TOTAL COST OF OWNERSHIP STUDY FOR DISTRIBUTED 5G/EDGE DEPLOYMENTS

Introduction

Network architectures need fundamental restructuring at the 5G Edge to achieve the ultra-low latencies and terabytes of throughput required by the rapid rise of data-rich and business-critical applications. Resource use efficiency must steeply increase to accommodate the expanded selection of applications.

The supply of network services is not keeping pace with application demand, unable to grow in concert with compute and storage. Current network architecture combines physical appliances and software-defined networking platforms deployed on x86 servers. It is hobbled by inflexibilities in proprietary network equipment and redundancies in packet movement between servers and switches. As more vendors and service providers enter the 5G ecosystem, the risk of being overwhelmed by data flows around their silos only grows.

5G-powered applications have raised the bar for performance, whichever scalability attribute is used as a measure - throughput, latency, jitter, or the number of sessions. The performance levels would need to rise further with the emergence of new frequency ranges (up to 1000 GHz) and the deployment of massive MIMO as dense as 64T64R. The number of data-rich applications at the edge is growing fast, running into network roadblocks constraining capacity. These changes require real-time intelligence for rapid responses to chokepoints holding up the desired levels of network performance and cost standards.

5G, along with emerging applications, such as connected cars, industrial robotics, Augmented Reality, Virtual Reality, or interactive gaming, raise the bar for cost reduction, throughput, and latency that cannot be addressed with existing infrastructure. Most vendors have been implementing 5G without changing the network technologies, with the cost points unchanged. For example, cameras used in smart city applications generate only $1 or $2 of monthly revenue, rather than the $50 that the typical smartphone user brings but generates far more data. A new or different approach is needed to reduce the cost of ownership to process data.

Fortunately, several opportunities exist to achieve the performance and total cost of ownership goals. Open networking creates an environment for innovating with diverse software and hardware. By reconfiguring and calibrating them, redundancies can be removed, and network functions consolidated to operate more applications with a shared infrastructure.

Traditional network switches cost increases disproportionately with dynamic data traffic growth

Networks use switches that operate with embedded software. They have little scope for programming to increase their data processing capacity or adapt to the changing traffic mix. Programmability is not just the ability to provision the switch but the ability to reprogram for new types of encapsulations without a complete hardware redesign. In the absence of programmability, ASICs had to be changed frequently to meet the evolving needs beyond the basic support of VXLAN.

In addition, the networking industry, except for the top CSPs, continues to rely on traditional and proprietary hardware for packet processing, significantly impacting the overall cost due to the lack of effective competition.

The capital costs decline sharply when hardware is separated from software. Commodity hardware from competing choices offered by Original Design Manufacturers costs an estimated 40-60% less than proprietary alternatives. Additionally, users have the flexibility to choose their preferred network operating system that is installed with the hardware.

Similarly, the bundling of network operating software and fabric controller with proprietary hardware keeps operating costs high. Moreover, it does not evolve with new services or updates past the warranty period.

Traditional fabrics add cost due to complex data flows

At the cloud edge, the redundant movement of traffic creates congestion. Traditional switches do not have in-built computing capability. As a result, the compute-intensive tasks are directed to servers elsewhere to perform compute-intensive network services such as load balancing, denial-of-service attack detection, intrusion detection, and setting up virtual private networks.

The back-and-forth data movement between switches, routers, and other network functions adds to capital and operating costs as the number of servers increases while network latency increases. For example, return traffic accounts for 46 percent of the traffic within the Facebook network. Likewise, networks isolated from compute and storage increase the complexity of data flows between them. Instead, users can consolidate data processing for multiple network functions on shared resources, significantly lowering costs and network latency.

Current static forwarding mechanism cannot support dynamic data traffic

Current mechanisms for forwarding packets, such as OpenFlow, are limited in adjusting rules to correct for imbalances such as data overflows at switches. New methods for higher-level data plane programming methods, such as P4, are needed as network heterogeneity expands. P4 language has greater latitude to adjust when its headers include IP addresses and physical address information to coordinate with physical network resources. With IPv6, it has room to factor in a much larger collection of addresses.

Furthermore, networks need to exercise strict control over the quality of service for mission-critical applications. They need forwarding rules not only for the IP layer but also for the underlay to allocate physical resources to achieve the desired quality of service. The enforcement of the quality-of-service rules are reinforced with hop-to-hop telemetry for In-band Operations, Administration and Maintenance (IOAM). Service quality can then be monitored and controlled at a granular level.

SRv6 with IOAM lays the basis for network slicing to achieve the goals of 5G communications to provide service quality for mission-critical applications.

Granular optimization of space and power is needed at the 5G edge

5G networks are coping with a tsunami of packet flows choking available network capacity, especially at the edge. Network capacity at the edge is also constrained by space, power, and cooling availability. As more computing is distributed to the edge, power, space, and cooling costs become acute. A survey by 451 of telecom operators found that:

- 94% of respondents indicated that 5G will raise overall energy costs
- 45% of respondents ranked site acquisition and availability of high-quality connectivity as most important to success.
- 43% of telcos worldwide use new cooling techniques to save on energy costs; this number is expected to spike to 73%.

As data processing needs at the edge rise exponentially, telecom operators face an elevated risk of prohibitive power costs putting the deployment of 5G applications in jeopardy. More 5G User Plane Functions (UPF) are placed closer to customer locations to meet the demand. However, software implementations of UPFs are quickly becoming chokepoints, putting new digital services' quality and overall growth at risk. Higher investments in general-purpose compute servers with monolithic software are wasteful as the investment in resources is disproportionately more than the gain in performance. Resource optimization at a granular level is needed for the proportionate use of resources for the outcomes achieved.

A recalibrated network architecture is warranted for microscopic visibility into resource use with telemetry and precise and flexible deployment of resources to achieve the desired outcomes. The visibility helps to pinpoint inefficiencies at a granular level and the precise deployment of resources to overcome them. Modularized container-based software helps to assign resources proportionate to the desired level of corrections. Furthermore, hardware acceleration helps to close shortfalls in software-defined alternatives.

Changing the architecture with programmability

Programmable network equipment changes the equations for costs incurred. Yesterday's capital costs are turned into operating expenses with on-demand deliveries. The disaggregation of hardware and software opens new vistas for innovation and competition. Customers have opportunities for cost reduction as they don't have to purchase bundles that include redundant components.

Kaloom embraces a brave new world of programmable cloud-native networking to improve cost-efficiencies at the edge

Kaloom believes that data processing in the 5G world calls for an entirely new network architecture that disaggregates software from the hardware of telecom equipment, is cloud-native, fully programmable, self-forming, and automated.

Kaloom's mission since its inception has been to reduce costs per connected device, per gigabit of traffic, and labor costs to run and manage distributed edge networks by reducing complexity. It wrings out inefficiencies by leveraging industry standard P4 programming language, operations on commodity hardware, and granular telemetry to achieve its high-performance goals at the lowest possible cost.

Unified Edge Fabric on programmable white-box switches enhances the cost-performance ratio

Kaloom's Unified Edge Fabric is a highly automated and virtualized network fabric. It uses Intel's programmable Tofino switch, replacing traditional switches for packet forwarding. It realizes high throughput of 100 Gbps per port with in-built computing for data processing. It adds programmability to silicon chips instead of accepting fixed ASIC functions.

The use of P4 programmable white-box switches lowers the fixed costs of hardware and increases the overall performance, thus requiring less hardware and power.

Network consolidation and heterogeneous hardware acceleration reduce CAPEX and OPEX

The programmable data plane enables Kaloom to consolidate multiple network functions such as Switch, Router, Gateway, and even network applications like the 5G UPF in a shared software-defined fabric to reduce capital of deployed dedicated hardware and operating costs of managing them. As a result, data-intensive packet processing is possible with various chipsets, including P4 programmable ASICs, FPGA, and Smart NICs, often referred to as Data Processing Units capable of handling Terabits of data that the CPU alone cannot achieve. In addition, different types of equipment used for purposes they serve best lower capital and operating costs.

Cloud-native containerization optimizes hardware footprint and operational agility

Unified Edge Fabric embeds the Red Hat Open Shift Container software, including the Linux OS and Kubernetes for orchestration. It significantly raises space utilization at the edge using common Kubernetes masters and worker nodes for network operations and applications. Embedding Red Hat OpenShift in Unified Edge Fabric potentially saves separate licensing and hardware costs for customers. Furthermore, network, compute, and storage nodes share the same underlying container-based execution environment with a standardized Linux operating system – bringing agility and OPEX savings to the data center operations.

Additionally, containers make more modest demands on hardware resources compared to alternatives like virtual machines. By containerizing network functions, Kaloom maximizes resource usage and reduces operating costs

Unified Edge Fabric with Embedded UPF offers disruptive TCO with 10x Savings

Figure 1 compares a typical edge data center, orchestrated by a Kubernetes container, with a Kaloom Unified Edge solution based on Red Hat OpenShift. The usual edge data center configuration would likely comprise the following: management switch, fabric & network overlay controllers, fabric switches, Kubernetes masters, and servers for 5G UPF to manage mobile data traffic – utilizing at least 11 Rack Units (RU) space.

Kaloom provides a 3GPP Release 15 compliant, P4 programmable, multi Tbps UPF fully integrated into its Unified Edge Fabric. Designed with the most mission-critical workloads in mind for 5G, it offers the most scalable, lowest latency, and highest performance UPF solution in the industry. Kaloom UPF is ideal for new emerging 5G Cloud Edge, Hybrid 4G/5G, and 5G Packet Core

By contrast, Kaloom's Unified Edge incorporates a management switch (1RU), and two or more switches with one a UPF (4 RU space) that are Open Compute Project (OCP) certified white boxes. Additionally, they have low-latency access to a failover site with the same capacity. Each white-box switch is equipped with Intel XEON processors and Intel Tofino P4 programmable ASIC. They are configured to run Red Hat and Kaloom Cloud-Edge fabric software components with containerized network functions (vSwitch, vRouter, VXLAN Gateway, MPLS/MPLS VPN, RedHat OpenShift Masters, as well as Kaloom UPF).

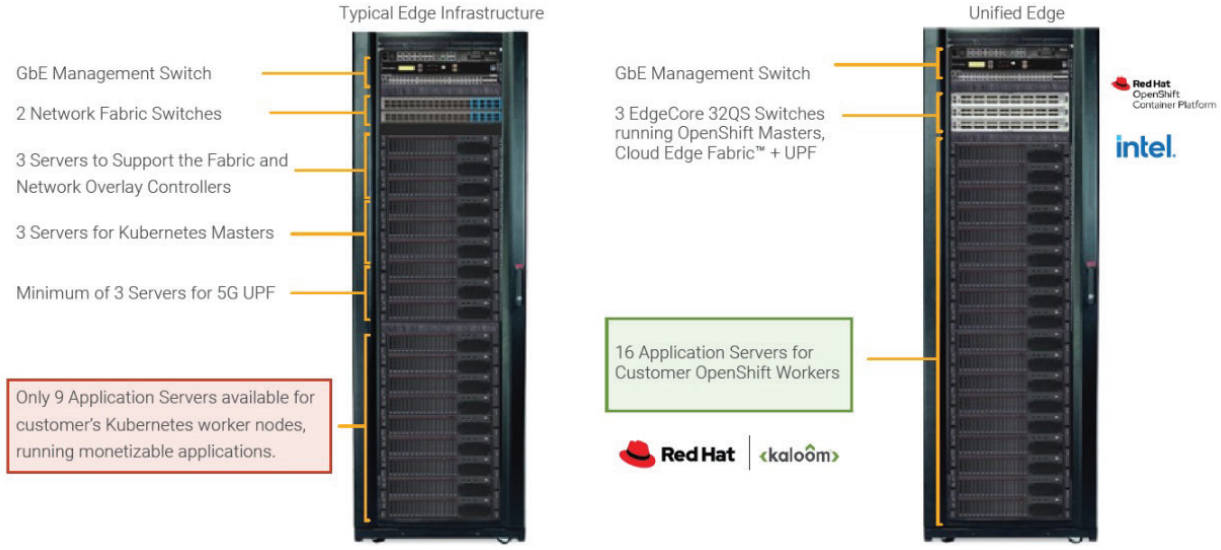


Figure 1: Typical Edge Infrastructure vs Kaloom's Unified Edge

The Unified Edge solution's optimized configuration with a Red Hat OpenShift enveloping the operations of switches, network devices such as the User Plane function, application, and storage servers, lowers capital and operations costs. The fixed cost of the Unified Edge solution is the three fabric switches (3 RU), compared to the cost of 11 RU servers of an alternate solution. It also offers a much lower recurring software cost for data processing on UPF and significantly reduced (> 5X savings) recurring expenses for power and cooling, as attested by its test results. The key outcomes were as below:

By deploying Kaloom's Unified Edge solution, mobile network operators and telco cloud service providers can benefit from TCO savings of >10x versus typical alternatives to implement 5G network at the edge.

- The baseline power consumption of Kaloom's 5G UPF solution is 959W
- The tested maximum traffic through our 5G UPF of 2.16 Tbps resulted in a measurement of 1068W or an increase of 109W or 0.05 W/Gbps
- Comparable solutions with a throughput of 2.16 Tbps, need between 5.3kW and 7.0 kW of energy which is 5x to 6.5x more than Kaloom. In another case, they need between 3.1 kW and 3.6 kW of energy which is 2.9x to 3.4x more than Kaloom
- For X-86-based UPF, the power consumption is 701.6 kW of energy or 655.7x more than Kaloom

In addition, with the transition to a Unified Edge solution, the additional nine freed-up rack unit slots for application processing earn more revenue for users without a corresponding increase in space or power consumed.

The solution yields additional benefits as follows:

- A single execution platform, under RH-OCP, fusing multiple EMS, M&O & execution platforms
- Heterogeneous hardware, each with control and data plane optimized for the business requirements
- Business and application agility realized by transformation from virtualized to containerized digital infrastructure
- A single platform for multiple networks and cloud connectivity
- Hybrid 4G and 5G can be supported with a common network fabric that supports both virtual network functions (VNF) and cloud-native network functions (CNF)
- Common network fabric for Data Center and Distributed Edge
- Converged network fabric for fixed and mobile networks

The Kaloom Unified Edge is designed for multi-tenant operations, providing services in isolated network slices. Once instantiated, the Unified Edge creates a self-forming virtualized fabric (vFabric) that runs IPv6 natively and is optimized for edge servers. Multi-tenant operations are facilitated via Unified Edge Fabric's fabric slicing capability, which ensures the complete separation of the services to fit the multi-tenant needs of the service provider, yielding additional cost savings.

Unified Edge TCO Case Study

Unified Edge - TCO Calculator

We used inputs from customers for a central office use case. The following parameters are used in the model.

- The number of central offices
- The project's depreciation cycle
- Number of PFCP Sessions (Sessions between 5G control plane (Session Management Function) and user plane (UPF))
- 5G UPF Throughput (Gbps)
- Electricity costs
- Installation costs of a Kaloom solution and a typical solution
- Software – one-time purchase or subscription

In the interest of accuracy, the hardware and software costs were adjusted for pre-defined costs built into the calculator.

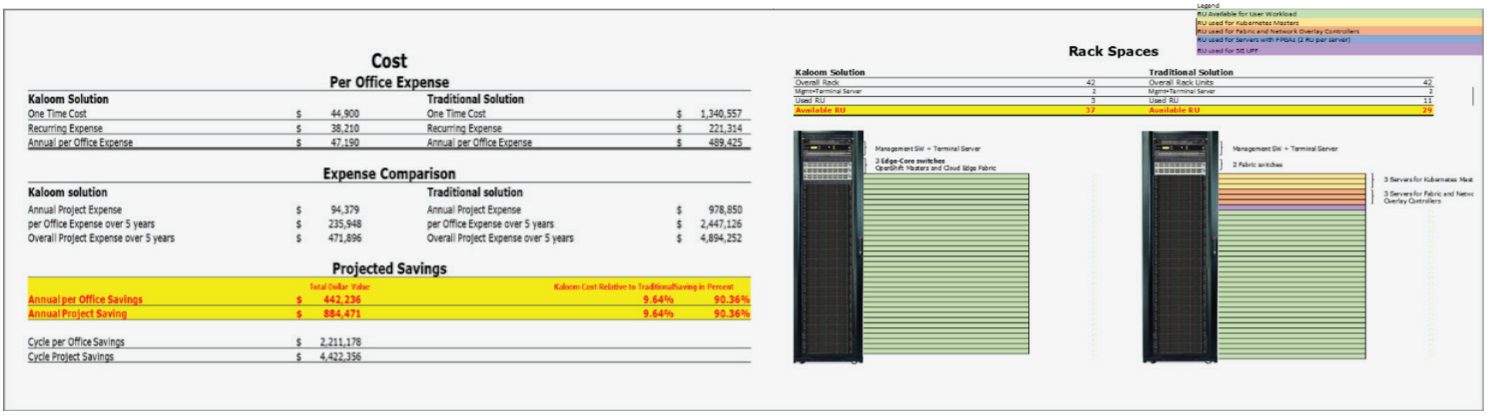
As output, the TCO calculator tool calculates summarized information regarding:

- One-time and recurring costs on a per office basis for a Kaloom solution and that of a typical solution
- Overall project costs for a Kaloom solution and that of a traditional solution
- Comparative projected savings of a Kaloom solution versus that of a traditional solution
- Comparison of the useable RUs in the selected rack-mount type for a Kaloom solution versus a traditional solution.

To compare the TCO, we have assumed:

- A deployment of UPF at two central offices (each processing the same data throughput)
- A 5-yr depreciation cycle for the TCO
- Similar cost for the installation of Kaloom and a traditional solution
- One full rack model with 42 Rack Units
- Electricity cost of $0.09 per kWh
- A software maintenance cost of 18% YoY
- List prices for hardware and software
- 200,000 PFCP sessions
- The subscription-based software pricing model for the Kaloom UPF

Figure 2 illustrates the summarized results of the exercise with a scenario as input in the TCO calculator for 600 Gbps throughput. **The overall cost savings of Kaloom's Unified Edge Fabric solution with embedded UPF is over 10X.**



Cost

Per Office Expense

Kaloom Solution		Traditional Solution	
One Time Cost	$ 44,900	One Time Cost	$ 1,340,557
Recurring Expense	$ 38,210	Recurring Expense	$ 221,314
Annual per Office Expense	$ 47,190	Annual per Office Expense	$ 489,425

Expense Comparison

Kaloom solution		Traditional solution	
Annual Project Expense	$ 94,379	Annual Project Expense	$ 978,850
per Office Expense over 5 years	$ 235,948	per Office Expense over 5 years	$ 2,447,126
Overall Project Expense over 5 years	$ 471,896	Overall Project Expense over 5 years	$ 4,894,252

Projected Savings

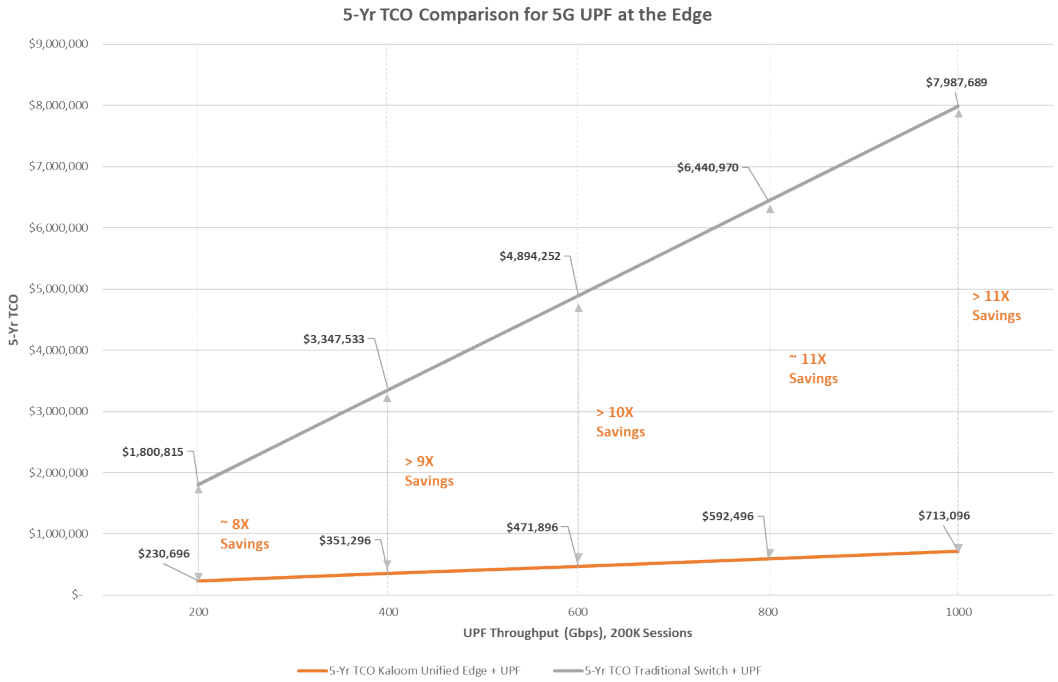
	Total Dollar Value	Kaloom Cost Relative to Traditional	Saving in Percent
Annual per Office Savings	$ 442,236	9.64%	90.36%
Annual Project Saving	$ 884,471	9.64%	90.36%
Cycle per Office Savings	$ 2,211,178		
Cycle Project Savings	$ 4,422,356		

Rack Spaces

Kaloom Solution		Traditional Solution	
Overall Rack	42	Overall Rack Units	42
Mgmt+Terminal Server	2	Mgmt+Terminal Server	2
Used RU	3	Used RU	11
Available RU	37	Available RU	29

Figure 2: Kaloom's TCO Tool

We also considered various scenarios with a low and high level of data throughput on UPF, scaling from 200 Gbps to 1 Tbps to support the growing 5G edge applications. The graph below demonstrates that the overall savings range from ~8X to over 11X. The savings get better as the data throughput increases.



The key contributing factors to this striking cost reduction are:

1. Reduced CAPEX by the consolidation of multiple cloud-native network functions on programmable switches
2. Reduced OPEX from highly competitive software cost of the network fabric and UPF
3. Reduced OPEX from significant power efficiencies – 4 to 5X better than competing solutions

Conclusion

Programmability of the data plane is the holy grail to achieve terabytes of throughput despite the stringent quality of service required at the 5G edge. Kaloom achieves the seemingly conflicting goals of cost-performance economies with a mix of container software for precision, a choice of hardware for accelerating packet processing, and consolidated networking operating environment for shared use of resources removing duplication and friction in their synchronization.

[1]https://onug.net/blog/whitebox-is-so-much-more-than-capex-reduction/

[2][2107.13694] P4COM: In-Network Computation with Programmable Switches (arxiv.org)

[3]https://networkbuilders.intel.com/solutionslibrary/intel-kaloom-create-p4-programmable-network-solutions

[4]From Energy Costs to edge computing transformation, 451 Research 2019

[5]https://www.servethehome.com/intel-tofino2-next-gen-programmable-switch-detailed/

[6]https://www.kaloom.com/download/a-unified-solution-for-the-distributed-edge

For more information please visit: www.kaloom.com or contact our sales representatives at sales@kaloom.com