INSTITUTE FOR COMMUNICATION SYSTEMS

5G INNOVATION CENTRE

5G Whitepaper:
The Flat Distributed Cloud (FDC) 5G Architecture Revolution

January 2016

Introduction

This 5GIC whitepaper proposes a disruptive change in architecture for next generation cellular networking that enables a user experience that is perceived as always sufficient for their current context. In order to meet this perception the network is designed so as to always make best use of the resources available at the time of each new communications request applicable to the context of the user at the time.

Traditional networks make use of network configuration, bearer and QoS information to satisfy user requests, but the FDC architecture additionally employs user and network context information such as where, when, why who and what is being requested as well as the users location and location type (home, work, out-and-about) to service requests. The network is also able to make use of learned intelligence gathered from these additional resources both at the device and in the network.

The FDC network will provide communications connection using both fixed and wireless bearers where available and will enable interconnection with traditional wireline, internet, cloud and new content distribution networks.

The vision is of a more connected experience over a dynamic and distributed cloud based architecture that separates the user and control planes and is flatter than the Long Term Evolution network (LTE), further reducing the network hierarchy from 3 layers today to 1 for most cases and two for an estimated 30% of cases.

Amongst the design objectives is a more Context Aware (CA) network that can identify and predict popular content, collate group content and according to mobility context, get the user data/ connection ready 'just in time' by harvesting user profile information that is communicated between user and service and/or network provider.

The proposed architecture provides native support for the Internet of Everything (IoE), in an Internet of Things (IoT) class based manner and proposes enhancing legacy mobile IoT support to add advanced SCADA-like (Supervisory Control and Data Acquisition) control system capabilities to its mobile IoT capabilities by significantly enhancing latency response to enable support for millisecond control loop response time operation over 5G (drones and auto applications).

The network is designed to employ the best of evolving Network Function Virtualisation (NFV) and Software Defined Networking (SDN) implementations and features to address known shortcomings of today's 3GPP based architectures and is able to inter-connect with and support different Radio Access Network (RAN) deployment configurations such as Cloud –RAN (C-RAN), traditional Distributed-RAN (D-RAN) or hybrids of both (H-RAN) according to available transmission options.

Connected World 2015

Communications

Globally, over the last 25 years telecommunications traffic has grown by an astounding amount.

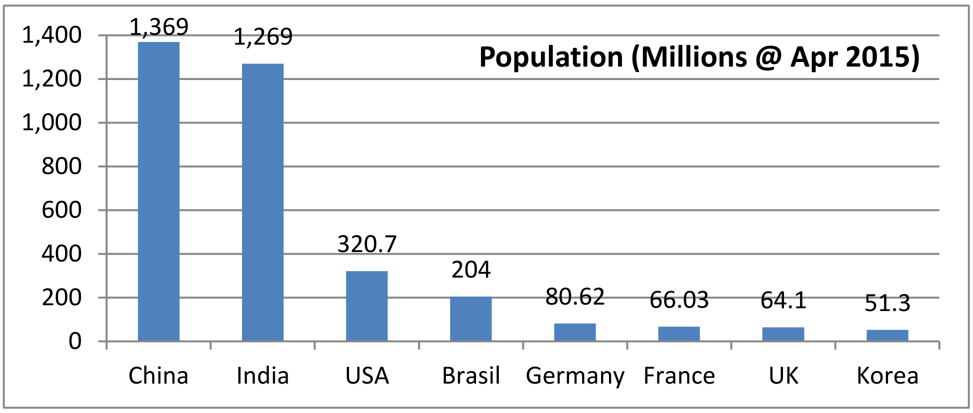

Figure 1: The world population by selected countries (as of 2015) (for sources see Ref-07)

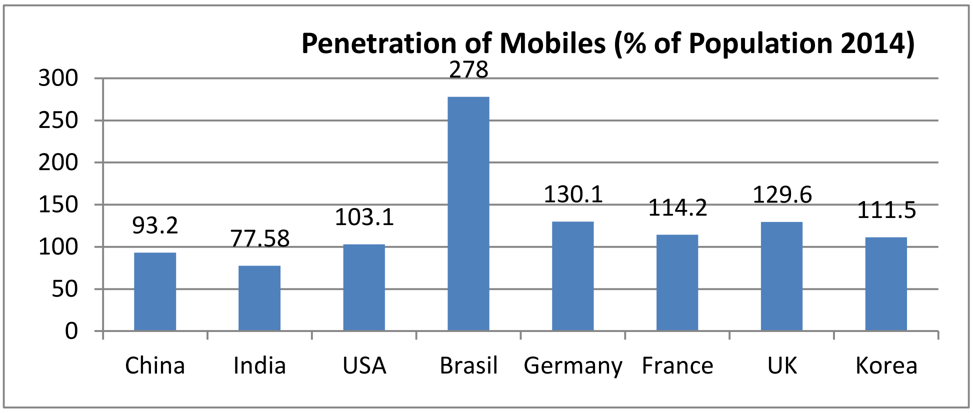

Figure 2: The mobile penetration (for sources see Ref-08)

The population growth and corresponding increase in mobile device penetration as depicted in fig. 1, 2, indicate that mobile phone usage has grown from almost zero to over 100% penetration across most of the world regions, and in approximately half of this same time period, broadband support in the home has grown from almost zero to the stage where mean download rate for home owners for >50% of the population of the world, as shown in Figure 3.

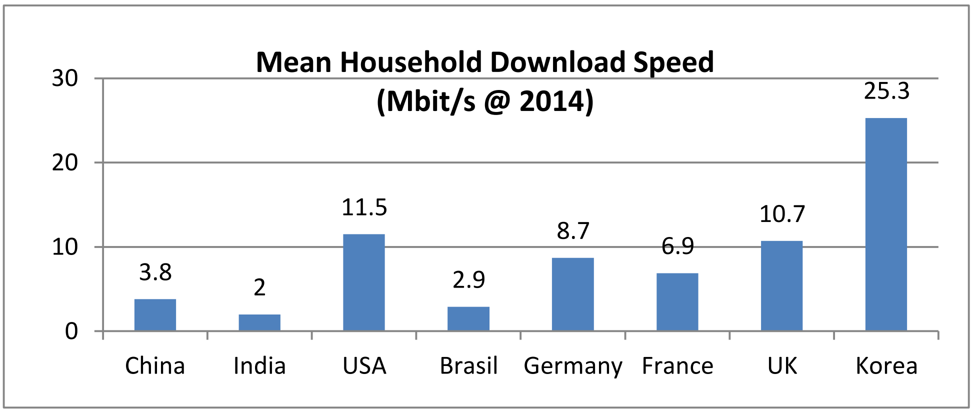

Figure 3: Mean household download speeds (for sources see Ref-09)

Home broadband subscriptions include an approximate distribution as follows: 20% Fibre, 20% Cable, 80% Copper. Leading global fixed broadband providers estimate that between 1 and 5% of the homes in the world being connected to fixed broadband every year, so that by 2020 even more of the world population will have access to home broadband.

World mobile broadband usage (by all types of subscription) is growing much faster than fixed at a CAGR of 20% [Ref-09]. Mobile broadband has taken longer to take-off but is also is expected to near 100% penetration by user mobile devices by 2020 [The Broadband Commission 2014/ITU].

In Europe today the average consumption of data per month per domestic residence is 20Gbytes per month and rising, whereas mobile connections typically download between 500Mbytes and 1Gbyte per month as is borne out by UK OFCOM statistics and many other similar reports from mobile operators. However, these download volumes are subscription limited not usage limited, so that if subscriptions change (e.g. more generous data caps) most users would download more. Cisco reported that globally, the data volumes over mobile connections grew 69% in 2014. The same report predicts that in 2020 mobile broadband downloads could exceed 20Gbyte per month per subscriber [Ref-01].

In, summary the statistics presented for the connected world 2015 indicate that across the globe:

1) Population: Many countries are already saturated with mobile coverage, and even the largest markets are nearing saturation independent of continent.
2) Wired broadband coverage is largely driven by direct or indirect government influence on the local telecommunications industry so that fixed broadband rates are disparate across the globe but do not correlate with mobile penetration.
3) Broadband home rates in most of the world are <20Mbit/second

This leads us to several conclusions for 5G:

1) Mobile penetration is likely to be >100% in nearly all markets in 2020
2) Rates on mobile broadband are potentially likely to be higher than home
3) Unless there are major changes in fixed broadband then mobile broadband is likely to dominate by 2020.
4) There is an opportunity to make the next generation mobile network the de-facto communications system

Heterogeneity of devices, diversity of applications and services, maturity of cloud computing technologies together with the need for optimized content delivery, and support for QoS as well as MTC/IoT, present a set of drivers for the design of the next generation highly flexible, intelligent and scalable mobile network architectures.

Devices

Today it is widely accepted that PC sales are declining at an estimated 11% year on year, whilst tablet sales are growing at a rate of >70% [see Ref-10]. Also out of these PC sales most new business and home PCs are now laptop devices, with USB3-docking and Wi-Fi connection to more ergonomic desk-based Human Interfaces (HI) for business use. Clearly more and more, people today prefer a nomadic and/or mobile approach where devices are concerned. Bring Your Own Device (BYOD) has become the standard way we work, and share information, applications and media today.

Clearly a network that enables true device mobility and supports the user of the device to be able to declare the context they are operating the device with, is what is required, whether the device is a BYOD, fixed connected, mobile connected or home, office or other environment based at the time of a communications request.. The FDC does this by adding user profiles to the architecture that can be operated by the user and the network to trade information that is selectively enabled by the user to inform the network of parts of their context that improve their experience. Addition of network profile information can then similarly be used by the devices and their resident application(s) to further improve their operation and thus further improve the user's experience.

Media/Applications

The way we watch Video and Audio is also changing. We seem to have an insatiable desire for more colour gamut, larger screens and image depth (3D) in our home video habits and global TV sales show no signs of slowing, despite the relative increased costs over the last 25 years. To move these new HD and 4K formats around we have grown to accept that compression is essential and that some quality loss is inevitable (particularly in our Audio habits) yet broadband video distribution continues to thrive. Today 50% of all mobile traffic volume is video and 15% to 20% of traffic volume is social media and photos [see Ref-11].

The FDC network assumes caching at the edge of the network as a built in feature in order to support extensive use of high capacity media caching of common content and frequently shared content.

Cloud and Mobile/ Nomadic Computing

In the Cloud business area growth is also strong. Computer world reports that in the enterprise computing business, hardware sales were down 24% in 2014 as compared to growth in all of the enterprise and business computing areas put together.

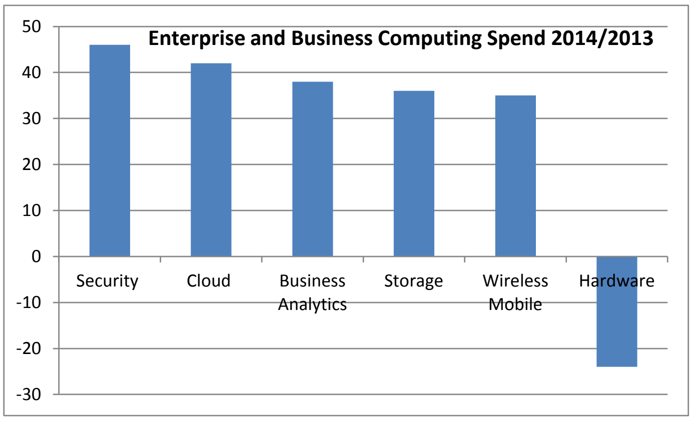

Figure 4: Year on Year Enterprise Business Spend Dynamics (%age cost change)

If we take a look at the evolution of cloud computing further, Cisco [Ref-03] report that Storage and Software as a service (SaaS) are booming with infrastructure (IaaS) and platforms (PaaS) lagging behind. Cisco further predicts that this trend will continue to at least 2018.

To meet the growing needs of cloud usage, the FDC network includes a user profile that can support an enhanced cloud experience by supporting frequent and recently used lists for an optimised cloud over 5G experience. In this manner, the FDC includes the capability to support a user's 5G briefcase over the network architecture and allows the user to control the management of the briefcase and their chosen cloud provider(s) via the network.

Content Delivery

Today most global mobile content delivery is handled by a small number of players. All of these players operate cloud-like content systems that are globally or wide-area regionally geographically distributed. However, these systems rarely closely connected with our local communications systems. Some communications operators have taken steps to seek methods to bring these content systems closer to the mobile edge but these methods are currently proprietary. The ETSI, Mobile Edge Computing (MEC) [Ref-05] group is just starting to investigate formalised methods to effect localisation and flattening of content to the edge but few decisions or requirements are published at this stage. This paper proposes that this kind of work is taken much further in order to create a content environment that meets the delay and latency targets that 5G communications systems are seeking to achieve.

Current global content players include Akamai, Google, iPlayer and YouTube, from producers Netflix, BBC, Sky and others. Most of the video content that is distributed over today's end-user communications networks has a common 10% that most people download and then a very long tailed type/tag based variability in the remaining 90% of available content thus making it difficult for mobile operators to operate local content caching to any notably useful degree. [e.g.: Ref-12]

Content Centric Networking (CCN) and Information Centric Networking (ICN) techniques are common in the academic literature, but practical approaches for implementation are less well developed, for instance in terms of i) a new system for advertising content by content ID ii) realising cheap content control at mobile nodes, iii) integrating a standard content control protocol that works well with mobile control protocols. What is needed for content control is an architecture that inherently supports mobile content control and enables developers to build the practical content control functions that are needed to make it commercial. The FDC separates out traditional Control plane and User Plane but also adds a User Plane Control protocol to do this.

Quality of Service (QoS)

Previous attempts to add QoS controls to a mobile network have been elaborate by design, painful in equipment design and implementation and unfortunately then have proved largely irrelevant or unused in operation. This has been largely because QoS controls have been seen as either too complex or insufficiently supported on an ETE basis to be workable. The IETF have been slightly more successful in their Differentiated Services Code Point (DSCP) approach although once again a small subset of controls is actually operated in most cases. 3GPP have four generations of experience in setting QoS for mobile networks, but provision and exposure of QoS controls to the device/application has proved poor. So it is proposed here that for 5G, each communications 'request' is made more descriptive of the request type, content and performance requires and the network selects the best available QoS controls in the network accordingly.

IoT

The Internet of Things represents the networking of sensor and actuator devices to the internet for improved discoverability, connectivity, for metering, measurement and monitoring applications and to ultimately enable support for control systems applications to manage these devices. However, public communications today is usually too expensive and slow reacting to effect much more than simple metering support and it does not currently support high end IoT systems for control loops. Current generations of wireless cellular communication systems remain unsuited for SCADA-like control systems because their response time is currently too slow. Note that SCADA-like control systems include IoT based systems for utility distribution and in the future would include Auto-Drive control loops.

The 5G Vision

The 5GIC vision document [Ref-06] sets out an ambitious goal of enabling a world where services are provided wirelessly to the end device by a fixed and mobile (converged) infrastructure that functions across the whole geography, including indoors and outdoors, dense urban centres with capacity challenges, sparse rural locations where coverage is the main challenge, places with existing infrastructure, and also where there is none. The foremost requirement is that the 5G infrastructure should be far more demand/user/device centric with the agility to marshal network/spectrum resources to deliver "always sufficient" data rate and minimal user plane (UP) latency (subject to use-case) so as to give the end-user the perception of an infinite capacity environment. Thus a new architecture is expected to address enhancements in terms of:

1. *Flexibility*: it should be easy to introduce new services, software upgrades and change traffic management policies and systems
2. *Complexity*: should be reduced in terms of implementation, deployment and costs structures
3. *Performance:* should be scalable, routing unlimited, UP and CP latency according to use case and traffic management made simple to set, monitor and .adjust

These enhancements are further mapped to the following five generic targets of the 5GIC vision, which are set to provide a route towards much higher performing networks and to a far more predictable Quality of Experience for users.

1. Minimise ETE latency for services and use cases that need this to execute successfully and/or safely.
   (e.g.: <=1ms IOT control activation turnaround such as Drone direction control adjustment)
   (e.g.: <=1s for IoT meter report update)
2. Maximise Quality of Experience
3. Enable new Applications
4. Make more efficient usage of all available communications resources at time of service request
5. Continue to minimise network OPEX and CAPEX

The 5GIC vision document does not explicitly declare targets for 5G expected throughputs, but in the design of the FDC, we have assumed that each cluster needs to be able to scale to support up to N x 10Gbit/s per cell in ultra-dense environments and at least 1Gbit/s in suburban areas per cell. It is further assumed that for network nodes providing cluster wide scope, they should be to be able to provide user plane capacity of between 1Gbit/s and N x 10Gbit/s per cluster and ~150-300kbit/s per cluster for the control plane.

Requirements for 5G Networks – Operator, Standards & industry Perspective

An important start in defining 5G has already been made in the Next Generation Mobile Networks (NGMN) [Ref-03][Ref-04] where 25 use cases have been identified, grouped into eight use case families, and serve as input for specification of requirements (as foreseen by a significant number of global mobile network operators) and defining the building blocks of the 5G architecture.

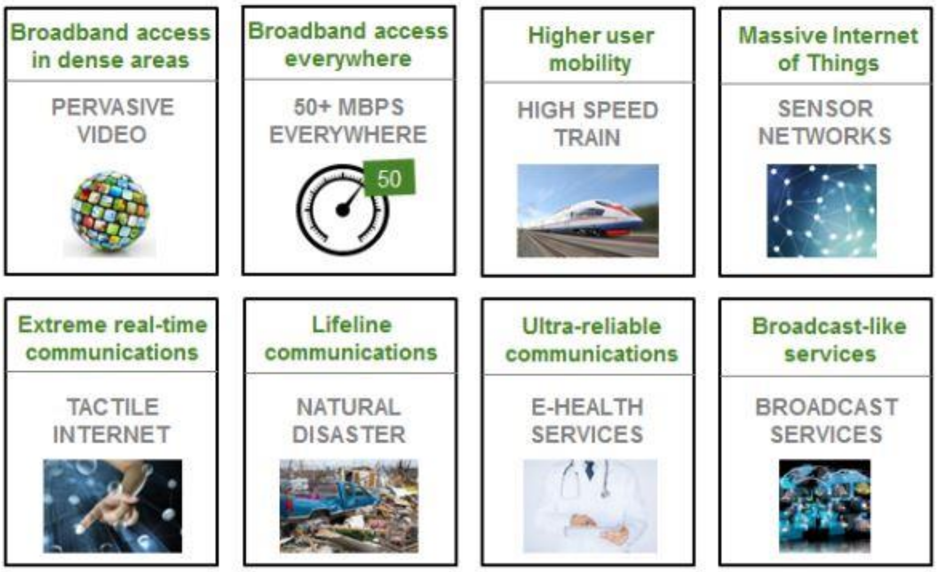

Figure 5: The eight use case families identified in the NGNM 5G White Paper

The 3GPP “SMARTER” study of next generation Market and Service Enablers [Ref-13] identifies 59 key use cases for 5G. The document builds on the success of mobile broadband provision afforded by LTE and includes approximately 10 use cases that necessitate ongoing provision of 4G/LTE and then adds new use cases that can only be provided by an enhanced mobile network.

The key difference in approach here is that the authors envisage that text generation mobile networks should be able to present multiple slices of the network to different users and/ or different services for same user depending on their usage context. This “Network Slicing” approach is a key tenet of SMARTER, that 5GIC has analysed as necessitating for ~ 60% of the use cases defined in the document. 5GIC supports the Network Slicing tenet and is working to enable this network architecture feature by supporting with methods to make the network Context Aware in order to support.

The 5GIC has identified three key tenets that a 5G Network architecture should provide, as below:

i) “the perception of infinite bandwidth”,
ii) “always connected capability”,
iii) “tailored context awareness”.

In considering NGMN, 3GPP and 5GIC key tenets, the 5GIC Architecture team has put together a high level set of network architecture requirements to support documented targets, use cases and business models as following.

Architecture Requirements

In order to support 5G, the 5GIC proposes a revolutionary architecture much flatter, cloud based but distributed. We’ve termed this the Flat Distributed Cloud (FDC) Network architecture. It is designed to meet the following key requirements:

Table 1: 5G Architecture Requirements

	Requirement	FDC support

	Requirement	FDC support
01	The FDC shall provide **distributed cloud based services and architecture** flexibility/evolution in a timely manner	The network shall be extremely flexible such that variable horizontally sized **clusters** of cells are overlaid onto a variety of sizes of datacentres according to number of cells supported and span of connected cells and available transmission. Clusters shall be able to be re-organised on a regular basis as service and load requirements change over time. The network shall also be able to be sliced into N vertical slices across the clusters in order to support different services/service groups requiring different performance targets at the same time. The Network therefore needs to be implemented in a soft manner employing NFV and SDN techniques in order to be able to swiftly reconfigure the scope and span of each cluster and vertical slice. All network nodes must be equipped with service compute power as well as supporting communications processing. All network nodes must be equipped with service and communications storage capabilities NFV techniques should be employed to build the infrastructure to support flexible Clustering (size of cluster and number of clusters) SDN support should be employed whenever highly dynamic bearer control is required. For example in dense urban areas with highly dynamic traffic patterns and many different user types in a concentrated area (therefore cost effective to deploy) The network architecture shall be flattened in order to keep the cloud network as low latency as possible
02	The FDC shall support **content centric networking**	Mobile ICN support must be in-built with the ability to employ User Profiling and Network profiling to tune content optimisation algorithms
03	The FDC shall support **integrated IoT** across a number of different IoT system types.	A number of IoT classes shall be supported in 5G. Legacy IoT class systems shall be able to integrate with the FDC network architecture through the use of mobile L2/3 gateways. Access times for certain classes of IoT devices shall be significantly enhanced in order to support a SCADA-like/Auto-drive control system class. A static IoT device class shall be supported in order to significantly reduce communications maintenance for low power devices.
04	The 5G Architecture shall provide support for **Context Aware networking** supporting both user and network 5W's to optimise use of available network resources and the service experience delivered to the user.	The FDC shall operate extensible, secure and selective user profile dialogue between user and network and service providers as information object classes that enable various extension features to support optimal slicing, clustering and QoS delivery. User profiles shall provide full '5W' capabilities (what, when, where, why, who) User profiles shall be selectively enabled for application/ network usage using secure certificates issued by the user/ users security broker
05	**Low latency services** shall be supported and	Separation of user and control plane shall be

	Requirement	FDC support
	download times significantly reduced in order to support Cloud computing and fast access IoT devices	implemented on an End to End (ETE) basis in order to minimise setup times for each new connection and/ or service request. The FDC shall operate predictive signalling in order to reduce signalling load on the network. The FDC shall support a softRRC in the form of a Common Resource Connection Control Protocol (CRCP) approach in order to provide: - Multi-RAT/Fixed Access (FAT) capabilities using whatever communications bearers are available to each user device (what) at each time (when) they access the network. - N-RAT/ FAT (multiple bearers at a time) per device connection in order to maximise available communication capacity - Group Connections may be employed for signalling - Dedicated and/ or group connections may be employed for user access. (Note: today, many cloud services are usually fast on a local basis with a good communication link, but remotely are impacted by poor remote communications services with highly variable latency) - typically <10-30ms Connection setup time - connection reload time <5ms - control response time ~1-2 ms
06	The user shall be provided with a **universal network capability** over as large a coverage area as possible.	The FDC shall support multiple communications transmission options in order to maximise provision of a common look and feel network architecture The FDC shall support C-RAN, D-RAN and hybrid –RAN approaches) according to available transmission in the area where the communications access points are provided. - Control plane coverage of at least 10Mbit/s per cell geographically - User plane coverage of between 10Mbit/s per cell to M x Gbit/s per cell and N x Gbit/s per user according to context (M >N)
07	**User plane control** shall to be more efficient than the GTP/ESM approach of LTE.	The FDC shall provide a user plane control (UPc) protocol that is context-aware and allows bearer service negotiation as follows: i) by service description and user context rather than by QoS class request ii) the network maps the service/context request to available network QoS controls (provided by FDC network elements) iii) FDC supports a mix of fixed and extensible service and context descriptors which can be used to describe each service as a UPc protocol. iv) Service descriptions are evolved by the operator over time and are managed to be a small number matched to current subscriber types and numbers. v) … whilst still maintaining control of Administrative functions such as Charging, Policy Management and Lawful Intercept The FDC does this with a new protocol called the Meta-

	Requirement	FDC support
		Data Protocol (MDP).

State of the Art Architecture Propositions for 5G

Considering the existing state of the art for 5G Architecture., there have been several references that provide use cases and performance requirements for a next generation mobile network over the last few years, for example: (Ref-01) NGMN, (Ref-13) SMARTER, (Ref-14) METIS and (Ref-17: IEEE 5G Architecture Requirements. All of these references propose significantly improved performance over that already afforded by LTE-A.

Also, it is noted in the technical papers that high levels of mobility signalling due to ultra-dense deployment of small cells and the massive connectivity requirements of devices/things expected for 5G will place undue demand/load on the core network nodes in the current long-term evolution (LTE) architecture. In fact (Ref-16) NSN on signalling in LTE, notes that signalling is growing faster than user data at present.

To reach the levels of scalability, flexibility and efficient use of resources together with reductions in costs expected in 5G, the main enablers i.e. NFV and SDN, are expected to play an important part. The motivations for introduction of NFV and SDN principles are manifold. NFV represents a framework for soft implementation of virtualized network functions (VNF). Virtualization is applicable to any data plane packet processing and control plane function in mobile networks. The VNFs can be deployed anywhere in the network and the trend towards virtualization of the core network nodes and sub-systems is continuing apace in the form of virtualised evolved packet core (vEPC) or virtualised IP multimedia sub-System (vIMS) offerings already available from a number of equipment vendors today. SDN makes data networks simpler to manage, configure/re-configure, and optimize. Complex control plane (CP) functions are abstracted away from forwarding elements so simplifying the user plane (UP) operations by relocation of control logic to physical or virtual servers running in a data centre. In addition, SDN enables Agile Innovation (operators can literally program their network in real time and when deployed together with NFV, introduce new services a in a matter of hours to days) and Intelligent path management, based on service needs rather than routing configuration and this is important where end users are mobile and bandwidth demands vary widely. SDN is mutually beneficial to NFV but the frameworks are not dependent on each other i.e. network functions can be virtualised and deployed without SDN and vice-versa.

An XaaS (everything as a service) model based on cloud computing (as opposed to relying on resources available locally) and employing SDN and NFV can bring agility to the delivery of services - including infrastructure, applications, software or platforms – over a platform that allows businesses to respond quickly to market changes as recommended in most references in this area. Other works e.g. Ref-17 and Ref-18 have advocated adoption of XaaS based models together with a transition to cloud-based architecture to enable 5G advanced infrastructures to be able to offer and deliver "anything", or "everything" as a service.

Lastly it has been noted that efficient design of NAS, service-dependent location of CP protocols and orchestration could also help reduce E2E latency as in Ref-17.

However blind adoption of SDN, NFV and Cloud principles as proposed in many of these papers without placing them in the context of some of the very specific challenges of mobile architectures.

Mobile systems require IP addressing in a mobile context and the mapping of the network across a flexible cellular system whilst handling soft radio scope and resolution onto a group of Cells or Radio units, whilst maintaining administrative functions such as accounting information generation and lawful intercept capability. In fact some initially attractive and inventive mobile network initiatives such as SIPTO and ANDSF stream breakout and interoperability have failed to become widespread network optimisation solutions entirely due to their incomplete mobile awareness definitions.

A recent initiative that is being addressed in 3GPP, starts to blend practical evolution with mobile awareness and is called CUPS “Control and User Plane Separation” (Ref-019/ Figure 6) for advanced LTE architectures. In CUPS, the LTE entities the Serving Gateway (SGW) and Packet Gateway (PGW) are being evolved so that their Control and User plane parts are separated out into their constituent User Plane Node (UPN) and Control Plane Node (CPN) parts in order to provide more flexibility and independent scalability suitable for NFV/SDN implementation whilst maintaining the mobility control afforded by the GPRS Tunnelling Protocol (GTP) which is retained between the evolved nodes.

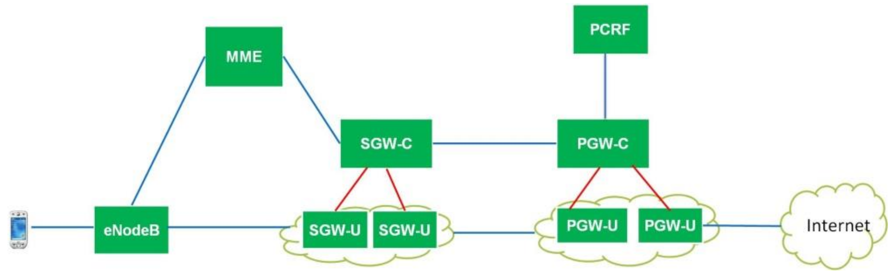

Figure 6: 3GPP, Evolved LTE-A CUPS Architecture (LTE-Pro)

Note: The 5GIC is currently working with ETSI to investigate next generation protocols as part of a new Industry Standards Group that is being setup called “NGP” to investigate how we can evolve current networking protocols to better support mobile networks and this work is expected to feed into 3GPP work such as “CUPS“ and “NexGen” in their Systems Architecture working group 2 which is looking at next generation architecture.

The FDC Network

Architecture

The Flat Distributed Cloud (FDC) architecture is proposed in this document for 5G to meet the requirements outlined in the previous section. In this architecture all nodes are both service and communications enabled with suitable processing and storage. The network is arranged in dynamic virtual Cell Clusters that are in turn overlaid onto Hardware Clusters that are located at key Datacentres of various sizes according to location type and available transmission support.

The architecture is very close to that of the CUPS architecture evolved from LTE, but the Control Plane nodes are mapped directly to MACRO/Umbrella cells as a single function per cluster of cells called the Cluster Controller (CC) and the user plane nodes are mapped to Small Cells as User plane functions called Cluster Member functions whilst the user operates dual connectivity to both the Control-Plane associated Macro-Cell system per Cluster and the most appropriate User Plane Cell within the cell-cluster. So the FDC approach combines Dual connectivity from 3GPP Rel-12, CUPS from Rel-14 and adds Context awareness and clustering to the architecture approach.

It should be noted that there are circumstances where the CC also provides local CM functionality e.g. in the case where it supports a mobile travelling at high speeds and a user plane function at the Macro-Umbrella level is more applicable to support this specific user context type.

The term "function" is used on purpose as these functions are intended to be implemented as soft entities over an SDN/NFV implemented set of equipment across variously sized datacentres in order that the cluster may be re-organised over time according to service evolution requirements, network slicing configuration and demand load. Re-configuration of this network architecture is envisaged with a suitably named Automatic Cluster Organisation (ACO) algorithm.

The architecture is envisaged as being implemented across a cloud based architecture but unlike the huge centralised clouds that the likes of Google, Apple, Amazon and Microsoft operate that are distributed physically according to used storage demands, the FDC architecture is distributed according to mobile demands.

The FDC also embraces the MEC initiative from ETSI (Ref-20) which adds hardware and software resources to enable caching either at the edge of the network or point of aggregation across a number of cells in order to provide computing and storage enhancement to the mobile functions. In an NFV/SDN flat distributed cloud scenario this means that the CC, CM functions when providing user plane service operate local caching and application processing capabilities.

The FDC operates as follows: a user device connects to a Cluster, umbrella Cell, which supports a Cluster Controller (CC) and directly connects to it to setup the network's communications connection signalling using a contended group connection system.

Once the device has a Control Plane communications connection to the cluster at the Cluster Controller it establishes a group NAS (Network Access Stratum) connection to the Cluster at the Cluster Controller virtual node. The user is then setup a default UP bearer according to their Device/ user profile and mobility (e.g. connect to umbrella id operating with high mobility) by connecting to the most appropriate CM, user plane node functionality.

Significant use of a user profile is made to optimise default bearer setup as a dialogue between the User and Network. The Cluster Member (CM) node provides UP control separately to the Cluster Controller functionality. As this network is context-aware, for fast moving mobiles, connection to the Cluster Controller itself is possible in the UP for this type of user context. Intra-cluster communications connections may be connected or 'Parked' to minimise connection overhead and then reloaded for UP service. However the signalling connection and bearer connection are managed independently with local network signalling between each instance to minimise over-the-air usage.

The architecture assumes an underlying NFV based implementation architecture with a 2-level Orchestration Controller approach. One controller is located at the CC level and one at the inter-CC level. This approach allows algorithms for topology optimisation to operate at the inter-CC level and then direct the periodic re-organisation of the clusters at the CC orchestration level according to information such as load vs time and user dominant usage type(s) collected at the cluster level. This allows periodic re-organisation of the clusters and their number and sizes at the inter-cluster topology level to realise Automatic Network Optimisation (ANO).

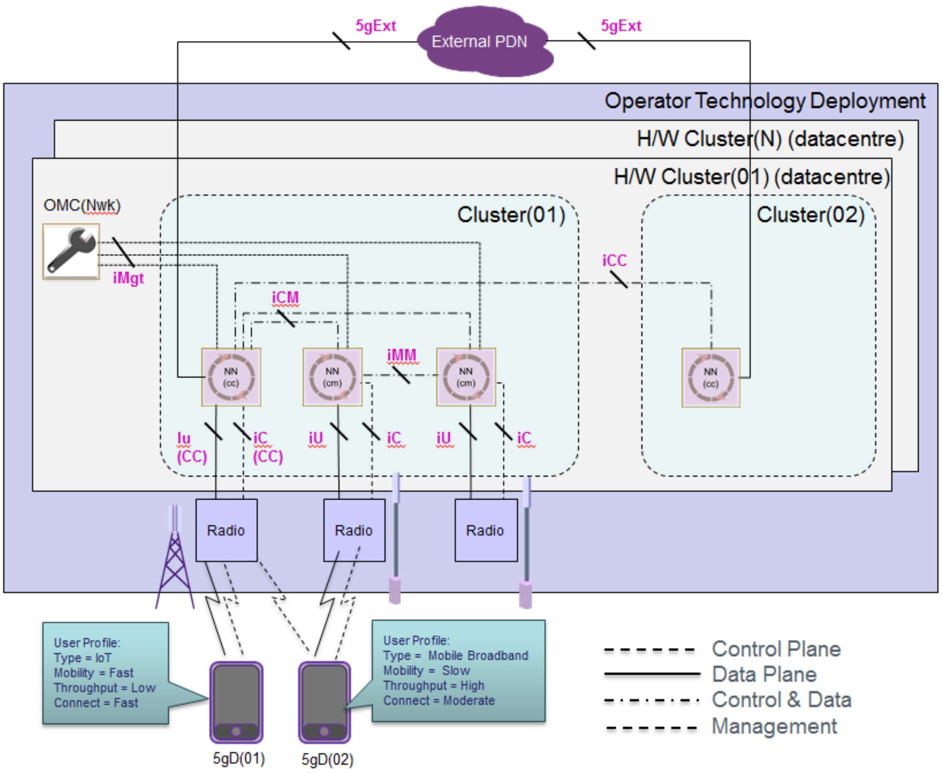

Figure 7: Proposed FDC Network Architecture

The Interface names for the FDC are as detailed in Appendix A The iC interface is for intra-cluster handover only.

Certain key aspects of the proposed FDC architecture are further outlined below:

Context Aware Networking

The FDC network will enable contextual information to be exchanged by the network and between different users of the network in a controlled and secure manner such as using brokered certificates per group of information. The Subscriber Data Management (SDM) system that in previous mobile generations has remained HLR centric is envisaged to be extended to support user profiles, so that the user may share information selectively with the network in order to improve their user experience. It is proposed that each user will have a **User Profile (Upr)** that will be extensible in a negotiable manner to provide information keys into new applications and network SON algorithms to improve the network and user experience.

Whilst some basic and essential information is certainly required to be known between the device and the network, say as a soft form of the current SIM (to provide base network, user addresses and security information) a new user profile for 5G potentially offers much more. User profiles are provided to enable the provision of information object classes (IoC) to which the user may selectively allow secured access by the network, in order to drive new SON functions in the network or beyond to improve the users experience.

This capability enables secure user profile IoC trading/information enablement between the device, the operator network, device applications and service interfaces to a number of players including: operators, service providers, content providers, application developers and cloud providers. This capability will significantly enrich service provision/evolution in 5G.

Connection Protocol Flexibility (Common Resource Connection Protocol)

The FDC network provides a soft-connection between itself and devices in the form of a Common Resource Connection Protocol (CRCP). The CRCP, as an evolution of RRC allows multiple bearer types from multiple technologies to be combined to support one common, dynamic, virtual connection from a device towards the FDC network. The CRCP connection is managed across the available bearer types depending on user context and available communications bearers. The connection may at any one time involve simultaneous paths across one or more available communications technologies under the common control of the CRCP.

In this manner the FDC network enables the potential to dynamically operate simultaneous multi-radio and multi-fixed access technology pooling, on a per user basis (multi-RAT/ FAT). The resultant available bearer pool formed is then operated using one or more of these bearers at a time (n x RAT/FAT) to best support the connection to each user locale potentially operated at the same time, as required, to provide always sufficient capacity for the communications task(s) in hand.

CRCP also provides group connections for signalling to make the signalling management across a group of users with a common context more efficient. In the user plane CRCP may also operate group connections for multiple devices at a time. One example of this behaviour is for a group of low capacity IoT devices, where a user dedicated connection is not essential. CRCP supports multiple levels of contention handling for 5G, including RACH, frame and message levels according to the service type requested (e.g. SCADA-like IoT is a special case that needs fast access but requires low throughput). Signalling is a different case again, often requiring fast operation but not necessarily requiring high volume through-put and usually requiring a highly robust bearer.

Broadband user access is a different case again, here we can afford slower access, but need very high throughput once connected and we can afford to be less robust as we are sending so much data volume that there is the opportunity to operate multi-level HARQ/ARQ mechanisms. Support for all of the above cases should be provided in 5G to enable a flexible communications network driven by contextual information from for example Extensible user profiles. Also there should be a dialogue between the mobile device and mobile user to keep the actors updated so that they can make best use of the available options on a well-informed basis.

For example when a user starts an HTTP connection the user is informed that they have for example one of the following options. The user or user proxy can then either choose a presented option or wait till they move or their communications environment changes to see if they can get better options.

Case0: e.g.: Poor RF mobile wireless, no Wi-Fi (expected speeds <300kbit/s, SMS, Voice and low speed internet access, expect some wait time or move north 100 yards)

Case1: e.g.: Moderate RF mobile wireless, but no Wi-Fi (expected speeds 10Mbit/s to 100Mbit/s, web page response for 100 objects ~ few s)

Case2: e.g.: Moderate RF mobile wireless, excellent Wi-Fi (expected speeds 100Mbit/s to 1Gbit/s, web page response for 100 objects ~ <0.5s

Case3: e.g.: Good mm-wave mobile wireless, excellent Wi-Fi (expected speeds 500Mbit/s to 3Gbit/s, web page response for 100 objects ~ instantaneous)

The CRCP protocol also responds to the other Upr context information, such as current device usage type (context) with relation to required mobility and current/ recent usage pattern. For instance if the user type is given as say 'work based, mobile broadband' then their mobility is likely to be low and their demand for bandwidth relatively high and so the CRCP protocol will at CP connection to the CC, assign the user to its closest small cell CM for UP connection. Alternatively, for a fast moving vehicle (FMV) such as a train, the user profile mobility information is updated so as to inform the network via and it may choose to relocate the UP connection to the CC itself in order to support the fast moving nature of the users connection at a the umbrella cell/ CC level.

Content Optimisation

The FDC operates an enhanced form of Mobile HTTP2.0 which examines each page object count and type before download, to establish the best settings to download the page. The FDC also operates a mobile-network optimised version of ICN and opens interfaces to deploy algorithms at each node that cache content based on shared user profile data and previous download history for overall efficient cluster-based content caching per user and per group and pre-loading control.

The goal of the FDC is to get user's data ready 'just in time' or with assistance from the user 'just for my today and now' focus, to alleviate some of the known limitations of cloud computing/ storage services such as delay. So the FDC operates its distributed service cloud so that it appears centralised to the user, but is intelligently distributed to keep user data close to the communications access anchor in each cluster as user moves across and within clusters.

Group Broadcast services

The FDC enables extended user profiles that provide the capability to securely and selectively operate a context aware network that can identify a) common events and b) common usage patterns and short term trends. With this kind of localised user and grouping of interest information, a network can quickly identify when a group of people all want to use/watch a particular dynamic content stream and dynamically collapse a number of identical unicast streams into one multicast stream periodically to meet a common crowd demand.

In this manner user profiles and context aware networking enablement can be used to improve some of the inefficiencies of previous, versions of MBMS (Multimedia Broadcast Multicast Service) and accurately trigger short term multicasting group setup. In this manner, on-the–fly broadcast resources may be setup and torn down according to this extra information, for example for sport and impromptu social gatherings and information/content sharing and broadcasting by web based dynamic postings employed to advertise using extended social media interfaces.

IoT Classes

The FDC supports a graduated IoT approach that acknowledges the massive installed base of IoT expected in 2020 and connects it into the evolving 5G infra-structure and peripherals. Today many high speed scanning telemetry systems (e.g.: SCADA systems) operate multiple static stations over unlicensed bands and use low cost equipment to operate. These systems often require multiple updates to a central control point within a minute and then require millisecond response time actuations once the reports are processed. At present, LTE is simply neither cost effective nor agile enough for this kind of application.

A new 5G SCADA-like enabled IoT framework is envisaged for the 5G FDC with fast control execution latency (few ms) using the enhanced connection management protocol CRCP. Enabling this kind of control system over 5G also enables the required IoT framework for evolving mobile IoT cases such as car and drone control systems such as Auto-drive, Auto-fetch and Auto-monitoring.

Predictive Signalling

The FDC uses location and user profiling to operate predictive signalling and signalling optimisation in the form of "memory-full" NAS messaging to reduce signalling load. In particular mobility updates and paging are minimised in CRCP idle/parked mode(s) and a reduction in measurement reporting during connected mode can be realized where the mobile is empowered to make their own handover decisions after demonstrating good handover prediction behaviour.

Legacy Support

The FDC control plane anchor is operated at the CC function, which is instantiated on a per cluster basis and realises the 5G CP features detailed in this document in association with other CM nodes that primarily support UP and user UPc functionality. However, a 5G network still needs to operate all of the traditional basic mobile services of an LTE network and so when the current radio and network conditions are not able to support all of the 5G enhancements, the FDC is able to operate as follows:

i) As performance degrades due to localise network or RF issues, the UE and FDC network revert to LTE-like basic network signalling (EMM for mobility and ECM for network connectivity) and UP services are scaled back using MDP, UPc control. This happens as long as there is still 5G coverage.
ii) If the device eventually moves into an area that does not have 5G coverage then it is assumed that the mobile can detect underlying LTE coverage and performs a handover from 5G to 4G.
    - for the CP, handover is effected using network controlled enhanced EMM procedures between a CC and an MME.
    - for the UP, handover is effected using UPc and SDN flow updates that are exchanged between the 5G UPc at the CC and/ or CM nodes towards the EPS Session Management (ESM) protocol at the MME and SGW, which are evolved for LTE to include a UPc network interface extension to SDN to interwork with 5G/MDP.

Mobility managed legacy interconnection to earlier 3GPP generations such as GSM and UMTS are seen as costly in terms of both multi-RAT measurements and interworking in the network. In the 2020+ network timeframe, it is assumed that LTE will have rolled-out sufficiently to provide complementary alternate wide area underlay coverage for 5G. It is assumed that if an area has 5G coverage with only underlying GSM and/or UMTS coverage then a re-selection is envisaged by the mobile, rather than a handover, but details are FFS.

Security Framework

A 5G Network such as the FDC will inherit the best of the security aspects of LTE and its predecessors such as: device, SIM and message authentication, IMSI privacy and message integrity checking. However, the biggest enhancement for a 5G security framework is enabled by flattening the network to a single layer hierarchy in the FDC network architecture enabling ETE stream encryption between the UE and the UP connection point in the network (CM or CC). This is also true for the CP which is also single layer in the form of the CC. It is further envisaged that each IoC of the Upr is able to be certified individually and brokered to other parties such as the operator network or another service provider under the control of the user by a certificate broker or agent. This capability provides the user with selective context and profile information control for all extension information used to drive 5G algorithms in the network.

RAN Flexibility (ability to support different RAN architectures)

The FDC assumes that according to deployment locale, the FDC will need to support multiple RAN architectures. The FDC assumes a predominately CP point of attachment to Umbrella Cells overarching clusters of cells or TRX points with predominately the UP provided through the underlying smaller-cells and/or TRX points. The FDC needs to support all three types of RAN architecture (C-RAN, D-RAN and H-RAN, see Figure 8) for good geographical coverage across rural, urban, suburban and dense urban topographies. This is because fibre connectivity to the home or even to the kerb/cabinet, is not ubiquitous in most (say) EU countries and is unlikely to become cost effectively so by the 2020 timeframe.

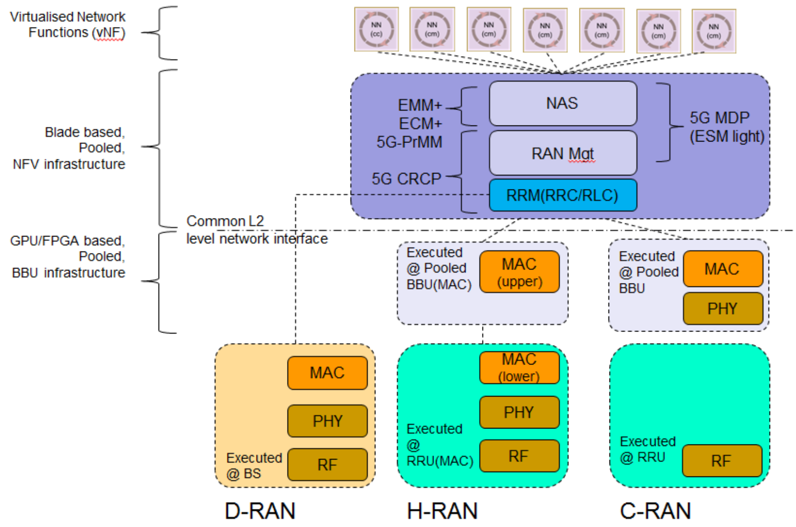

Figure 8: C-RAN, D-RAN, H-RAN Architectures Supported by FDC Network

As such we note that the FDC should have a common RAN recombination point for all types of supported RAN. It is envisaged that this common RAN connection point will always be the MAC/PDCP level. It is assumed that conventional D-RAN base station type of 5G evolved Radio Equipment will still need to be operated in very rural environments and that they will pass through full MAC/PDCP level SDUs to the FDC network. It is assumed that Cloud-RAN RRU/BBU types of Radio Equipment (capable of waveform combining) will be operated for 5G in dense urban environments.

It is also envisaged that 5G will operate Hybrid forms of Cloud RAN that do not perform waveform combining but operate some form of hierarchical BBU based MAC/PDCP level combining towards the FDC. As such the FDC is envisaged as a "flat", 5G architecture that will sit on top of this plethora of different RAN architecture options but that they will all provide a common logical MAC/PDCP level point of attachment once the potentially multi-point, multi-cell radio signals have been combined.

User Plane Split

In the network layer we are proposing that there are likely to be 2 of these RAN recombined connection points. Most current thinking on operating these two network connection points is that one is supported by umbrella cell connection exclusively for CP network services and the other is for serving the small-cells or RAN-points providing exclusively UP network services. Alternatively, in our proposed cloud based network architecture, we argue that in most cases (as often in our business) what is required for any practical usage of such a system is to support a softer approach where exclusive CP and UP is evolved to a softer set of assumptions which means that the umbrella cell is predominately used for the CP and the smaller cell is predominately used for the UP. We also appreciate that there are service use cases where it is advantageous for, for instance, the umbrella cell to support UP traffic albeit at a potentially lower data rate and with a higher mobility.

Transmission considerations

RAN inter-connection

It is envisaged that the FDC will connect to the different supported RAN architectures using several different front-haul and traditional backhaul options. For C-RAN in dense urban areas, common single mode fibre is physical medium and CPRI the L1 de-facto format standard for today's technologies. However with significant expected reductions in frame length and waveform combining techniques and a TDD based radio system anticipated for 5G , it is expected that cells (or multi-point deployments) will be deployed as clusters in dense urban areas and brought together to a local central point to be all sent over the same DWDM fibre back via OADM/ROADM infrastructure to the same BBU element for combining at the target datacentre in order to minimise delay variation and enable constraints on BBU buffer implementation which is likely to be requires as frame duration moves towards the us level.

It is expected that governments wishing to stimulate their economies will invest in significant fibre network extension programs for the 2020 timeframe. Economies are now all notably driven by their communications infrastructure and with a move to ubiquitous wireless rather than wired access, provision of sufficient fibre backbone services in a country is now critical to a nation' economic success.

For D-RAN/H-RAN it is expected that fibre will again be used where available, but where it is not cost effective it is expected that millimetric backhaul is likely to be employed for the last 200m or so of transmission as one or more hops, employing some of the recent and evolving backhaul techniques from the IEEE for example one of their stated targets is to investigate the possibilities of 30Gbit/s over 60GHz with again local concentration back to the nearest fibre connection point.

D-RAN in areas where it is not cost effective to deploy Fibre or mm-wave is likely to adopt the same overall backhaul topology approach using available wired line connection or pairing to existing connections and adopt G.fast similarly over ~ 200m to 500m with expected rates of a few hundred Mbit/s to 1Gbuit/s to a central office or new mm-Wave concentration point then back into fibre.

Network Inter-connection

Datacentres are expected to be intra-connected at L2 with 1, 10 and 100Gbit/s Ethernet based technologies and evolutions. IP transport between servers is envisaged as being maintained at a minimal level to reduce transmission and session overheads and wherever essential operating UDP as the prime transport with ARQ strategies built in only as necessary. Traditional, usual VLAN and default gateway approaches from the IT world are also minimised wherever possible operating dedicated dynamic L2 paths as required. Inter-Cluster connection is optimised so that clusters with the largest inter-cluster traffic are instantiated at the same datacentres using automatic network optimisation (ANO) strategies. It is envisaged that all connections between datacentres will be optical, with dynamic L2 path control.

Strategies to optimise transmission efficiency in the FDC era are anticipated as follows:

i) Seek to keep transmission at L2 wherever possible
ii) Operate UDP with simple ARQ mechanisms at datacentres
iii) Operate dynamic L2 path control for all bulk persistent paths for intra and inter-datacentre communication
iv) Where necessary, over the L2 paths available, operate selective soft session and flow management between server blades, using SDN and its evolved versions
v) Operate SON-Automatic Transmission Optimisation (ATO) algorithms and SON Automatic Network Optimisation (ANO).
vi) Minimise the number of direct requests to the internet/ external PDNs through extensive caching of content, IoT, management and profile data for maximum localisation.
vii) Minimise inter-Cluster Handovers and repeat internet/ external PDN request, through regular ANO re-organisations of the cluster and inter-cluster topologies and ATO to drive the config of dynamic, optimised bulk transport L2 paths according to diurnal load, mobility and context usage changes.

In addition to these strategies for transmission, it is expected that mobile SDOs will need to work closer with Network SDOs in order to shape and evolve the usage of the internet and IP protocols to better interwork with 5G mobile systems as network latency is increasingly an equally important part of the ETE performance as the actual Radio system itself.

Initial Analysis Results for the FDC Concept

The 5GIC architecture research team have analysed most of the fundamental principles through simulation studies with some encouraging results on performance enhancements over previous generations. Some of the key high-lights are listed below:

Table 2: Initial FDC, 5GIC Analysis

Aspect Researched	Improvement
Signalling reduction in RRC Idle mode through the application of Predictive signalling (PrMM)	80% reduction in paging and mobility updates
Signalling reduction in RRC Connected mode through the application of Predictive signalling (PrMM)	40% reduction in Measurement reporting
Evolution of Attach to Association	Adds context awareness to the mobile network to decide the best connection point for the current context of the user (current: mobility-entropy and services required, recent behaviour …).
Adoption of optimised CRCP and moveable User Plane connection point	Reduction of connection from >100ms in LTE to <30ms in FDC Reload approach affords reconnection in 1-5ms Much faster Connection handling of this architecture and operation of CRCP enables fast IoT use cases to realise ms speed actuations.
Application of memory-full NAS IE caching.	Up to 50% reduction in mean message sizes for 70% of all cases
Edge Caching	10% of all content can be predictively preloaded based on medium term download history tracking, resulting in significant latency improvements for Video content downloads. This accounts for 90% of all content requests to some sources we have analysed.
Mobile HTTP page download latency reduction	30% faster than HTTP1.0 (greater gain for most commonly requested pages than measured for SPDY analysis)
IoT latency Reduction: - Operation of Group Signalling for NAS connection - Operation of Group IoT UP connections - Static IoT profiling	All following 80% faster than LTE: - from Idle to operational user plane - IoT report time - IoT activation time
Flexibility afforded by End to End separation of User Plane and Control plane elements as the CC and CM over a soft NFV/SDN implementation Architecture enables Automatic Cluster Re-organisation according to service and demand loads	i) enables slicing in the vertical axis ii) enables dynamic clustering in the geographic (horizontal) axis (see Ref-21)
Flattening of the network hierarchy and flexibility to deploy tenets of 5G SGW, PGW functionality either separately or in integrated form at CP or UP nodes	Multiple collapse of GTP tunnels affords ~ 35% tunnel overhead reduction end to end in the user plane Enables moveable breakout options either at CC or CM level for user plane enabling advantage to be taken of local breakout options whilst maintaining full admin control via CC.

From Architecture back to Visions

To make this new architecture tangible for the consumer we take three illustrative use cases to demonstrate the transformation of user experience the new architecture can deliver.

Use Case 1: "a more connected/ uncluttered computing experience".

Today

At work, most users connect a laptop or desktop to a screen, keyboard and mouse and operate applications that are installed onto that machine. This mode of computing requires constant software updates, careful management of backups of critical data and solving frequent IT issues with IT support staff for communications and hardware conflicts.

The monitor is LED based but still quite power hungry and not touch screen, the mouse is wireless but battery operated and the keyboard gets dirty quickly. A typical computing worker frequently has to switch between Wi-Fi and wired connectivity as they move between meeting rooms and they still find themselves carrying quite a few cables and adaptors for all eventualities. At home the worker typically has another machine and has to adapt the work laptop to the UI equipment they have at home. The home equipment is usually different to that at work and does not always connect without some re-configuration. Similarly when the worker travels they have to adapt to many different environments and interface equipment (e.g. projector, monitor, display screen).
Most workers use some form of cloud but these are currently highly centralised and often the communications from a remote point is not as good as at work and the worker finds themselves often waiting for downloads when at a non-work location.

Future: 5G Nomadic Cloud Computing (home, office, 'out and about')

The future should see much improved Human Interface (HI) equipment pods, with more universal communications interface(s). This 5G enabled HI equipment will operate over a number of RAT/FAT interfaces that 5G supports and will be conversant with a CRCP operating in the 5G-FDC. With the net effect that frequent docking/cabling swaps per location will cease whilst still remaining secure.
It is envisaged that Mobile computing interface devices (currently either tablets and/ or mobiles) will seamlessly connect to a variety of HI equipment operating security tokens wherever the users HI pods are located. Additionally when travelling or moving around the workplace, for a small fee or for a secure free issue system from a hotel or conference/meeting room the user can obtain a security token (potentially also wirelessly/ or by landing web page) to connect to a visited HI pod. So, wherever a user takes their chosen device they can simply and securely connect to their chosen network provider and use the local HI pod and access their applications, services and information anywhere.

Once online at the users owned or visited HI pod, their device (which probably already has working copies of all most recent documents) checks that they are the most recent copies using the 'recent information list' on their 'user profile' with their distributed cloud reference store and they start working. Their social network, email, documents and personal 'most recently used stuff' is at their device when they need it, where they need it.

Their device has a medium amount of processing and storage and all demanding computing is operated at a local piece of their chosen supplier with their selected applications (here, the cloud provider may or may not be the users communications provider). The users deep store information can be accessed on-line at a moderately slower but useful pace.

The index of all the users stuff is stored securely by their provider cloud and/ or communications provider as a User Profile with their preferences and tokenised classes of usage information that they can securely and selectively chose to share with them depending on benefit/ value to them. When a user moves on from a given location/ HI pod they just take their device with them. This is just the same as for a mobile device today, but the difference is that now the mobile device is also the gateway to all of their connected stuff, their preferences and stored regular behaviour whilst they are connected.

The user no longer needs to have multiple devices, the costs for hardware are lower and maintenance is managed centrally with online support, at all times. The user employer provides the HI pod at work and the user provides their own HI pod for their home physical location(s) or customisable vehicle interface and in both cases the users device seamlessly connects wirelessly to the user's BYOD (Bring Your Own Device). The user profile may be moved in a secure manner to a number of devices by paired security tokens/certificates and further security tokens are used to secure interaction with the users personal and work distributed cloud(s). A user's communications provider may be two fold (the user and their employer) or funded by the employer alone. The Communications providers are supplied most of the employer owned Work User Profile on a secure trust basis and the users select what of their Personal User Profile they chose to share with their providers.

The communications provider uses the User Profile(s) to decide the users '5W's,

What is stored/operated/ executed?

Where it is stored/operated/ executed?

When the user needs the information?

Why the user needs information?

Who/what the user's information is provided to?

The holistic distributed cloud environment is serviced with multiple- communications services that are managed into one connection towards the user's device that is dynamically managed according to their 5Ws and their User Profile(s). Information is always there just in time using selective edge caching.

Use Case 2: "More useful, integrated and extended IoT environment".

Today

Many utility companies want to put "a small box in a consumer's house" or "a small controller" in their house that gives them very small insights into how they use their utilities that they probably already know about and only really save them money if they connect up many more devices in some form of control loop that will not give them any real payback for 5 years.

They could buy the latest thermostat controls integrated into Apple ™ or Android ™ or suchlike equipment, but it's quite expensive, doesn't do much and is not that well integrated into their devices unless they sign up with a big player who really just wants their behaviours for their latest marketing campaign so they can learn how to sell the consumer more stuff. Again this costs quite a lot and does not integrate easily with their typically existing 10+ year old home. Scanning Control and Data Acquisition (SCADA) systems are usually operated over wired Ethernet on a campus or some form of proprietary wireless bus or low power unlicensed radio to control a closed loop plant feedback system. These systems work well but are proprietary in nature and have to be cost effective to operate or the control system stays manual. Some of these slower IoT systems are deployed over custom implementations of mobile networks.

Future

Metering applications are integrated with the 5G network cloud and may be securely accessed on a selectively enabled basis via a consumers User Profile (be they corporate, machine or individual human). The access time on 5G is much shorter than previous generations so that SCADA-like applications and Auto-Drive applications may now be integrated with ones connected lives with extended User Profile, Information Object Classes to support them. Legacy IoT devices may also be addressed via IoT communications gateways to the legacy systems from 5G and their data appended/linked to ones cloud in a secure manner. Also with the advent of wireless integrated Wi-Fi to the 5G holistic communications interface the higher layers of the 5G network protocols enable these devices to be seen as authenticated 5G devices through a legacy gateway.

Use Case 3: "My content, faster"

Today

A consumer can download what a regular content producer/ provider partner pair advertises on an internet page over their subscriber mobile network or download the same over their home network. A consumer adds their selection to the download queue and waits their turn for their download from the central store. If this is not very local then they wait quite some time.

Future

A consumer select what they want to download from a Content Centric Network (CCN) or Information Centric Network (ICN) advertisement page that does not operate centrally, but is aware of their preferences, typical behaviour and knows where they move between in terms of communications Wi-Fi access points, mobile base stations, and their home location. The consumer's profile knows where they are and searches for content according to 5Ws knowing where they are and if the content has recently been downloaded in their area, it downloads from there and not centrally which is a faster approach than today.

Over time the 5G content and Context aware network learns where a consumer is in a typical day (and by their location if this is not a typical day or one of their typical days) and if it is a typical day and they are close to a convenient and high speed download point it downloads their content when they are close to this optimum download point. If they want the content now and the download point is not good they can still insist they get the content, but the network tells them it will be slightly slower and offers them alternative approaches like move 20 metres towards the canteen and it will be faster etc. The network is in effect in active dialogue with them and so it can negotiate for an optimum 5G experience. But overall their content download experience is faster than without ICN/CCN and Context Awareness.

Conclusions

The drivers and key requirements for a 5G architecture evolution have been described and a new model to meet these requirements has been presented based on a flat distributed cloud approach (the FDC).

The proposed architecture is designed to be able to adapt to user service requests according to user and network context and be able to respond to demand such that user perception is always managed to best meet each request with the available resources.

The architecture has built in support for advanced content management features and IoT, both of which are major drivers for the 5G timeframe.

All of the RAN architectures envisaged in a 5G timeframe are supported in any combination by the proposed common FDC network architecture and it is able to operate communications connections across multiple bearer types and access technologies using a new, flexible and context aware protocol (the CRCP).

5GIC Research

The 5GIC will continue to develop the FDC network architecture as we evolve our simulations into emulations and confirm the key tenets of the architecture on 5GIC Network Test-bed at the University of Surrey.

References

Ref-01: Cisco, 2014 'Global Mobile Data Traffic Forecast Update 2014–2019 White Paper'
Ref-02: The Broadband Commission, 2014, 'The State of Broadband 2014'
Ref-03: "NGMN 5G whitepaper", NGMN, February 2015.
Ref-04: © 2015 Next Generation Mobile Networks Ltd. Photos: © Shutterstock.com/ Leigh Prather, Alexey Stiop, SOMMAI, Syda Productions, flydragon, BackgroundStore, Ortodox, tovovan
Ref-05: MEC: http://www.etsi.org/technologies-clusters/technologies/mobile-edge-computing
Ref-06: 5GIC Vision Document
Ref-07: Source: World Bank, Official Population clock
Ref-08: Source: Mobilthinking, government census statistics and operator final year reports
Ref-09: Source: Akamai and Ookla
Ref-10: Sources: CISCO VNI, Statista, Gartner and Forbes
Ref-11: Sources: CISCO VNI, ERICSSON
Ref-12: Analysis of anonymised BBC iPlayer data by 5GIC
Ref-13: 3GPP FS 22.891 SMARTER, Feasibility Study on New Services and Markets Technology Enablers;Rel-14
Ref-14: ICT-317669 METIS project, "Scenarios, Requirements and KPIs for 5G Mobile and Wireless System" METIS Deliverable D1.1, May 2013.
Ref-15: "IMT for 2020 and beyond", ITU, Switzerland
Ref-16: NSN, "Signalling is Growing 50% Faster than Data Traffic," White Paper, 2012.
Ref-17: Patrick Kwadwo Agyapong, Mikio Iwamura, Dirk Staehle, Wolfgang Kiess, and Anass Benjebbour, "Design Considerations for a 5G Network Architecture", IEEE Communications Magazine, November 2014
Ref-18: D Soldani, A Manzalini, "A 5G Infrastructure for anything-as-a-service", Telecommunications System & Management, 2014.
Ref-19: 3GPP TR 23.714 "Study on control and user plane separation of EPC nodes"
Ref-20: MEC, Mobile-Edge Computing – Introductory Technical White Paper, ETSI, September 2014
Ref-21: IEEE: A two-dimensional architecture for end-to-end resource management in virtual network environments, Network, Volume:26, Issue:5

Glossary of terms

IoE	Network of physical objects or "things" embedded with software, sensors and connectivity, covering a variety of protocols, domains, and applications, collectively termed the "Internet Of Things."
iU	User plane interface from the mobile to the Cluster member for a user with a slow moving user context
Iu(CC)	User plane interface from the mobile to the Cluster member for a user with a fast moving user context, potentially with a slower throughput but requiring faster connection and handover than on the iU
iC	RAN control plane interface for user plane connection setup to the Cluster Member
iC(CC)	Control plane for all NAS signalling towards the CC and RAN setup of the CC connections for control and user plane at the CC
iCM	Control relay between the CC and CM for bearer coordination and handover between user context states e.g. when mobile is fast moving mode or slow moving mode for example
iCC	CC to CC control plane and temporary bearer relay during inter-Cluster handover
iMM	CM to CM control plane and temporary bearer relay during intra-Cluster handover
iExt	Breakout point for the Cluster from the CC towards external Packet Data Networks, such as the Internet or a private intranet.