BT Group

Telco cloud evolution - what really matters?

Introduction - Characteristics of Telco Cloud

Since the advent of Network Functions Virtualisation and Software Defined Networking (NFV/SDN) [1], telecoms operators like BT have followed a route to adoption for building and controlling their own on-premise cloud infrastructure, particularly suited for telecoms workloads and applications. The main drivers for this hinge around the core principle of disaggregation, whereby hardware and software components can be provided by separate best-of-breed providers, and softwarised network functions can be provisioned and lifecycle managed in a much more agile and flexible fashion than their hardware-based predecessors.

Across the telecoms industry as a whole, the term “telco cloud” often refers to such deployments in general terms to distinguish them from data centres used to host an array of enterprise applications and workloads. While sharing many of the key underlying building block components like commodity servers, storage and switches, there is a wide range of other criteria that must be factored in when hosting telecoms workloads on such cloud-based infrastructure: high availability, resiliency, security, performance and observability to name a few. Indeed, the workloads running on a telco cloud are specialised tasks and services that telecommunications companies deploy and manage within a cloud environment, tailored for their needs. These workloads can include gateway functionality, which facilitate communication between different network domains, and are not limited to traditional client-server models. They often utilise specialised protocols beyond standard HTTP/S to handle the unique requirements of telecom operations (voice, mobile payload and signalling, multicast, video & TV streaming). Additionally, telco workloads frequently consist of complex and compound microservices architectures designed for specific functions, such as a 5G standalone (5G SA) core. To ensure optimal performance, these workloads require dedicated resources, high input/output (I/O) capabilities, and measures to avoid “noisy neighbour” interference that could degrade performance. Security is also paramount, with stringent regulations like the Telecoms Security Act in the UK mandating robust protections to safeguard sensitive data and maintain network integrity and availability.

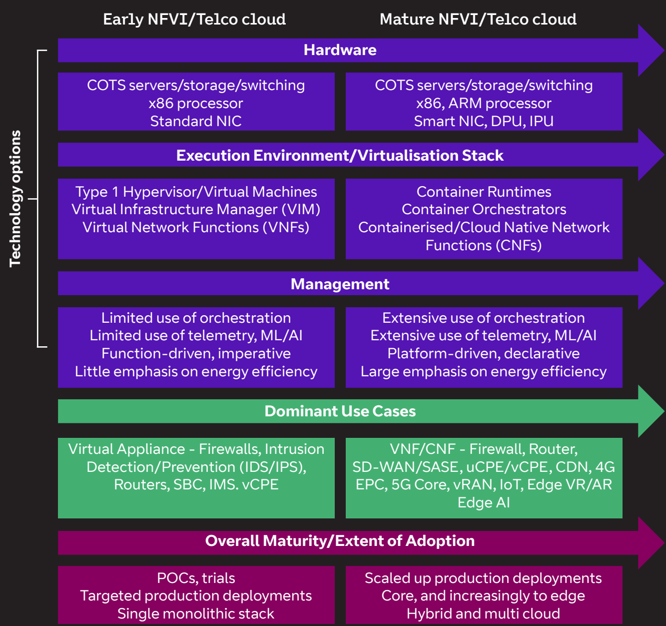
BT ’s Network Cloud has been operational for some years [2] and is now a mature and scaled-up platform delivering a range of telecoms-oriented workloads including the 5G standalone core [3]. Since the inception of NFV/SDN to the current time, there has been a wide range of technology changes and trends impacting on the way in which telco cloud infrastructure and the services that run on top, are managed. As the technology landscape continues to shift, this article summarises some of the key trends which will influence operator decision-making in the coming years: hardware heterogeneity, cloud-native workloads and operations, network edge ramifications and sustainability. For each trend, we identify *three things that really matter*. In drawing conclusions, we also reflect on *what it means for our customers*.

Inception of NFV/SDN to Present Day Telco Cloud

There are a number of useful references [4-6] which outline the trajectory of telco cloud evolution from the very early days of NFV/SDN (the original ETSI white paper that set the wheels in motion for the ecosystem to evolve, was published in 2013 [1]) to the present state-of-the-art. The diagram of **Fig. 1** aims to illustrate some of the main features of telco cloud evolution. The shift from left to right indicates a high-level snapshot “then and now” overview of technology options, use cases and overall maturity and adoption levels. In broad terms, we see an evolution from fairly homogeneous hardware to more diverse hardware options, from hypervisors and Virtual Machines/Virtual Network Functions to container runtimes and orchestrators, and from function-driven management with limited automation to a more platform-driven framework assisted by high degree of automation.

In terms of use cases, we also observe the shift from early quick-win implementations of “virtual appliances” for fairly standalone network functions like routers and firewalls (where a monolithic piece of software/OS is essentially re-factored to run on commodity hardware inside a Virtual Machine via a type 1 hypervisor) to the emergence of VNFs and latterly Containerised/Cloud Native Network Functions (CNFs) which cater for some level of software decomposition applicable to the network function(s) in question. The remainder of the article will expand on many of these points.

Fig. 1: Early and mature telco cloud characterisation



1 https://portal.etsi.org/NFV/NFV_White_Paper2.pdf
2 https://www.telecomtv.com/content/cloud-native/how-bt-is-reaping-the-benefits-of-its-network-cloud-50676/
3 https://canonical.com/blog/bt-group-and-canonical-deliver-5g-to-uk-stadiums
4 https://www.etsi.org/images/files/ETSIWhitePapers/ETSI-WP-53-In-the-Light-of-Ten-ears-from-the-NFV-Introductory-Whitepaper.pdf
5 https://lfnetworking.org/etsi-nfv-a-decade-of-transformation/
6 https://blog.huawei.com/2023/06/27/nfv-telco-cloud-next-decade/

Drivers For Change

Hardware heterogeneity

There is perhaps no better illustration of how technology trends go in cycles, than the comparison of hardware aspects of “early” versus “current” telco cloud. It is highly likely that many of the initial deployments of NFVI/telco cloud would be based on COTS (Commodity-Off-The-Shelf) servers with CPU-based architectures and fairly standard Network Interface Cards (NICs). The key point was that there was a generation of processors performant enough to satisfy demanding workloads linked to telecoms network functions like routers, firewalls, gateways, etc. This coincided with the emergence of system-level performance tuning capabilities like Data Plane Development Kit (DPDK), Shared Root I/O Virtualisation (SR-IOV) and Vector Packet Processing (VPP) [7]; distinct, and in some cases complementary approaches to expediting packet transfer from NICs through the host Operating System to the guest or “user space” application. Such capabilities, along with other set-up tweaks in the Linux network stack [8], or the way in which CPU and memory resources are allocated (e.g. CPU pinning), could ensure high packet throughput performance of network-oriented applications without the need for any specialised hardware (note: this is not limitless however, there are still cases like very high throughput backbone routers where more customised flavours of silicon for packet forwarding are needed).

At the present time, there are potential drivers for operators of telco cloud infrastructure to consider additional hardware types. As well as alternative CPU architectures (like ARM versus x86-based), the potential use of *specific* hardware for *specific* workloads has emerged. In some ways contradictory to the original “NFV/SDN” ethos, the use of custom hardware like accelerator cards to handle some of the more process-intensive aspects of a virtualised Radio Access Network (vRAN) has been proposed by a range of silicon vendors. Such dedicated acceleration cards are mostly housed within the parent server and are connected to the server’s motherboard via PCIE slots. The use of Graphics Processing Units (GPUs) has also been proposed in the context of vRAN, with such technology being also suitable for video rendering and AI-oriented use cases.

Another flavour of specialised hardware has emerged around Data Processing Units (DPUs) and Infrastructure Processing Units (IPUs). Although slightly different terminology, these are essentially geared towards similar use cases, aimed predominantly around “network offload”. In simple terms, DPU/IPU cards – hosted on a parent server via PCIE slots, but acting in many ways like a self-contained “mini server” – can permit direct processing of certain workloads associated with network processing tasks like virtual switching and routing, without requiring attention by the CPU. This in turn frees up CPUs for the “higher value” processing tasks directly linked to the guest workloads resident on the server; this is one reason why hyper-scalers have an interest in this technology, as the CPU footprint is directly monetisable. For telcos and their on-premises telco cloud, DPU/IPU technology *could* be considered for network offload use cases, but additional use cases should also be explored. For example, since DPUs/IPUs operate with a self-contained OS, there may be potential opportunities around security and tenant isolation.

Table 1 summarises three items that matter most around the telco cloud trend of “hardware heterogeneity”.

Table 1: Hardware heterogeneity

Hardware heterogeneity - 3 key things	Why does this matter?
1. Cost benefits validation	Any new hardware deployed on host servers to provide accelerated performance and/or offload benefits, must realise a net techno-economic benefit compared with using standard CPUs and system-level performance tuning. This should equate to whole life costs which factor in aspects like power consumption.
2. Interoperability and re-usability	The introduction of new hardware should avoid vendor-specific application support; e.g. RAN accelerator card from vendor A only works with vRAN software from vendor B, or DPU/IPU card from vendor C only supports guest applications from vendor D. Open access to hardware offload via suitable APIs in as flexible a manner as possible is crucial, and “single use” hardware should be avoided.
3. Management	In the case of DPUs/IPUs running their own operating system, the “server within a server” architecture creates extra management challenges for the network operator. Consistent tooling and instrumentation to enable suitable DPU/IPU system management in tandem with that of the parent server, is necessary.

7 “Characterizing the Performance of Concurrent Virtualized Network Functions with OVS-DPDK, FD.IO VPP and SR-IOV”, N. Pitaev et al, 2018 ACM/SPEC International Conference
8 https://access.redhat.com/sites/default/files/attachments/20150325_network_performance_tuning.pdf

Towards cloud native network functions and workloads

In relation to the virtualisation stack and execution environments of telco workloads running on on-premises NFVI/telco clouds, early deployments relied heavily on hypervisor/VM technology, while in more recent years there has been a major shift towards container-based technology (Fig. 1). While still possible to host containers inside VMs (so the technology stacks co-exist), the real point is that Network Functions (NFs) have been developed by vendors to use container run-times (like docker, containerd, cr-io) and be managed at scale by container orchestrators (like Kubernetes). This technology trend goes well beyond a "different flavour of virtualisation" as it also aligns with the principle of much more modular and portable software implementations, and how such workloads are managed throughout their lifespan.

It is helpful to directly cite the Cloud Native Compute Foundation (CNCF) definition of "cloud native" [9] as; "*Cloud native practices empower organizations to develop, build, and deploy workloads in computing environments (public, private, hybrid cloud) to meet their organizational needs at scale in a programmatic and repeatable manner. It is characterized by loosely coupled systems that interoperate in a manner that is secure, resilient, manageable, sustainable, and observable. Cloud native technologies and architectures typically consist of some combination of containers, service meshes, multi-tenancy, microservices, immutable infrastructure, serverless, and declarative APIs — this list is non-exhaustive*".

As can be seen, this describes an overall approach and characteristics, without being prescriptive around specific technology choices. These definitions are useful as they should be generically applicable to any workload or application that runs on cloud infrastructure - whether they be Network Functions (NFs), IT applications or AI workloads. That said, a sticking point in the telco cloud landscape has been to marry together some of the more generic properties of cloud native as per the above outline description, with the very specific – and in some respects demanding – properties of network-based functions and workloads required by telcos (high availability, resiliency, security, performance and observability, to name a few). A working group within the CNCF (now called CNTi - Cloud Native Telecom initiative) has undertaken studies to bridge this gap. As well as detailing the preferred approach to lifecycle management of cloud native network functions in relation to validation, inter-dependencies, automation, security and resilience, a number of "factors for compliance" for NFs from vendors to be considered truly cloud native are proposed [10]. A further industry position around the topic of cloud native methodology from the perspective of the network operator, was published by the NGMN (Network Generation Mobile Networks) alliance in autumn 2023 [11].

Essentially, cloud native can be viewed as a property of the software/application/NF that runs on cloud infrastructure and in that context we outline the three crucial items from a telco perspective (**Table 2**).

Realisation of these properties will help telcos achieve a platform-oriented approach to telco cloud evolution and more seamless horizontal scale-out of the cloud infrastructure footprint, avoiding so-called "siloed" stacks. The Linux Foundation (LF) "Sylva" project [12] has many aims that align with this aspiration.

Table 2: Cloud native network functions

Cloud native network functions – 3 key things	Why does this matter?
1. Portability	Software functions should not be bundled and tied into specific operating systems, hardware or other such system dependencies; certifications should also not be overly restrictive and inhibit flexible technology choices. They should run on any modern generation of x86-based CPU and standard NICs, while in relation to storage should not presume any particular brand of storage array or disk.
2. Observability & Reachability	Software functions should enable pluggable observability agents without being overly prescriptive around bundled stacks preferred by the vendor. They should presume to work in an air-gapped deployment without recourse to "calling home" to a supplier's network at any point during deployment.
3. Performant	Software functions should ship with fairly generic performance tuning profiles. Solutions that work with good performance across a wide range of hardware types are generally preferred to solutions highly tuned for performance on one very specific hardware variant, but potentially detrimental on others.

9 https://github.com/cncf/toc/blob/main/DEFINITION.md
10 https://github.com/lfn-cnti/bestpractices/blob/main/doc/whitepaper/Accelerating_Cloud_Native_in_Telco.md
11 "Cloud Native Manifesto - An Operator View", Version 1.0, September 2023. https://www.ngmn.org/wp-content/uploads/NGMN_Cloud_Native_Manifesto.pdf
12 https://sylvaproject.org/home/

Towards cloud native automation, orchestration and operations

The broad trend of "cloudification" of telco infrastructure has brought with it significant opportunities to deliver services in a more flexible and agile fashion. Indeed, while ideas around NFV/SDN were formulated and brought together in standards fora such as ETSI, a golden objective was to facilitate high degrees of automation and orchestration of VNFs residing on NFV Infrastructure. After all, "softwarisation" of network functions running on a de-coupled infrastructure layer, offered the promise of (and lends itself to) a very flexible and agile mode of operation.

The earliest incarnations of orchestration systems for workloads running on NFVI/telco cloud were often limited in scope, due to specific requirements of specific VNFs and their associated approach to element management. These factors could often inhibit the level of automation achievable in practice. Furthermore, the essentially hierarchical nature of the NFV MANO (Management and Orchestration) framework presents challenges around placement and ownership of "state" in relation to how a particular VNF (or set of VNFs comprising a so-called Network Service) consumes resources from the underlying infrastructure. The use of so-called "imperative" actions which focusses on how to do something in a *prescriptive way* can also be viewed as a fairly rigid methodology.

Cloud-native automation [13] is a promising architecture embedding principles like intent-based declarative orchestration of resources (which focusses on *what* to do rather than *how*, by leveraging suitable APIs) into the framework. This framework also moves from the hierarchical model of early NFV MANO, to a decentralised model with accurate state information applicable to distinct layers of the architecture held more locally at that layer.

A key part of the cloud-native automation concept is the way in which it is supported by tooling, and much of that has stemmed from, and is built upon, the "DevOps" approach to software development which embodies the principles of Continuous Integration, Continuous Deployment and Continuous Testing (CI/CD/CT). By leveraging GitOps repositories for ensuring "single source of truth" version control of software, data models, and so on, a consistent methodology for lifecycle management of each layer in the technology stack – infrastructure, platform, services/applications – can be adopted [14]. The application of intent-driven declarative data models leveraging Infrastructure-as-Code (IaC) can further ensure both accuracy and a high degree of flexibility when co-ordinating telco cloud resources. Another vital piece of the puzzle (see also **Fig. 2**- future telco cloud vision) is the role of telemetry within the broader scope of observability frameworks. To achieve efficient and performant allocation of resources in a telco cloud environment, the collation and intelligent insights gleaned from data metrics is paramount and will also be a major facet of "AIOps" practices.

Table 3 summarises three items that matter most around the telco cloud trend of "cloud native operations". Regarding the last point in Table 3, the concept of "build-to-order" platform environments for telco cloud is not new, a Horizon 2020 project "next generation platform-as-a-service" explored and progressed many of these concepts [15]. More recently, the Linux Foundation "Nephio" project [16] is aiming to cover some of the items listed above.

Fig. 2: Future Telco Cloud Vision

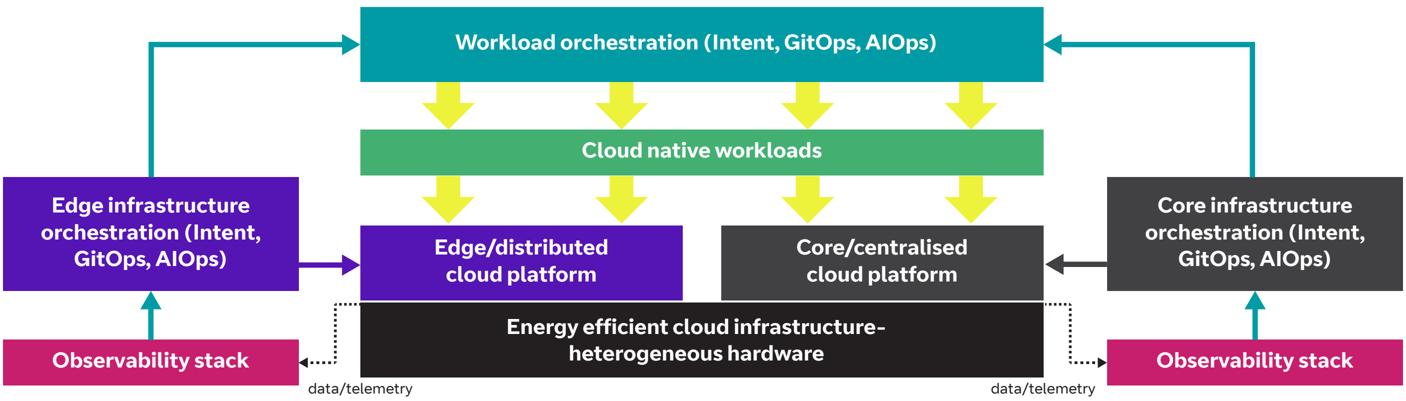


Table 3: Cloud native operations and automation

Cloud native automation and operations – 3 key things	Why does this matter?
1. Platform-oriented	The increased adoption of cloud native automation and operations fit well with the aspiration for a platform approach to telco cloud, where infrastructure can be scaled horizontally and siloed, application-centric stacks can be avoided.
2. Intent-oriented	Intent-driven and declarative data models used in conjunction with GitOps can marry together required accuracy of the state of any given resources of a particular layer of the telco cloud stack, with a high degree of repeatability and flexibility.
3. Singular/Unified approach to automated lifecycle management across disparate hardware types	The aspiration to manage a potentially large number of disparate compute instances in multiple locations, demands high levels of automation. Hence, cloud native automation principles are equally important for bootstrapping and life-cycle managing the underlying *infrastructure* – possibly leveraging disparate flavours of hardware, operating system, etc - used to host cloud native workloads.

13 "Cloud-native automation: the transformation of CSP networks", A. Gaili, Analysys Mason report, May 2023.
14 "NFV evolution: Towards the Telco Cloud", ETSI White Paper No. 65, March 2025.
15 "A Vision for the Next Generation Platform-as-a-Service", S. Van Rossem et al, IEEE 5G World Forum, July 2018.
16 https://nephio.org/

Towards the edge

The deployment of computing resources at the network edge to host an array of workloads and applications has had widespread discussion and debate in the telecoms industry, much of it tied to the inception of 5G networks. Opinions across industry still vary in relation to the most credible use cases that make sense in an edge compute context. Moreover, "where exactly is the edge?" is context-specific, but if being discussed in relation to BT's network fixed footprint as an example, the "near" edge could be metro locations (of which there are about 100) and the far edge could be at selected local exchange locations (of which there are around 1000). It would be possible to extrapolate the definition of far edge to thousands of mobile Radio Access Network (RAN) locations, but this would still be considered an extreme variant of *network edge*.

When the edge moves closer to the data source and reaches into the *customer environment*, there are two further categories of edge to consider. Firstly, the *customer premises edge* within retail stores, factory floors, and vehicles; this may focus on extremely low latency use cases like Augmented Reality/Virtual Reality (AR/VR), industrial automation and control, and autonomous vehicles. Secondly, the *device edge* consisting of the communications endpoint; this could encompass a range of scenarios where data collection and pre-processing is done on end user computing hardware, for example a 5G User Equipment (UE) handset.

From a telco perspective, there are three main credible drivers for considering the deployment of cloud computing resources at the telco's network edge [17]. The first would be driven by performance-demanding workloads that require very low latency, for example the Distributed Unit (DU) component of vRAN, AR/VR services, or possibly some AI-oriented applications. The second would be to help deliver content in the most efficient way possible by more localised/deeper offload, for example by pushing Content Delivery Network (CDN) instances deeper to reduce core network traffic [18]. The third reason would centre around *how* and *where* certain data gets processed, with a view to ensuring security and data privacy of certain applications by striving to keep data "on-net", and/or within certain geographical domains which perhaps could not be guaranteed by sending such traffic "off-net".

While the case for edge compute is still subject to ongoing scrutiny, we can outline the three principal considerations from a telco perspective (**table 4**).

Table 4: Towards the edge

Network edge - 3 key things	Why does this matter?
1. Techno-economics	Edge compute for workloads with very low latency should only be considered if such requirements cannot be met by existing telco cloud footprints and the associate round-trip times (RTTs) achievable with current architectures. Also, hosting certain services on network edge compute resource to achieve more localised traffic offload should realise a net techno-economic benefit compared with current practices where compute infrastructure is generally much more centralised. This should equate to whole life costs which factor in aspects like power consumption, and available accommodation to host edge compute infrastructure.
2. AI at the edge – hype or reality?	While there is much industry speculation and hype around "AI at the network edge", operators of telco cloud infrastructure must better understand the practical ramifications of this scenario actually playing out based on realistic and credible assumptions. In some respects it will boil down to simple techno-economics in terms of "*where in the end-to-end continuum*" does it make most sense to process such workload types.
3. Scaling out - operational requirements	Since the number of edge compute instances could (hypothetically) become very large, and mixed workloads like vRAN, AR/VR, and AI could dictate specialised hardware to ensure performance criteria are met, the considerations in earlier sections around hardware heterogeneity and cloud-native automation/operation, could become particularly pertinent. Streamlined management of diverse hardware on the one hand, and diverse workload types on the other, is a major and complex challenge. The LF "Sylva" project cited earlier [12] has a specific focus to cater for "edge" scenarios and aims to avoid siloed application-specific technology stacks which run on dedicated/ring-fenced hardware.

12 https://sylvaproject.org/home/
17 European Commission Edge Observatory for the Digital Decade Case Study - "*Edge, virtualisation and the future of telecom network*", https://digital-strategy.ec.europa.eu/en/policies/edge-observatory
18 https://newsroom.bt.com/bringing-the-action-closer-to-the-customer-with-new-content-delivery-technology/

Energy efficiency

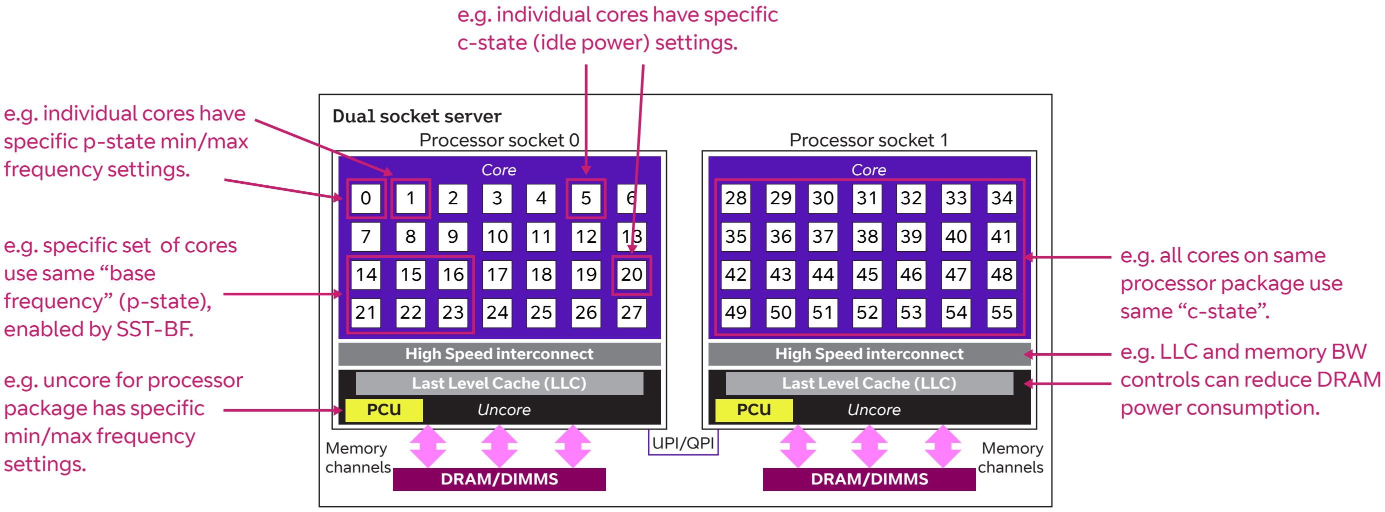
The need to drive down power consumption is of particular importance to telcos. In parallel with initiatives to migrate and switch off energy-hungry legacy infrastructure [19], the equipment that comprises "telco cloud" infrastructure (being made up of commodity servers, switching and storage) is expanding. While such infrastructure can be power hungry (even in idle states with low utilisation), there are a range of features that allow optimisation and tuning of power settings. With regard to commodity servers using Intel x86 processors, tuning can be applied on a per CPU core basis to c-state and p-state settings; the former affects how cores may enter sleep states when idle, while the latter affects the tuning of core frequency (clock speed) while cores are active [20]. Additional manipulation of p-state settings for Intel processors is feasible using so-called Speed Select Technology [21].

The important point here is that different methods of manipulating core processor settings affect the corresponding processor power consumption, as well as the performance of workloads that run on those processors. There is generally "nothing for free" and power and performance will always be traded off against each other. The clever part of some of these advanced features is that different settings can be linked to distinct cores (within a "multi core" processor- see **Fig. 3**), and correspondingly, different workloads associated with those cores. So, it is possible to cater for completely distinct workloads in terms of their required performance and associated contribution to overall power consumption [22].

Successful exploitation of processor tuning to balance performance and power efficiency relies upon accurate and real-time data on resource utilisation across a wide range of metrics, often realised by telemetry-based data collection and pipelining. Indeed, this intelligence-driven approach to telco cloud resource management and visibility can extend to other component parts of infrastructure (which also contribute to overall power consumption) like processor "uncore" [23], as well as system memory [24]. More generally, the collection of real-time resource utilisation insights at the physical layer of the telco cloud infrastructure will also be a key "enabler" for cloud native automation and orchestration as discussed earlier (**Fig. 2**).

Looking forward, newer generations of server processors with much lower "Thermal Design Power" (TDP) envelopes are being developed, including ARM-based processor architecture, as well as more energy efficient x86-based solutions. The potential introduction of heterogeneous hardware (DPU/IPU/GPU as discussed earlier) will also bring with it specific considerations around power consumption. Further ahead still, more radical solutions like neuromorphic computing – of interest already for AI use cases [25] - may present a more transformative opportunity to drive down computing power consumption, although how this technology may fit into a telco cloud context is still at a very early stage of exploration. Table 5 summarises three items that matter most around the topic of telco cloud energy efficiency.

Fig. 3: Potential Server Processor Power Configurability (Available in Certain Intel x86 Xeon Processors)



19 https://www.bt.com/bt-plc/assets/documents/about-bt/tech-fellowship/towards-a-more-energy-efficient-future-telecoms-network.pdf
20 "Understanding the Performance and Power Saving Trade-offs of Server Sleep States", A. Griffiths, A. Morsman and P. Veitch, IEEE CloudNet, Nov 2023.
21 https://docs.kernel.org/admin-guide/pm/intel-speed-select.html
22 "Balancing NFV Performance and Energy Efficiency", P. Veitch, C. Macnamara, J.J. Browne, 25th Conference on Innovation in Clouds, Internet and Networks (ICIN), March 2022.
23 "Uncore Frequency Tuning For Energy Efficiency in Mixed Priority Cloud Native Edge", P. Veitch, C. Macnamara, J.J. Browne, 2024 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), June 2024.
24 "Memory Bandwidth Throttling to Maximise Performance and Reduce Power Consumption", P. Veitch, C. Macnamara, J.J. Browne, 27th Conference on Innovation in Clouds, Internet and Networks (ICIN), March 2024.
25 "An Overview of Neuromorphic Computing for Artificial Intelligence Enabled Hardware-Based Hopfield Neural Network", Z. Yu et al, IEEE Access Vol 8, April 2020.

Table 5: Energy efficiency

Energy efficient telco cloud – 3 key things	Why does this matter?
1. Performance v power	Improving telco cloud infrastructure efficiency by leveraging power-saving measures, is only feasible if service agreements and performance KPIs are not jeopardised. Ideally, future processor generations to underpin telco cloud should combine very low Thermal Design Power (TDP) characteristics with high performance and mitigate the need for heavily customised tuning of processor resources where possible.
2. Insights & Observability	Fine-tuning of processor-oriented settings to better manage power and performance in the telco cloud environment is heavily reliant on accurate and real-time intelligence; collection of accurate and *meaningful* metrics is paramount.
3. Holistic approach	Telco cloud power consumption should be viewed from all angles. For servers, all aspects of the processor (e.g. core and uncore) as well as memory components should be factored in, alongside external hardware, which could be NICs, GPU and IPU/DPU cards. Storage and switching components to be found in the telco cloud data centre environment should not be overlooked.

Closing Remarks (Including Customer Perspective)

This article has summarised some of the key telco cloud trends which will influence operator decision-making in coming years: hardware heterogeneity, cloud-native workloads and operations, network edge ramifications and sustainability. For each of these areas of interest, we have distilled three key aspects that matter most from a telco perspective.

The crucial aspect that must not be lost within an interesting technical discussion is "*what does it all mean for the customer?*". A notable observation here is that early positioning of NFV/SDN was very much anchored in achieving much greater operational flexibility and portability, with accelerated deployments of network functions due to their "softwarisation". This was, and probably still should be viewed, as an *enabler* for the way in which services are delivered by telcos, and hence the claimed benefits to end customers would be similar to many of those linked to "digital transformations" (simplified customer journeys, speed of delivery, self-serve, etc). Mostly, however, customers do not need to know or care about the technology telcos use within their core infrastructure to deliver the services and applications that run on top; they expect brilliant and seamless service delivery and rightly expect the telco to make decisions around technology needed to meet those expectations.

It could be stated that the first wave of telco cloud implementations and their associated technology choices (e.g. "pre cloud native") were not best equipped to realise many of the latent benefits. More mature telco cloud, and the direction of travel for which these deployments are likely to evolve – highly automated, programmable and adaptable – should hopefully unlock many of the fundamental operational enhancements, from which end customers should also profit.

Author bios

Paul Veitch
Paul Veitch leads the Cloud Infrastructure research centre of excellence and is a Distinguished Engineer at BT. He has published widely and his current research activities include telco cloud resource optimisation, energy efficiency and novel compute architectures. Paul holds M.Eng. and Ph.D. degrees from the University of Strathclyde, Glasgow.

Nick Heatley
Nick Heatley is a Principal Network Architect and Distinguished Engineer at BT with focus on Infrastructure topics. He is currently working on the evolution of BT's Network Cloud, a private cloud computing platform, and has 25 years' experience working on IP and Mobile Core infrastructure, and IPv6 deployment for EE/BT.

Michel Grech
Michel Grech is a Senior Manager in the Network Strategy team supporting BT in shaping and optimising its network infrastructure. He has had several roles in the telecoms industry, having previously worked for network equipment vendors and TMT-focused consulting firms . He is an alum of London Business School, Aston University and Essex University.

Glossary of Terms

AI	Artificial Intelligence
API	Application Programming Interface
CDN	Content Delivery Network
CI/CD/CT	Continuous Integration/Continuous Deployment/Continuous Testing
CNCF	Cloud Native Compute Foundation
CNF	Cloud Native Network Function
CNTi	Cloud Native Telecom initiative
COTS	Commodity-off-the-Shelf
CPU	Central Processing Unit
DPDK	Data Plane Development Kit
DPU	Data Processing Unit
DU	Distributed Unit
EPC	Evolved Packet Core (4G core)
ETSI	European Telecommunications Standards Institute
GPU	Graphics Processing Unit
HTTP	Hyper Text Transfer Protocol
HTTPS	Hyper Text Transfer Protocol Secure
IaC	Infrastructure as Code
IoT	Internet of Things
IPU	Infrastructure Processing Unit
KPI	Key Performance Indicator
LF	Linux Foundation
MANO	Management and Orchestration
ML	Machine Learning
NF	Network Function
NFV	Network Functions Virtualisation
NFVI	Network Functions Virtualisation Infrastructure
NGMN	Next Generation Mobile Network
NIC	Network Interface Card
OS	Operating System
PaaS	Platform-as-a-Service
PCIE	Peripheral Component Interconnect Express
RAN	Radio Access Network
SA	Stand Alone (5G core)
SASE	Secure Access Service Edge
SDN	Software Defined Networking
SD-WAN	Software Define Wide Area Network
SR-IOV	Single Root I/O Virtualisation
TDP	Thermal Design Power
uCPE	Unified Customer Premises Equipment
vCPE	Virtualised Customer Premises Equipment
UE	User Equipment
VIM	Virtual Infrastructure Manager
VM	Virtual Machine
VNF	Virtualised Network Function
VPP	Vector Packet Processing
VR/AR	Virtual Reality/Augmented Reality
vRAN	Virtualised Radio Access Network

BT Group