Over the last five years, AT&T has led the industry as a disruptive innovator of Network Function Virtualization (NFV) and Software Defined Networking (SDN) by executing a network transformation initiative that’s spanned from the network edge to the core. In addition to virtualizing 75% of its network functions by the end of 2020, AT&T embarked on one of the largest OpenStack-based cloud infrastructure buildouts in the world, and disrupted the SDN industry by open sourcing its innovative ECOMP platform – Now called Open Network Automation Platform (ONAP). AT&T reached additional major milestones at the end of 2019: 100% of the data traffic running through the multiprotocol label switching (MPLS) tunnels, which connect the core elements of our network together, is now under SDN control.

It also activated the first 400-gigabit SDN-enabled optical connection carrying live internet traffic across its network for customers. The connection - between Dallas and Atlanta - is the first in the industry and is built on flexible, low-cost White Box hardware supporting OpenROADM standards. Just how big is a 400G pipe? Basically, a customer could download all 70 high definition episodes of Game of Thrones in less than 8 seconds (400-Gigabit Network Connection, 2019).

Figure 1.0 The Domains of AT&T’s Network Transformation

Not surprising, AT&T’s velocity of network innovation is speeding up. One recent example is the first deployment of the Network Cloud that enabled the 2019 5G launch. This is part of the evolution towards a distributed, high-performance network edge where customers and application developers can build, integrate, and run highly-optimized applications directly with AT&T’s 5G network while running their application’s on public cloud infrastructure. Additionally, the industry’s growing ecosystem of open disaggregated software and hardware, SDN automation augmented with Machine Learning (ML), Artificial Intelligence (AI), and network fabric automation have already become key facets that are necessary to meet the demands of future traffic growth.

It is estimated that 70% of the global population will have mobile connectivity by 2023 while IoT will account for 50% of all networked devices (Cisco, 2020). More pronounced for AT&T has been the 470,000% of mobility data growth since 2007 (Andre Fuetsch, Opening Up for 5G)
5G has reached 190 markets covering over 120 million subscribers by April of 2020. And it is not stopping there - AT&T Plans to provide nationwide 5G coverage in the Summer of 2020. 5G capabilities will also continue to evolve and will be one critical factor in enabling future Enhanced Mobile Broadband (eMBB) use cases like immersive AR/VR experiences that have a minimum control plane latency (idle state to continuous data transfer) of 20ms (AT&T Labs & AT&T Foundry, 2017). The user plane latency minimum requirements are 4ms for Enhanced Mobile Broadband applications and 1ms for Ultra-reliable low-latency communications (URLLC) like self-driving cars or mission critical applications.\(^1\) The “connection density” for quality of service (QoS) in the 4G RAN can support thousands of mobile devices within .38 miles or one kilometer today – 5G “connection density”-required to support 1 million (ITU, 2017).

So, what’s next in AT&T’s network transformation? There are seven principles that can answer that question. While they are not altogether comprehensive, they will illuminate what’s transpiring today, the network vision of the future, and the relevancy to customer needs of tomorrow.

1. **Network growth necessitates economies of scale that can only be achieved from interoperability and open disaggregation**

This ecosystem is the genesis for innovative network solutions that meet the large scale and fast-evolving feature requirements for MAN/WAN networking - encompassing cell-site routers, metro Ethernet service and RAN backhaul routers/switches, Internet and VPN service edge routers, carrier intra-PoP interconnect fabric, backbone core routers. There are three essential elements to AT&T’s network transformation ecosystem:

- The open dNOS, like AT&T’s Vyatta NOS, that’s disaggregated from the underlying hardware (chassis, controller, line cards, etc.)
- Well-defined standard interfaces and APIs enabling modularity with the ability to mix and match applications
- Independent scaling of control and user/data plane functions to drive an open ecosystem. AT&T is leading efforts in industry bodies like the Open Compute Project (OCP), Open Networking Foundation (ONF), Open-RAN (O-RAN), OpenROADM, and the Linux Foundation to commoditize interoperable network functions and compute modularity without being cost prohibitive.

A significant impediment to NFV and SDN scale is the current Virtual Network Function (VNF) and Containerized Network Function (CNF) variability across carrier cloud infrastructure. The Common Network Function Virtualization Infrastructure Telecom Taskforce (CNTT) is tackling this challenge by addressing:

---

\(^1\) User plane latency is the contribution of the radio network to the time from when the source sends a packet to when the destination receives it in ms. Requirements stated specify latency for unloaded, single user, and small IP packets (0byte payload + IP header)

AT&T Intellectual Property. All rights reserved.
Heavy operational overhead integrating VNFs/CNFs with diverse underlying infrastructure
Increased complexities resulting from varying versions of network function applications
Minimal VNF provider testing and certification requiring lengthy carrier test cycles
High carrier development costs due to multiple custom platforms for VNFs/CNFs
Slow adoption of cloud-native NFV applications

Overall, the CNTT mission is to minimize variability of architectures in use by carriers in order to accelerate time-to-market. Through the GSMA and Linux Foundation strategic partnership, the Open Platform for NFV (OPNFV) Project will be able to deliver a community reference model and a normalized set of community accepted architectures.

White Box hardware/software disaggregation using open dNOS, such as AT&T’s Vyatta NOS, is proliferating from the network edge to the core

The flexibility, scalability, and operability benefits of the White Box enables AT&T to decrease its cost of operating the network while being able to iteratively refresh the technology base at scale. One example of AT&T’s White Box and open NOS design contributions are contributions to the Cell Site Gateway Router (CSGR) specifications with the Open Compute Project (OCP). As a result, Original Design Manufacturers (ODMs) have adopted and built the CSGR White Box switch to AT&T’s specifications with AT&T’s Vyatta NOS.

An emerging AT&T led White Box initiative is the Distributed Disaggregated Chassis (DDC) White Box design that was submitted to the Open Compute Project (OCP) in September of 2019. In contrast to traditional high capacity routers using modular chassis designs with separate equipment and fabric line cards, the DDC White Box is a much smaller compute form factor that contains the integrated fabric cards in addition to equipment cards – The backplane in the traditional line card chassis is simply replaced with external cabling to the White Box itself. This DDC design allows for faster and larger scale-out in comparison to traditional individual line card chassis’ limited capacity (AT&T Submits Design for Service Provider-Class Routers to the Open Compute Project, 2019).
A second example is the next generation of Converged Access which includes the 5G Fronthaul Gateway (FHG), SDN-Enabled Broadband Access (SEBA), and Converged Access Switch (CAS); a White Box an initiative based on OCP, Open Networking Foundation (ONF), ONAP, and Akraino open source software. FHG will enable support of Radio over Ethernet (ROE) for 4G and Time Sensitive Networking (TSN) for real time 5G traffic with aggregation of LTE CPRI and 5G eCPRI signaling at the cell site. With ROE technology, line code aware and structure aware modes can help achieve significant fronthaul bandwidth savings of 20-40%. Lastly, a 70% fronthaul bandwidth reduction is possible from moving the Low PHY function from the BBU to the FHG at the cell site locations. An opportunity to unify the Baseband Unit (BBU) architecture within the Centralized RAN (C-RAN) Hub is possible by allowing legacy 4G LTE radios and 5G New Radios (NR) to be served by a common virtualized Distributed Unit (DU) Pool.

A final example is an SDN-controlled “open hardware” solution, or “Gray Box”. This platform is developed and supplied by a single supplier but conforms with open-interoperability requirements and standards. For example, a pure White Box solution isn’t plausible for optical components due to its inherent analog nature. Simply put, a pure white box solution isn’t plausible for optical components. What is plausible is the disaggregation of these fiber optic transport systems and the separation of data and control plane. The AT&T led OpenROADM standards body, has produced standards-based APIs and interoperability consistent to open white box initiatives, OCP and O-RAN. The OpenROADM MSA has close to 20 members which gives Network Operators, constituting 50% of the membership, the ability to work with equipment suppliers in close collaboration. The OpenROADM MSA publishes optical specifications for data plane interoperability as well as open APIs to expose data models in YANG for model-driven design.
Network edge densification is critical for low latency, near-real time connections, high-speed requirements for 5G and low latency enterprise applications. Individual users and business customers will have many different needs, requirements, and varying levels of mobility (stationary, nomadic or full mobility). Because of this, AT&T has been focusing on building 5G and edge technologies together to give customers a choice when it comes to cloud and edge services. Multi-Access Edge Compute (MEC) will create localization for edge computing, data traffic routing, and policy management using edge compute hardware for a private 5G experience. Complimentary to MEC, Network Edge Computing (NEC) allows for the secure co-location of public cloud partners with AT&T’s 5G network edge to create powerful applications. A solution running customer premise MEC-User Plane Forwarding Functions (UPF) while deploying the MEC/Network Edge Compute (NEC) controller in the AT&T Network Cloud for applications requiring different latency and performance conditions.

The next generation of the 5G RAN will decouple the network control plane from the user plane with open interfaces for interoperability and ML/AI tools based on the Open-RAN (O-RAN) and ONAP standards. The disaggregation of the RAN on White Box and AT&T’s Network Cloud will include the Radio Units (RUs at the cell sites), Baseband Units (BBUs), the DUs/CUs for pooling/acceleration, and a RAN Intelligent Controller (RIC). The RIC will provide advanced control functionality, extend the Control Plane/User Plane split within the 5G C-RAN, and will further enhance the traditional remote radio management functions with embedded intelligence to manage non-real time and near-real time traffic. Through the O-RAN Alliance and ONAP, AT&T will be driving the open standards for non-real time/real-time radio resource management, automated policy management with AI/ML, open fronthaul interfaces, interface interoperability across RU/BBU and Control plane functions, and a stack reference design for O-CU and O-DU architectures.

AT&T’s Mobility Core and IP Communication Core are pivoting to be cloud native

Moving to the networks next transformation phase, our Mobility Core will begin to evolve to the Option 2 (5G stand-alone) Next Generation Multi-Access Core (NGxC), will be cloud native (containerized, require OVS-DPDK, necessitate High IOPS storage, etc.), will implement 3GPP
Control Plane User Plane Separation (CUPS) on a large scale to support 5G latency performance requirements, and will ultimately converge discrete physical Mobility Core instances in a common core (FirstNet, Consumer, Enterprise, and 5G cores). A significant advancement resulting from the NGxC will be the ability to implement Slice Selection Functions (SSF) to partition traffic based on service type in support of traffic separation, privacy, policies, APN management, IP addressing, WAN connectivity, and mobile traffic access controls.

IP Communications runs on AT&T’s IP Multimedia Subsystem (IMS) Core enabling Voice-over-LTE (VoLTE), FirstNet, AT&T Phone (CVoIP), and Business voice services (BVoIP) for millions of subscribers. A Next Generation IMS Core will anchor on design principles that include consolidating standard logical network functions into fewer manageable cloud native network functions (CNFs). First, it will also be decomposing CNFs into microservices to reduce network and OA&M interfaces. Second, adopt the 5G network slicing concept to assign dedicated packet core resources to IMS. And finally, improve underlying resiliency by decreasing the size of multiple deployments and load sharing across different modules.

Another major shift in AT&T’s cloud journey and the NFV roadmap is the evolution from VNFs on virtual machines (VMs) to Containerized Network Functions (CNFs). To accelerate adoption, AT&T has been progressing work to close gaps in the Cloud Native Computing Foundation (CNCF) deployment and management landscape through Airship: Airship is a collection of loosely-coupled interoperable open source tools that automatically deploys, manages and seamlessly upgrades AT&T’s virtualized network functions. It utilizes Kubernetes, as one of its tools, to provide container orchestration, resiliency, and production-grade Kubernetes clusters.

**ONAP and AT&T’s ECOMP SDN platform are evolving for diverse NFV instantiation orchestration models - APIs not GUIs**

Network functions and how they orchestrate are now diverse. Today, orchestration protocols and methods are extending beyond HEAT or TOSCA for virtualization on VMs. They require support for container instantiation using Helm, Azure Resource Manager (ARM), Ansible, and Kubernetes. Applications, microservices, OSS, Non-OSS, and vNFs require scalable, reusable user libraries, system libraries, and individual operations/methods from each component for specific customization as well. To enable this, ONAP and ECOMP are evolving to expose component functionality to provide flexible consumption of API-based service models, creation, instantiation, NFV control, inventory management, policy, control loop automation, and automation of operations.

**Cloud-Based Data Warehouse, Machine Learning, Artificial Intelligence, and policy-driven SDN-control are a powerful combination**

AT&T is embedding intelligence at the edge using string analytics, algorithms, automated policy, and ‘if-then’ statements with SDN-based control loop automation for traffic flow and performance management. Its augmented with data streaming to cloud data warehouses with
self-service SQL-accessible data storefronts. The end result is the ability to gain deep near real-time network insights not attainable before open-network disaggregation and virtualization.

**Acumos AI** - An industry-first AI platform and marketplace that provides enhanced user experience, model training capabilities and support for licensing and commercialization of AI models. Future releases will continue to build upon each other to make AI accessible to everyone.

**Practical Use Cases for an Intelligent RAN** - The scale of the 5G RAN, for example, will require automation of configuration and management of tens of thousands of cell towers and micro-cells that will be deployed for 5G. To support this in part, AT&T is contributing to Akraino Edge Stack for use cases like the Radio Edge Cloud (REC) and is developing the automation process for new launches with “zero touch” monitoring using ONAP building blocks (Configuration, SDN-control, and policy automation). Additional use cases of emerging automation in AT&T’s intelligent RAN include:

- Carrier Insertion Monitoring and Tuning enables enabling efficiencies across a landscape of varying vendors using Self-Optimizing Network (SONs) and automation of cell site configuration.
- Cell site energy management with cell site sleep/wake up based on KPI monitoring for high/low traffic conditions and eNodeB Closed Loop Repair
- The elimination of manual resolution of daily reboot instances during peak load periods using proactive automation and algorithms that optimize allocation of PCI/RSI values for 5G.

**ML/AI will improve Energy and Cooling Power Usage Effectiveness (PUE)** - Today, AT&T has thousands of Central Offices (COs), data centers, and Mobile Telephone Switching Offices (MTSOs). Reduction of energy consumption from the use of SDN control, ML/AI, and policy automation is a priority as compute demand increases. Currently, AT&T is using cloud analytics and machine learning to inject ML/AI policy-driven automation of chiller water set points. ML and policy automation provide recommendations based on feedback from AT&T’s cloud-based Enterprise Building Management System (EBMS) and data collectors from the local Data Center control systems – Neural ML and policy automation dictates control logic to each of the local cooling systems.

**7 AT&T network security paradigm is rapidly changing from the customer premise, to the network edge, and at its core**

Traffic flows no longer necessarily pass through a perimeter which serve as a significant location for many of our security controls. As a result, new means for achieving the desired security controls are evolving. Sometimes via new implementations and sometimes via a new security architecture or model. A distributed model with security localized and optimized to protect the resource helps in many aspects and represents a dramatic shift – one that SDN with its dynamic and intelligent capabilities can and will play a significant role.
Looking ahead at the next five years, the precept for AT&T’s network transformation will be anchored by three strategic principles:

**An open network ecosystem is critical for scale out, interoperability, and performance** – Insatiable customer video consumption, application “zero-tolerance” for 5G latency variance, dense penetration of cell sites/microcells, the speed of deployment is necessitating a shift in the network ecosystem. Interoperability and openness are critical to achieve not only a dense 5G edge but for economies of scale as well.

**Convergence and simplification of network function integration for improved network operability** – Converging and aggregating best of breed components by working with ODMs obtaining better insight/input into merchant silicon roadmaps, and industry supported open interface specifications give AT&T the flexibility to produce a ubiquitous model for automation.

**Network disaggregation and open source is paramount to innovation** – To ensure innovation can continue for the customer, the disaggregation of the software stack from hardware using standards-based APIs and broad adoption of open source must accelerate. There are two prominent areas of innovation: Edge-based services that can be “productized” and network data insights with Network AI that dramatically optimizes the customer experience and network performance.

AT&T’s network transformation through open disaggregation, cloud, and intelligent SDN is just beginning. 5G customer experiences, scale to meet unprecedented video consumption, Network AI-driven innovation, and sustaining a velocity of capital returns without compromising AT&T’s best-in-class network performance are just a few examples of impactful outcomes. As characterized by Andre Fuetsch, AT&T’s CTIO, at the Open Networking Summit in 2017, “This is more than just about lowering costs and achieving higher performance. Frankly that’s table stakes…”
REFERENCES


