Understanding Information Centric Networking and Mobile Edge Computing

December 2016

TABLE OF CONTENTS

1.0 INTRODUCTION

In the last decade, mobile data traffic has skyrocketed, a trend that's unlikely to subside in the foreseeable future. The use of the Internet, cloud and the ubiquity of the web as an application platform created an unprecedented increase in traffic due in part to the development of new web standards and the innovation in mobile wireless technologies. To meet this demand, the mobile industry is rethinking some fundamental aspects of network design.

With this growth in use, there's an increased amount of data being sent over multi hop public links. For example, the industry currently uses a client-server model for applications. The client typically resides on an end user device such as a smartphone or Internet of Things (IoT) module, while the server is hosted (resides) in a large data center using either private or public cloud infrastructure.

In order to protect the integrity and confidentiality of the online transactions, HTTP traffic can be secured using HTTPS transport layer security (TLS). HTTPS is great for e-commerce and banking, and most users understand the secure nature of HTTPS. However, using TLS and HTTPS for the majority of traffic has performance and functional drawbacks mainly because the HTTP session is encrypted as a whole. This makes optimization solutions hard to implement.

Modern web architecture involves sophisticated caching schemes that fetch various objects (e.g., images, libraries) from various locations in the path to avoid latency and improve the overall user experience while reducing bandwidth use. This is an important consideration where fast network infrastructure is not available.

Split browser architectures have pushed some of the processing from the client to the cloud. Content distribution networks (CDNs) are caching content on behalf of content owners.

New concepts such as cloudlets and FOG computing have also been developed. Cloudlets are components of the application functionality that are deployed in a more globally distributed fashion to reduce the end-to-end interactions between the client and the server, therefore reducing overall latency.

FOG computing takes a partial peer-to-peer approach to software design and addresses the latency reduction by pushing some of the server functionality to the client. This approach allows other clients, based on their network proximity, to avoid a full end-to-end round trip. These application development trends are multiplied with network virtualization and the migration to a software defined network (SDN).

Network features and functionality in 4G are increasingly implemented in software. In 5G, the non-software aspects are expected to be further minimized. The general assumption is that the network functions and applications in 5G will be designed as cloud native applications[1] to run in virtualized environments on massively distributed infrastructure. The infrastructure in which virtualized functions execute is generally referred to as the network function virtualization infrastructure (NFVI). The NFVI can be thought of as multiple, distributed cloud computing infrastructures providing computing storage and networking functions.

The infrastructure, network functions and applications, as well as end-to-end services, are all expected to be dynamically configured by software. Commercial drivers for services are expected to rapidly evolve in this context as 5G capabilities enable open innovation on a larger scale. Modern software development approaches emphasize agile methodologies to respond rapidly to these changing commercial priorities with

[1] See e.g., https://cncf.io/

iterations known as code sprint cycles (Figure 1). These processes are measured in weeks, thereby quickly producing deployable versions, albeit with constrained functionality.

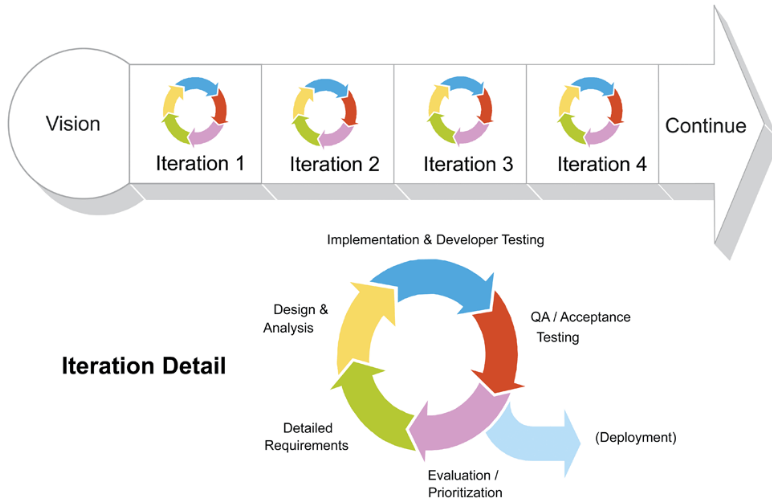

Figure 1: Agile Software Iteration cycles

A modern trend is to develop and or integrate some portion of the necessary software in the context of open source communities. OpenStack[2] is an example of a code development open source community supporting the evolution of a cloud computing management system. OPNFV[3] is an example of an open source community that is more focused on the integration aspects to ensure component interoperability (e.g., of SDN controllers and cloud managers).

Some of these emerging applications may require placement at the edge of the mobile or fixed infrastructure, such as in order to reduce the response time latency. Multi-tenancy aspects of the distributed NFVI may be particularly important for open innovation in services and for some regulatory requirements. For some of these applications, the content provides a particularly significant role, and content identifiers, rather than network addresses, may become the basis for routing, forwarding and storage decisions.

Whether deployed at the edge of the network or at other locations, the virtualized functions should be packaged with the same metadata for ingestion and lifecycle management. At the edge of the network, the NFVI must be adapted to support different I/O formats (e.g., radio or fiber interfaces) required for customer traffic access.

NFV and SDN also allow for rapid insertion of new routing techniques such as information centric networking (ICN), where network-layer functions are enriched with content awareness. This approach enables routing, forwarding, caching and data transfer operations to be performed on topology-independent content names rather than on IP addresses. ICN lends itself to 5G due to its inherent support for mobility and is discussed in detail in Section 3.

---

[2] www.openstack.org
[3] www.opnfv.org

5G may introduce new air interfaces. However, the promise of the next generation of mobile networks is as much about new architectures in support of diverse use cases from The Internet of Things (IoT) to high-definition video as it is about radio frequency (RF) parameters. The technologies discussed in this white paper support this diversity.

2.0 MOBILE EDGE COMPUTING

2.1 DESCRIPTION OF THE TECHNOLOGY

Mobile edge computing (MEC) is a form of edge computing that extends the client-server architecture by introducing an intermediary component. This intermediary component is located at the network edge to improve application responsiveness. Moreover, in wireless and mobile environments, it also helps in overcoming the harshness of the last hop radio link. There are currently various projects in the industry addressing this topic. In this paper, the term Mobile Edge Computing (MEC) refers to the general approach, and where a specific approach is used, it will be called out and explained explicitly.

The basic premise of Mobile Edge Computing (MEC), as well as edge cloud in general, is to place generic compute and storage close to the mobile network edge. This extends the cloud—typically a centralized, single resource—to the local environment. The MEC environment would be embedded in, and managed from, the operator environment. It is intended that the compute/storage resources will be exposed via a set of APIs such that application operators and developers can utilize their capabilities.

A loosely coupled approach to providing MEC benefits uses NFVI to distribute the mobile core to the network edge, which may be an aggregation node or even the base station/access point itself), and to collocate application hosting resources with the virtualized core instance. Current 4G standards support this approach.

This proximity to the user enables higher bandwidth and lower latency than would be possible in a centralized cloud environment, There are additional benefits, such as locally targeted service delivery through enhanced metadata (e.g., customer footfall in a retail environment).ETSI Multi-Access Edge Computing (ETSI MEC) – Exposing Radio Network Information Services (RNIS), as suggested in the ETSI-MEC work, provides a tighter coupling between the quality metrics available from the radio node (e.g. eNodeB) than is currently possible because this data is currently buried deep in the core network. This environment, whether created according to the ETSI MEC formulation or in a more loosely coupled way via NFVI, can enable entirely new service categories and capabilities:

- **Real-time:** enables real-time delivery of live and on-demand content. Provides robust low latency for critical communications.
- **Interactive:** Maximizes transaction rate between device and local application for a user-unique experience.
- **Analytics:** Allows real-time analytics at the point of capture, requiring minimum cloud ingress bandwidth.
- **Security and Privacy:** Localized communications to provide networks enhances performance, privacy and security.
- **Distributed:** Distributed computing capability for intense local tasks minimizes the impact on the network.

MCORD[4] is an ON.Lab research activity focused on Mobile Edge Platform for Central Office Re-architected as a Data Center (CORD). It extends and augments the use of the CORD with the corresponding XOS network operating system to mobility use cases such as augmented reality, IoT and caching. It also provides network slicing and EPC disaggregation with ONOS control.

Open Edge Computing[5] is an open source initiative based on the initial work from Carnegie Mellon University on cloudlets. The initiative is now focused on driving technology improvements and prototyping applications that leverage edge cloud computing. It also focuses on building a developer ecosystem and collaborating with other industry communities such as OPNFV and standards development organizations such as ETSI.

FOG computing is a decentralized computing infrastructure primarily focused on IoT use cases. The OpenFOG Consortium, comprising nine industrial and eight academic partners, was created in 2015 to promote the idea of FOG computing.

2.2 HOW EDGE COMPUTING ADDRESSES THE PROBLEMS

As networks evolve, mobile traffic expands and the Internet of Things (IoT) devices become more widespread, issues surrounding available bandwidth and the need for low latency become more critical.

The MEC server can address these issues by providing cloud computing capabilities closer to the end user than currently provided by traditional centralized cloud-based computing systems. Applications and analytics at the MEC server are not impacted by congestion in other parts of the network. By performing analytics or caching content at the MEC server, the volume of data transmitted to the core for processing is reduced, and the impact of the data traffic through the network is minimized. This results in a more efficient use of existing network bandwidth. This establishes an ultra-low-latency environment while providing real-time user and application flow information (e.g., subscriber location, local conditions. This information can then be used by applications and services hosted by the MEC service to offer location/context-related services to subscribers.

Because these applications and services are found at the edge instead of within a centralized cloud, responsiveness can be improved, resulting in an enriched quality of experience (QoE) for the user.

2.3 EXAMPLE USE CASES

2.3.1 IOT

IoT devices are expected to generate a significant amount of data as their use becomes ubiquitous. Design considerations intended to improve power/battery life of remote IoT devices will constrain their processing and memory capabilities.

MEC can address these issues presented by remote IoT devices by serving as an aggregation point while also handling additional analytics and data-logging functions. Because the data processing is not handled by the IoT device, its power consumption can be greatly reduced.

[4] http://opencord.org/
[5] http://openedgecomputing.org

Additional processing by the MEC can handle any decision logic based on the analytic results and then access end devices with instructions for any necessary actions. This approach still benefits from the low latency provided by aggregation at the MEC.

The MEC can also address the IoT device's limited memory storage by logging the data received from the end devices, aggregating the messages and logging them into a local database, to be distributed to downstream to operations/control centers for further collection and analysis. Additional security measures could be provided by having the MEC encrypt the data before being sent downstream.

2.3.2 VIDEO DISTRIBUTION/CONTENT CACHE

This use case focuses on improvements to over-the-top services by caching video or supplemental information at the MEC server to enhance the user experience. Live content and/or real-time information could be provided to subscribers without impacting the bandwidth available from the core to the access network.

Applications located on the MEC server would interface with smart devices to provide local object tracking and video content of interest to the user. Real-time information could be displayed quickly to registered users, without impacting other facilities deeper in the access network. Placing the MEC server at the edge minimizes round trip time at the same time maximizing throughput, which provides a better QoE for the user.

2.3.3 GENERIC COMPUTE RESOURCE

This use case focuses on cloud-based APIs stored within the MEC, available in a facility with a large user segment such as an arena, stadium or airport.

Connection to the MEC application from existing apps found on the user's smart device would be able to provide specific location information for the user. The information provided could manage foot traffic flow in an efficient manner by identifying safety, security or evacuation information, highlighting nearby facility services (e.g., dining, shopping) or providing directions and travel time within the facility.

A specific example would be in-transit passengers arriving at an airport that they're not familiar with. Upon reaching the arrival gate on an inbound flight, the user would activate the MEC application, which would connect to the user's smart device and identify connecting flight information. The MEC application would then contact the airline to get the connecting flight's departure gate and determine the quickest directions and time required to reach that gate. Passengers would then be guided to their gate using augmented reality, or via a facility-based map service. Other information provided could be dining or shopping options, or other general facility information such as baggage claim, restrooms, storm shelters or ticketing.

Placing the MEC server at the access network would leverage and enhance the value of local content and be perfectly adapted to the applications target users.

2.3.4 EDGE ANALYTICS

With the expected increase in remote-sensing devices, it becomes possible to make local decisions based on the data received from these devices. An analysis can be undertaken based on the input from numerous sensors and then be used to enhance local control systems.

An example is a traffic intersection. Currently traffic flow is set up based on historical time-of-day traffic patterns with a set amount of time for each direction to proceed. This approach does not always result in optimized traffic flow as local conditions/traffic patterns may fluctuate.

MEC could improve the efficiency of the traffic flow by analyzing the data from several sensors (e.g., connected cars, crosswalk occupancy sensors) local to the intersection and instruct the signal controller to change the traffic light pattern based on the analysis of the MEC. It also could possibly meet preset conditions for changing wait (red light) times at that location. By undertaking this analysis at the MEC instead of a downstream location, latency and reliability of the analysis and resulting actions can be significantly improved.

2.3.5 CELL LEVEL PERFORMANCE OPTIMIZATION USE CASE

This use case focuses on optimization of radio performance of a cell or a group of cells (Figure 2), using information from the ETSI MEC ISG work effort.

This ETSI MEC ISG application (for more information on ETSI MEC ISG, see Section 2.4.1) connects to the platform's interface to receive real-time information about radio and protocol performance of individual subscribers, such as user location, signal measurements and data rate. This information will be delivered to the centralized optimization function and used for detection and resolution of radio performance problems such as inter-cell interference, cell congestion and poor coverage.

Rationale for proposing edge compute centers around real-time network data collection to improve performance of third-party optimization tools (SON).

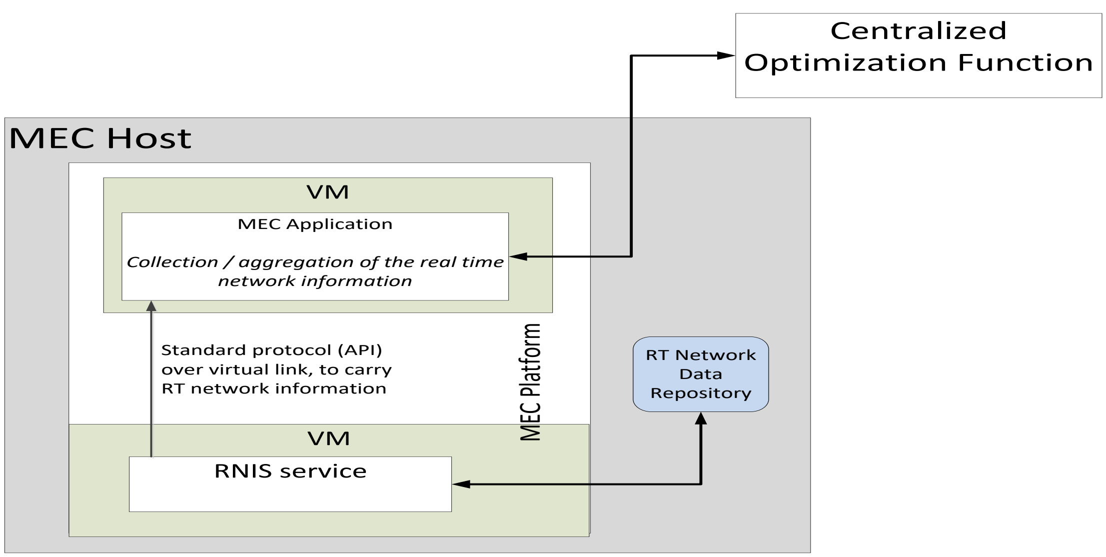

Figure 2: Cell level performance optimization use case

2.3.6 LATENCY-SENSITIVE APPLICATIONS

Latency is one of several parameters that are important system requirements. It directly affects the user experience on mobile wireless networks. Low latency and ultra-high reliability are critical for applications such as:

- Automatic driving, traffic control and V2X.
- VR applications.
- Mission-critical use cases such as public safety communications.
- Remote health care (g., remote surgery).
- Extreme real-time applications such as tactile internet.
- Real-time HD video sharing
- Industrial and manufacturing applications that require real-time remote control and operations (e.g., robotic controls).

The NGMN white paper[6] stipulates that 5G systems should support E2E latency (latency perceived by the user) of 10 ms in general and 1 ms for extremely low-latency applications. The NGMN white paper (section 4.1.5) also provides user experience KPIs requirements for a variety of latency-sensitive use cases.

E2E latency is affected by a host of factors, including application processing, RAN, core network and application servers. In the context of this white paper, the focus is on the latency between the user and the network, specifically the network/cloud edge. Using MEC and related technologies can help reduce this latency and help meet E2E latency requirements of the above use cases.

2.4 MOBILE EDGE COMPUTING – STANDARDIZATION, RESEARCH AND INDUSTRY COLLABORATION OVERVIEW

There are edge computing-related activities in progress in numerous academic, industry and standardization fora. The following shortlist highlights a sampling of these.

2.4.1 ETSI

The ETSI Mobile Edge Computing Industry Specification Group (MEC ISG) was formed in December 2014. Its purpose is “to create a standardized, open environment which will allow the efficient and seamless integration of applications from vendors, service providers, and third-parties across multi-vendor Mobile-edge Computing platforms.”[7]

Through July 2016, the MEC ISG has published the following specifications:

- Foundation specification GS MEC 001 –MEC Terminology (2016-03)

[6] http://ngmn.org/fileadmin/ngmn/content/downloads/Technical/2015/NGMN_5G_White_Paper_V1_0.pdf
[7] http://www.etsi.org/technologies-clusters/technologies/mobile-edge-computing

- Foundation specification GS MEC 002 –MEC; Technical Requirements (2016-03)
- Foundation specification GS MEC 003 –MEC; Framework and Reference Architecture (2016-03)
- GS MEC-IEG 004 –MEC; Service Scenarios (2015-11)
- GS MEC-IEG 005 –MEC; Proof of Concept Framework (2015-08)

In addition to these published specifications, the MEC ISG is currently working on an additional set of 16 specifications that includes a series of MEC API specifications.

The ISG recently received a two-year extension in September 2016 (beginning in March 2017). The ISG has modified its terms of reference to include the investigation of the use of edge computing beyond a “mobile-only” access environment, such as areas with Wi-Fi and fixed access technologies. As a result, the ISG will be changing its name from “Mobile Edge Computing” to “Multi-Access Edge Computing” to reflect the change in the terms of reference, while still maintaining the widely recognized “MEC” acronym.

2.4.2 3GPP

3GPP has recently published their work plan[8] toward 5G, and more specifically toward meeting IMT-2020 requirements and proposal deadlines. Figure 3 outlines 3GPP’s work plan. The first part includes study items and development of technical reports on the scenarios and requirements that will drive the 5G New Radio (NR) architecture and interfaces. 3GPP TR 38.913 captures the current view on scenarios and requirements for 5G NR, while 3GPP TR 38.801 has begun to capture some of the NR architecture and interfaces aspects for 5G.

[8] http://www.3gpp.org/news-events/3gpp-news/1787-ontrack_5g

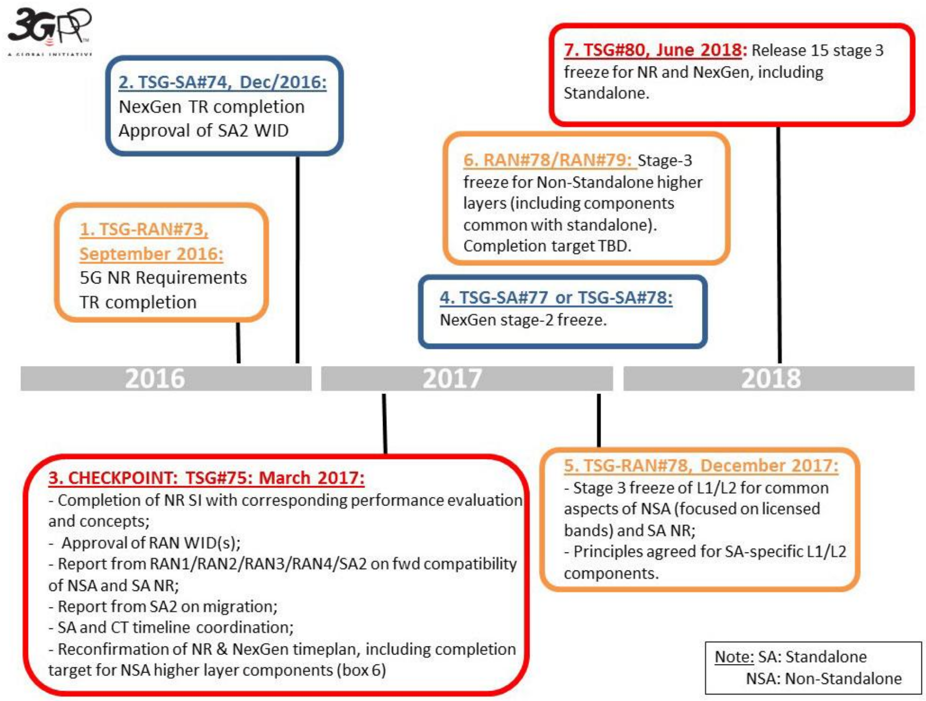

Figure 3: 3GPP work plan for 5G

TR 38.913[9] contains some clear requirements from 3GPP about the 5G NR architecture that promote the evolution to architectures that can leverage the benefits of edge computing. Specifically:

- Different options and flexibility for splitting the RAN architecture shall be allowed.
- The RAN architecture shall allow for deployment flexibility, such as to host relevant RAN, CN and application functions close together at the edges of the network when needed, such as to enable context-aware service delivery and low-latency services.

These requirements have driven 3GPP to examine various 5G NR architecture options as shown in Figure 4. This figure shows that 3GPP will be looking at pretty much all options for splitting the RAN functions so certain groups of RAN functions can be deployed in a more distributed manner. They could potentially be combined with certain CN and application function that can then take advantage of the benefits of edge computing technology.

[9] 3GPP TR 38.913, "Study on Scenarios and Requirements for Next Generation Access Technologies", Rel-14, V0.3.0, March 2016.

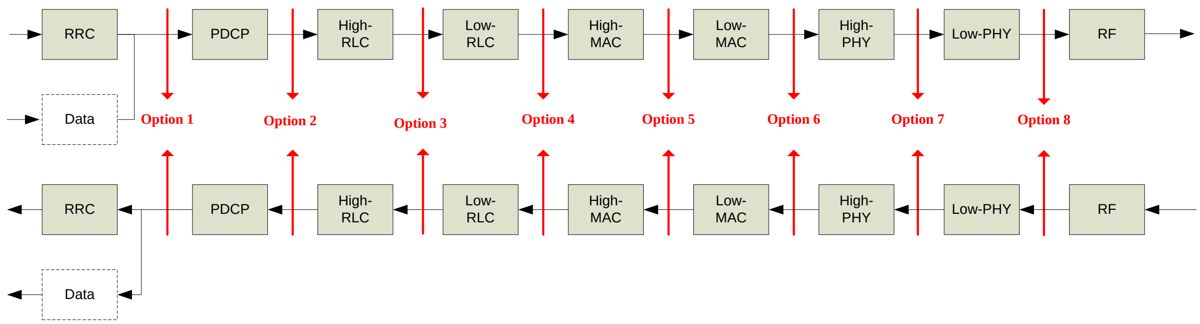

Figure 4: 3GPP study on 5G NR RAN architecture options

2.4.3 OTHER MOBILE EDGE COMPUTING OPEN SOURCE INITIATIVES

Started a few years ago as an academic initiative out of Carnegie Mellon to promote the concepts of cloudlets as described in the initial paper[10], Elijah[11] was the first attempt to demonstrate edge computing using OpenStack and by leveraging open source to bring the cloud closer to the mobile edge.

The Open Edge Computing project[12] was created in 2015 to facilitate prototyping applications that can take advantage of edge computing, building an ecosystem and engaging with relevant development communities. The project is expected to synchronize work with other efforts in the standards and open space, including ETSI ISG MEC and OPNFV.

2.5 DEPLOYMENT STRATEGIES

MEC supports multiple deployment scenarios, including at a multi-radio access technology (RAT) cell aggregation site, at an aggregation point, at the cloud RAN and at the edge of the core network. This flexibility allows placement where it can provide the greatest benefit to the network.

3.0 ICN

3.1 DESCRIPTION OF THE TECHNOLOGY

The basic idea of ICN is to enrich network-layer functions with content awareness so that routing, forwarding, caching and data-transfer operations are performed on topology-independent content names rather than on IP addresses. Data is divided into a sequence of chunks uniquely identified by a name and permanently stored in one or more servers. Naming data chunks allows the ICN network to directly interpret and treat content per its semantics without the need for deep packet inspection (DPI) or delegation to the application layer.

The naming convention is application-specific and not defined by the protocol architecture. The architecture simply requires the name to have a hierarchical structure, like that HTTP already uses. More precisely, a

10 http://elijah.cs.cmu.edu/DOCS/satya-ieeepvc-cloudlets-2009.pdf
11 http://elijah.cs.cmu.edu/
12 http://openedgecomputing.org/

name is composed by a variable number of components (not necessary human-readable as URIs), organized in a hierarchical structure. The use of hierarchical names facilitates scalability by allowing prefix aggregation in name-based routing tables.

Authenticity - One of the most important differences between name-centric and traditional host-centric networking is that data is retrieved by name rather than location. Hence, in ICN architectures, the data authenticity verification (i.e., the verification of the publisher of a named data object) is an important challenge. Data authenticity is achieved by applying a digital signature (hash of name, plus data object through publisher's key) to named data object with hierarchical naming scheme.

Update and versioning - In ICN, routable data object names are globally unique. Hence, updating an object or creating a new version of an object corresponds to the creation of a new object. With the hierarchical naming schema, a component of the data object's name can be considered as its version.

Name encoding - As previously mentioned, ICN names can be potentially unbounded. An important challenge is to define an efficient name-encoding scheme to achieve: i) fast name parsing ii) reduction of the space needed for carrying the name in ICN packets and iii) flexibility. Encoding proposals satisfying such requirements exist and leverage type length value (TLV) encoding with component offset encoding, highly flexible, compact and faster to parse.

Name-based forwarding and routing - Named data networking[13] (NDN) network routers process user requests (interests) by name in a hop-by-hop fashion toward a permanent copy of the content. To this goal, every router has a name-based routing table storing one or more potential next hops towards a set of content items. A set of dynamic forwarding algorithms has been developed to select a given next hop per specific metrics (e.g., time-monitored delay) and with the objective of achieving optimal throughput while minimizing network cost. NDN routers also keep track of received Interests to return content chunks to the user following the reverse request path (symmetric routing).

The content delivery process is driven by three basic communication mechanisms: name-based request routing, pull-based connectionless transport and in-network caching.

Upon reception of a request packet from an input interface, intermediate nodes perform the following operations:

1. **Cache lookup** to check if the requested data chunk is locally stored. In case of a cache hit, the data is sent through the interface the request is coming from.
2. **Pending interest table (PIT) lookup** to verify the existence of a pending request for the same data chunk. If yes, the interest is discarded (and the interface through which the interest arrived is added to a list of interfaces awaiting that data chunk) because a pending query is already outstanding. If not, a new PIT entry is created for the interest, listing the interface on which it arrived, and the interest is forwarded toward the content according to the forwarding information base (FIB (see 3, below).
3. **Forwarding Information Base (FIB) lookup** via longest prefix matching returns the interface where to forward the interest (selected among the possible ones).

[13] Named Data Networking and Content Centric Networking (CCN) are two representative flavors of ICN, and are used interchangeably with ICN in this document.

FIB entries are associated to name prefixes. Data may come from a server, or from any intermediate cache along the path with a temporary copy of the data packet. Forwarding operations are illustrated in Figure 5.

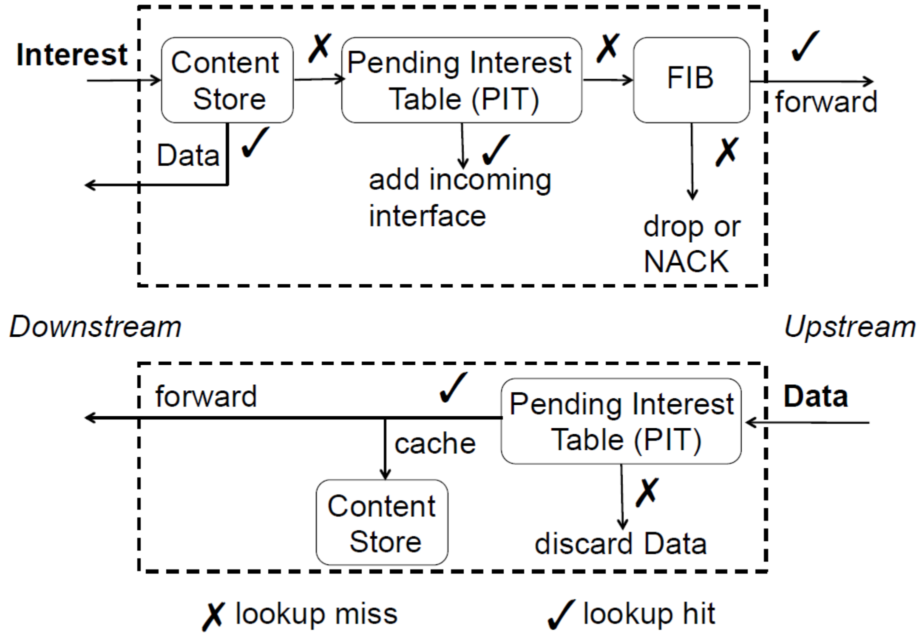

Figure 5: NDN Forwarding Engine

3.2 HOW ICN ADDRESSES THE PROBLEMS OF TRADITIONAL IP NETWORKS

3.2.1 PULL-BASED CONNECTIONLESS TRANSPORT

In contrast with the current sender-based TCP/IP model, the NDN data transfer process is triggered by user requests addressed to chunks of the requested content item (i.e., pull-based model). Rate and congestion control is performed at the user by means of a connectionless yet stateful transport protocol with the following characteristics:

- No connection instantiation and support for user/content mobility.
- Support for retrieval from multiple sources, a priori unknown at the user (e.g., intermediate caches).
- Support for multipath communication to improve user performance and traffic load balancing.

NDN transport protocol is proven to optimally allocate bandwidth resources among users in a fair and efficient fashion, jointly with request routing.

3.2.2 MOBILITY MODEL

In the ICN architecture, interfaces do not have network addresses, so a change in physical location does not imply a change of address in the data plane. Support for consumer mobility emerges naturally from the architecture because of the connectionless, symmetric transport model. With ICN's pull-based communication model, the consumer expresses interest packets that are routed in the network toward the data. The data is returned toward the client following the paths traversed by the interests. In a case of a move before data in flight is received, the consumer may simply re-express the interests for those data objects. The network may now be able to fetch it from local caches filled by the data in flight.

Producer mobility and real-time group communication are more challenging to support, depending on the frequency of the mobility and on the content lifetime. Again, the basic interest/data exchange mechanisms provide a means to rapidly update local routing tables to ensure continued reachability of mobile content. The distributed in-network caching of ICN allows to smooth handoffs and to prevent service quality degradation

3.2.3 STORAGE MODEL

Network nodes temporarily store content items in order to serve future requests for the same content.

Whenever an interest is received at a NDN node, it first checks if the requested chunk is present in the local cache. If so, the content is returned to the user. Otherwise, the request is forwarded to the next hop by the NDN request routing.

In-network caching allows the network to exploit current buffers in routers, possibly enhanced by additional memory blocks.

The content-awareness provided by names to network nodes enables a different use of buffers, not only to absorb input/output rate unbalance but for temporary caching of in-transit data packets. Even without additional storage capabilities in routers, the information access by name of NDN allows two new uses of in-network wire-speed storage:

- **Reuse**: Subsequent requests for the same data can be served locally with no need to fetch data from the original server/repository.
- **Repair**: Packet losses can be recovered in the network, with no need for the sender to identify and retransmit the lost packet.

Simple cache management policies and coordination techniques allow an efficient allocation of distributed in-network storage resources at very low computational overhead and without requiring the complex management of today's CDN.

The presence of distributed in-network storage and of name-based lookup automatically distributes copies of popular content closer to the users as demand materializes.

3.2.4 SECURITY MODEL

Current internet security is made available by means of ad-hoc protocol extensions such as DNSsec, IPsec and TLS. TLS provides web security by encrypting a layer 4 connection between two hosts. Authenticity is provided by the web of trust (certification authorities and a public key infrastructure) to authenticate the web server and symmetric cypher on the two end points based on a negotiated key. In presence of TLS, many networking operations become unfeasible, including filtering, caching, acceleration and transcoding.

The NDN security model is radically different. Instead of securing by encrypting simply connections, the NDN object-security model allows the separation of security actions regarding privacy, data integrity and data confidentiality, all of which leverage an existing web of trust based on certification authorities and a public key infrastructure. The security actions are performed directly at network layer with content identification provided in data names. All data is integrity protected, whereas confidentiality (via data encryption) is optional. Integrity protection guarantees the authenticity of the data bound to the name by including the producer signature of the data plus its name.

The atomic security service provided by NDN guarantees that the producer has published a piece of data with the name available in the packet. This service enables location-independent secured content access. Denial-of-service attacks based on cache poisoning can be blocked using signature verification techniques. However, the cost is not negligible, and some recent work has started to build network layer trust management that does not required in-network signature verification by using the concept of interest-key binding.

The NDN security framework enables content networking, several services require redesign. Access control, for instance, requires managing and distributing keys to the group of users with granted access to the controlled data. Also, content revocation requires data version management and policy enforcement to delete obsolete content from the network when needed.

3.3 MOBILE NETWORK USE CASES

3.3.1 IOT

The range of IoT devices and applications is vast: from periodic updates where small measurements are sent once every day to haptic feedback between an actuator and robot arm for medical procedures.

Optimizing the last-mile connection where over 51 billion devices will be connected requires that more efficient use of the air interface is made, duplicate data is removed and transport-layer addresses are minimized.

IoT is ideally suited for a name-based protocol such as ICN. Sensor names can translate directly to ICN URIs, and use of caches at the ICN forwarders can remove duplicate data.

Also, the self-learning nature of ICN through the flooding of interests to locate a device means that the current issues of scaling routing protocols can be alleviated. Mobility is inherent in ICN, as is support for embedded security.

3.3.2 MEDIA DISTRIBUTION

5G holds much promise in becoming a ubiquitous access and core technology to support massive media up- and downloading, as well as IoT, not in the least through 5G's projected improvements in communication capacities and latency. 5G will be compatible with technologies such as LTE and its evolution, and it will coexist well with Wi-Fi and other radio access technologies.

The 5G ecosystem is expected to provide even more affordable and sustainable mobile wireless technologies for anything from massively deployed machine-type communications to video distribution toward 4K TV sets and beyond. 5G is projected to become a technology to support massive video downstreaming.

ICN in 5G could be enabled by the following solutions:

- **ICN connectivity all the way to the end user devices:** In order to unleash ICN's full capabilities, potentially leading to the paradigm shifts, it is expected to have to go all the way to the end user devices and be visible to the applications. This doesn't exclude hybrid scenarios, particularly in the initial phases of ICN deployment, but stresses an important target with ICN value creation. With 5G, and indeed other 3GPP technologies, running ICN natively over the air on, for example, a separate radio data bearer, is not specifically challenging technically. The technical challenge is comparable to having parallel connections between network and UE for best effort IP traffic and VoLTE.

- **Parallel support for ICN and IP networking over 5G:** The 3GPP bearer concept is very flexible and a powerful tool to control connectivity and its quality E2E. The flexibility means, for example, that ICN at the 5G base station can break out the user plane ICN frames from a UE and feed those to an ICN network that is logically separate from the IP-based mobile backhaul network. A similar process works in the opposite direction for downlink traffic toward the UE. This should not be construed to necessarily mean that there is a need for parallel physical networks in future mobile backhaul solutions. Indeed, it is envisioned that both ICN and IP based mobile backhauling operate over the same physical infrastructure, presumably based on Ethernet switching enhanced with both IP routing and ICN forwarding capabilities. In the age of SDN, it isn't foreseeable to expect the need to switch new types of Ether types or frames to and from radio base stations as a show-stopper for implementing sustainable and affordable media-optimized solutions.
  ICN as a separate 5G network slice: From a core perspective, one can view ICN connectivity as a separate slice, such as separate ICN frames as discussed in the previous paragraphs. The control of the ICN slice(s) will still be under the jurisdiction of the EPC control plane, but the ICN frames will receive special treatment compared to IP packets, as discussed above.

3.4 ICN - STANDARDIZATION, RESEARCH AND INDUSTRY COLLABORATION OVERVIEW

There are ICN-related activities in progress in numerous academic, industry and standardization fora. The following shortlist highlights a sampling of these.

3.4.1 3GPP

3GPP recently started the NextGen study item, which aims at defining the 5G system architecture.

One goal for 5G is to define a generic architecture allowing for various types of payload to be transported via the 5G radio, including IP, transparent data (referred to as non-IP traffic) and potentially also Ethernet frames.

ICN is not yet being explicitly addressed as a dedicated payload type in the 3GPP system. This is partly due to the fact that the related work on defining the ICN protocol and the related mechanisms in IRTF and IETF are still ongoing.

However, many of the system attributes required for efficient operation of ICN over a mobile network are not specific to ICN. The key enabler for many ICN use cases is the ability to deploy ICN routers close to the radio network. From a 3GPP system perspective this translates into terminating sessions very close to the radio network (e.g., by selecting the user plane entities for a session closer to the base stations serving a given subscriber).

Similar functionality is also required to enable mobile-edge computing scenarios (where applications are hosted closer to the RAN) and also as an enabler for very low latency scenarios. 3GPP's NextGen study is therefore addressing those requirements by studying efficient user-plane architectures, which allow for flexible allocation (and relocation, as needed) of the 5G system's user-plane entities.

Once the IETF has progressed the definition of the ICN protocol and related mechanisms, further 3GPP system enhancement supporting ICN traffic as a specific payload type in the 3GPP user plane can be envisioned. This may include, for example, broadcast and multicast support for ICN traffic.

In addition, future architectures may also go beyond supporting ICN as a pure payload type. This could include leveraging some of ICN's inherent features to address mobile network specific challenges, such as handover support and paging.

3.4.2 IETF, IRTF

In April 2012, the Information-Centric Research Group (ICNRG) was formed to identify outstanding research challenges for ICN, and to couple ongoing ICN research with solutions that are relevant and appropriate for evolving the Internet at large. Research challenges include:

- Naming schemes for ICN.
- Scalable routing schemes.
- Congestion control, QoS approaches and caching strategies.
- Metrics that make it possible to consistently evaluate ICN implementations.
- Security and privacy.
- Application/application-protocol design and APIs.
- Business, legal and regulatory frameworks.

The work in the ICNRG (in the form of Informational RFCs, meeting contributions, etc.) is completely documented and accessible through its website at https://trac.tools.ietf.org/group/irtf/trac/wiki/icnrg.

In July 2016, an ICNRG study team was assembled to address harmonization of the various ICN approaches (the CCN and NDN implementations). This work is currently underway and is viewed as laying the groundwork to be carried forward by an IETF working group in future.

3.4.3 ITU-T

The IMT-2020 focus group was created to study how emerging 5G technologies will interact in future networks. The group's final output included studies on high-level network architecture, an E2E QoS framework, emerging network technologies, mobile front haul and back haul, and SDN. The group created a report[14] on "standards gaps" in the five focus areas.

In December 2015, the group received an extension with new terms of reference to engage open-source communities, influencing and taking advantage of their work by introducing them to the challenges that telecoms players must overcome in the development of the 5G ecosystem. One of the specific tasks will be to enhance aspects of SDN and ICN.

There currently is not sufficient participation to produce additional documents beyond those from the proof of concepts (PoCs). Therefore, there are no regular conference calls for ICN.

There is interest from Q15 Data Aware Networks to get input from the IMT-2020 focus group on Y.DAN-reg-arch, which is currently under development. Based on its needs, the group may produce specific input for that document.

The group will begin work on consolidating the IRTF standards document in an ITU-T document for the focus group for possible consideration by study group 13 (SG13) next year if there is no progress in forming an IETF work group.

Currently, this effort is not a direct standardization of ICN, but it could result in the development of a question to study group 13. The main objective is to provide an ICN PoC at the fall focus group meeting in Geneva, Switzerland.

3.4.4 ATIS

The ATIS board initiated an ad-hoc committee to investigate the Evolution to Content Optimized Networks (eCON). Although not exclusively focused on ICN, most work addressed the overall ICN opportunity space from a network operator's perspective. The group outlined the opportunity for content optimized network technologies including:

- Linear TV / IPTV Replacement.
- Efficient and automated edge caching for internet content.
- Seamless "session" management and continuity across different access types (e.g., a device moves to/from a private/unlicensed network node to a licensed network).
- Increased access efficiency through integral multi-path support.
- New internet broadcast/multicast services (cost-effective Internet-based broadcast/multicast).
- Application of object-based communication models for IoT devices.
- New ways of more effectively managing security and information related to security.

[14] http://www.itu.int/en/ITU-T/focusgroups/imt-2020/Documents/T13-SG13-151130-TD-PLEN-0208!!MSW-E.docx

3.4.5 5G AMERICAS

5G Americas' efforts are not a direct standardization effort, but rather a standards influencer by developing a 5G white paper.[15] The paper describes in some detail the ICN architecture, along with ICN benefits and use cases.

ICN is presented as a potential technology for consideration as 5G. The paper suggests that 5G should be based on new network architectures and protocols designed specifically with support for mobility, security and content caching as fundamental design criteria. ICN as realized in the named-data networking (NDN) and content-centric networking (CCNx) programs is described as a leading architecture that can meet such design criteria.

3.4.6 NGMN

NGMN published a 5G white paper[16] in March 2015. The paper provides requirements for 5G and encourages the adoption of new emerging technologies. ICN is one of the technology building blocks considered by NGMN. ICN is described as having the potential to migrate from a host-centric and node-centric model to a content-centric, data-oriented and information-centric model with an intrinsic focus on named information objects in-network caching and name-based routing.

3.4.7 OTHER RESEARCH PROJECTS

Over the past decade, there have been numerous nationally funded research efforts to address various research topics in ICN. In the U.S., these programs fell largely under the NSF Future Internet Architecture initiative (e.g., CCN, NDN, Mobility First, Xia). In Europe, they're under the EU Framework programs (through H2020).

4.0 RECOMMENDATIONS AND CONCLUSIONS

MEC and ICN are complementary concepts, yet they are by no means mutually inter-dependent. Both MEC and ICN solutions may be deployed independently of one another, and both could add value. For example, a 5G mobile network may make use of ICN for content distribution, may take advantage of ICN for transparent mobility among multiple technologies and leverage ICN's ability to retransmit lost packets over an unreliable radio link. None of these capabilities requires MEC. Similarly, one might envision MEC employed to reduce latency for Virtual/Augmented Reality (AR/VR) applications, or to perform distributed data-reduction and security functions for an IoT network, without requiring the use of ICN. However, there are certainly synergies that can be exploited when the two technologies are deployed cooperatively.

To realize the full benefits of ICN, an end-to-end ICN framework is desirable. However, it is not necessary to discard and/or replace existing networks. This framework can be implemented as an overlay on existing IP networks (ICN over IP). It can also be implemented using IPv6 addresses as content object names (ICN within IP), which would require a dual forwarding capability in the network routers. And as has been stated in section 3 of this document, ICN can be deployed within a 5G network slice. Until applications are ICN aware, protocols such as HTTP can be supported transparently to existing applications via the use of

[15] http://www.5gamericas.org/files/2414/4431/9312/4G_Americas_5G_Technology_Evolution_Recommendations_-_10.5.15_2.pdf
[16] http://ngmn.org/fileadmin/ngmn/content/downloads/Technical/2015/NGMN_5G_White_Paper_V1_0.pdf

appropriate proxies, making it possible to evolve applications gracefully from today's networks to ICN, and reaping greater benefits with each successive step.

ACKNOWLEDGEMENTS

The mission of 5G Americas is to advocate for and foster the advancement and full capabilities of LTE wireless technology and its evolution beyond to 5G, throughout the ecosystem's networks, services, applications and wirelessly connected devices in the Americas. 5G Americas' Board of Governors members include América Móvil, AT&T, Cable & Wireless, Cisco, CommScope, Entel, Ericsson, HPE, Intel, Kathrein, Mitel, Nokia, Qualcomm, Sprint, T-Mobile US, Inc. and Telefónica.

5G Americas would like to recognize the significant project leadership and important contributions of Michael Recchione of Cisco as well as representatives from the other member companies on the development of this white paper.

APPENDIX A: ACRONYM LIST

Abbreviation	Meaning
3GPP	3rd Generation Partnership Project
4G	4th Generation
5G	5th Generation
API	Application Program Interface
ATIS	Alliance for Telecommunications Industry Solutions
CDN	Content Distribution Network
CN	Core Network
CORD	Central Office Re-architected as a Data center
DNS	Domain Name System
DPI	Deep Packet Inspection
E2E	End-to-End
eCON	evolution to Content Optimized Network
EPC	Evolved Packet Core

ETSI	European Telecommunications Standards institute
ETSI MEC	ETSI Multi-access Edge Computing
FIB	Forwarding Information Base
GS	Group Specification
HD	High Definition
HTTP	Hyper Text Transfer Protocol
ICN	Information Centric Networking
ICNRG	ICN Research Group
IETF	Internet Engineering Task Force
IMT	International Mobile Telecommunication
I/O	Input/Output
IoT	Internet of Things
IP	Internet Protocol
IPTV	IP Television
IRTF	Internet Research Task Force
ISG	Industry Specification Group
ITU	International Telecommunication Union
ITU-T	ITU Telecommunication Standardization Sector
KPI	Key Performance Indicator
L1	Layer 1
L2	Layer 2
LTE	Long Term Evolution
MCORD	Mobile edge platform for CORD
MEC	Mobile Edge Computing
NDN	Named Data Networking
NFV	Network Function Virtualization
NFVI	Network Function Virtualization Infrastructure
NGMN	Next Generation Mobile Networks
NR	New Radio
NSA	Non-Stand Alone
ONOS	Open Network Operating System
OPNFV	Open Platform for NFV
PIT	Pending Interest Table
PoC	Proof of Concept
QoS	Quality of Service

RAN	Radio Access Network
RAT	Radio Access Technology
RFC	Request for Comments
RNIS	Radio Network Information Services
SA	System Architecture or Stand Alone
SDN	Software Defined Network
SON	Self Optimizing/Organizing Network
TCP	Transmission Control Protocol
TLS	Transport Layer Security
TLV	Type Length Value
TR	Technical Report
TSG	Technical Specification Group
UE	User Equipment
URI	Uniform Resource Identifier
V2X	Vehicle to anything
VM	Virtual Machine
VoLTE	Voice over LTE
VR	Virtual Reality
WID	Work Item Description