The New Middle Mile 2023

A Next-Generation View of Middle Mile Infrastructure

RESEARCH BRIEF

Table of Contents

Research Briefs are independent content created by analysts working for AvidThink LLC. These reports are made possible through the sponsorship of our commercial supporters. Sponsors do not have any editorial control over the report content, and the views represented herein are solely those of AvidThink LLC. For more information about report sponsorships, please reach out to us at **research@avidthink.com**.

About AvidThink®

AvidThink is a research and analysis firm focused on providing cutting-edge insights into the latest in infrastructure technologies. Formerly SDxCentral's research group, AvidThink launched as an independent company in October 2018. AvidThink's coverage includes 5G infrastructure, private wireless, edge computing, SD-WAN, SASE, SSE, ZTNA, cloud and containers, SDN, NFV, and infrastructure security. Our clients range from Fortune 500 enterprises and hyperscalers to tier-1 communications service providers, fast-growing unicorns, and innovative startups. AvidThink's research has been quoted by the Wall Street Journal, Light Reading, Fierce Telecom, Fierce Wireless, Mobile World Live, and other major publications. Visit AvidThink at **avidthink.com**.

Zero Touch Edge Orchestration

Aarna Edge Services (AES) is here

AES solves the complexity of edge orchestration while supporting multi-cloud use cases across the edge and public clouds — all with user-friendly management from a single pane of glass.

aarnanetworks.com/products/aes

AARNA NETWORKS
AES

Discover ACE:
One scalable architecture from Edge to Multi-cloud

ACE

Arrcus Connected Edge™

Network Different™

arrcus.com

RESEARCH BRIEF

The New Middle Mile

A Next-Generation View of Middle Mile Infrastructure

Executive Summary

The middle mile holds a unique position in the communication infrastructure. This strategic Goldilocks location is the right distance from end-users and data centers, allowing it to facilitate faster communication and host next-generation services that demand a better experience. Given the importance of achieving application performance goals and the increasingly tight coupling between computing and networking, we propose examining the middle mile through a joint networking and computing lens, redefining the New Middle Mile (NMM).

The New Middle Mile is transforming from a traditional thick-to-thin, fat-tree distribution network into a more dynamic and interconnected web. This shift is driven by the changing needs of consumer and enterprise workloads, including the increasing demand for low-latency applications, real-time analytics, artificial intelligence/machine learning (AI/ML), and edge computing requirements. By adopting a more flexible architecture across all layers, from physical to network to application, the NMM can provide more flexible and resilient connectivity, better accommodate the diverse requirements of modern workloads, and deliver a superior end-user experience.

By adopting a more flexible architecture across all layers, from physical to network to application, the New Middle Mile can provide more flexible and resilient connectivity, better accommodate the diverse requirements of modern workloads, and deliver a superior end-user experience.

Part of the drive towards new flexible and performant architectures involves the continued upgrade of the underlying networks from 400Gbps to 800Gbps to 1.6Tbps, and increased convergence of metro networks (unified fabric across access types and converged IP/optical management). Disaggregation and software-defined techniques will play an essential role in improving the responsiveness and reconfigurability of the middle-mile architecture.

Despite its importance in the digital pipeline, the ecosystem has overlooked this NMM (at least compared to the last mile) and is potentially underinvested. We anticipate increased private and public funding playing a crucial role in developing and expanding NMM infrastructure comprising networking and computing. As more value creation and value creation happens inside the middle mile, that will attract more competitors. The resulting competitive landscape will improve the quality and efficiency of middle-mile services.

Capturing value from the NMM necessitates observing, understanding, and acting rapidly based on changing needs for networking and computing. Monetization in this space will rely on extracting customer surplus by efficiently allocating computing and network resources and intelligently assigning workload components for new value-added services.

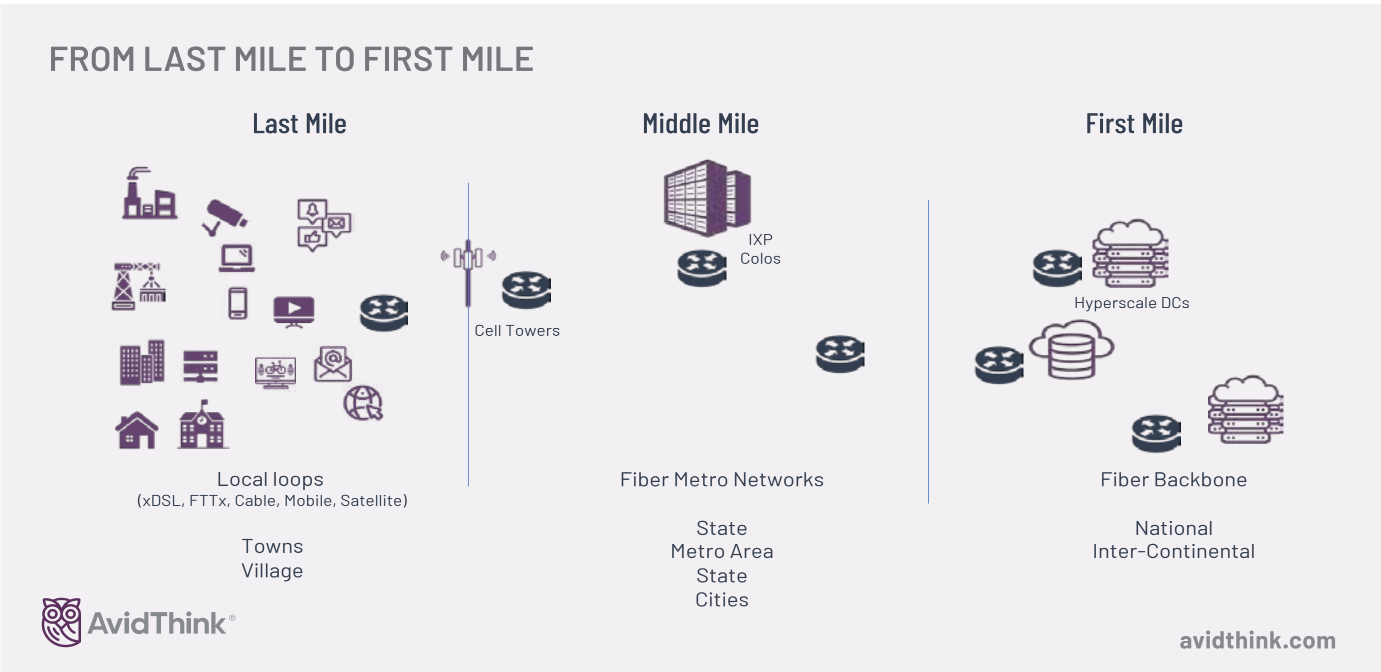
Orchestration will play a pivotal role in this process, enabling the coordination and management of diverse middle-mile assets, optimizing network performance, and facilitating the seamless integration of new application technologies and services. Today's middle mile is characterized by high search and coordination costs across a diverse ecosystem and a fragmented marketplace. By investing in advanced orchestration capabilities, middle-mile operators can unlock significant revenue opportunities and ensure a competitive edge in the market.

Introduction

The rapid growth of digital services and increasing demand for seamless connectivity have led to significant advancements in the telecommunications industry. The expansion of fiber (FTTx) connectivity, upgrades to cable infrastructure with DOCSIS 4.0, and 5G mobile access, along with the roll-out of 5G fixed wireless access (FWA), are driving **last-mile** access.

First-mile backbone infrastructure connecting large regional data centers operated by hyperscalers, co-location providers, cloud service providers, and carriers continues to grow in capacity and interconnection density, driven by cloud adoption and ongoing digitization.

The **middle mile** region that connects the first and last miles increasingly represents a strategic stage in the end-to-end digital pipeline, serving as a vital link between end-users and cloud regions. The middle mile plays a crucial role in today's workloads by facilitating faster communication and hosting next-generation services demanding superior quality of experience.



Meanwhile, at the overlay layer, we see a surge in software-defined wide area networking (SD-WAN), secure access service edge (SASE), multi-cloud network (MCN), and next-generation virtual private networks (VPNs) powered by alternatives to IPsec that combine zero-trust security with network access (ZTNA).

This industry report analyzes the evolving middle mile landscape through a combined networking and computing lens. We will explore the new definitions of the middle mile, identify key players in the ecosystem, and examine the workloads driving investment in the NMM. Additionally, we will delve into the technologies that impact and can improve middle-mile connectivity and computing, such as virtualized and disaggregated networking, orchestration, and automation.

Evolving the Middle Mile

Before we discuss the middle mile, let's define related terms — **first mile** and **last mile.** The terms first mile and last mile are subjective, based on which entity is transmitting and generating data, but we'll use typical industry definitions for our report.

The **last mile** includes wireline and wireless infrastructure providing connectivity to business and consumer users in offices, homes, public venues, and on-the-go individuals (and devices). The last mile includes fiber-to-the-premises (FTTx), cable connections, digital subscriber line (xDSL), low-earth orbit (LEO) and other satellite connections, point-to-point wireless, fixed wireless, and mobile wireless. It includes less common lightwaves and free-space optics technology.

Conversely, the **first mile** includes backbone infrastructure and long-haul links (global, cross-country) connecting large regional and national data centers operated by carriers, web giants, hyperscalers, governments, and other data center operators.

Now, we examine the traditional network-centric definition of the middle mile, which broadly refers to the portion of the network that connects network aggregation points for first-mile connections to the nearby points of presence (POPs) on an operator's core network.

Middle-mile infrastructure (consistent with US government and state definitions and not substantially different in other countries) includes:

- Leased dark fiber, interoffice transport, backhaul, carrier-neutral internet exchange facilities, carrier-neutral submarine cable landing stations, undersea cables, transport connectivity to data centers, special access transport;
- Wired or private wireless broadband infrastructure, including microwave capacity, radio tower access, and other services or infrastructure for a private wireless broadband network, such as towers, fiber, and microwave links.

Based on ongoing conversations with enterprises, telcos, hyperscalers, and networking vendors, AvidThink believes in expanding the Middle Mile concept. We suggest redefining the New or Evolved Middle Mile as additive to the traditional Middle Mile by including the following:

- Computing and storage infrastructure and facilities co-located with the networking terminus of the middle mile. We include local data centers, internet-exchange points (IXP), co-location facilities, telecom regional data centers, or other carrier-neutral multi-tenant facilities located in semi-permanent structures near fiber aggregation points or radio towers.

We could also include mobile switching offices, central offices, and cable headends as part of the definition. There's overlap with the termination of the last mile, but they do represent on-ramps to the middle mile.

Before we delve deeper into the NMM, let's explore recent business, societal, and technology trends driving change in the Middle Mile infrastructure.

New Consumer and Enterprise Demands

The demands placed on the last and first-mile infrastructure also affect the middle mile, which historically has served as a web of static connections carrying traffic to and from last-mile devices. The ongoing digitization of business and consumer activities drives middle-mile workloads. The top activities driving middle mile requirements include:

Media and entertainment

- Digital on-demand entertainment
- Increase in the creation and consumption of high-resolution digital media
- internet-based gaming
- Augmented reality and virtual reality for entertainment, travel, and other consumer visualization

Collaboration and networking

- Social networking growth
- Real-time communication and collaboration accelerated by work-from-home, work-from-anywhere trends, and increased use of telemedicine

Productivity

- Dependence on web-based SaaS applications for work, education, government services, and entertainment
- Adoption of cloud-based infrastructure services for building and hosting public and private applications

Industry 4.0 and IoT

- Migration of industrial processes with industrial IoT
- AR and VR for work-related use cases, including medical and industrial
- Increased reliance on surveillance and video capture for security, safety, and productivity
- Proliferation of intelligent and connected devices with consumer IoT, connected vehicles, and the internet of medical things (IoMT)

Cybersecurity

- Increasing digitization and connectivity of users and devices enlarge attack surfaces and increases the complexity of software and hardware stacks
- Digitization of consumer and business workflows and processes creates more cybersecurity attack targets for criminals and rival nation-states

Workload Implications for the Infrastructure

Readers are likely familiar with the use cases described above. Common to all the use cases is an increased reliance on connectivity to remote and cloud-based resources. SaaS applications, cloud-hosted storage and computing, and digital communications for work, play, education, and governance mean secure and reliable connectivity is critical.

In addition, as part of the telco move to hybrid cloud environments (mix of public clouds, public edge, and private clouds), some telco workloads will be spread across various locations, connected via middle mile networks. Disaggregated 5G RAN and core, SD-WAN, SASE, and ZTNA gateways for security, IoT aggregation hubs, and other telco-provided services will require the placement of compute and storage elements closer to subscribers and devices.

During the recent pandemic, lack of and poor connectivity impacted the standard of living for citizens everywhere. The digital divide between the connected and unconnected (and under-connected) continues to be a global issue with visibility at the highest levels of governments worldwide.

To improve the quality of experience across all these use cases, we need to scale up computing, storage (cost-effectively, sustainably), and network capacity to handle more bits while improving application responsiveness by reducing the round-trip time between data generation, processing, and consumption. The first, middle, and last-mile networks all play a role but may have different impacts.

First-mile expansion decisions are made by a few large organizations, each with extensive data on traffic patterns and sophisticated forecasting capabilities.

Meanwhile, the last mile exhibits static topologies and tends to be dominated by a few large regional providers. At the last mile, any inadequacies in networking are keenly felt by users. During the recent pandemic, lack of and poor connectivity impacted the standard of living for citizens everywhere. The digital divide between the connected and unconnected (and under-connected) continues to be a global issue with visibility at the highest levels of governments worldwide.

This awareness has led to a substantial focus on the last-mile divide. In the US, there's $65B allocated under the Infrastructure Investment and Jobs Act (IIJA) and Broadband Equity, Access, and Deployment (BEAD) programs, plus $20B under the Rural Digital Opportunity Fund (RDOF), and another $10B as part of the US Treasury's Capital Projects Fund (American Rescue Plan). Additionally, there are USDA Reconnect, the Tribal Broadband Connectivity Program, and the Broadband Infrastructure Program.

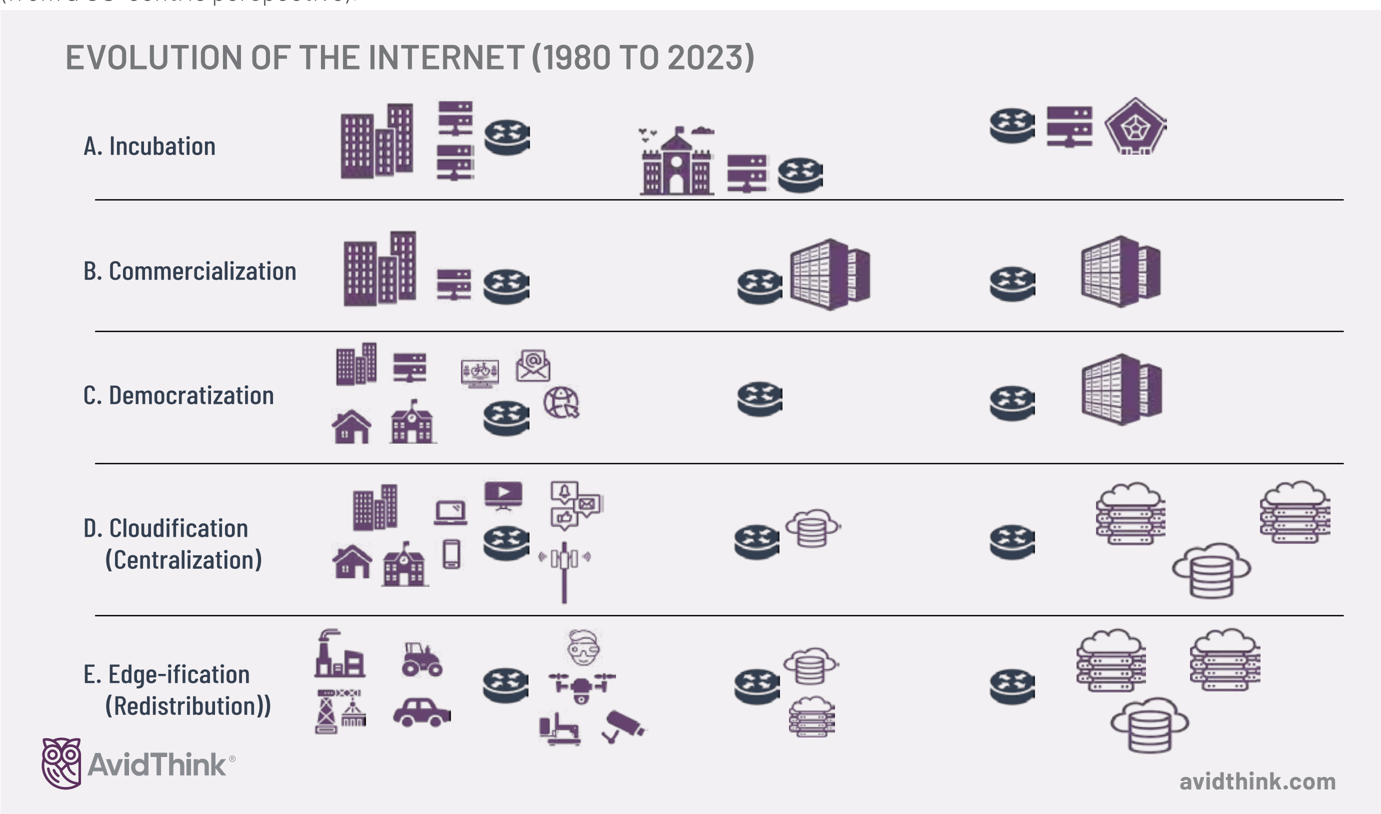
Many of these programs target the last mile. Unfortunately, the middle mile may have been overlooked despite its strategic importance. Even considering that last mile buildout is expensive, particularly in rural geographies, only $1B of the US IIJA is designated for developing the middle mile.

Given the relatively fixed nature of first and last-mile topologies, the opportunity for infrastructure optimization lies in the middle mile. However, the industry needs to approach the middle mile differently.

Revisiting the Middle Mile

The middle mile is a crucial component of the digital pipeline, as it connects last-mile access networks to global backbone networks and the data center resources they contain.

To understand the middle mile's role, let's take a brief look at the evolution of the internet from the original ARPANET to today (from a US-centric perspective):

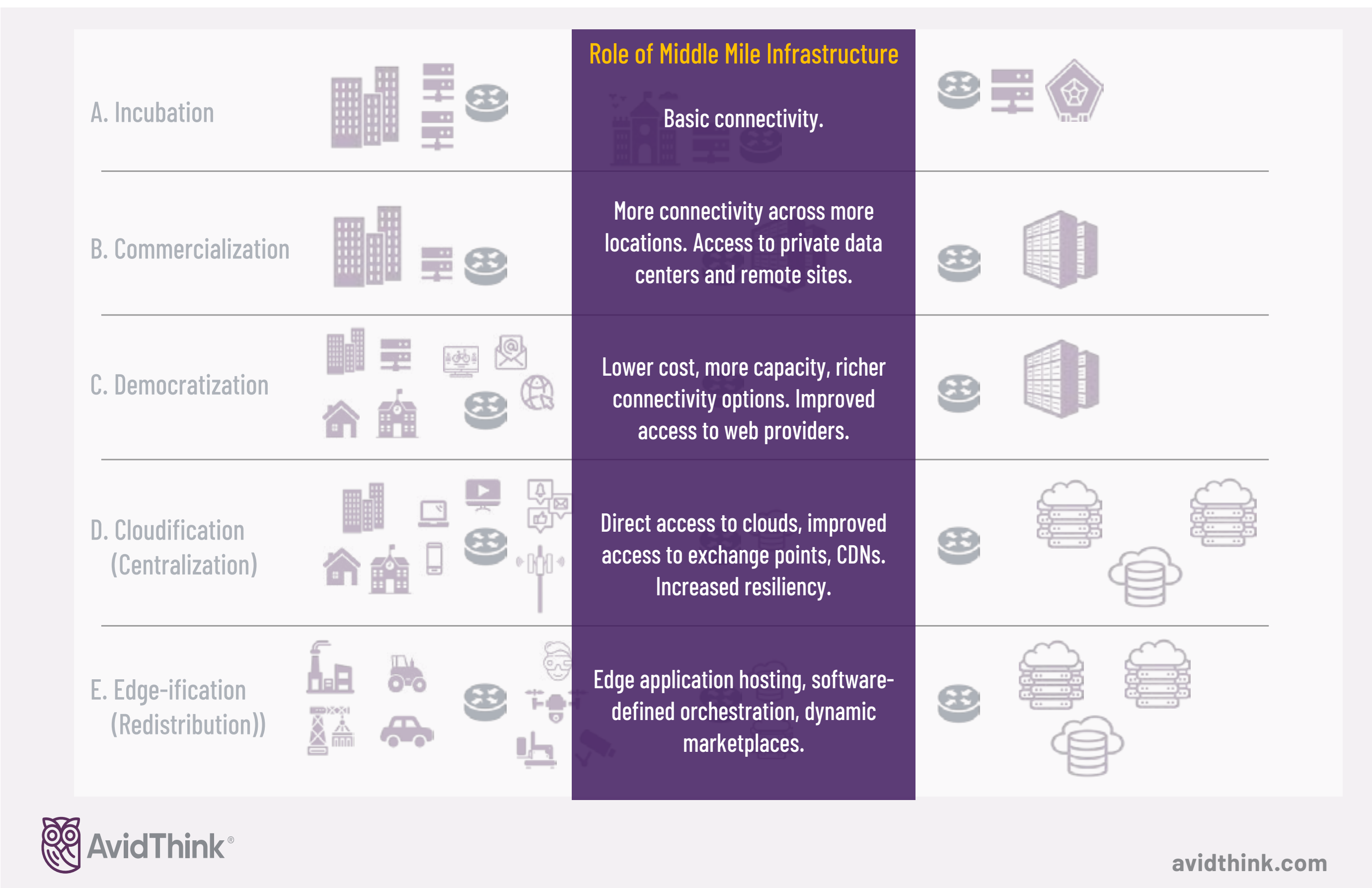


Our depiction of the journey consists of five major phases:

- **Incubation** — In the 1980s, the early days of the internet saw ARPANET and the NSF's NSFNET transition into more public networks. TCP/IP became the standard, research networks flourished at universities, MILNET evolved, and early businesses began utilizing the internet for communication.
- **Commercialization** — During the 1990s, the widespread use of email services introduced the internet to more users, predominantly students. The World Wide Web replaced early information systems like Gopher and surpassed dial-up walled gardens such as Prodigy, AOL, and home-grown Bulletin-Board Systems (BBS). Mosaic gave way to Netscape, and the initial browser wars ensued. Public key encryption with a secure sockets layer (SSL) facilitated early e-commerce.

- **Democratization** — By the 2000s, internet access was considered essential in many developed and developing countries. E-commerce gained widespread acceptance, businesses established websites and email addresses, and early adoption of online entertainment took place.
- **Cloudification/Centralization** — The 2010s saw an explosion of over-the-top services. Most services were delivered via the internet, with social and streaming media dominating. Mobile internet usage grew significantly. Cloud platforms from AWS, Microsoft, Google, Alibaba, Baidu, and Tencent, along with major web and cloud services from Facebook (Meta), YouTube, Netflix, and Apple, became the predominant technology platforms for consumers and businesses.
- **Edge-ification/Redistribution** — From 2020 onwards, it was recognized that centralization and cloudification, while beneficial, limited the types of workloads that cloud technologies could support. Building on lessons from the cloud and improved orchestration, multiple ecosystem players began promoting early edge services for specific use cases. This redistribution isalso driving enterprises to turn to multi-cloud, and hybrid or distributed clouds, spreading workloads and application components across the spectrum of regional clouds to on-premises edge platforms.

With the evolution of the internet, the middle mile continued to play a vital role (see figure for details). As a crucial intermediary, the middle mile aggregates traffic from multiple last-mile networks, offers efficient routing, and ensures a seamless experience for end-users. This reality has persisted since the internet's commercialization. With the arrival of the edge-ification phase, the middle mile infrastructure is essential for latency-sensitive applications, such as gaming, video conferencing, and real-time data analytics, which demand rapid and reliable data transmission. Likewise, the middle mile will be responsible for enabling high performance access to cloud platforms and SaaS services that consumers and businesses rely on for everyday tasks.



Why Include Computing Infrastructure?

Historically, discussions around the middle mile focused on network topology and capacity. Today, the business and consumer conversation is not merely about connectivity but the end-user application experience.

In today's outcome-driven world, the middle mile infrastructure that delivers content to the last mile is not a passive carrier of bits but actively generates, transforms, accelerates, encrypts, and protects content. Content processing requires computing resources, hence the convergence of computing and storage with the middle-mile networking infrastructure.

The middle mile's unique position allows it to host next-generation services and edge computing resources, empowering businesses to deploy applications closer to the end user, resulting in improved response times, reduced backhaul costs, and enhanced scalability.

As part of the edge-ification of our infrastructure, the middle mile can host distributed mini-clouds or intelligent content caches for applications or power security gateway functions. Its Goldilocks location — proximity to end-users coupled with sufficient aggregation — allows for low latency in AI/ML inferencing, robotics, and other control applications while benefiting from scale economies.

Why isn't this Edge Computing by Another Name?

Astute readers will note that our pitch above looks like more edge computing hype. While we agree that NMM overlaps with edge computing, there are significant differences.

Edge computing focuses on fixed locations for compute placement, such as on-premises or network edge locations, including far-edge cell towers, near-edge carrier aggregation points, and hyperscaler local clouds.

Our lens focuses on a joint view of computing and networking that's network-centric — a pliable middle-mile infrastructure coupled with the ability to dynamically locate or utilize compute within this middle-mile fabric.

Our NMM lens focuses on a **joint** view of computing and networking that's network-centric — a pliable middle-mile infrastructure coupled with the ability to dynamically locate or utilize compute within this middle-mile fabric. We can bring new locations into play by forming new network connections to adapt to changing application requirements. We believe this holistic approach is necessary for the industry to meet the unique performance demands of modern applications.

New Middle Mile Ecosystem

Key players in the middle mile include a wide range of organizations: communications service providers (regional and national, wireless, wireline, satellite), data center operators, large technology companies (web properties), hyperscale cloud providers, co-location providers, public and investor-owned utilities, grid operators, cooperatives, and academic institutions. We'll touch on a few key ones below.

Datacenter Operators

Datacenter (DC) operators manage the infrastructure, operations, and maintenance of global, regional, or local data centers. This broad description covers the following:

- Hyperscalers (AWS, GCP, Azure)
- Cloud providers (Oracle Cloud, Akamai/Linode) who host cloud services
- Global data center operators selling a mix of co-location, powered-shell, wholesale, and retail offerings
- Regional and smaller data center operators selling a similar mix of services

DC operators play a crucial role in hosting computing and storage services and contribute to the stability and performance of networks by maintaining infrastructure and providing services like cross-connects and peering, enabling traffic exchange between different networks.

As players in the evolving middle mile control more assets in both connectivity and compute hosting, data centers will be utilized by hyperscalers (who lease capacity from them), carriers, and other businesses for strategic expansion.

Co-location Providers

A sub-category of DC operators, co-location providers (colos) offer data center space, power, cooling, and connectivity services to organizations that want to house their IT equipment and network infrastructure in a shared facility. In the middle mile ecosystem, colos like Equinix, Digital Realty, CoreSite (part of American Towers), Cyxtera, CyrusOne, and DataBank play a crucial role in enabling interconnectivity between different networks, including those owned by CSPs, hyperscalers, and other service providers. This interconnection allows for efficient routing, reduced latency, and improved end-user network performance.

Further, many internet exchange points (IXPs) (both non-profits and for-profits) have established points of presence (POPs) in co-location facilities, increasing the importance of the colos.

Colos have expanded services to include bare-metal server hosting and providing direct connectivity into hyperscaler clouds. As the middle mile evolves, colos will continue to play a pivotal role given their unique location in internet real estate (yes, it's also about location, location, location in the virtual world). Forward-thinking colos, besides adding compute and storage services, are also providing dynamic upper-layer network security services (firewalls, SD-WAN, SASE) and looking at ways to make their connectivity fabric more programmable.

Hyperscalers

Hyperscalers like AWS, Microsoft Azure, and Google Cloud Platform are large cloud service providers offering massive computing resources and infrastructure. As DC operators, they have vast global data center footprints and strong interconnectivity between their data centers. Many hyperscalers have investments or strong partnerships in subsea cables and satellite networks. Hyperscalers play a vital role in the middle mile ecosystem as both the source and destination of a large proportion of business and consumer traffic. They also operate local edge locations and the middle-mile networks that connect to those.

The challenges with many CSPs are their lack of agility, programma-bility, and automation, which limits their ability to serve modern consumer and enterprise workloads requiring on-demand connections and tunable quality of service.

Communication Service Providers (CSPs)

CSPs offer a range of telecommunication services to consumers, businesses, and government entities. They own and operate extensive network infrastructure across the first, middle, and last mile and have in-house capabilities to manage, maintain, and grow these wired, mobile, and satellite networks. They are both bound and protected by in-country regulations.

In the evolving middle mile ecosystem, CSPs are vital in ensuring seamless connectivity and have critical network assets they can bring to the table. The challenges with many CSPs are their lack of agility, programmability, and automation, which limits their ability to serve modern consumer and enterprise workloads requiring on-demand connections and tunable quality of service.

Nevertheless, CSPs are looking to use their facilities and networks to serve modern network workloads (security, SD-WAN, SASE) and provide cutting-edge computing capabilities.

Tower Companies

Tower companies like American Towers, Crown Castle, Cellnex, Indus Towers, Vantage Towers, and edotco/Axiata Group own, operate, and lease communication towers and other supporting infrastructure for wireless service providers. While they are critical in supporting last-mile wireless connectivity, they play a role in the middle-mile ecosystem by providing infrastructure for backhaul and fronthaul networks. These networks are essential for connecting mobile access points to aggregation core sites and the middle mile and backbone networks, ensuring seamless connectivity for mobile and wireless users.

As the demand for wireless connectivity and 5G services grows, tower companies will continue to invest in and expand their infrastructure to support the evolving middle-mile ecosystem. We've already seen tower companies like **American Tower expanding its footprint** by acquiring data centers and expect to see more cross-domain plays in the space.

Content and Application Delivery Networks

Another class of organizations with incentives to play in the middle mile includes evolving content distribution networks like Akamai, Cloudflare, and others that now provide a wide range of services, including security, content distribution, distributed databases, name resolution, and edge workload hosting in serverless form factors. Their role in optimizing and securing content distribution involves managing and navigating the first, middle, and last mile of the internet.

Networking and Infrastructure Equipment Vendors

Vendors that provide infrastructure equipment and software to these players above (networking, computing, storage vendors, and ISVs — independent software vendors) are part of the ecosystem as well. Traditionally, networking vendors associated with the middle mile are transport providers (IP, optical, converged switching, and routing) like Arista, Ciena, Cisco, Huawei, Infinera, Juniper, and Nokia.

With the rise of virtualized networking and disaggregation, other vendors have entered the market, including upstart DRIVENETS (almost $600M raised), which is pursuing a disaggregated cloud-native approach to routing that supports the entire range across core, aggregation, and edge with a scalable software platform running on white box equipment.

Another notable software-centric vendor in the space is Arrcus (about $130M raised), which is focused on building the Arrcus Connected Edge (ACE) — a scalable software platform that addresses end-to-end network routing. And with FlexMCN, a multi-cloud networking solution based on ACE, Arrcus aims to also solve the critical issue of unifying distributed workloads across different clouds (hybrid/public) and enable edge to multi-cloud use cases. Arrcus has also been championing new networking standards like segment routing (SRv6) that enable end-to-end service quality from the access network through the middle mile into the core data centers.

At the overlay layer, SD-WAN, SASE, and MCN vendors play a role, as well as a new class of network-as-a-service (SaaS)/SDN solutions like PacketFabric and Graphiant whom we'll touch on later in the report.

Beyond networking, providers of server equipment lay a role in providing computing and storage infrastructure for the nodes within the NMM. Well-known vendors include Dell, HPE, and Lenovo, along with white box server counterparts from Celestica, Inventec, Quanta, Super Micro, and Wistron.

Orchestration and Virtualization Vendors

Given that we've incorporated edge computing into our NMM definition, it's hard to avoid including vendors like VMware and Red Hat, who, in addition to shipping private cloud platforms, provide distributed and hybrid cloud capabilities for hosting edge workloads. They compete with hyperscalers edge strategies encompassing hyperscaler-designed hardware and running hyperscaler software and orchestration stack on third-party servers.

A new breed of orchestration vendors who are targeting managing distributed workloads for enterprises (industry 4.0, retail) and telcos (5G core, 5G RAN, SASE) have entered the market. Aarna Networks, which is now leveraging an open-source orchestration project from Google called Nephio (hosted by Linux Foundation), has seen interest in such capabilities from its ecosystem partners like Equinix. Aarna also incorporates modules from other Linux Foundation projects (e.g., ONAP, EMCO) into its open-source offering and is exploring the best integration points for extending their platform to Nephio.

A new breed of orchestration vendors who are targeting managing distributed workloads for enterprises (industry 4.0, retail) and telcos (5G core, 5G RAN, SASE) have entered the market.

Other Organizations

The middle-mile market structure differs across countries. Depending on the country, state and local governments, electric and other utility cooperatives, railroad operators, research entities, and other organizations will have influence and assets within the middle mile.

For example, state and local (or national) governments may choose to invest or participate in open-access middle-mile networks to spur last-mile development. Electric cooperatives, encouraged by incentives, regulatory nudging, and increased revenue opportunities, might do the same to help drive broadband offerings in unconnected or underserved rural areas.

Another less obvious set of organizations that provide value to the middle mile ecosystem is geographic information systems (GIS) companies that offer detailed connectivity information within served and unserved regions.

And finally, a unique company we were referred to by multiple parties as we researched the middle mile ecosystem, Connectbase. Connectbase runs the Connected World platform, a marketplace platform for network providers to buy and sell connectivity. Powering the marketplace is a database comprising data for about 5,000 carriers that shows on-net and near-net options across locations worldwide (with high accuracy for the US and other select markets). Such a database would provide value in any middle-mile planning and automation.

As the last mile networks grow in capacity and coverage, middle mile capabilities will have to step up, and in any evolutionary market, there will be jostling, new entrants, and opportunities for different roles for existing players.

Distribution Networks and the Messy Middle

To understand how the middle mile might evolve, we should examine other types of similar networks. Traditional communications infrastructure has analogs in different distribution architectures, such as the power grid, municipal utilities, transportation systems, and physical goods distribution systems.

Characteristics of Distribution Networks

Many of these networks share common traits, including:

- **Tree-like structures**: Longer-distance and larger capacity cores or trunks, a mid-distance and mid-capacity distribution tier, and a shorter-distance and lower-capacity access tier.
- **Hierarchical management**: A hierarchical topology provides easier management, optimization, and standardization. This hierarchical approach mirrors governance and geopolitical infrastructure, with countries organized around national, regional/provincial/state, and local/city constructs. This alignment of distribution infrastructure with governance supports the span of control and helps simplify permitting and scope regulations.
- **Greater economies of scale in upper tier**: Per unit of item transported (goods, electricity, digital bits), there's greater amortization of physical costs across upstream large 'pipes.'
- **Diversity at edge**: Less diversity in the core in terms of the 'technology,' with a focus on speed, cost-effectiveness, and capacity. Diversity increases as we get to the access or edge layers to accommodate different pricing sensitivities, real-world physical limitations such as terrain, and end devices or equipment needs.
- **Different update rates**: The rate of change of distribution architectures has historically seen faster innovation at the edge, which also results in greater diversity and lesser change in the core due to the impact of changes and the need to evolve or adapt interfaces between the layers. Core changes take longer and are executed as a step function, while edge changes are incremental. However, upgrades in the network's core have much larger ramifications and higher leverage than changes at the edge.

Analog to Physical Distribution Chains

The retail distribution network in developed countries presents an analog for our digital distribution network. In the physical world, the first mile would be the long-haul routes between factories to warehouses and large distribution sites; the last mile would be the routes from neighborhood retailers to consumers or businesses; and the middle mile would be the primary distribution sites to local warehouses and retail stores.

We've seen innovation in the last mile, with various types of delivery services evolving with the rise of e-commerce. Consider the rise, fall, and consolidation of next-day, next-hour, and sub-hour delivery systems across the globe.

There's recognition in retail distribution that many new opportunities for cost and delivery time optimization will **come from the middle mile**. From the use of autonomous vehicles between fixed routes for multiple daily runs, to **advanced routing and optimization algorithms with detailed telemetry** combined with big data analytics and AI/ML, and a rethinking of warehousing locations and inventory management, there's innovation and sizable investment in R&D from major retailers like Amazon, Walmart, and logistics companies.

This evolution parallels the efforts in our communications infrastructure revamp.

Evolution of Middle Mile Topology

Over time, the shape of a distribution network will shift as distribution and communications patterns change.

For example, in our delivery network example above, when there's an increased load on the part of the distribution network, new hubs will form to facilitate distribution in regions of high traffic. Local warehouses and second-tier distribution sites show up in neighborhoods with great demand.

In electronic communication networks, a recent example is the data center infrastructure. For a long time, data center networking was dominated by a fat-tree architecture with a thick trunk at the core and smaller pipes at the access layer.

However, starting in the early 2010s, with the need for more east-west traffic communication versus north-south, the spine-leaf architecture that prioritized consistent, low-latency, high-capacity paths between any two server nodes became popular, migrating the fat-tree into a multi-stage Clos architecture.

Communication Networks vs. Distribution Networks

In the metro area networks, where the middle mile sits, we see new traffic patterns emerge (perhaps less dramatic than the rise in east-west versus north-south flows in data centers). Data flow from the edge is increasing with more digitization. Historically, wide area networks look like distribution networks where people download more over-the-top (OTT) video (which makes up a significant proportion of network traffic) than they upload. However, with the rise of video-centric social media like TikTok, high-resolution 4K video cameras, and social gaming, the end nodes in communication networks will generate more upstream data.

Current indications are that downstream will still dominate, but upstream data needs will grow due to increasing OTT entertainment at higher resolutions (4K). **AT&T indicated in their 2021 investor conference** that the ratio of downlink broadband usage to uplink broadband usage in 2015 for their customers was almost 10:1 and dropped to 7:1 in 2020, and they expect the monthly total bits consumed to go from a 10:1 (download/upload) ratio in 2020 to 5:1 in 2025. In their **2022 investor conference**, the ratio was back to 9:1 (showing downstream still dominating). Meanwhile, an NCTA-commissioned report on the cable industry shared that **downstream to the upstream ratio for US customers was 14:1 in 2021 compared to 4.5:1 in 2010, even with the pandemic and work-from-home initiatives**.

Regardless, there's a recognization that upstream needs will continue to increase as consumers depend on the cloud for storage of their captured photos and videos and generate more data at the edge. The increase in upload traffic will be more pronounced in the business sector, as the increased use of video surveillance (security, safety, marketing) and IoT drives more upstream data.

Emergence of New Hubs and Scale-Free Topologies

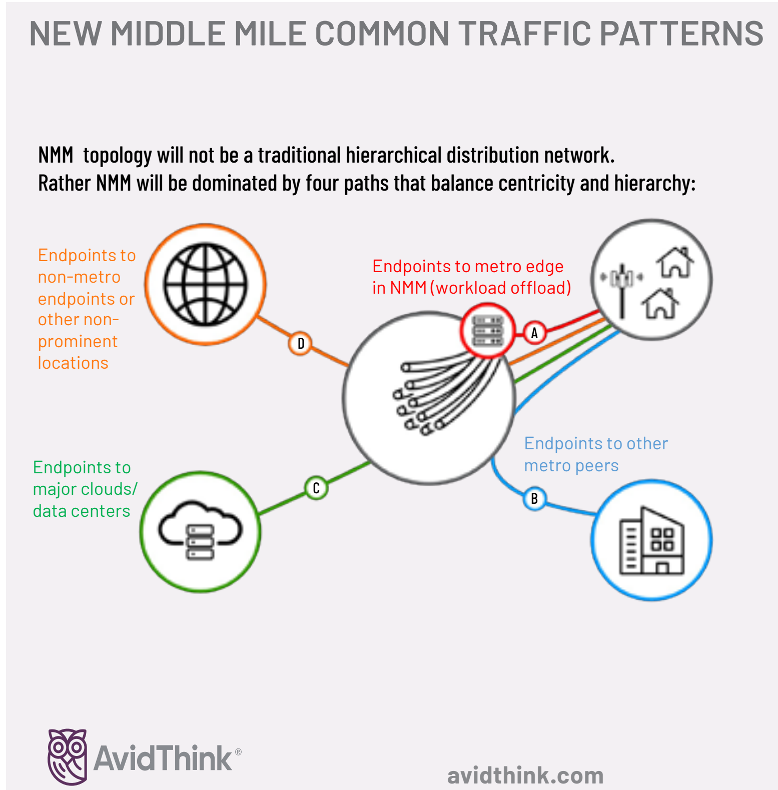
If we look more at the traffic patterns for the middle mile, there are a few significant paths over middle-mile networks that will be of importance in the next few years:

A. Endpoints to major clouds and data center

B. Endpoints to the metro edge (cloud offload, caching, edge computing)

C. Endpoints to other peers within the same metro area (localized traffic)

D. Endpoints to non-metro endpoints or less-popular locations



Today, paths A and B are dominant between direct-to-cloud and CDNs for cached video traffic and SaaS services. Path B will multiply as the industry finds new ways to reduce latency and bandwidth costs. And this is where the NMM will come to play a more prominent role.

We expect the emergence of hubs in more distributed locations to improve the experience, lower latencies, and adapt to local and regional needs as the services are increasingly consumed and become popular.

The rise of preferential nodes on the middle mile network will shape its topology. Just as airports act as hubs in transportation networks or influencers play the same role on social media, we will see new key nodes appear, and the potential rise of decentralized Web3 (while early) will reinforce the topology — commonly known as a **scale-free network**.

Scale-free networks grow when new nodes added show preference in connecting to a select number of hot nodes. They exhibit different properties than a randomly connected network. For example, scale-free networks provide more resiliency than random networks. This evolutionary pattern means that hubs of importance will continue to emerge and become critical control and value capture points in the NMM topology — existing IXPs are an example of these hubs, but as the middle mile evolves, other hubs could emerge, changing ecosystem dynamics.

Technology Trends Impacting the Middle Mile

As the NMM topology shapes itself, the architecture and capabilities of the middle mile are intertwined with emerging technology trends that include:

Virtualization and disaggregation: Virtualized and disaggregated services, including virtual and cloud-native network functions (VNFs/CNFs), enable network operators to deploy software-based network functions on standard servers. This reduces reliance on specialized hardware, increases flexibility, and enables greater scalability in response to changing demands. It means more flexibility in placing different components of these services that both (a) power the middle mile (e.g., routing and switching) and (b) can be hosted within the NMM. The NMM can host parts of the 5G Radio Access Network (RAN) and central office functions like virtual Broadband Network Gateway (BNG). Additionally, the computing facilities located in the NMM can host components of the 5G Core. The unique location and performance characteristics of the NMM can bring improved scale and performance for 5G networks.

Software-Defined Networking (SDN) and Programmable Networks: SDN today is a catch-all for centralized management of networking elements that exhibit separation between the management, control, and data planes. SDN is also used to denote programmable networks — networks that can be configured with programmable APIs in real-time or near-real-time. NMMs will need fast reconfiguration for optimization, performance, resiliency, and security, and SDN and network programmability is critical.

Cloudification and Application Rearchitecture: As more enterprises adopt distributed cloud architectures and microservices for their applications, the middle mile infrastructure must evolve to support these new models. Fast access to interexchange services and direct peering paths to hyperscaler data centers and those of major SaaS and cloud services (e.g., Office 365, Google Workspace, SalesForce, Dropbox, FaceBook, YouTube) bring value to consumers and businesses. The NMM can also cache and host application components for these popular applications. By enabling seamless integration of applications and workloads across distributed cloud environments, the NMM can better support the needs of modern businesses.

AI/ML and Big Data Analytics: Hosting data analytics and AI/ML capabilities (particularly the inferencing workloads) within the middle mile infrastructure allows for faster processing of large amounts of data closer to where it is generated or needed. This can enhance the performance of data-driven applications and enable real-time decision-making for applications like IoT and industrial automation. The other impact of AI/ML is the use of AI/ML to assist with workload placement and optimization of middle-mile resources. Given the complexity and messiness of the NMM — network topology, computing resources, multi-organization, and multi-domain coordination, AI/ML could help lower OpEx when coupled with an orchestration solution.

Next-Generation Networks including Optical Advancements (400Gbps to 800Gbps to 1.6Tbps): Underlay speeds limit how fast traffic can be carried across the middle mile. The NMM will require continue equipment upgrades (along with deployment architecture changes at nodes) to increase the carrying capacity and programmable capabilities. NMM is dominated by fiber connectivity and advances in optical like compact pluggable optics (e.g., 400ZR and 400G ZR+). Converged IP/optical platforms are essential to increasing capacity and control, reducing latency, improving resiliency, and supporting sustainability. Leading networking vendors like Cisco and Juniper have recognized the importance of the metro network, with Juniper pushing new offerings in their Cloud Metro lines and Cisco demonstrating intent-based and improved assurance via the Cisco Crossworks management platform coupled with Cisco ASR/NCS platforms. More recently, Ciena unveiled WaveRouter, a coherent router with converged IP/optical multi-layer capabilities and a new architecture different from traditional chassis, spine-leaf, and distributed disaggregated white-boxes.

Orchestration and Automation: Managing workloads across the cloud, local edge in the middle mile, and on-premises infrastructure requires advanced orchestration tools. Edge orchestration technologies, which we touched on earlier in the ecosystem players section, provide seamless integration and management of these diverse workloads, optimizing resource utilization and ensuring consistent performance across the entire network. Similarly, to control operational costs in this complex multi-domain environment, automation technologies, augmented by intent-based frameworks and AI/ML will be key. Given the diversity of players across the middle-mile ecosystem, some organizations look to standards as a coordination mechanism to define inter-organization interfaces and APIs. The standards organization, MEF, has been working with a number of CSPs, hyperscalers, and colos to help standardize these interfaces for service orchestration. The MEF LSO standards focus on lifecycle management of end-to-end connectivity services across multiple network service domains.

Edge Computing and Infrastructure Acceleration: While not delving too deep into the details here (readers are welcome to read our Telco Cloud and Edge and SmartNICS and Infrastructure Acceleration reports for more details), edge computing infrastructure is part of the NMM and advances to host and accelerate workloads (AI/ML, graphics rendering for VR/AR, telco network services) are critical to NMM's development.

Collectively, these technologies enable a more dynamic, flexible, and scalable NMM infrastructure that can better support the needs of an increasingly connected and data-driven world.

New Middle Mile Use Case Benefits

We've touched on the importance of the NMM to leading use cases for consumers and enterprises. In this section, we'll categorize and explicitly state the value that NMM (as powered by the technologies just discussed) brings.

- **Interactive and Collaborative Applications** — With edge computing sites integrated into the NMM infrastructure, processing and storage resources are located closer to the end users. This proximity allows for faster content delivery and real-time data processing, resulting in a more responsive and seamless experience for end-users. For example, streaming services can leverage these edge computing sites to cache popular content, minimizing buffering times and delivering higher-quality video streams to subscribers. Similarly, augmented reality (AR) and virtual reality (VR) applications can provide more immersive experiences by offloading processing tasks to nearby edge computing resources, reducing the load on end-user devices.
- **Data-Centric Workloads** — Integrating edge computing sites into the middle mile infrastructure allows businesses to scale their IT resources more efficiently. Companies can manage traffic and workload spikes without overburdening their core infrastructure by offloading processing tasks from centralized data centers to distributed edge sites. This distributed approach not only improves resource utilization but reduces backhaul costs, as data can be processed and analyzed at the edge before being transmitted to centralized data centers. Furthermore, the scalability offered by edge computing sites enables businesses to expand their services to new markets with minimal investment in additional infrastructure. The NMM plays a role not just in computing but also in ensuring optimal routes between data generation and data processing.
- **Peer-to-Peer Applications** — Time-sensitive applications such as online gaming, financial trading, music collaborations, industrial control systems, surveillance, and safety applications require low-latency networks to function. By incorporating edge computing sites into the NMM infrastructure, data processing and routing can be performed closer to the data source, reducing round-trip times. This results in reduced latency and improved performance for these critical applications. For instance, multiplayer gaming platforms can use edge computing resources to host game servers, ensuring faster response times for player actions and minimizing lag during gameplay.
- **Telco Workloads** — Edge computing sites within the NMM provide CSPs and other players in the connectivity ecosystem with greater flexibility in deploying and managing network functions and services. By leveraging edge computing resources, CSPs can offload network functions such as virtualized firewalls, security (SASE/SD-WAN) routing, 5G RAN, 5G core, and application load balancing to the edge, streamlining their core network operations and reducing the complexity of managing these functions centrally. Additionally, using edge computing sites enables CSPs to rapidly deploy new services and applications, ensuring they can quickly adapt to changing market demands and customer needs.

In addition to application benefits, the NMM can address additional business, consumer, and government concerns:

- **Privacy and Sovereignty** — The NMM can assist in the placement logic of workloads, ensuring that data traverses links within a specific geographic boundary and that computing for the data occurs in a location compliant with data sovereignty laws while optimizing for costs and performance metrics. Keeping data transmission and computing within a metro region will likely help with data sovereignty obligations. Given the cloud-centricity of many enterprise workloads, there will be coordination of the use of NMM resources to augment regional cloud resources.
- **Security and Resiliency** — In the last mile and middle mile boundary, the NMM can include security capabilities at aggregate sites and IXPs to screen and protect traffic (bi-directionally protecting last-mile assets and business cloud assets) from DDoS and other malicious profit-driven organizations and state actors. The aggregation of statistics at the middle-mile level makes spotting unusual patterns across disparate networks more accurate, and economies of scale in enforcement and blocking can reduce the cost of security. Furthermore, the NMM can be dynamically reconfigured to act as caches and buffer zones that provide resilience in infrastructure due to climate or other disaster-triggered damage. Buffer capabilities can include caching critical network and application services (domain lookup, caching, application databases).

Middle Mile Value Stack Evolution

The middle mile value stack comprises diverse players working together to deliver efficient and high-performance connectivity services. Each of the players we described plays a critical role in the delivery of services, from providing network connectivity and infrastructure to offering cloud and edge computing.

The multiparty environment in the middle mile necessitates the brokering of connectivity and reduction in search costs of finding network peering or hosting partners. IXPs/DC operators, tower companies, and others in the space are well-positioned to capitalize on this unique characteristic of the middle mile.

In addition, digital infrastructure players like DigitalBridge that **owns numerous assets across the digital pipeline**, including edge, enterprise, hyperscale data centers, fiber, small cells, and towers, may be able to improve coordination across portfolio assets to optimize the performance of each segment.

Beyond end-to-end players like DigitalBridge, there are other trends that will impact the value stack in the middle mile:

- **Hyperscaler, colos, telco interactions** — As the demand for cloud services and high-speed connectivity increases, hyperscalers, and colos, telcos are increasingly collaborating to optimize infrastructure and provide seamless end-to-end services. This interaction involves leveraging each other's strengths, such as telcos' extensive network coverage, hyperscalers' massive computing resources, and using colos to provide on-ramps for telco customers into cloud platforms or SaaS services. As hyperscalers and telcos build out the backbone and interexchange offerings, and colos build out more hosting capabilities, market power will be rebalanced.
- **Hyperscaler edge presence** — Hyperscalers are expanding their presence beyond traditional data centers by deploying smaller, distributed edge data centers called cloud satellites. These satellite locations bring cloud computing resources closer to end-users, resulting in lower latency and improved application performance. In many locales, this is done with colo partners; in others, with telco partners (mobile edge); and in other regions, by directly leasing data centers.
- **Transformation of colos** — Co-location providers are evolving to accommodate the changing needs of businesses and the growing demand for edge computing. They are offering flexible and scalable infrastructure solutions, such as modular and micro data centers, to support the deployment of edge computing resources. Colos have expanded into providing computing resources and networking services, sometimes in partnership with other innovative companies. For instance, Equinix provides a marketplace platform that third-party software vendors can engage with to provide value-added

Is the Middle Mile Underfunded?

The middle mile is crucial in reducing the costs associated with last-mile broadband provisioning for non-incumbent operators. By leveraging economies of scale, these operators can purchase internet bandwidth at lower prices when procured in bulk from major internet peering points. Additionally, accessing end-customer ports within an incumbent operator's local distribution plant (where telecom regulators may mandate local loop unbundling) is relatively inexpensive compared to typical broadband subscription costs.

Nevertheless, acquiring middle-mile access from incumbent operators can be a significant expense for non-incumbent broadband ISPs, often exceeding internet bandwidth costs and access to end-customer ports. Yet, constructing their own middle mile networks is capital-intensive and may not be feasible for new entrants. To address this issue, governments can contribute to the development of open-access middle-mile networks, promoting a diverse range of last-mile providers and reducing barriers to entry in the market.

However, during a recent Broadband Communities Summit, **panelists shared that the IIJA deemphasizes middle-mile infrastructure** — IIJA will provide $42.5 billion to improve broadband infrastructure, but only $1 billion is designated to be spent on the development of the middle mile, and the **NTIA has already received over $5B in requests**.

Regardless, some states are pushing ahead with their middle-mile initiatives. For example, California signed a **$3.25-billion Middle-Mile Broadband Initiative** (bill SB 156) into law in July 2021.

There are critics of using public funds to build the middle network (or overbuild in areas with a commercial footprint). They argue that having a middle-mile network doesn't imply that last-mile providers will show up.

As always, the right solution per metro region depends on the costs of middle mile versus last mile build (terrain-dependent), end-user density, and nature and value of workloads. Even as we debate whether the middle mile is underfunded, given the bright spotlights on last-mile, we believe that, at the very least, the middle mile is overshadowed.

solutions — e.g., startup InsidePacket leverages Equinix Fabric to offer a multi-cloud solution. These colo-offered solutions may overlap with products from their telco and hyperscaler partners, who have tolerated these moves to date and who continue to partner closely with colos in the NMM ecosystem.

- **Network-as-a-Service and IXP alternatives** — While co-location providers (and other IXPs) have traditionally provided interexchange services between enterprises, service providers, major web services, SaaS providers, and hyperscaler, there are new entrants in the market taking a software-defined approach to interconnectivity. These companies, including PacketFabric (over $100M raised, Cisco's investment partner Digital Alpha is an investor) and Graphiant (raised almost $100M, Sequoia Capital is an investor, founders from Viptela, which Cisco acquired), aim to bring on-demand flexibility and QoS to private enterprise WAN connections. Compared to their SD-WAN and SASE counterparts, who tend to take an overlay-centric approach, this solution category provides more underlay control, enabling tighter control over QoS. While working with co-location providers as partners, these upstarts will disrupt a segment of the value that the colos provide today — the ability to interconnect securely to other enterprises, private application services (e.g., financial services), data centers, and hyperscale clouds.
- **Satellite last mile and satellite ground stations** — The development of satellite last-mile connectivity and satellite base stations is transforming the communication landscape by providing internet access to remote and underserved areas. These satellite-based solutions can complement existing terrestrial networks, improving network coverage and reliability. As satellite technology continues to advance, satellite ground stations will become part of the NMM ecosystem, with the option to use hyperscaler or other data center facilities to host edge computing and improve application performance.

As the digital landscape transforms, players within the middle mile must adapt and innovate to remain competitive and seize new opportunities. The continuous evolution of the middle-mile value stack will play a critical role in shaping the future of the global digital ecosystem, ultimately driving innovation, fostering inclusivity, and enabling the next generation of digital services.

Value Creation and Capture in the New Middle Mile

The middle mile landscape is undergoing significant transformation, giving rise to new opportunities and challenges in value creation and capture. As the NMM evolves, stakeholders must adapt and collaborate to maximize value across the digital landscape and meet the market's diverse needs. We see the following capabilities as vital in capturing value in the NMM ecosystem:

- **Deep-Pocketed Financing** — The NMM will require substantial funding for capability updates and capacity upgrades. Ongoing last-mile funding and buildout will push new capacity requirements to the middle mile. New capabilities for orchestration and flexible networking and computing in the NMM need to be funded.
- **Broad and Deep Visibility** — The fragmented ecosystem and lack of available information around middle-mile routes and connection points increase search costs for companies in the ecosystem looking to connect. It also means that today's insiders have an advantage, and any private information they have can be leveraged to spot early trends they can capitalize on. This is not always market-efficient, as existing middle-mile players might not have the DNA or agility to address market needs.
- **Cross-Domain Understanding** — Related to visibility, the complexity of the NMM, from physical network connectivity to compute hosting, to facilities build and management, local and regional permitting laws, data center operation, to software-based networking management, and application orchestration, make this a hard-to-understand game. Existing players with the right connections and understanding of relationships already in play can better analyze and predict market direction.

- **Agility and Fast Execution** — The middle mile is where quick-to-execute players like hyperscalers meet historically slower-to-act CSPs, with IXPs sitting somewhere in the middle. As the jostling over the middle mile continues, savvy insiders must act quickly to spot, grab, and capitalize on emerging control points. Companies that can orchestrate or influence different players across the value chain can reconfigure end-to-end offerings to match market needs faster.

In the evolution to the NMM, infrastructure companies, IXPs/colos, CSPs, hyperscalers, utilities, and open-access cooperatives have roles to play and opportunities to benefit. However, whether they can develop the capabilities described is still being determined, given that some organizations have failed to commit to transforming themselves.

Wrapping Up the New Middle Mile

In conclusion, the middle mile represents a critical component of the end-to-end digital pipeline, offering unique opportunities for enhancing connectivity, computing capabilities, and the overall quality of experience for end-users. The strategic location of the middle mile, combined with the evolving needs of modern workloads, calls for a more dynamic and interconnected infrastructure that leverages both network pipes and computing resources. As the communication and cloud computing landscape continues to transform, driven by trends such as remote work, AR/VR, IoT, and edge computing, the middle mile will play an increasingly important role in enabling new services and supporting the growing demands of businesses and consumers alike.

Private investment, collaboration among key players, and the adoption of innovative technologies and solutions will be crucial for unlocking the full potential of the middle mile. The development of advanced orchestration capabilities and the efficient allocation of network resources will facilitate value capture and drive monetization in this space. As we move forward, it is essential for stakeholders across the middle mile ecosystem, including CSPs, hyperscalers, co-location providers, data center operators, and tower companies, to work together and adapt to the rapidly changing landscape. By embracing new technologies and fostering innovation, these players can ensure the continued growth and evolution of the middle mile, delivering a superior experience for end-users and driving long-term success in the industry.

AvidThink, LLC
1900 Camden Ave
San Jose, California 95124 USA
avidthink.com