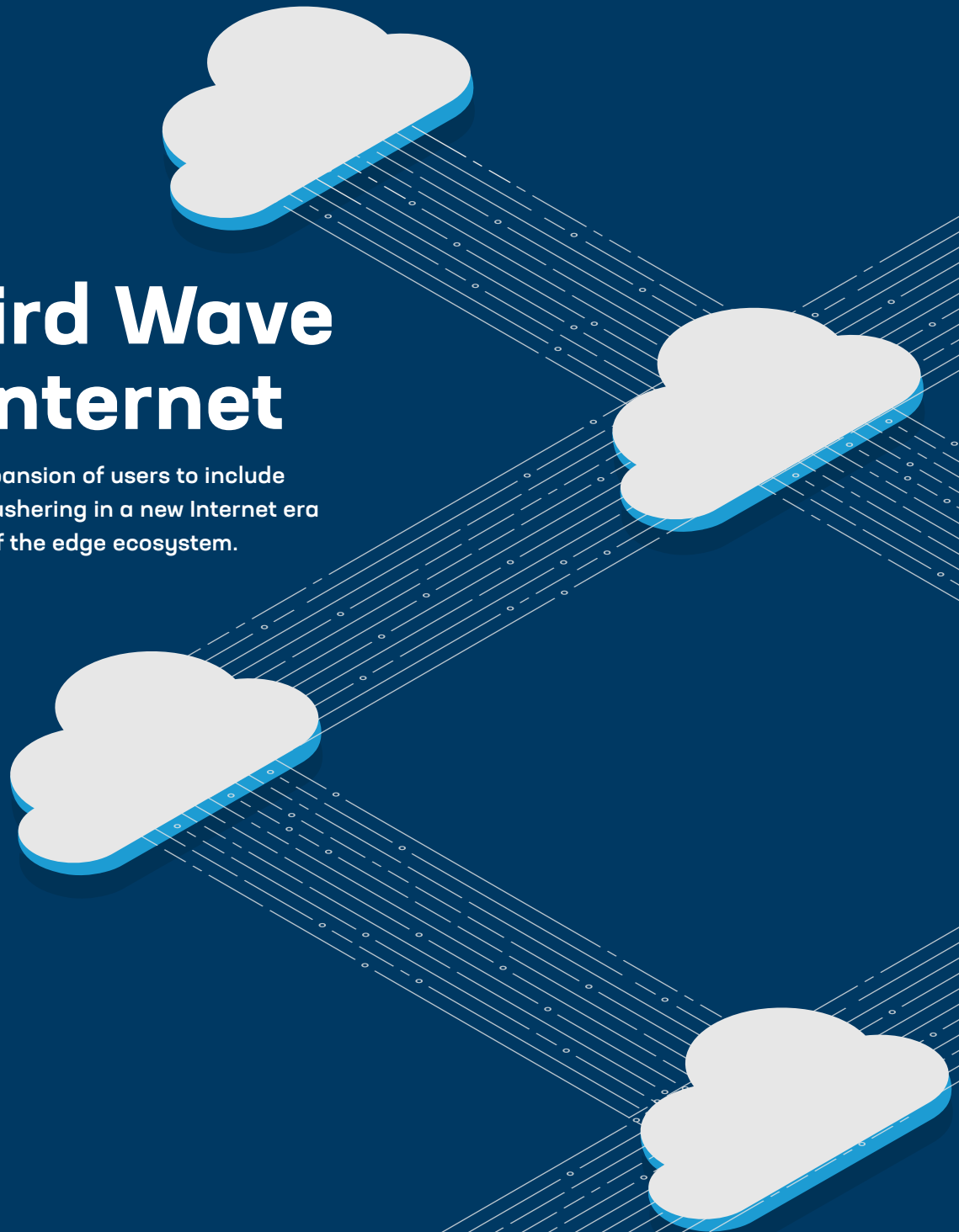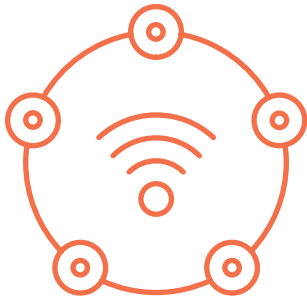# The Third Wave of the Internet

Rapid digitization and expansion of users to include devices and machines is ushering in a new Internet era that is forcing evolution of the edge ecosystem.

By Geng Lin, F5 CTO

The COVID-19 pandemic that shocked the world has brought greater volatility, but the world is not reacting unexpectedly in the face of crises and opportunities. We have seen that COVID has exponentially accelerated the pace of digitization. Satya Nadella, CEO of Microsoft, has famously stated that we have witnessed multiple years' worth of Digital Transformation being accelerated in just a few months.

Today, the world is inarguably going digital. One of the consequences of digitization is more data. IDC, in its "Data Age 2025" report, predicts that the world's data will grow to 175ZB by 2025[1]. This data will be stored in the core (traditional and cloud data centers), at the edge, and on edge endpoints like PCs, smartphones, and IoT devices. Moreover, 30% of this data will be consumed in real-time.

This is, in part, due to advances in technology. Data transfer speeds with 5G are up to 100 times faster than previous wireless generations, and latency typically decreases from 20ms to 1ms[2]. These new capabilities will increase data generation velocity and the ability to process it in real-time.

Much of this real-time data is generated and consumed by stationary devices: light bulbs, security cameras, home appliances. One-third of homeowners have increased usage of devices during the pandemic, including nearly half (46%) of smart door lock owners[3]. But a significant percentage are mobile: wearables in healthcare, connected vehicles, sensors that track and monitor supply chains. As of November 2020[4], in the U.S. alone, 45% of web traffic originated from mobile phones.

This explosive growth of devices has radically changed the definition of a user, with machines and scripts and software now acting in a role once delegated only to human beings. This growth is expected to continue.

At the same time, the number of people using this technology continues to expand. In 2019 there were 4.9 billion Internet users. By the end of 2022 that is expected to grow to 6 billion. And by 2030, experts predict that 90 percent of the projected world population—8.5 billion— six years of age and older will be digitally active[5]. Many of these users now rely on digital services in almost every area of their lives. For example, telemedicine use grew a staggering 6000% during the pandemic[6].

The pressures and demands of a distributed, digitally active society combined with explosive growth of devices signal the start of a third Internet era.

Edge Evolution Is Driven by the Third Wave of Internet

We believe that the rise and evolution of edge computing will inevitably follow the arrival of the Third Wave of the Internet. As we know, Internet transformation drove the world towards the PC and Internet era. Cloud computing and adoption of smartphones introduced the Mobile Internet era. Now we are entering a third era, the Internet of Moving Things.

The challenges emerging in this era are driving change into the edge ecosystem, moving from static and closed Edge 1.0 to open and autonomic Edge 2.0. This process is like the evolution from single-celled organisms to complex organisms.

THE EDGE ECOSYSTEM HAS EVOLVED
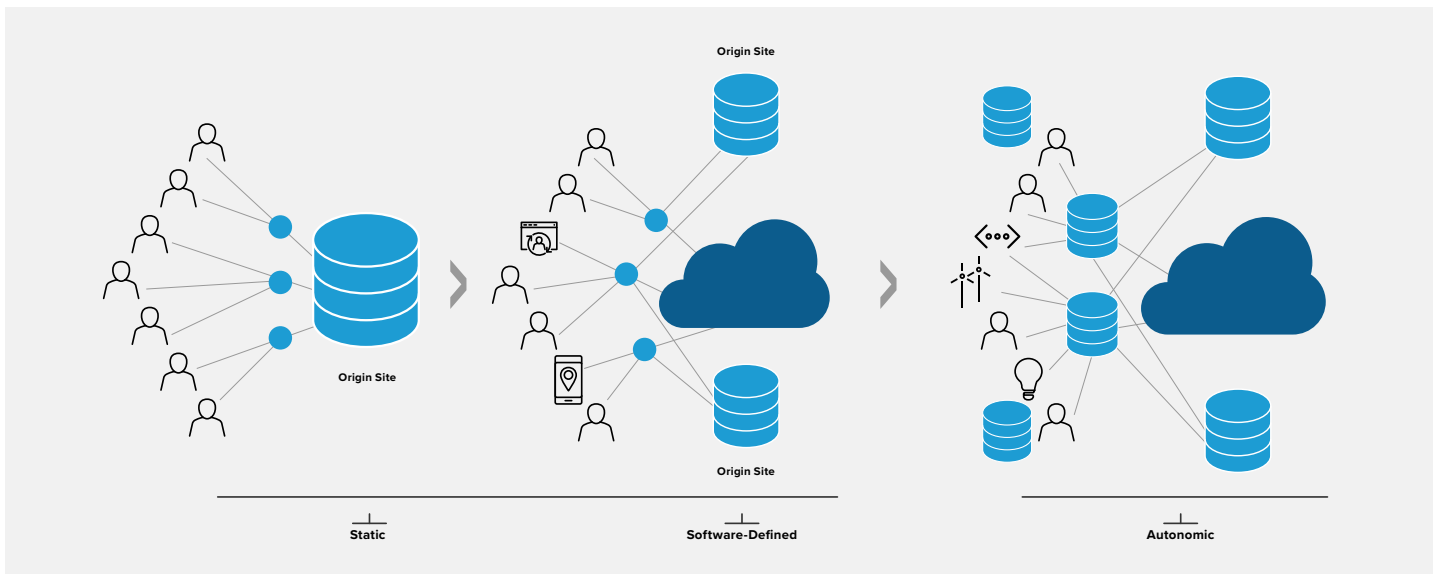
For example, in the early stages of the cloud movement, a few large public clouds and content delivery networks (CDNs) dominated the Internet application delivery and digital service distribution. These providers functioned as the centralized control points for the Internet application ecosystem, somewhat similar to the way the twelve cranial nerves function in a human body. As the cloud use case and ecosystem grows, we are seeing the need for digital services to make real-time decisions based on localized knowledge at the "edge" of the internet, akin to the way the autonomic nervous system behaves in human bodies.

This is the evolution driving the edge into a new autonomic era. This is not surprising. Each wave of the Internet brought with it challenges which were addressed in part by edge computing.

FIRST WAVE : EDGE 1.0

Tim Berners-Lee, the inventor of the World Wide Web, foresaw the congestion challenge related to the transfer of large amounts of web content over slow links that the Internet users would face; he called this issue the "World Wide Wait." The focus of the dominant paradigm at the time was, appropriately, on distributing the relatively static web content or web applications to get them closer to users to address the need for speed and redundancy. That need led to a set of key architecture tenets including physical Point of Presence (PoP) close to end users, content caching, location prediction, congestion avoidance, distributed routing algorithms, and more.

SECOND WAVE : EDGE 1.5

The advent of Web 2.0 coupled with the emergence of public clouds and SaaS solutions introduced new architectural tenets. Applications became the primary form of content over the Internet. As such, the distributed edge could not persist in its nascent form: it had to evolve along with the application architectures it delivered while under increasing pressure to secure a growing digital economy. With so much of the global economy now highly dependent on commerce-centric applications, security services quickly became an add-on staple of CDN providers, whose existing presence around the globe stretched closer to the user—and thus resolved threats earlier—than the cloud and traditional data center. These services were built atop the infrastructure put in place to distribute content and therefore represent closed, proprietary environments.
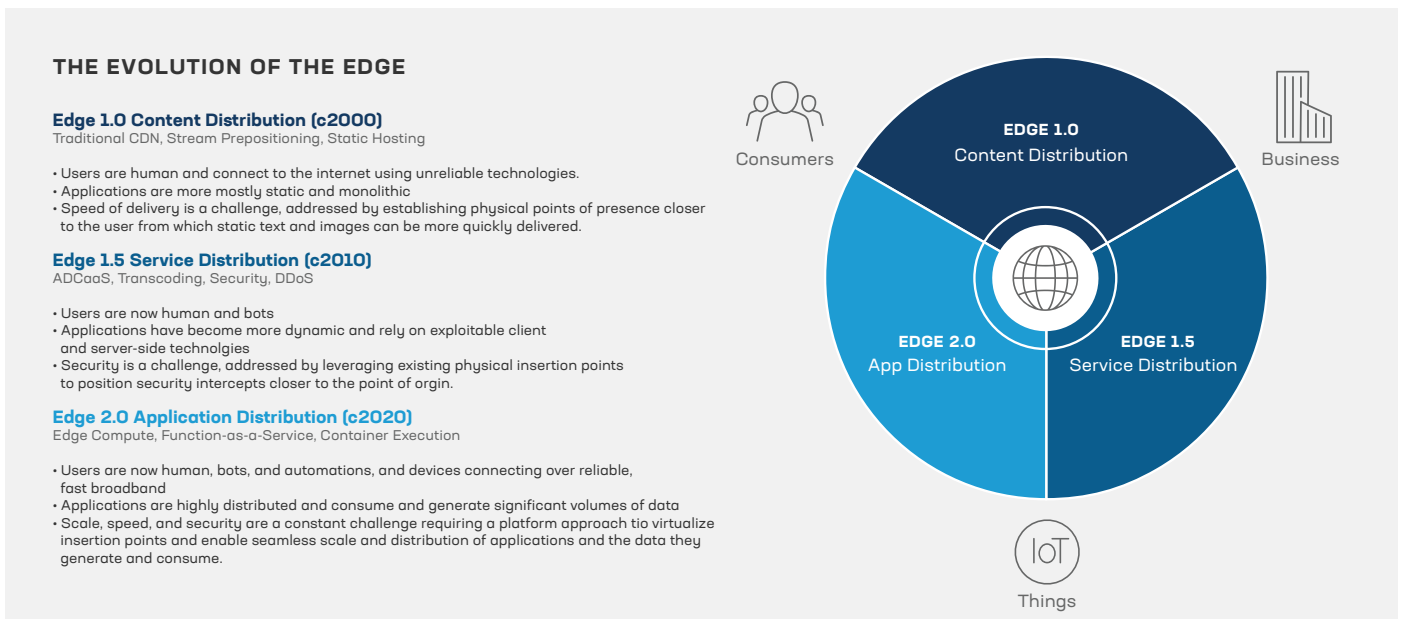
THIRD WAVE : EDGE 2.0

Today, applications are no longer the "passive" routing destinations of the delivery network but are instead active participants. For example, with Kubernetes-based distributed applications, the application logic—packed inside a container—can dynamically move to any appropriate compute location with a supporting Kubernetes stack. This is in direct contrast to the architecture principles upon which early edge solutions were built. That is, they are rooted in a time when contents (or applications) were static entities associated with the physical locations. Such edge solutions presume that the content delivery network alone functions as the "intelligent platform" to connect users to applications, while the applications (and users) remain passive "endpoints" to the "intelligent platform." This approach is no longer the best architectural way to connect users to content or applications.

Users, as well, have evolved. Not only is their digital sophistication and appetite for digital engagement light years ahead of where they were when the first CDN started in 1998, but

technology has forced a change in the definition of what they are. Today, a "user" might well be a machine, a script, or an automated service acting on behalf of a human. It might be a sensor collecting critical data from a manufacturing plant or a farm field. On one hand, these "users" continue to carry their human counterparts' desires for speed, security, and privacy. On the other hand, these new "users"—intelligent IoT endpoints alike with embedded application stacks—often participate in the dynamic processing of application logic and data analytics to deliver secure and optimal user digital experiences.

The Emergence of Edge 2.0

The core application challenges that the Edge emerged to address—speed and then security—still exist today. What has changed is the definition of application (from a static instance residing in a fixed location to "movable" container units), user (from a human user to an intelligent "thing"), location (from an IP-address to a logical identification), and the use cases that the edge aims to support (from content delivery to dynamic application distribution and real-time decision making at Edge).

THE EVOLUTION OF THE EDGE

Edge 1.0 Content Distribution (c2000)
Traditional CDN, Stream Prepositioning, Static Hosting

• Users are human and connect to the internet using unreliable technologies.
• Applications are more mostly static and monolithic
• Speed of delivery is a challenge, addressed by establishing physical points of presence closer to the user from which static text and images can be more quickly delivered.

Edge 1.5 Service Distribution (c2010)
ADCaaS, Transcoding, Security, DDoS

• Users are now human and bots
• Applications have become more dynamic and rely on exploitable client and server-side technolgies
• Security is a challenge, addressed by leveraging existing physical insertion points to position security intercepts closer to the point of orgin.

Edge 2.0 Application Distribution (c2020)
Edge Compute, Function-as-a-Service, Container Execution

• Users are now human, bots, and automations, and devices connecting over reliable, fast broadband
• Applications are highly distributed and consume and generate significant volumes of data
• Scale, speed, and security are a constant challenge requiring a platform approach tio virtualize insertion points and enable seamless scale and distribution of applications and the data they generate and consume.

Consumers

Business

EDGE 1.0
Content Distribution

EDGE 2.0
App Distribution

EDGE 1.5
Service Distribution

IoT

Things

Digital transformation and IoT is driving new requirements for digital experiences that result in the need for application distribution, real-time intelligence and decision making at the Edge.

As such, edge computing is becoming a key enabler of digital transformation in the industry. According to the 2021 State of Application Strategy report[7], 76% of organizations have implemented or are actively planning edge deployments, with improving application performance and collecting data/enabling analytics as the primary drivers.

Furthermore, a huge number of "things" have been incorporated in the latest round of digital transformation. Cisco's Annual Internet Report[8] predicts that "by 2023, there will be more than three times more networked devices on Earth than humans. About half of the global connections will be machine-to-machine connections and the M2M space will be dominated by consumer-oriented 'things' in smart homes and automobiles."

Due to the past separation of IT and OT (operational technology), although cloud computing has brought about a great increase in computing power, the addition of "things" still introduces challenges to the network architecture under the cloud model. In the mobile IoT environment of the Edge 2.0 era, IT and OT will be converged and have more powerful intelligent sensing and automation capabilities. In other words, in addition to the centralized data processing enabled by cloud computing, the network edge will bring together an abundance of devices and data, and provide tremendous computing power close to the endpoint, thereby unleashing great business value.

For organizations to take advantage of Edge 2.0 and reap its rewards, one of the things they will need is an application distribution platform centered around holistic application distribution and based on a different set of technology design principles.

An Edge 2.0 application distribution platform must be based on the following key design principles:

Unified control plane

- For an Edge 2.0 application distribution platform, "edge" can be any physical or logical environment ranging from user endpoints to public clouds. A unified application control plane ensures the common definition of security policies, data location policies, and user identity management across different environments and enforces execution via integration with automation and orchestration tools.

Application-oriented

- An Edge 2.0 application distribution platform will integrate fully with application life cycle management tools. Application security policies, data location policies, identity management, and resource orchestration are "declared" via the unified control plane and enforced in any environment the platform runs inside. The Edge becomes a "declared property" of the target application (resulting in every application having its own "personalized" edge) and is "executed" by the platform without required manual provisioning. Developers can simply focus on the application logic, application interactions (APIs), and business workflows without worrying about managing infrastructure or locations.

Distributed security embedded in the platform

- In an Edge 2.0 application distribution platform, application security policies are defined in a common way via the unified control plane. They are distributed for enforcement in every environment the application runs inside. Security capabilities embedded directly in the platform (for example, encryption and best-of-breed bot detection) allow these security functions to move with the application by default.

Distributed data processing and embedded analytics

- An Edge 2.0 application distribution platform becomes a global fabric for application logic and a global fabric for data processing and analytics. Any digital service requires both data and application logic, but the location for storing, processing, and transforming data should not be required to be the same as where the application logic resides. The data location should be specified independently as a set of platform level policies determined by factors such as data gravity, regulations (PCI, GDPR, and others), and the relative price/performance of processing. Similar to security policies, data location policies should be "declared" by the users via the unified control plane and enforced by the platform in any environment. An Edge 2.0 platform has a role to play in other data management policies, such as data lineage—storing details on data as embedded attributes—in addition to having a set of built-in operational capabilities for observability, telemetry streaming, ML tools, and ETL services.

Software-defined elastic edge

- For an Edge 2.0 application distribution platform, the "edge" is no longer defined by physical PoPs in specific locations. Instead, it is defined dynamically by the Edge 2.0 control plane over resources that exist anywhere a customer may desire: in public clouds, hypervisors, data centers or private clouds, or even physical machines in "remote" locations unique to their business. The connection network capabilities are also delivered in a software-defined way overlaid atop private or public WAN infrastructure without herculean effort to assemble and configure. It will respond to the target application's "declaration of intent" by delivering a customized, software-defined elastic Edge at an application's request. "Edge"—and all that it establishes for the application as part of that declaration—will become a simple, easy-to-use property of the application.

Hardware-optimized compute

- The advancement in processor and chipset technologies—specifically GPU, DPU, TPU, and FPGAs that are emerging in capability and capacity—has made it feasible for specialized compute to significantly optimize resource use for specific workload types. An Edge 2.0 application distribution platform will interface with systems that possess this special hardware to target, land, and run specific application workloads that would benefit from this assistance. For example, it will find and configure GPU resources for AL/ML intensive workloads or locate and integrate a DPU for special application security and networking services that an application needs. Hardware awareness in an Edge 2.0 application distribution platform offers intriguing benefits for the creation of special-purpose application-facing intelligent "industrial systems" and thus boundless possibilities for establishing attractive IoT solutions where real-time processing is needed locally. For example, an EV charging station can serve as a data aggregation point of presence for the copious amounts of data generated from the EV sensors, or an autonomous vehicle powered by Android OS can behave like a mobile data center running continuous hardware-assisted self-diagnostics.

Navigate the Third Wave of Internet with Innovative Application Delivery

The distributed and real-time intelligence enabled by the Edge 2.0 application platform will play a central role in tomorrow's digital world. For most enterprises, this means to reimagine their application delivery models. Current application and service delivery is built around a centralized model where application logic is centrally located. Application logic is hosted in public cloud or private data centers. In the Edge 2.0 era, the infrastructure, data, and application architecture will be more distributed and adopt peer-to-peer approaches. However, we envision this transition will be an evolution, namely an augmentation, from today's application delivery technologies, rather than a revolution.

With years of experience in multi-cloud application security and application delivery technology, F5 has always served the needs of applications—the core asset of the organization in the digital era. In the Edge 2.0 era, the edge is changing from a closed model to an open one. With the recent acquisition of Volterra, F5 is in the perfect position to lead the creation of this Edge 2.0 application distribution paradigm.

Today, all industries are accelerating their digital transformation journey. We know that edge deployment is gradually becoming a part of our customers' application strategy, and we look forward to working with them to navigate this new wave of the Internet.

Resources

1 IDC Data Age 2025

2 https://www.gsma.com/wp-content/uploads/2019/04/The-5G-Guide_GSMA_2019_04_29_compressed.pdf

3 Park Associates

4 Statista, "Share of Traffic Coming from Mobile Devices"

5 Cybersecurity Ventures "How Many Internet Users Will the World Have in 2022 and 2030"

6 Definitive HC, "COVID-19 Drives 6000 percent growth in telemedicine use"

7 F5 2021 State of Application Strategy report

8 Cisco Annual Internet Report