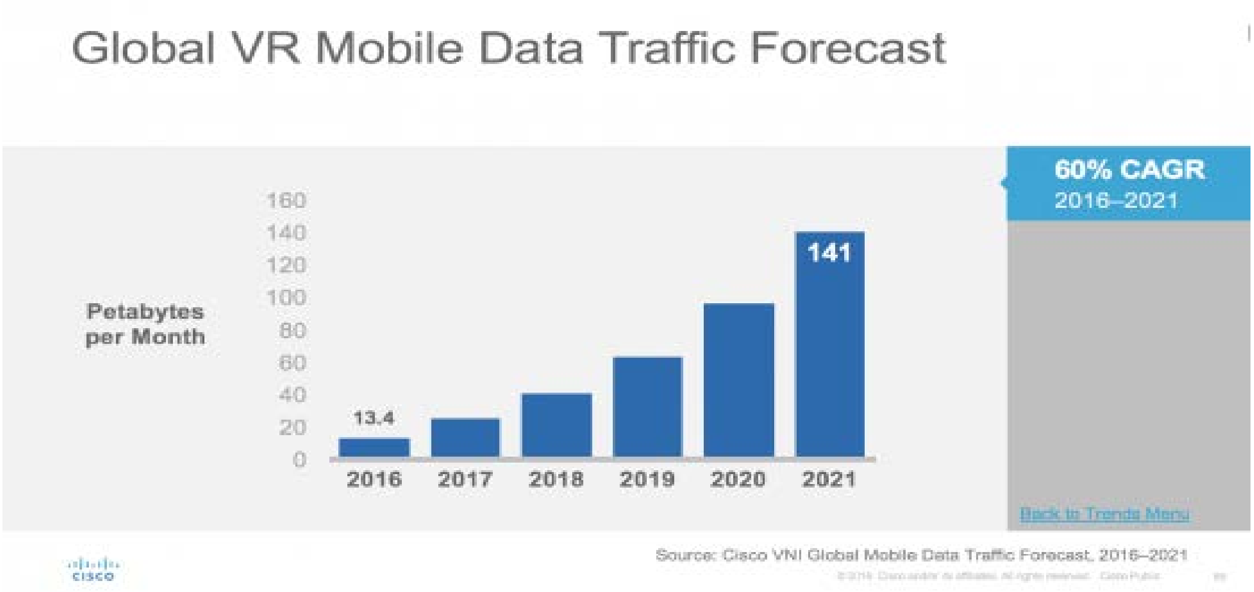


Figure 8: Global VR Mobile Data Traffic Forecast

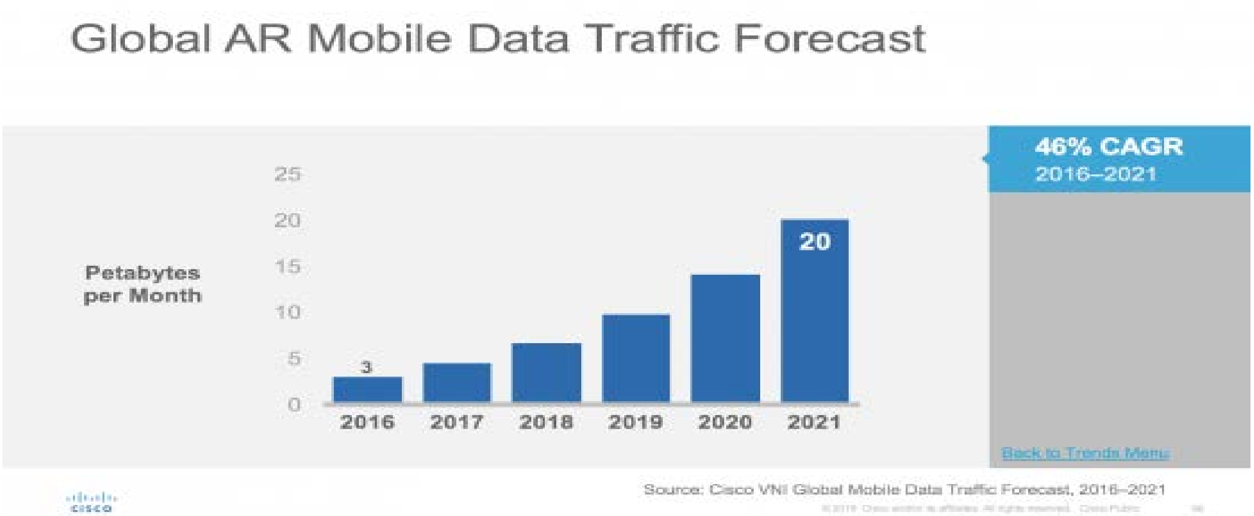


Figure 9: Global AR Mobile Data Traffic Forecast

The model for AR/VR content distribution content is still unclear. However, most of the content is expected to be dynamic and accessed by pull from end users. As in the previous use case, the content cannot be pre-provisioned over a known channel.

Currently, most of this content is delivered to a fixed endpoint, such as a desktop PC with a connected AR/VR display. The content is first downloaded locally onto the PC, and then the user can interact with the content using the display. This model will soon shift to one where the display itself is the end-user device accessing content over the mobile network, through different access technologies, and in different locations.

The evolution in architectures and mechanisms described in Section 2 and in the previous use case are still valid here. There is an evolution toward greater, dynamic distribution of (virtualized) network functions and compute/caching capabilities closer to the end user.

As in the linear video distribution use case, the new CDN mechanisms that have evolved to improve content distribution in general. Video distribution in particular is expected to be leveraged to distribute AR/VR applications.

The following sections describe the issues surrounding an AR/VR application, including the shortcomings of the existing solutions. It explores how incorporating ICN principles in the transport layer of these networks could enhance evolving networks and CDNs to improve performance and user experience while lowering costs.

3.2.2 Description

The specific use case to be considered is *AR/VR live event distribution* over heterogeneous network technologies. These can be a combination of mobile/4G/5G, wireline, Wi-Fi, and other access technologies.

The goal of this use case is to:

1. Transmit a live event to remote viewers so they can be immersed in the event experience. This corresponds to an immersive video streaming/VR experience.
2. In a later phase, layer this view with an AR layer to enhance the user experience.

In the case of a live sports event, the AR layer may superimpose statistics or the box score of each performer in the event. "Live" means viewers must be able to see the event within small delay bounds. However, the event is not necessarily interactive, meaning that the delay can be around a few seconds.

The event is distributed to an audience that can be relatively large, such as millions in the case of a typical sports event. The content distribution must be scaled up to at least this order of magnitude.

The audience view must be customized to each user. For example, suppose that some audience members are viewing the event with a VR headset. Customization means their view must be react to their movement and adapt to their field of view.

The desired features of a system to distribute a live event in AR/VR format are similar to those named in the previous use case. It should:

- Support seamless mobility among multiple access mechanisms (e.g., wireline, 4G/5G, and Wi-Fi):
  - Minimize handover latency between technologies.
  - Support minimally-interrupted content delivery to the end user.
- Take advantage of all access paths to the user simultaneously, ideally balancing the load among them dynamically (determined by policy) to deliver the highest available bandwidth to the end user while using network resources most efficiently.
- Correct transmission errors/losses locally to the extent possible, thereby reducing overall traffic and server loading, and improving the end user experience.
- Make the most efficient use of transport within the network by dynamically balancing loading over available routing/forwarding paths.
- Deliver the highest possible video quality to the end user within the constraints imposed by network bandwidth and policy.
- Minimize start-up latency.
- Minimize signaling and control overhead.
- Manage potentially large, dynamic, and non-stationary loading scenarios.
- Require the lowest cost to deploy and operate.
- Scale gracefully to accommodate projected future growth in video traffic.
- Provide security even when content is located in different and multiple locations within the network.
- Allow different layers to be delivered in a synchronous manner to the users.
- Allow potentially large number of users to participate concurrently in the AR/VR experience.
- Support a control layer that can manage the sessions of these users and manage session start-up for users joining the events, users' mobility while watching the event, and session termination when users leave the event.

3.2.3 Delivery in the Multi-Access Network

It is expected that the delivery of AR/VR applications will leverage current content delivery technology, such as the CDNs described in detail in Section 3.1.3.

Two key differences are:

- Multiple servers and databases will provide the content in a quasi-synchronous manner. The content is a composition of a 360-video stream for the live event with superimposed metadata requested by a specific user. The AR layer may be fetched from a location other than the live event feed.
- Content that is distributed should match the user's field of view, and content outside of this field of view may either not be distributed or be distributed at a lower resolution.

3.2.4 Architecture Incorporating ICN

As in the previous use case, an ICN enhancement or overlay can be integrated within the current IP architecture. Namely, ICN does not have to be present at every node in the network, but only at a subset of nodes that support ICN in order to deliver network-wide ICN characteristics. Figures 5 and 6 still apply for this use case. The dual-mode forwarding in ICN[13] allows ICN and IP to co-exist.

3.2.5 AR/VR Challenges and Solutions

Some broad categories of challenges presented by AR/VR live event distribution are similar to the linear video distribution use case:

- Reducing latency, network traffic, and server loading.
- Increasing reliability (loss recovery), particularly over wireless and highly non-stationary paths.

---

13 "Supporting dual-mode forwarding in content-centric network," Ravishankar Ravindran; Guoqiang Wang; Xinwen Zhang; Asit Chakraborti, IEEE ANTS 2012.

- Using all access paths optimally, such as via load balancing, access aggregation, seamless mobility (including inter-access), and multisource, including multiple data layers to be rendered on the display.
- Improving end-user QoE.

These are discussed qualitatively below.

3.2.5.1 Reducing Latency, Overall Traffic, and Server Loading

AR/VR is extremely latency sensitive, especially when consumed on a head-mounted display. The application is highly bandwidth intensive because it requires a 360-degree view.

Modern 360-degree video distribution takes advantage of different encoding options for the different fields of view, providing a higher resolution to the one in front of the user, and lower resolutions to the rest. Some techniques separate the fields of view into different tiles, which are combined to create the display.

The request aggregation mechanism described in the previous use case still applies. It provides the mechanisms to filter interests and forward only one interest to the next tier when several concurrent requests are issued.

In AR/VR, ICN allows the naming of the different tiles to be exposed at the network layer. Therefore, the distribution of the tiles can be shared between multiple users accessing an overlapping field of view. Tiles can be cached within the network at the ICN nodes and made available when a new request comes.

3.2.5.2 Increasing Reliability (Loss Recovery)

The packet loss discussion of Section 3.1.5.2 applies as well here. Copies of the data cached near the users can be leveraged in response to a packet loss.

3.2.5.3 Support for Access aggregation, Multipath, Dynamic Load Balancing, Multi-Server, and User Mobility.

The support for access between heterogeneous technologies applies as in the previous use case. The view on the display is a composition of data from multiple sources, including a 360-video stream of the live event and AR layered over this stream. As a result, the support for multisource distribution is even more relevant in this use case.

3.2.6 AR/VR Live Event Summary and Discussion

The AR/VR live event use case shares many similarities with the linear video use case, because it includes a video stream at the core of its user experience.

As in the previous use case, it is important to note that ICN can be deployed within an IP network and does not require a wholesale replacement of the underlying infrastructure. Several solutions exist to ensure that ICN can coexist within IP.

Similarly, the desired characteristics of a linear video distribution overlap with those of an AR/VR live event distribution. Both systems require a service that operates over a distributed network and that delivers the highest quality video experience to users who are mobile, and who access the network via multiple access paths and technologies, sometimes simultaneously.

The AR/VR use case highlights another benefit of ICN: the naming features that allow different field-of-view tiles to be identified at the network layer. While each user in such an application may have its own customized view, it will be composed of different tiles. These tiles can be shared between users, and ICN provides the mechanisms to support the caching and the sharing of these tiles transparently.

The benefits in reliability (loss recovery), multipath transport, multisource delivery, access aggregation, and load balancing still apply. Multisource delivery is a stronger requirement in an AR/VR application because multiple layers compose the view rendered onto the display.

3.3 IoT – Smart Mobility

3.3.1 Introduction

One of the domains most affected by the increasing "smartness" of endpoints is transportation and Intelligent Transport Systems (ITS), which is expected to become a significant market, as shown in Figure 10. ITS aims at providing multimodal transportation, embracing public and private municipal, regional, national, trans-national vehicles, and fleets comprised of both autonomous and non-autonomous vehicles.

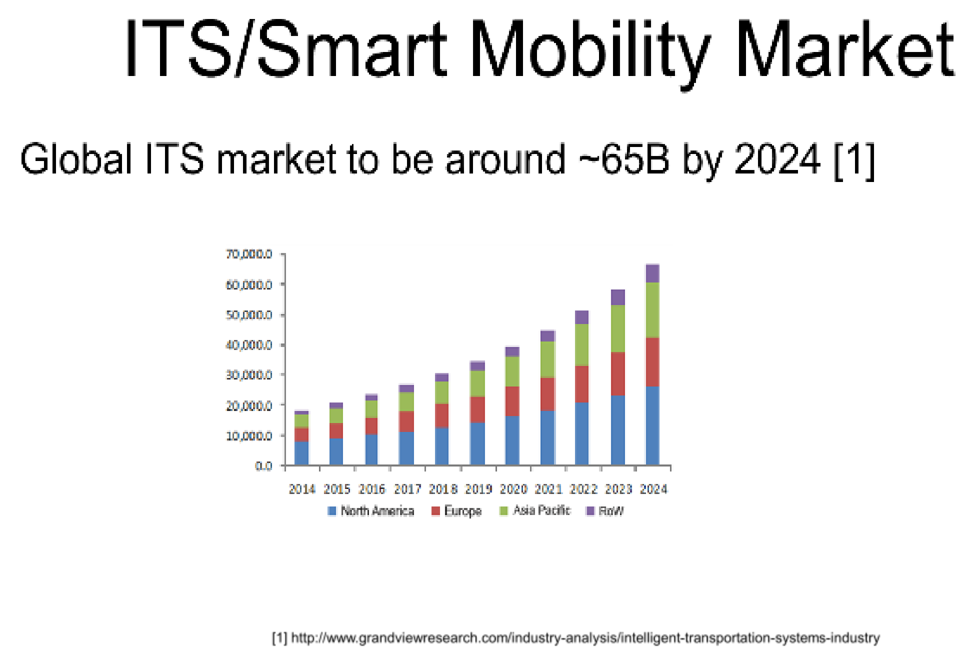


Figure 10: Smart Mobility Market

This extremely heterogeneous transportation ecosystem is made available to users and citizens through advanced services aided by vehicular ad hoc networks that enable several V2X modes of interaction, where X represents vehicles, pedestrians, roadside units (RSUs), Internet, etc. These services fulfill usability requirements while pursuing system-level objectives, including:

- Reduction of the CO2 footprint.
- Real-time delivery of specific goods.
- Reduction of traffic within urban areas.
- Provisioning of pleasant journeys to tourists.
- General commitment of satisfactory travel time and experience[14].

In this context, IoT technologies can play a pivotal role. In particular, traffic management systems (TMS) aided by IoT technologies are creating novel approaches to traffic

[14] Papadimitratos, P., La Fortelle, A., Evenssen, K., Brignolo, R., and S. Cosenza, "Vehicular communication systems: Enabling technologies, applications, and future outlook on intelligent transportation," IEEE Communications Magazine, vol. 47, no. 11, pp. 84-95, November 2009.

modeling[15]. Moreover, such features enable advanced design paradigms (e.g., mobility as a service (MaaS)) with huge implications for systems architectures[16]. As a consequence, smart mobility support can be considered a significant use case for ICN-IoT.

3.3.2 ITS Applications and Challenges

An ITS framework includes many network and system components in support of diverse heterogeneous applications with varying requirements. These applications can be classified as being part of traffic management, safety, environmental applications, or infotainment, and have diverse requirements, as shown in Figure 11. These requirements include requiring high-throughput and/or low-latency, along with very high-reliability and positional accuracy.

To meet the varying requirement challenges, a system infrastructure should ideally offer the following features:

- Support ad hoc communication.
- Self-configuration.
- Multihoming capability.
- Operation over heterogeneous radio interfaces, such as 4G/5G and DSRC.
- Design considering security and privacy.
- Enable efficient edge computing.

[15] Project, BonVoyage., "European Unions - Horizon 2020, http://bonvoyage2020.eu," 2016.
[16] Melis, A., Pardini, M., Sartori, L., and F. Callegati, "Public Transportation, IoT, Trust and Urban Habits," Internet Science: Third International Conference, INSCI 2016, Florence, Italy, September 12-14, 2016, Proceedings.

Use case	E2E latency (ms)	Reliability	Data rate (Mb/s)
CCAS	10	10^{-5}	Less than 5
BEVS	50	10^{-3}	40
AR	100	10^{-2}	100
INS	100	No	50
4K live video	500	No	40 (per video)

CCAS: Cooperative collision avoidance system
BEVS: Bird's eye view system
AR: Augmented Reality
INS: Intelligent Navigation System
From: 5g-ppp, "5G Automotive Vision", https://5g-ppp.eu/ white-papers/

Figure 11: ITS Application Requirements

3.3.2 ICN Suitability to Smart Mobility

ICN is uniquely positioned to serve as a unified network layer that can be adapted to operate in an ad hoc support-constrained IoT network and wired scenarios. A smart mobility infrastructure over ICN satisfies the requirements presented in Section 3.3.1, as discussed below.

Support of Ad Hoc Communication: Ad hoc networking is challenging because the devices are in constant motion, resulting in dynamic topology changes. Protocols designed for fixed-host networking, such as IP, are not suitable in such an environment. ICN's nature as a receiver-oriented session-less transport with the ability to operate in broadcast mode makes it a good candidate protocol to operate in ad hoc conditions.

For example, in a set of moving vehicles equipped with ICN stacks having applications expressing interests, these interests could be broadcast over each vehicle's radio to the neighboring vehicles. These neighboring vehicles could then rebroadcast those interests to other vehicles in proximity. In a normal mode of ICN operation, the data object is expected to retrace the path to the requesting vehicle, but in dynamic conditions this path may no longer be valid. An ICN layer optimization proposed in *Vehicular Inter-*

Networking via Named Data[17] is the notion of data muling, where vehicles listen for broadcast data and cache it opportunistically, resulting in self-replication of content objects, increasing availability and decreasing the latency for a valid response to reach consumers.

Self-Configuration: This is required for ad hoc networks due to difficulties building any control infrastructure to configure each vehicle or supporting control plane configurations. ICN achieves this as its name-based APIs allow applications to bootstrap and begin operation without relying on any network-based configuration, as an IP network would today. However, network-based configuration is required if a vehicle interfaces with a formal cellular infrastructure in order to enable access to ICN services.

Multi-Homing Capability: This allows the vehicles to use multiple interfaces with similar capabilities (e.g., Wi-Fi, LTE) to increase the overall throughput of consuming applications. ICN's receiver-oriented consuming nature and session-less transport allow applications to benefit from this feature.

Operation over Heterogeneous Network Interfaces: Vehicles can be equipped with multiple heterogeneous radio interfaces such as 802.11p (also known as DSRC) and Wi-Fi/LTE, designed for different operational scenarios. In this situation, the ICN protocol can use adaptor functions to map Interest/Data PDUs over these interfaces, while adapting to any communication mode (broadcast, anycast, or unicast) based on the underlying link or ICN network properties. The adaptation layer can also handle transport functions related to reliability (retransmission and loss detection) to improve efficiency to the above ICN layer. Current approaches achieve this through application level gateways on the vehicle or by using external entities to handle this interworking function.

Design for Security and Privacy: Security is one of the primary design considerations in ICN. All data objects in ICN are authenticated using signatures of the producing entity and can be optionally encrypted based on consumer application requirements. Similarly, interests can also be authenticated or can request that the network handle the

---

[17] Giulio Grassi et al, "Vehicular Inter-Networking via Named Data," http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.740.9872&rep=rep1&type=pdf

authentication functions and offload related high computational tasks. This requirement for provenance may be less critical in ad hoc situations[18].

Edge Computing: Smart mobility applications should be developed to take advantage of the edge resources to develop smarter ITS applications requiring data analysis, offloading computing tasks, leverage caching, and storage resources, etc. ICN can also offer an efficient substrate to host services and handle routing logic to interconnect efficiently to service-specific application requirements. Furthermore, ICN introduces new APIs[19] to conduct efficient service execution in the network. One example is by allowing computation logic to be migrated or replicated in different points based on the data location to satisfy user requests efficiently.

3.3.3 5G/ICN Architecture

The next-generation 5G core network architecture (5GC) offers many features designed to simplify networking such as:

- Espousing NFV principles of system implementation to allow separation of the control and user plane functions for better scalability and flexibility.
- Enabling easier plugging-in of multiple heterogeneous radio access technologies (RATs) that can be managed by a common control framework.
- Allowing co-existence of multiple logical network or network slices through formal support of associated context, such as slice identifiers in the control and user plane designed to enforce appropriate traffic management policies. The network slicing paradigm allows the introduction of new network architectures that can coexist with current IP networking[20].

Enabling ICN in 3GPP's 5G NextGen Core Architecture[21] offers architectural discussions on supporting ICN formally in 5GC. This architecture extends the 5GC presented in 3GPP's

---

[18] Mario Gerla et al, "Internet of Vehicles: From Intelligent Grid to Autonomous Cars and Vehicular Clouds," IEEE, WF-IoT, 2014.
[19] M.Sifalakis et al, "An information centric network for computing the distribution of computations," ACM, Sigcomm, 2014.
[20] R. Ravindran, A. Chakraborti, S. O. Amin, A. Azgin and G. Wang, "5G-ICN: Delivering ICN Services over 5G Using Network Slicing," in IEEE Communications Magazine, vol. 55, no. 5, pp. 101-107, May 2017. doi: 10.1109/MCOM.2017.1600938
[21] Ravi Ravindran et al, "Enabling ICN in 3GPP's 5G NextGen Core Architecture," IETF/ICNRG, 2017

TS.23.501[22] and TS.23.502[23]. Figure 12 shows ICN-specific control and user plane functions (in green) to support ICN-UE authentication, routing, forwarding, and interfacing with the 5GC user and control plane to allow efficient handling of ICN PDUs.

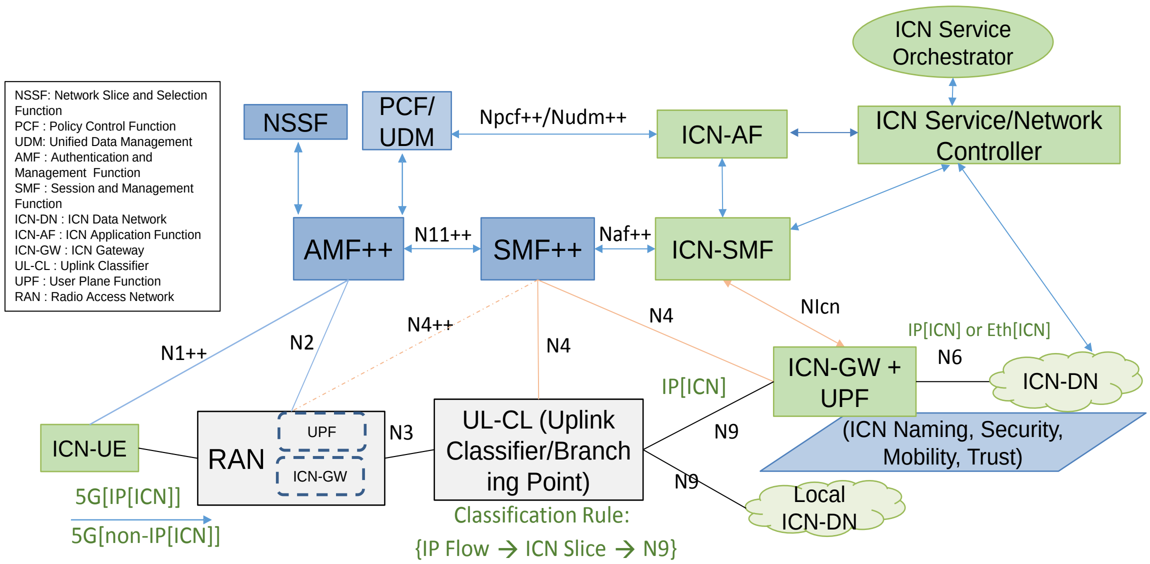


Figure 12: 5G/ICN Architecture

The ICN control plane is orchestrated using a logically centralized ICN service orchestrator to handle service specific requirements related to bandwidth, caching/storage, and computing resources. The orchestrator interfaces with the ICN application function (ICN-AF), which interfaces with the PCF/UDM function of the 5GC to authenticate ICN-UEs into the network and authorize them to access ICN services.

The ICN-AF also interfaces with ICN's session and management function (ICN-SMF) to provision forwarding policies in the ICN gateway (ICN-GW), which is the service gateway to the ICN data networks (ICN-DN). The ICN-SMF interfaces with the 5GC's extended session and management function (SMF++) function to provision ICN session flows in

[22] 3GPP TS 23.501, "Technical Specification Group Services and System Aspects; System Architecture for the 5G System (Rel.15)", 3GPP, 2017.
[23] 3GPP TS 23.502, "Technical Specification Group Services and System Aspects; Procedures for the 5G System (Rel. 15)," 3GPP, 2017.

the UPFs of the 5GC network. Appropriate radio resources are provisioned during the session management procedure using the extended authentication and management function (AMF++) which interfaces with the SMF++ function. Further details on the 5GC and ICN user and control plane functions and extensions have been presented in the *Enabling ICN in 3GPP's 5G NextGen Core Architecture* document.

The architecture offers two modes of ICN deployment:

1. In the integrated mode (differentiated from overlay because the 5GC is aware of the ICN service) in which the ICN in ICN-UE operates over the traditional IP layer. In this case, ICN awareness in the 5GC allows efficient handling of ICN traffic and fast handover to ICN edge services or to the ICN gateway (ICN-GW), which is also the IP anchor point for the ICN-UE.
2. Alternatively, ICN can be formally supported as one of the non-IP protocols. This mode allows for very efficient deployment because the ICN PDUs can be handled at the BS's ICN instance (shown as ICN-GW in RAN in Figure 12), allowing a flat ICN network over which even mobility can be handled. However, this deployment must address the challenge of handling functions such as policy, charging, and legal intercept in a distributed manner, which requires more investigation.

These options can be considered within the scope of realizing virtual networks.

This architecture is used as the basis to discuss the traffic management scenario discussed next.

3.3.4 Smart Mobility Traffic Management Scenario

The scenario shown in Figure 13 has vehicles interacting with each other and with the roadside units over DSRC (802.11p). The RSU's may also be connected over the 5G network traffic sensing and monitoring service hosted in the edge cloud. The vehicles themselves may also directly connect to the traffic management services over a cellular service. It is assumed the information from the localized edge cloud is processed centrally to offer near real-time traffic information to the vehicles for the driver's benefit.

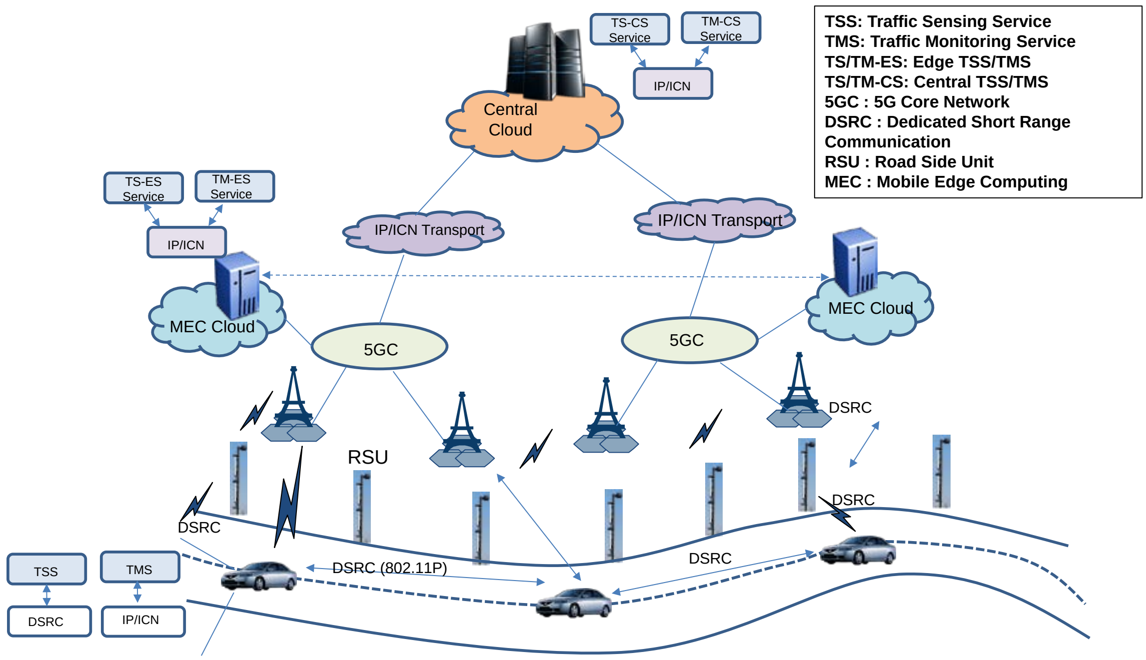


Figure 13: Traffic Management Scenario Over 5G

From the traffic management service perspective, it is assumed the vehicle, the edge cloud, and the central cloud have two components related to traffic management application: one for traffic sensing (TS) and the other for traffic monitoring (TM). The TS service (TSS) in the vehicle obtains traffic statistics such as vehicle position, estimated proximity to other vehicles, and speed through periodic exchange of information with other vehicles. Similar information can also be obtained by the RSU monitoring road conditions within its zone. The vehicle and the RSU share this information with the TS service component in the edge cloud instance (TS-ES). The information at the edge service is processed, and this scoped information is sent to the central TS-CS module to generate useable navigation information. This information can be periodically pushed to or pulled by the edge traffic monitoring service (TM-ES). The car's navigation system (CNS) uses this data from the edge traffic monitoring (TM-E) service instance to satisfy specific navigation queries or offer rich insights on the road conditions, such as real-time congestion assisted with media feeds.

The next discussion focuses on the architectural view along with discussion of this scenario when developed over IP versus ICN with assistance from edge cloud over 5GC.

3.3.4.1 IP-Based Solution

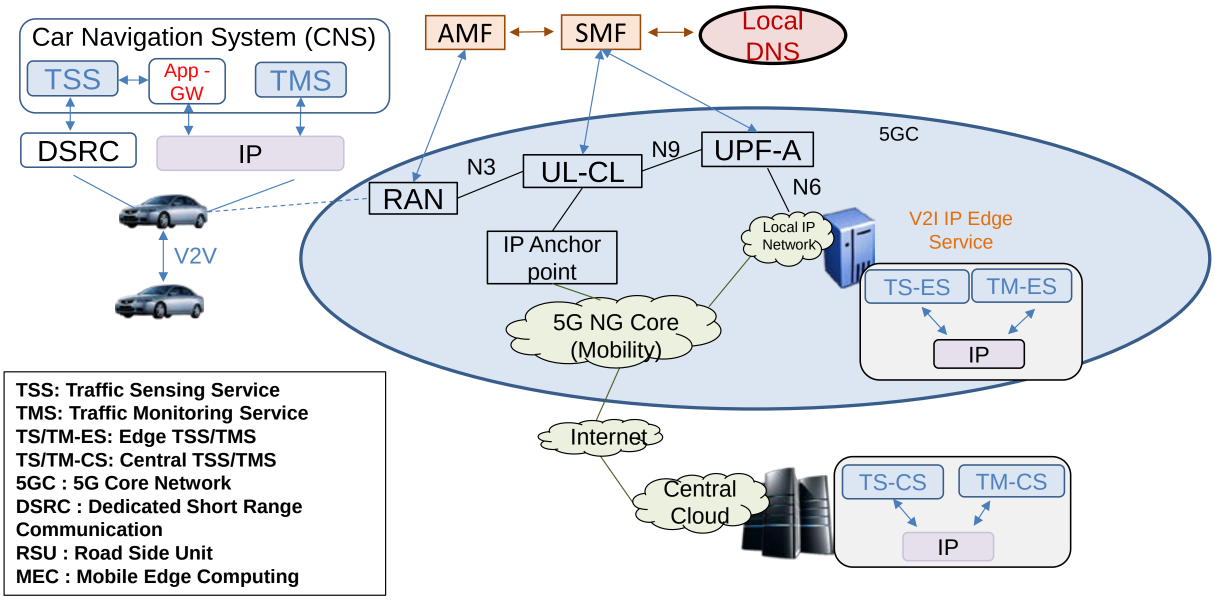


Figure 14: Realization of TM System over IP

As the vehicle's networking system comes online on an IP network (Figure 14), it first undergoes an attachment process with the 5G-RAN, which includes authentication, IP address assignment, and DNS discovery. The attachment process is followed by PDU session establishment, which is managed by SMF signaling to UL-CL and the UPF instances. When the CNS application initializes, it assumes this IP address as its own ID and tries to discover the closest TS/TM service instance. Local DNS then resolves the service name to a local MEC service instance. Accordingly, CNS learns the IP service point address and uses that to coordinate between traffic sensing and monitoring applications.

This design creates the following challenges:

1. At the CNS level, non-standardization of the naming schema results in introducing an application level gateway to adapt the sensing data obtained from DSRC systems to IP networks. This gateway becomes mandatory if the applications are from different vendors or derived from orthogonal standards.
2. As the mobility results in handover between RAN instances, service-level or 5GC networking-level mechanisms need to be initiated to discover a better TM/TS edge instance. This process may affect the service continuity and result in session reestablishment that introduces additional control/user plane overheads.

3. Data confidentiality needs, authentication, and privacy control are offered through an SSL/TLS mechanism over the transport channel, which must be re-established whenever the network layer attributes are reset. Also, the data obtained using channel-based security models lose validity outside their scope, making the data not shareable with other vehicles on the road.

3.3.4.2 ICN-Based Solution

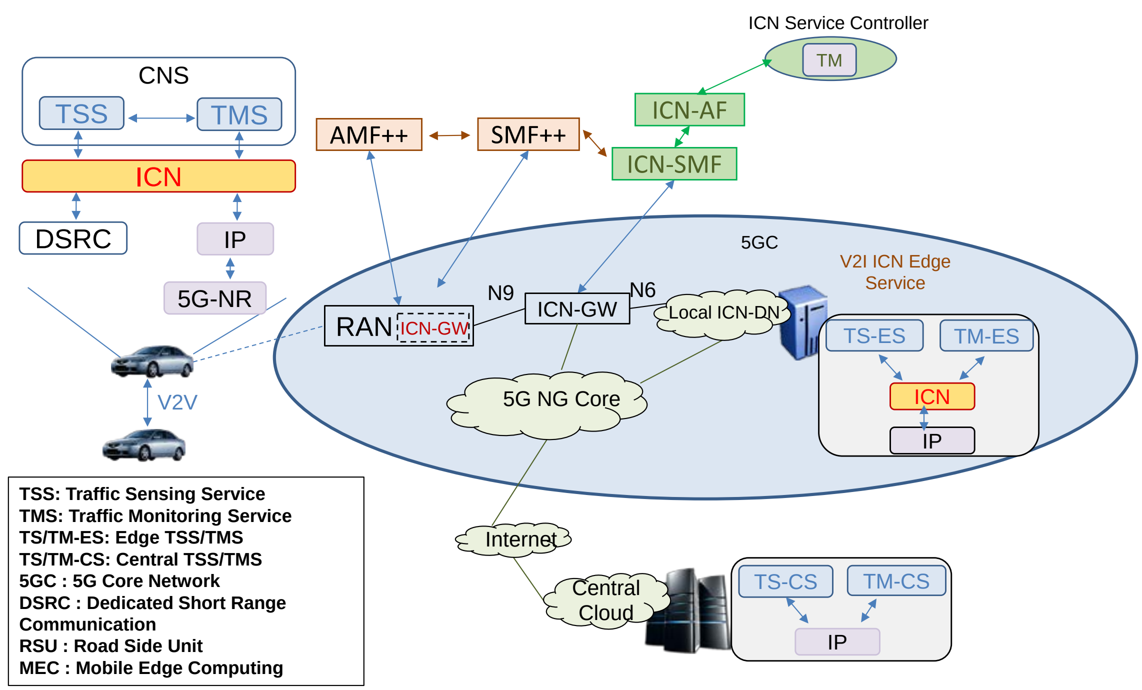


Figure 15: Realization of TM System over ICN

If the CNS application is developed over ICN (Figure 15), ICN allows the same named data logic to operate over heterogenous interfaces (such as DSRC and IP or natively over a 5G link), thereby avoiding the need for application layer adaptation. Here, adaptation is resolved by the named-networking architecture that allows the data to be seamlessly exchanged between distributed application functions.

The advantages of using ICN-based MEC are as follows:

1. As vehicles within a single road segment are likely to seek the same data, ICN-based MEC allows the leveraging of opportunistic caching and storage enabled at

ICN-GW or even at the BS, thereby avoiding service level unicast transmissions. Potentially, ICN and the radio MAC layer can be cross-optimized to enable multicast on the air interface, as well, saving significant radio resources.

2. Processed and stored traffic data can be easily contextualized to different user requirements leveraging the computing resources at ICN-GW or the BS very efficiently using ICN's native support for service functions.
3. Appropriate mobility handling functions can be used depending on mobility type (as consumer or producer). Specifically, when an ICN-UE moves from one RAN instance to another, the next IP hop, which identifies the ICN-GW function, must be re-discovered. Unlike the IP-MEC scenario, this association is not exposed to the applications. As discussed earlier, control plane extensions to AMF and SMF can enable re-programmability of the ICN layer in the vehicle to direct it towards a new ICN-GW, or to remain with the same ICN-GW, based on optimization requirements. These reset the tunnel interfaces between the UE and ICN-GWs. Alternately, when ICN operates as a non-IP protocol, an IP tunnel no longer needs to be managed between the ICN-UE and the ICN-GW.
4. As ICN offers content-based security, produced content can be consumed while authenticating it at the same time, allowing any data produced to diffuse to its point of use through named data networking.

3.3.5 Quantitative Support

The use of ICN for smart mobility is rationalized based on the qualitative and quantitative metrics that affect end-user experience. These include throughput, latency, and overhead due to ICN's data object-based security feature.

3.3.5.1 High Throughput

Studies such as *ICN-MCN: An Architecture for an Information-Centric Mobile Converged Network*[24] have shown a benefit of ICN is in offering higher throughput in wireless conditions. This benefit comes from the ability of the ICN end points to leverage their multi-homing feature, along with multipath content retrieval. As a result, they can potentially achieve the aggregate of the throughput typically achieved over each individual path. High throughput is also aided by strategic layer policies that can be

[24] Reza Tourani et al, "ICN-MCN: An Architecture for an Information-Centric Mobile Converged Network," IEEE Communication Magazine, Sept. 2016

applied over the faces in the ICN layer to enable dynamic forwarding decisions, derived from the response-time statistics, channel conditions feedback, and loss rate accounting over the multiple faces. The study referenced above particularly focuses on benefits gained by a UE operating under multi-radio access scenarios toward achieving high-bandwidth and node mobility. High node mobility is possible because ICN end points do not operate under session semantics and avoid challenges related to session setup, teardown, and repair phases.

Figure 16 shows the simulation setup for the study, where the UE is operating in a multi-homed scenario comprised of Wi-Fi, LTE, and ad hoc wireless access. The interests from the applications are divided among these faces based on a latency learning scheme derived from exponentially weighted moving averages over the response time of the content requests measured and normalized over all the faces. Figure 16 also shows this dynamic latency learning and interest request adaptation over the available faces resulting in higher throughput than when a single interface or other random schemes are used.

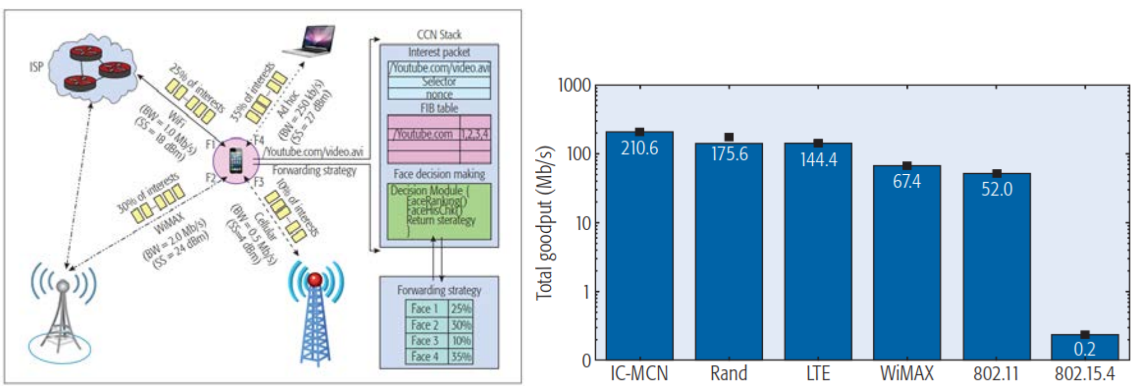


Figure 16: Simulation Setup and Goodput Performance

3.3.5.2 Latency Performance

Although ICN aids latency through multiple features such as caching, multi-homing, and multipath routing, the focus here is on how ICN's in-network mobility function fares with mobility. Two kinds of mobility must be considered here: the consumer mobility scenario and the producer mobility scenario. As ICN consumers operate in a session-less manner, consumers can recover the interests expressed from their previous point of attachment

(PoA) and re-express the interests from their new PoA. But as shown in *Experimental Evaluation of Consumer Mobility on named data networking*[25], this default mechanism may not serve the needs of all applications. With the increasing user speed, the ability to recover the pending interests expressed when attached to the previous PoA may suffer because of multiple factors such as:

1. The ICN layer's ability to express those interests in a timely manner depends on the signaling it can receive from the lower link layer about the transient changes such as change of PoA.
2. The engineered caching resources in the network to aid consumers with quick recovery of content after an interest re-expression.

Producer mobility has more challenges in its default mode of operation if only routing updates are used to track its current PoA[26]. Scalable approaches to handle producer mobility have been found to handle seamless producer mobility[27,28]. The *Seamless Producer Mobility as a Service in Information Centric Networks*[29] study shows the effectiveness of introducing the notion of application and network identifier split in ICN, aiding to seamless mobility experience, where the loss of data is observed to be less than 50ms during each handover.

3.3.5.3 Security Overhead

Object-based security is an important feature that enables multi-sourcing of content and leverages caching through name-based routing primitives. Costs can vary based on the ICN design choice. The security overhead in this case versus HTTP/HTTPS has been

---

[25] Zongzhen Liu et al, "Experimental Evaluation of Consumer Mobility on named data networking," IEEE, ICUFN, 2014.

[26] A. Azgin, R. Ravindran, and G. Wang. "Mobility Study for Named Data Networking in Wireless Access Networks." In IEEE International Conference on Communications (ICC), pages 3252–3257, Sydney, Australia, June 2014.

[27] A. Azgin, R. Ravindran, and G. Wang. "A Scalable Mobility-Centric Architecture for Named Data Networking." CoRR, abs/1406.7049, 2014.

[28] R. Ravindran, A. Chakraborti, and A. Azgin. "Forwarding-Label Support in CCN Protocol" - draft-ravi-ccn-forwarding-label, IRTF-ICNRG draft, 2018.

[29] A. Azgin, R. Ravindran, A. Chakraborti and G. Wang, "Seamless Producer Mobility as a Service in Information Centric Networks," ICN, Sigcomm, 5G-ICN Workshop, 2016.

analyzed in *Forwarding-Label Support in CCN Protocol*[30]. As shown in Figure 17, the overhead varies depending on the implementation of CCN over the lower layers. Compared to CCN over Ethernet, where allowed payloads are restricted to 1300 bytes, CCN over UDP offers better performance as bigger ICN payloads were used to amortize the cost of the security overhead. This results in very few data packets used to transport the content to the application.

	HTTPS	CCN/UDP	CCN/ETH
App Payload	16944	16944	16944
Packets Out	16	5	16
Bytes Out	1548	629	1791
Avg. Size Out	96.75	125.8	111.94
Packets In	22	14	14
Bytes In	21232	18253	20910
Overhead In	25.31 %	7.73 %	23.41 %
Avg. Size In	965.09	1303.79	1493.57

Figure 17: Security Overhead Analysis

4. Key Findings

This report's findings were specifically designed to reflect the forward-looking use cases selected for this value assessment. Therefore, it was important to select use cases that represented market-relevant and network-impacting opportunities, as opposed to use cases that were specifically conducive to ICN-based solutions. Given this objective, use cases were selected from the following use case classes:

Linear video delivery: Represents the major consumer of network resources as an increasing amount of video content is streamed, including live content from premium services, social media, Internet services, and live events.

AR: This is a prime example of a content-rich set of applications that is already emerging at the device level and will introduce new requirements in networks, in terms of

[30] R. Ravindran, A. Chakraborti, and A. Azgin. "Forwarding-Label Support in CCN Protocol" - draft-ravi-ccn-forwarding-label, IRTF-ICNRG draft, 2018.

integrating pre-staged content with real-time information and metadata, to create a unique user experience.

IoT (Smart Mobility): Rapidly expanding set of applications that are characterized by many new sources of (rapid) mobility, sharing of real-time data, caching near the network edge, and frequent renewal of information.

In assessing each use case, it is acknowledged that video delivery represents an area of greater network design knowledge at this time and, consequently, greater access to actual experimentation results. The assessment of the remaining use cases, focused on AR and smart mobility, relied on research-level results and qualitative analysis to serve as the foundation for comparative assessments. In addition, the future introduction of 5G technology and new architectures, such as mobile edge computing, required a level of predictive analysis to evaluate the impact on each use case.

The following is a summary of key findings for the identified use cases.

Linear Video Delivery

- Separation has been made between the beneficial reduction of traffic volume and server loading (including avoiding the thundering herd problem) that accrue from increased distribution of caching capabilities and tailored cache management strategies, from the performance impact and additional capabilities that are intrinsic to ICN versus other state-of-the-art mechanisms for delivering content in the multi-access, dynamic network. In designing a CDN for linear video distribution, the architect must trade off the costs and performance impacts of providing additional caching capabilities in the forwarding layer (i.e., additional memory at a forwarding node) against the placement of a L4/L7 application-layer cache at a strategic (perhaps the same) location. This decision will generally be driven by the amount of content that needs to be stored in the particular location. The optimal overall solution will comprise a combination of distributed caches/proxies and a delivery mechanism that is suited to supporting the objectives stated in this report.
- Increased reliability can be achieved with ICN-IP by virtue of the fact that ICN forwarders deal with losses by serving missed packets from copies retained in the buffer of the nearest upstream forwarder. Current TCP/HTTP mechanisms rely on TCP for packet loss handling. Thus they can reduce latency at the expense of

reliability, and they perform poorly in the case of losses and highly dynamic channels.

- The ICN-IP approach has additional inherent capabilities that should improve performance for linear video delivery: better support for access aggregation, multipath, dynamic load balancing, multi-server, and user mobility. This is primarily due to the ability to take advantage of multiple delivery paths and can be further advantaged by dynamic discovery of paths and the most efficient server. ICN (and ICN-IP) supports bandwidth aggregation and load balancing not only from an end-to-end point of view, but also from hop to hop, making instantaneous routing decisions for interest packets on a per-packet basis based on estimates of instantaneous link quality.

AR/Live Event

- AR/VR is extremely latency sensitive and high-bandwidth intensive due to the requirements for a 360-degree view. The request aggregation mechanism associated with ICN enables interests to be filtered and forwarded more efficiently.
- ICN allows distribution of named tiles (exposed at network layer) to be shared between multiple users accessing an overlapping field of view. While each user may have a perceived customized view in the AR/VR application, ICN can promote the transparent sharing and caching of these tiles.
- From a reliability perspective, the packet loss discussion described under linear video is also applicable in the AR/VR use case because copies of the data cached near the users can be leveraged in response to a packet loss.
- Support for multisource distribution is especially relevant in the AR/VR application area because the view on the display is composed of data from multiple sources.
- Similar to the linear use case, it is expected that the AR/VR live event application could take advantage of ICN-IP solutions to avoid full network replacement scenarios.

IoT/Smart Mobility

- In many Smart Mobility applications, vehicles are likely to seek the same data. ICN-based MEC allows the leveraging of opportunistic caching and storage enabled at ICN-GW or even at the BS, thereby avoiding service-level unicast transmissions.

- Processed and stored traffic data can be easily contextualized to different user requirements leveraging the computing resources at ICN-GW or the BS very efficiently using ICN's native support for service functions.
- Appropriate mobility handling functions can be used depending on mobility type (as consumer or producer), avoiding the need to re-discover the next IP-hop.
- As ICN offers content-based security, produced content can be consumed while authenticating it at the same time, allowing any data produced to diffuse to its point of use through named data networking.
- In a broader sense, Smart Mobility can take advantage of some of the characteristics that are benefitted by ICN, including higher throughput (leveraging its multi-homing and multi-path content features), better latency performance (through multiple features such as caching, multi-homing, and multipath routing), and utilization of object-based security, an important feature that enables multi-sourcing of content.

5. Conclusions and Next Steps

This *eCON Value Assessment Report* and the previously published *Evolution to Content Optimized Networks Report* were developed to assist the industry in assessing the state of readiness and the value of ICN-based named content solutions. As content delivery needs continue to grow in the marketplace, and requirements for performance factors such as throughput, latency, and security become more challenging, network architects and planners will evaluate new content-centric architectures. This report's goal is to advise network decision-makers on the feasibility of ICN-based solutions by identifying the incremental and evolutionary benefits associated with approaches that are designed to meet future applications, such as streaming video, AR, and IoT.

One of this report's significant findings is that ICN-based solutions, which are supportive of specific content rich applications, will most likely be implemented as part of network slicing approaches over IP core networks. Therefore, there will be an extended period of co-existence and complementary strategies between ICN and IP approaches. This suggests a close pairing between ICN and the applications that will demand not only features such as in-network caching and multicast delivery, but also a long evolutionary path in terms of leveraging existing investments in IP networks. The introduction of 5G technologies will undoubtedly play a major role in assessing the incremental value between IP/MEC solutions and ICN solutions. In addition, the degree to which content-rich applications (that are dependent on edge processing, storage, and analytics) emerge in the marketplace will play a major role in driving name-based content solutions into next-generation networks.

As ICN-based protocols and inter-working standards work continues in the industry, ATIS will continue to promote the analysis and findings associated with eCON to the appropriate organizations. The two eCON reports represent a unique deployment-level assessment of ICN-based solutions and provide network architects and planners with valuable insight around content-centric solutions. ATIS will also leverage the output of this eCON work with other ATIS strategic and committee activities related to next generation networks, IoT, AR, artificial intelligence, identity management, Smart Cities, and related areas.

Annex A

Research Related to ICN-Based Solution for Linear Video Delivery Use Case

Cisco Labs conducted a series of experiments to study the network performance and user experience impacts of ICN-based solutions for linear video distribution applications. Cisco Hybrid ICN™, an implementation of ICN within IP, was used for all of the experiments discussed in this annex. This ICN-IP implementation also supports *Wireless Loss Detection and Recovery* (WLDR), a mechanism that provides sub-RTT lost packet recovery for ICN over wireless channels.[31]

The Cisco Hybrid ICN™ forwarder is an example of an IP forwarder augmented by an ICN module able to recognize IP packets carrying an ICN semantic and divert them from the regular forwarding pipeline to be processed through an ICN stack. ICN names are mapped into IPv6 addresses. If ICN-IP is enabled in the forwarder, each incoming packet is examined to determine whether it is an ICN packet. When ICN packets are identified, they are processed as ICN-IP, and then forwarded using standard IP mechanisms. In the forwarders used in these experiments, the ICN-IP functions could be turned off and on, allowing the same forwarders to be used for both the ICN and non-ICN measurements.

The results of these experiments are illustrated below for three types of network performance categories:

1. Reduced latency via in-router buffering at the edge.
2. Increased reliability at low latency via in-net control.
3. Reduced latency/load with dynamic multisource.

Low Latency Video Distribution via In-router Caching and Control

The purpose of this set of experiments was to:

- Demonstrate how ICN-enabled edge routers accelerate live distribution while offloading origin server/PoP caches without requiring any content placement optimization or request routing/transport operations to associate user requests to closest hitting caches.

[31] G.Carofiglio, L. Muscairello, M.Papalini, N. Rozhnova, X. Zeng, 2016, Op Cit

- Illustrate the latency reduction resulting from in-network control operations such as loss detection/recovery vs. end-to-end retransmissions.

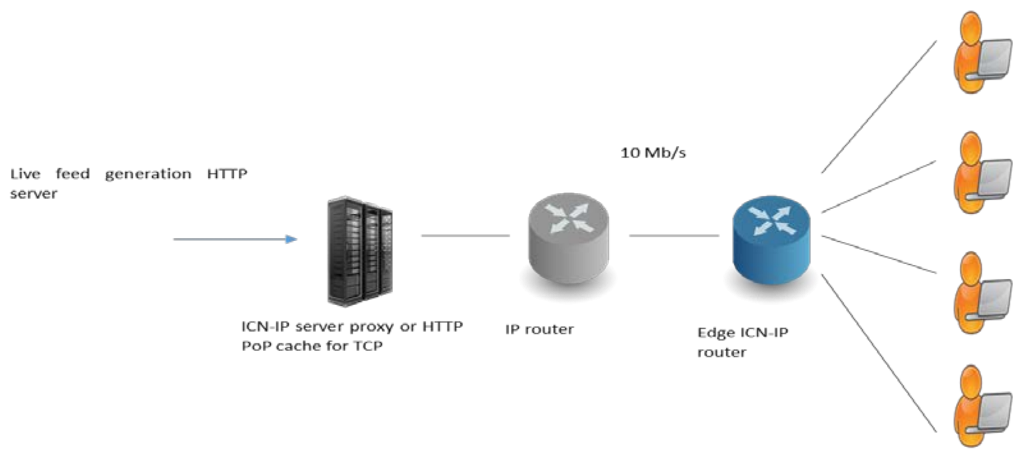


Figure A.1: Reduced Latency via In-Router Buffering at the Edge

In this experiment, the following set of conditions was applied:

- Four clients, configurable as TCP/IP or ICN-IP.
- Clients initiate sessions independently, asynchronously.
- 10 Mb/s link between IP router and edge router.
- Configurations tested included:
  - TCP/IP only.
  - ICN-IP at server and clients only.
  - ICN-IP at server, edge router, and clients.
- Measurements made of per-client throughput, latency, and total traffic bandwidth.

Figure A illustrates the results for a single-bit-rate video.

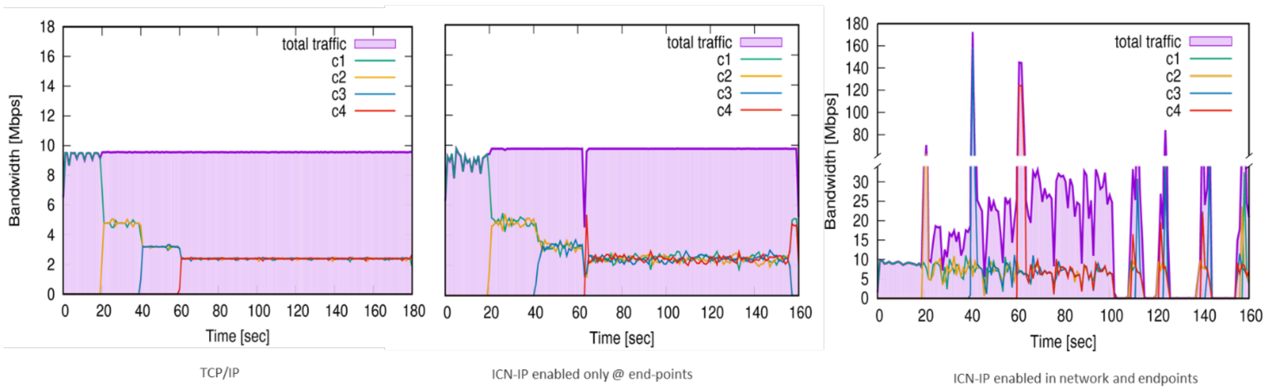


Figure A.2: Bandwidth (for Single-Bit-Rate Video)

- For a TCP-only case, the total bandwidth is limited by the 10 Mbps link. As each client joins, the per-user bandwidth decreases to a total of 10 Mbps, and latency increases. With four clients active, per-client bit rate is reduced to 2.5 Mbps.
- When ICN-IP is enabled only at the client, performance is slightly better than TCP only, but is very similar.
- When ICN-IP is enabled at the edge router and the clients, total average traffic bandwidth is reduced, and average per-client throughput is increased to close to 10 Mbps (almost 4x improvement over TCP/IP).

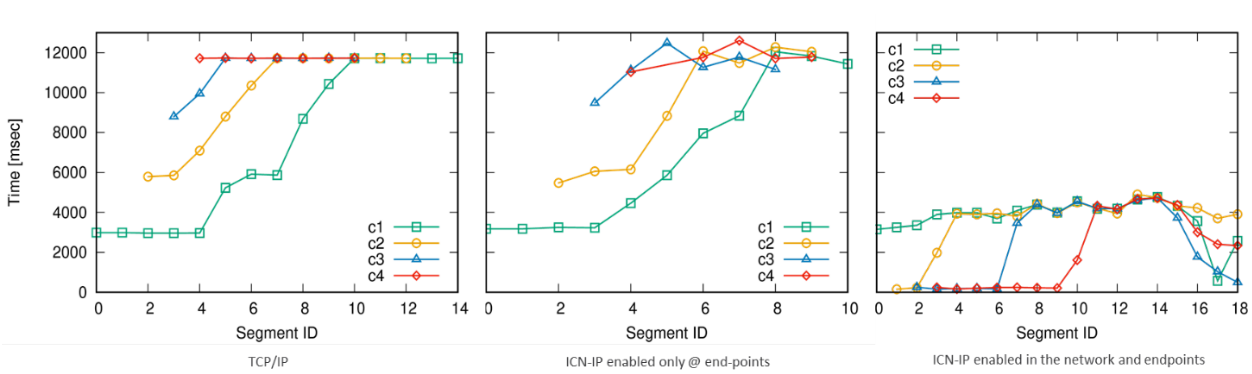
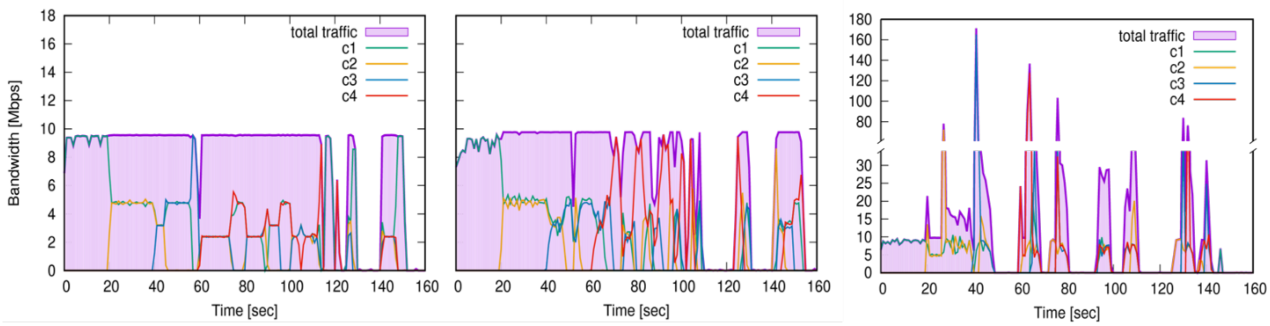


Figure A.3: Time Required to Download a Segment (Latency)

- For both the TCP/IP and ICN-IP at server/client-only scenarios, average time to download a segment (latency) increased as new clients joined, settling at approximately four times the single client value.
- For the ICN-IP enabled at the edge router and at the server/client case, average time to download a segment did not increase as new clients were added and remained at

about the single-client value. This amounts to a factor of N latency improvement over TCP/IP, where N is the number of clients.

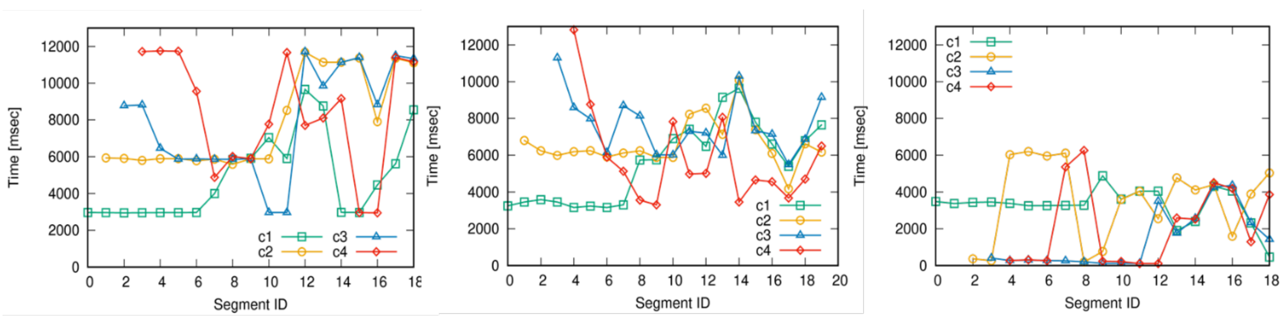
A more challenging experiment was conducted for multiple bit rates, where each client may independently request a different video quality.



TCP/IP ICN-IP enabled only @ end-points ICN-IP enabled in network and endpoints

Figure A.4: Bandwidth (For Multiple Bit Rates)

- This is more challenging for edge-router caching because each client may independently request a different video quality.
- Results are similar to single bit rate case: For TCP/IP, once all the clients have joined, per user bit rate is close to 2.5Mbps. When ICN-IP is enabled in the server, client, and edge router, the average per-user bit rate climbs to close to the 10Mbps limit.

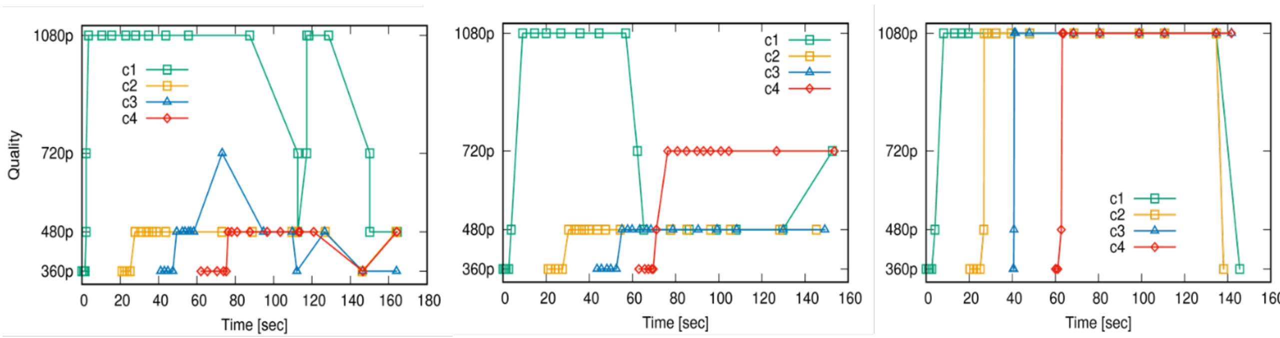


TCP/IP ICN-IP only enabled @ end-points ICN-IP enabled in the network and endpoints

Figure A.5: Time Required to Download a Segment (Latency)

- Similar to the single-bit-rate case, latency improved with ICN-IP enabled.
- In the TCP/IP scenario, peak latency was 4x the single active client latency, and the average was more than double that of the ICN-IP enabled scenario.

- With very few exceptions, the per-client latency with multiple clients active never exceeded the single active client latency when ICN-IP was enabled.



TCP/IP ICN-IP enabled only @ end-points ICN-IP enabled in the Network and endpoints

Figure A.6: Video Quality

- In TCP/IP only scenario, only client c1 gets the highest quality.
- When ICN-IP is also enabled in the network instead (right-most box in Figures above), all the clients are able to get the highest video quality.
- Besides the unfairness, the gap in terms of video quality is of two video quality levels gained with ICN-IP.

Live Distribution over Mobile and Heterogeneous Network Environments

The objective for this set of experiments was to demonstrate:

- Advantages of managing mobility at the client in ICN (fast, anchorless mobility).
- Benefits of exploiting multiple paths in parallel by making the network and video aware of load-balancing decisions (bandwidth aggregation over multiple access paths).

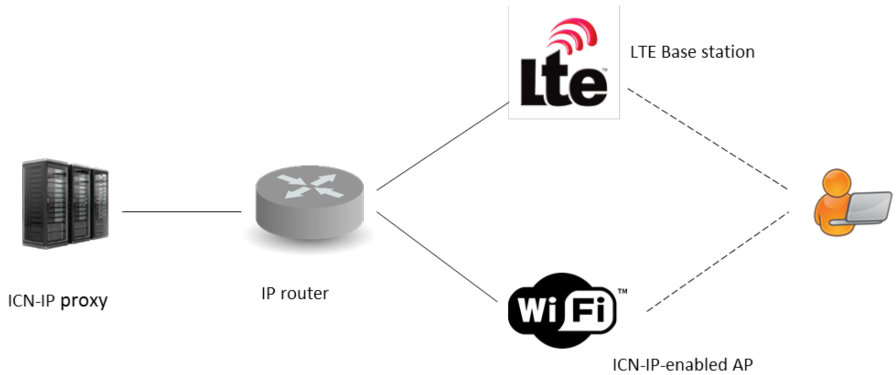


Figure A.7: Live Distribution over Mobile and Heterogeneous Network Environments

In this experiment, the following set of conditions exists:

- Both APs (Wi-Fi and LTE) are ICN-IP enabled.
- Live, ABR video streamed from proxy, as in previous experiment.
- Radio channel conditions varied to force handovers at pre-determined intervals (1, 2, and 5 seconds).
- Loss-recovery mechanisms demonstrated in the previous experiment are turned off to highlight only ICN-IP's inherent mobility support.

Figure A8 shows the results for the single inter-access mobility experiment.

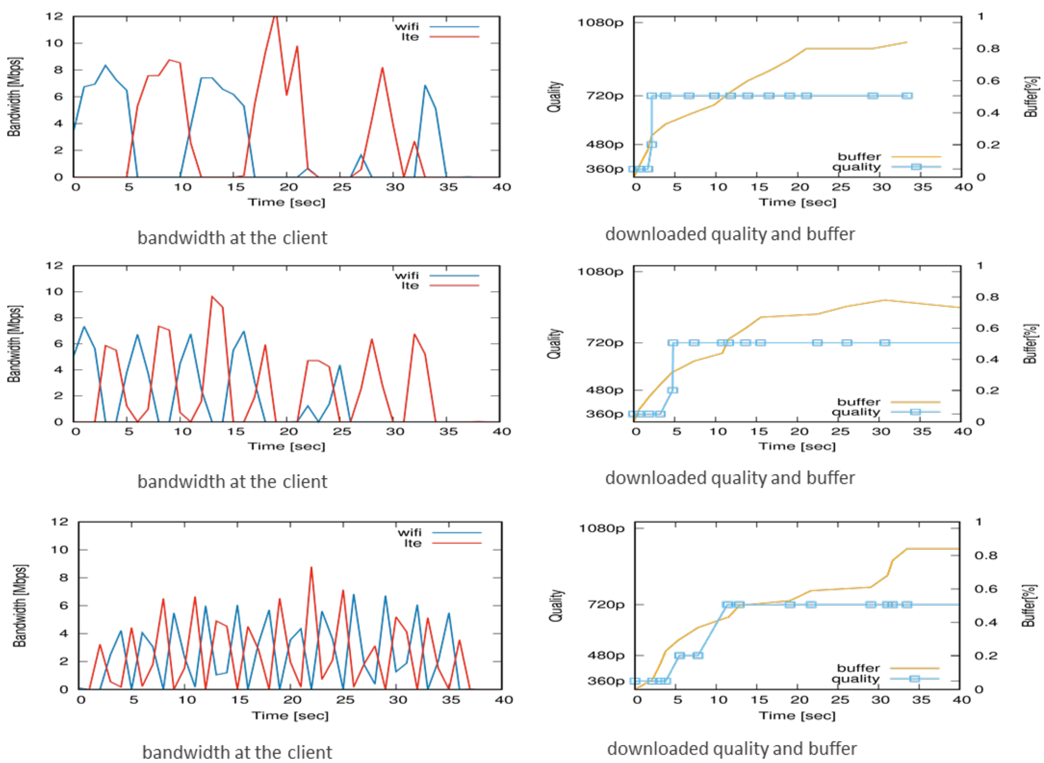


Figure A.8: Single Inter-Access Mobility

- Radio link quality is varied to force handovers at 5, 2, and 1 second, respectively.
- Paths are not known *a priori*.
- Session-less, client driven.
- With 5 second handover frequency, performance is the same as with no mobility at all: builds to 720p as soon as the buffer reaches threshold.
- Even with 1 second handover frequency, video quality builds to 720p and remains there for the remainder of the trial.

In order to demonstrate seamless bandwidth aggregation over multiple access channels, the following topology was used.

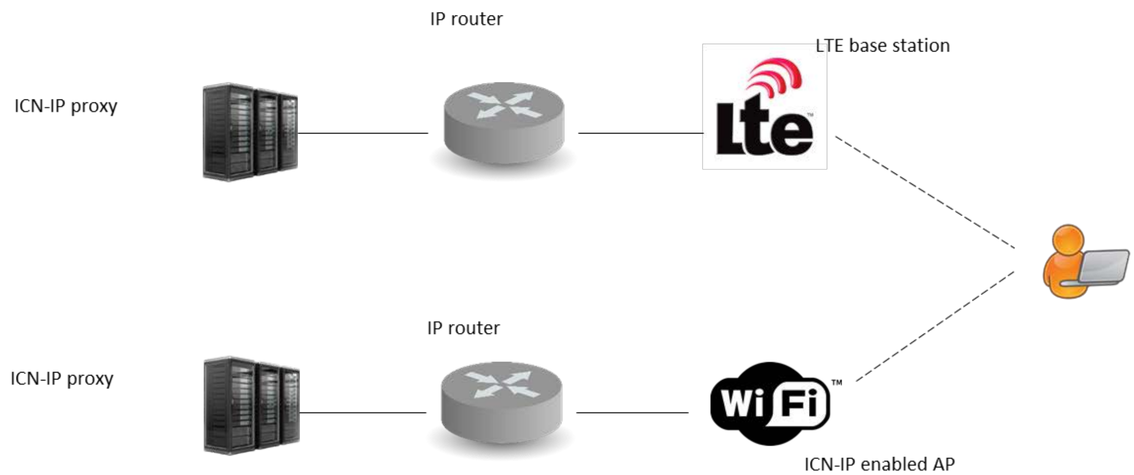


Figure A.9: Demonstration of Seamless Bandwidth Aggregation Over Multiple Access Channels – Multiple Video Sources

Figure A.10 shows the results from this experiment.

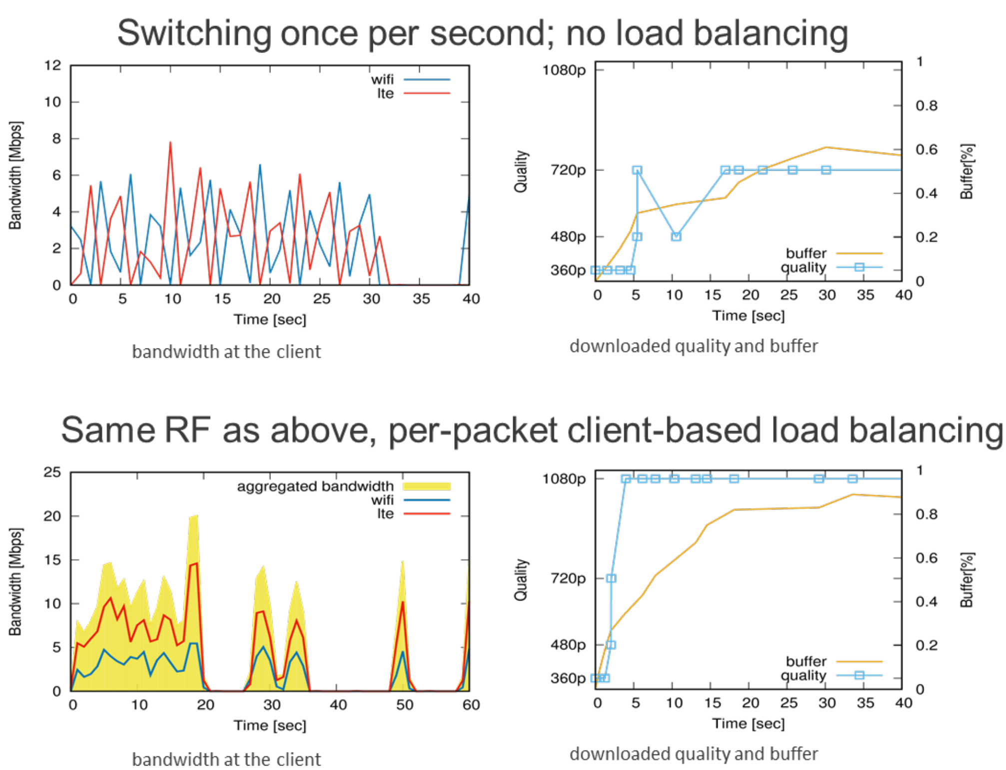


Figure A.10: Dynamic Inter-Access Bandwidth Aggregation

- Per-segment load balancing performs exactly like best single-path selection (720p), because of the rapid variations in RF link quality (compared to DASH segment length).
- Per-packet load balancing makes the best use of all the BW available on both channels, and results in the highest video quality at the client (1080p).

Live Distribution at Scale and Multicast

The following experiment was conducted to demonstrate the scaling advantages of ICN with respect to existing alternatives. The following is the list of objectives:

- By addressing content by name rather than hosts, ICN makes video distribution scale with live-sessions rather than with the number of users.
- Beyond the latency and bandwidth efficiency benefits illustrated in the previous experiments, the state at intermediate caches or at the server proxy is considerably reduced because there is no need to keep track of user sessions.
- ICN removes the scaling limitations at PoP caches or at the server in many OTT delivery systems (e.g., Facebook, PBS).

The following topology was used to demonstrate scaling and multicast efficiency:

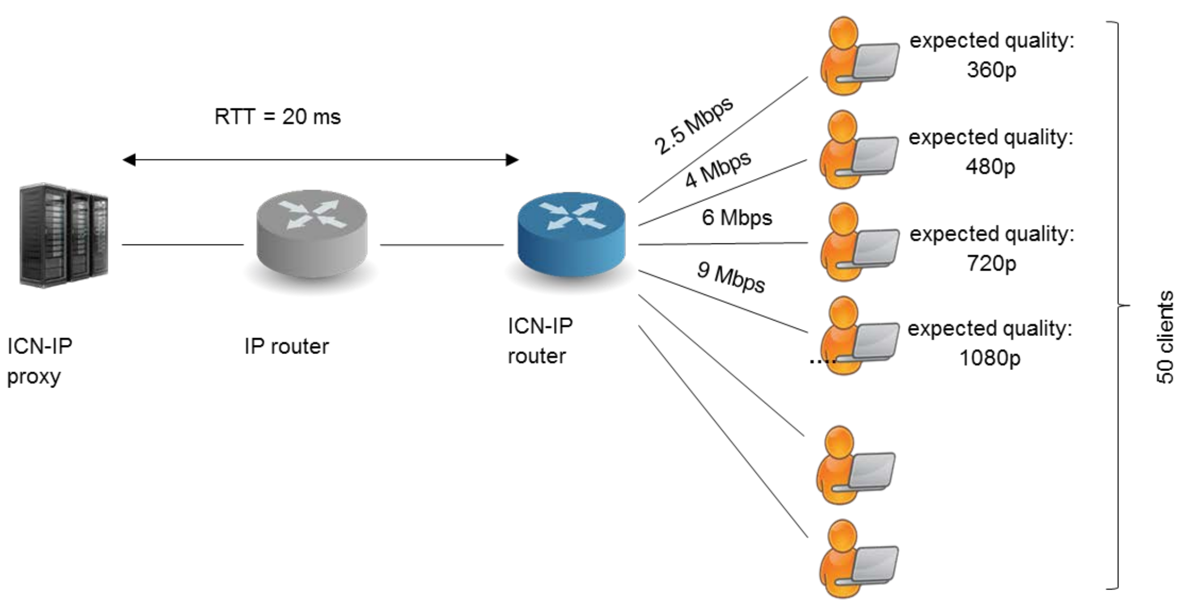


Figure A.11: Demonstration of Scaling and Multicast Efficiency

The following conditions apply to this topology:

- ABR video source producing four video qualities.
- Single ICN-IP edge router (e.g., a BNG router).
- RTT from ICN-IP proxy to ICN-IP edge router = 20 ms.
- 50 clients served by the ICN-IP router.
- Multicast is ad-hoc and not provisioned.
- Bottleneck links to clients are uniformly distributed over four values, giving four "expected" video qualities.
- Multicast delivery efficiency is measured by the load seen at the server.

Figure A.12 illustrates the results for this experiment.

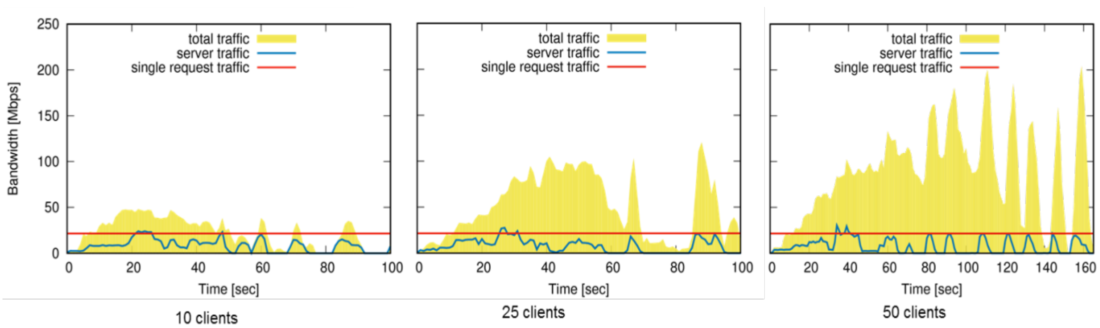


Figure A.12: Server Load Scaling with Increasing Number of Clients

- The yellow area in the graphs is the total load generated by the clients as a function of time. The blue line is the load seen by the server. The red line is the ideal server load for perfect multicast, which is the sum of the four supported video rates.
- The results demonstrate that interest aggregation caching at the edge router made a significant difference in server/proxy load, and verify the assertion that, for ICN, server load scales with the content, not with the number of clients.

Use Case Findings

For the low latency video distribution via in-router caching and control experiment, the following findings were observed.

ICN-IP offered reduced latency to users as a result of:

- Closer distance from clients (content is available in the access/backhaul vs. PoP).

- Lower load in the access than in the PoP (due to larger fan-out).
- L3 network latency (ms) is smaller than HTTP L7 latency (s).

It demonstrated the following additional advantages:

- **Low state overhead:** It scales with the number of content, while caches in the PoP (e.g., Facebook solution) scale with the number of users and content.
- **It is easy to integrate:** No requirement for any box in the PoP, or any change to HTTP/ABR (and thus compatible with HLS/DASH).
- **It has simpler configuration and operations:** The push/pull primitives are at the transport layer rather than at the application layer, and caching is reactive.

For the live distribution over mobile and heterogeneous network environments experiment, the following findings were observed.

The benefits of seamless, frequent inter-technology handover demonstration are:

- Reduced latency during consumer mobility due to immediate retransmissions at consumer (vs. timer expiration + ½ RTT).
- Reduced path stretch (=latency) during produced mobility (guaranteed in ICN/ICN-IP MAP-Me protocol).
- Low signaling overhead (no routing update involved, only to traversed routers in new to closest previous producer location).
- Better reactivity to frequent mobility (support for real-time group communications).

The benefits of seamless bandwidth aggregation over multiple access channels – multiple video sources demonstration are:

- **Higher throughput:** Reduced segment latency (e.g., startup delay) due to parallel downloading on a per-packet basis.
- **Reduced max link load/ congestion at server** by virtue of better load distribution.
- **Better reactivity to channel quality changes** by virtue of congestion-awareness in real-time.
- **Fine-tunable:** Capability to program application/user-based forwarding policies (which is not possible in L2 channel bonding).
- **Easier to insert than alternative mechanisms:** Does not require any proxy/tunnel or kernel modification at client/server, nor standardization, 5G/CUPS compatible.

For the live distribution at scale and multicast experiment, the following findings were observed:

- Reduced load on ICN-IP routers since each edge router has a smaller fan-out with respect to PoP. For example, a PoP cache for Facebook live can handle up to 200K client, while beyond that requires redundancy and load balancing[32].
- Reduced server load as the video server receives a few requests per content object, no connection to maintain (and not with users, which provide automatic traffic redundancy elimination).
- Smaller in-router caches that are distributed in the network, not all resources at the PoP, which simplifies implementation, deployability and maintenance.
- Request/reply vs. pub/sub (ad hoc).
- Traffic scaling with content.
- ABR support for multicast delivery.
- Expected gain: orders of magnitude higher scale and reduced servers' load for single rate and multi-rate video.

Overall conclusions across each of the comparative experiments include:

- Experiments focused on a large-scale, live video distribution application, and span the most challenging use cases encountered in that application:
  - Scalable HTTP ABR distribution everywhere, over wired and wireless networks.
  - With and without mobility support.
  - Built-in load balancing, multi-homing, multi-access aggregation.
- Results are based on Cisco Hybrid ICN™, an ICN within IP software addition to a virtualized IP router data plane.
- The HTTP adaptive video streaming application sits on top of a novel ICN-IP socket API which transparently provides support to web clients and servers.
- ICN-IP transport copes well with very fast handovers between different networks, while delivering high-quality video.
- The network is able to reduce redundant traffic by using (small) in-network memories and automatically aggregating redundant requests.
- The network adaptively selects the best network paths based on congestion, latency, and losses.

---

[32] Facebook live streaming: http://highscalability.com/blog/2016/6/27/how-facebook-live-streams-to-800000-simultaneous-viewers.html

- Best routes and server end-points are selected by smart network forwarding strategies.