

Arista 7050X3 Series Switch Architecture

The growth in adoption of high performance servers using virtualization and containers with increasingly higher bandwidth is accelerating the need for dense 25 and 100G Ethernet switching in both the leaf and spine tiers of modern Enterprise, Cloud and Carrier networks. Next-generation networks require systems that deliver a balance of higher performance, scale and efficiency with enhancements for architectural changes, requiring new tunneling and routing options, advanced monitoring, telemetry and programmability with no loss of existing functionality. The Arista 7050X3 Series are flexible datacenter switches with wire-speed layer 2 and layer 3 features combined with low latency and comprehensive and consistent features for software driven cloud networking that include innovations for load balancing, network tracing and scale:

- Optimized Path Selection for efficient utilization of multipath networks
- Network Address Translation (NAT) at line rate and low latency
- Enhanced network telemetry with triggered buffer capture and flow tracker
- Double the throughput and scale with a fully shared intelligent buffer
- Support for advanced features including segment routing and VXLAN routing

The 7050X3 support for a wide range of interface speeds including 10G, 25G, 40G, 50G and 100G, combined with Arista EOS™, delivers the rich features required for big data, cloud, virtualized and traditional network designs, accommodating the myriad different applications and east-west traffic patterns found in modern datacenters.

The Arista 7050X3 switches enhance the Arista 7050X portfolio with the addition of key new technologies and features with significant improvements in Layer 2 and Layer 3 scale.

The 7050X3 series introduce support for 25G and 100G uplinks to the 7050X portfolio. Operating networks at 100G is both a major increase in bandwidth compared to 40G, as well as a more efficient and cost effective way to scale total bandwidth. With consistent cabling for connections at 10G or 25G, and 40G to 100G, the 7050X3 provides an easy migration path to upgrade the network while protecting the investment in cabling and server infrastructure. This enables customers' networks of all sizes to transition to 25G server technology and get the full benefits of server performance and higher bandwidth. Other enhancements on the 7050X3 include:

- A fully shared 32MB packet buffer common to all ports. Intelligent dynamic buffer management handles speed changes, microbursts or sustained network congestion by allocating buffer fairly to all ports and reserving buffer for critical application traffic. In addition, with support for features such as PFC, ETS and RoCE, the 7050X3 enables lossless Ethernet for storage applications.
- Network scalability is directly impacted by the size of a switch's forwarding tables. In many systems a 'one size fits all' approach is adopted using discrete fixed-size tables for each of the common types of forwarding entry. The Arista 7050X3 platforms leverage a common Unified Forwarding Table (UFT) for the L2 MAC, L3 Routing, L3 Host and IP Multicast forwarding entries, which can be partitioned per entry type. The ideal size of each partition varies depending on the network deployment scenario. The flexibility of the UFT, coupled with the range of pre-defined configuration profiles available on the 7050X3, ensures optimal resource allocation for all network topologies and network virtualization technologies.
- The Arista 7050X3 series packet processor architecture is enhanced with a flexible packet pipeline. This allows for the addition of new capabilities to the forwarding plane of the switch through software upgrades without requiring change or replacement of the system. This enables rapid testing and deployment of new capabilities, avoiding costly replacements or waiting for network upgrades. Together with the flexible resource allocation provided by UFT, the flexible pipeline increases the versatility of the platform, allowing for broader use cases and ensuring investment protection.
- The Arista 7050X3 series architecture supports cut-through and store-and-forward switching. The platform delivers very low latency starting at 800ns with cut-through switching between any two ports of same speed or from higher-speed port to lower-speed port.
- The 7050X3 supports a consistent set of EOS features that are already supported on Arista X-Series systems including Smart System Upgrade (SSU), LANZ and advanced network telemetry as well as packet timestamping. Maintaining operational and feature consistency lowers the qualification time typically associated with introducing new products.

With increased system performance, scale, consistent features and innovations, the 7050X3 platforms are ideally suited for the evolution of large Enterprise datacenters, big data and machine learning environments, traditional and virtualized datacenters, and Service Provider edge networking roles.

Arista 7050X3 Series Model Choice

The 7050X3 Series are available in a choice of models to allow flexibility of interface type and system density.



Figure 1: Arista 7050X3 Series

The table below provides details on the 7050X3 Series models.

Table 1: 7050X3 Series – System Specifications		
7050X3 Models	7050CX3-32S	7050X3-48YC12
Switch Height (RU)	1	1
10G SFP+	2	--
25G SFP	--	48
40G QSFP+	--	--
100G QSFP	32	12
Maximum Density 10GbE ports	128	96
Maximum Density 25GbE ports	128	96
Maximum Density 40GbE ports	32	12
Maximum Density 100GbE ports	32	12
Maximum HW System Throughput (Tbps)	6.4	4.8
Maximum Forwarding Rate (Bpps)	2	2
Latency	From 800 nsec	From 800 nsec
Packet Buffer Memory	32MB	32MB

Arista 7050X3 Series Deployment Scenarios

Each of the 7050X3 models offers multiple connectivity options that provide flexibility in building scalable leaf and spine designs. The operational flexibility offered by the entire 7050X3 series ensures suitability for a variety of deployment scenarios. The following are a selection of use cases:

- Dense top of rack for server racks with both 10GbE and 25GbE systems:
- 10GbE to 25GbE Migration — IEEE 802.3by 25GbE and Consortium compliant for seamless transition to the next generation of Ethernet performance
- Grid / HPC — designs requiring cost-effective and power-efficient systems to enable non-blocking or minimal oversubscription for 10G and 25G Servers
- Leaf-Spine — two-tier designs with open standards based L2 and L3 with telemetry and visibility features
- 100GbE Scale Out Designs — Small to medium locations requiring power efficiency and high density compact systems
- ECMP designs up to 128-way — cost-effective multi-pathing using open protocols and the Arista 7320X and 7500R as 100GbE modular spine switches
- Large scale L2 environments — flexible resource allocations with UFT allow maximum scale

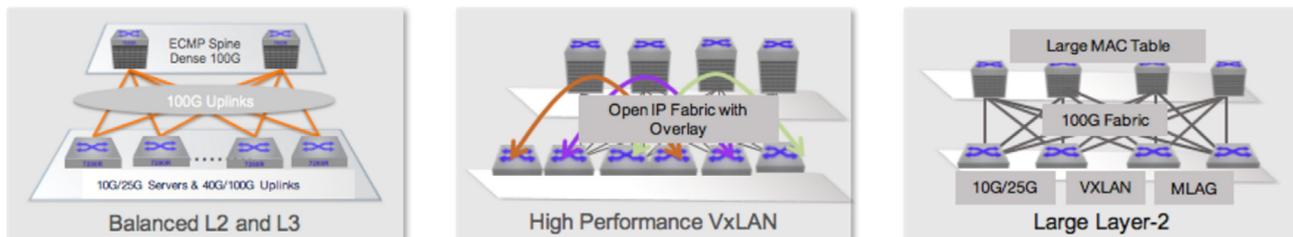


Figure 2: 7050X3 Deployment Scenarios

Arista 7050X3 Switch Architecture

All of the 7050X3 Series share a common system design built around a high performance x86 CPU and 8GB of system memory for the control plane. The CPU is connected to internal flash, bootflash, power supplies, fans, management I/O and peripherals.

The x86 CPU is also connected over PCIe to the Switch on Chip that runs all the data plane forwarding and has all the directly connected front panel ports.

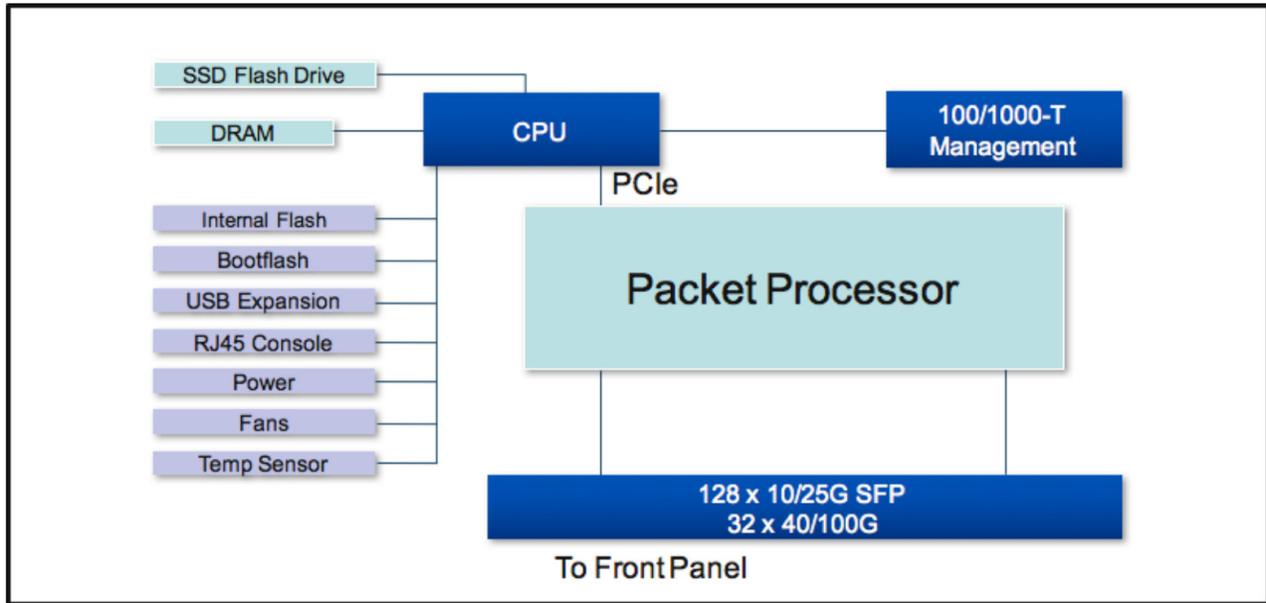


Figure 3: 7050X3 Architecture

Arista 7050CX3-32S

The 7050CX3-32S is a 1RU system with 32 100G QSFP ports offering wire speed throughput of up to 6.4 Tbps. The switch is optimized for high density 100G or 40G spine connectivity or high performance server and storage connectivity with support for 2x50G on each QSFP interface.

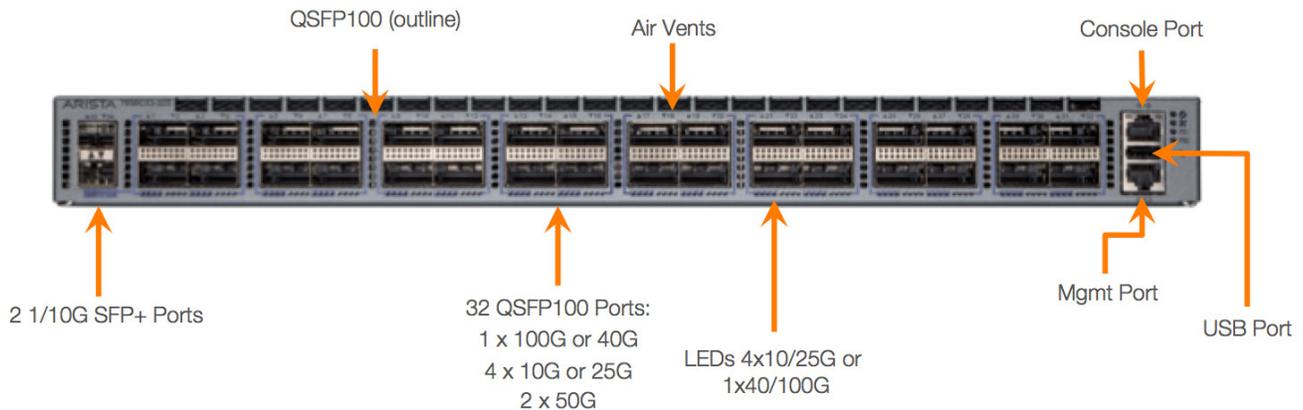


Figure 4: 7050CX3-32S Switch

The 7050CX3-32S is a high density 100GbE system that offers:

- A range of 5 speeds on all QSFP ports for flexible 10GbE to 100GbE with each port being configurable as 1x 40G, 4x 10G, 1x 100G, 4 x 25G, 2x 50G
- QSFP ports with IEEE 25GbE specification support
- Wire speed performance with 32MB of fully shared buffer
- Support for cut-through switching and low latency starting from 800ns

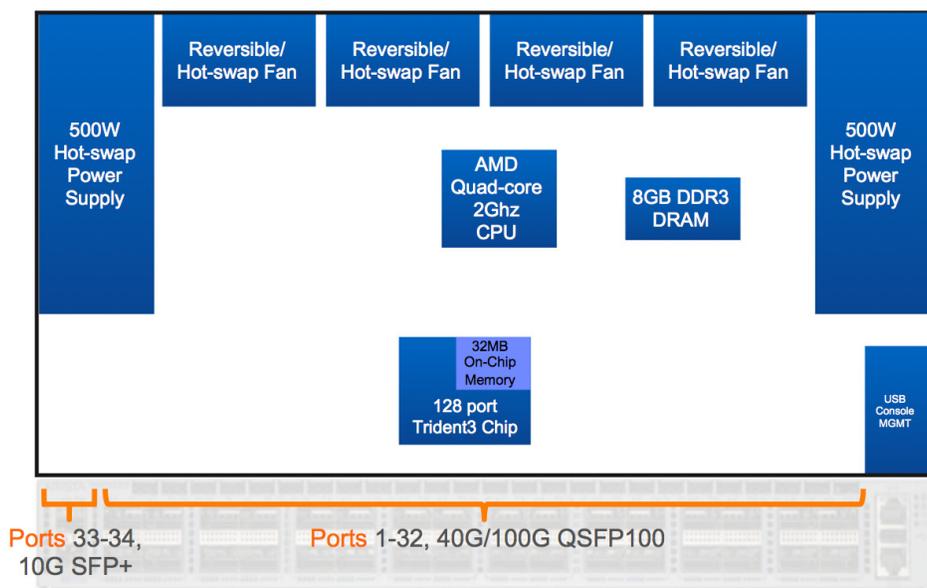


Figure 5: Arista 7050CX3-32S Architecture Block Diagram

Arista 7050SX3-48YC12

The Arista 7050SX3-48YC12 is a 1RU system with 48 ports of 25G SFP and 12 ports of 100G QSFP with an overall throughput of 4.8Tbps. The switch is designed for non-blocking designs with a choice of port configurations in both leaf and spine deployments, and for dense high performance compute and storage racks.

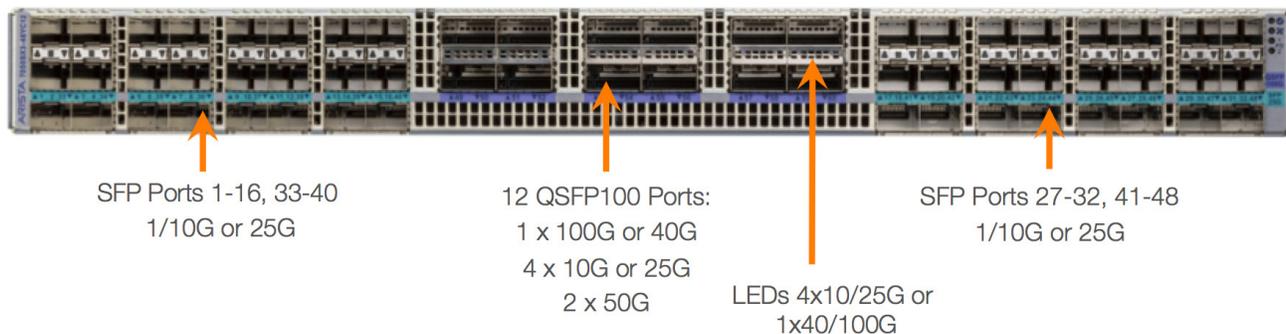


Figure 6: 7050SX3-48YC12 Switch

7050SX3-48YC12 is a high density 25GbE/10GbE system with 100GbE QSFP that offers:

- 48 wire speed 25GbE ports and twelve 40/100G QSFP ports for up to 96 total 25G or 10G ports when used with breakout cables and parallel optics
- Full IEEE 25GbE specification support to enable migration to 25G from 10G
- 48 high density SFP ports can be enabled in groups of 4 to run either at 25G or a mixture of 10G/1G speeds
- 12 QSFP ports each allow for a choice of individually configured speed choices including 100GbE, 40GbE, 4x10GbE, 4x25GbE or 2x 50GbE
- Easy migration from 1/10G to 1/10/25G using familiar SFP connections and cabling solutions
- Wire speed performance with 32MB of buffer
- Support for cut-through switching and low latency starting from 800ns

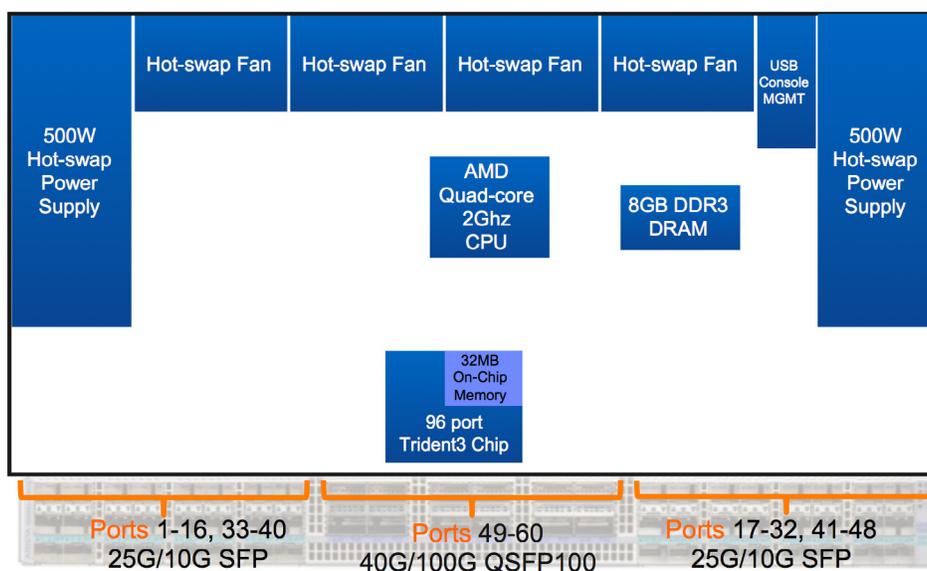


Figure 7: Arista 7050SX3-48YC12 Architecture Block Diagram

Datacenter Grade Availability and Redundancy

The Arista 7050X3 series switches are designed for high availability from both a software and hardware perspective. Key high availability features include:

- 1+1 hot-swappable power supplies and four N+1 hot-swap fans
- Color-coded PSU's and fans
- Live software patching
- Self-healing software with Stateful Fault Repair (SFR)
- Smart System Upgrade (SSU)
- Multi-chassis LAG for active/active L2 multi-pathing
- 128-way ECMP routing for load balancing and redundancy

Table 2: 7050X3 Series - Power and Fan Redundancy		
7050X3 Models	7050CX3-32S	7050SX3-48YC12
Power Supplies (Redundancy)	2 hot swappable (1+1)	
Fans (Redundant)	4 hot swappable (N+1)	
Airflow	Front to rear and rear to front	Front to rear



Figure 8: Arista 7050SX3-48YC12 Switch Rear View

Scaling the Control Plane

The central CPU complex on the 7050X3 Series switches is used exclusively for control-plane and management functions; all data plane forwarding logic occurs at the packet processor level.

Arista EOS®, the control plane software for all Arista switches executes on multi-core x86 CPUs with multiple gigabytes of DRAM. As EOS is multi-threaded, runs on a Linux kernel and is extensible, the large RAM and fast multi-core CPUs provide for operating an efficient control plane with headroom for running 3rd party software, either within the same Linux instance as EOS or within a guest virtual machine.

Out-of-band management is available via a serial console port and/or the 10/100/1000 Ethernet management interface. The 7050X3 Series also offer USB2.0 interfaces that can be used for a variety of functions including the transferring of images or logs.

Table 3: 7050X3 Series CPU Complex		
7050X3 Models	7050CX3-32S	7050SX3-48YC12
CPU	Quad-Core x86	
System Memory	8GB	
Flash Storage	8GB	

The Next Generation Packet Processor

The 7050X3 Series are built using a single Switch on Chip (SoC) silicon. The 7050X3 implements a new flexible pipeline, which delivers a number of advantages, while remaining consistent to the widely deployed 7050X Series. This flexibility enables the addition of new capabilities through modifications to the pipeline, implemented in EOS.

The following section describes various key components on the packet processor and functionality provided by these components.

Table 4: 7050X3 Flexible Forwarding Advantages	
Feature	Benefit
Fully Deterministic Pipeline	Line Rate Throughput and Low Latency
Large Lookup Stages	Higher Memory Efficiency and Reduced Power
Intelligent Parallelism	Enhanced Packet Processing Capacity in Fewer Stages
Configurable Special Functions	Advanced Features not in lookup engines
Run-time Programmable Tables	No Reboot Required to modify features

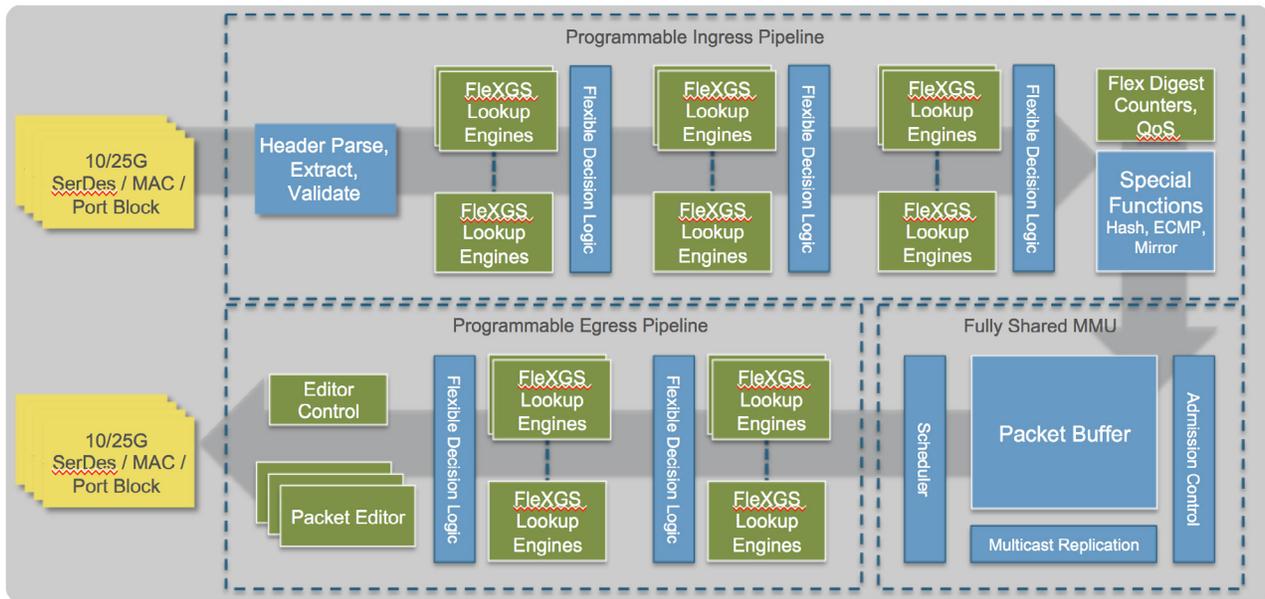


Figure 10: 7050X3 Series Packet Processor

Flexible Pipeline

The Arista 7050X3 series support an enhanced forwarding architecture with a flexible packet pipeline, enabling the addition of new capabilities to the data plane of the packet processor through software upgrades without changes or replacement of the underlying hardware. This allows for rapid testing and deployment of new capabilities such as new header types or packet lookups avoiding costly replacements or major upgrades. Together with a configurable lookup forwarding resource allocation provided by the Unified Forwarding Tables (UFT), the 7050X3 pipeline increases the system flexibility allowing for a broad set of use cases and enables investment protection.

Unified Forwarding Table

Network scalability is directly impacted by the size of a switches’ forwarding tables. In many systems, a ‘one size fits all’ approach is adopted using discrete fixed size tables for each of the common types of forwarding entry. The Arista 7050X3 switches leverage a Unified Forwarding Table (UFT) for the L2 MAC, L3 Routing, L3 Host and IP Multicast forwarding entries, which can be partitioned per entry type. The ideal size of each partition varies depending on the network deployment scenario. The flexibility of the UFT coupled with the range of pre-defined profiles available on the 7050X3 ensures optimal resource allocation for all network topologies and network virtualization technologies.

The UFT holds up to 8 banks of 32K entries each individually allocated to L2 or L3. In normal mode, Bank 2 is always used for L2/MAC entries, however banks 3, 4 and 5 can be individually allocated to either L2 or L3 (host routes).

Table 5: Arista 7050X3 UFT modes

UFT Mode	0	1	2 (Default)	3	4
MAC Addresses	288K	224K	160K	96K	32K
IPv4 Host Routes	16K	80K	144K	168K	16K
IPv4 Multicast (S,G)	8K	40K	72K	104K	8K
IPv6 Host Routes	8K	40K	72K	104K	8K

Additionally, it is possible to use the UFT in Algorithmic LPM (ALPM) mode, where all banks are allocated to the LPM enabling the switch to store up to 384K of IPv4 LPM routes.

Table 6: Arista 7050X3 ALPM mode

LPM Table Mode	ALPM	1	2	3	4
IPv4 LPM Routes	384K	32K	32K	32K	32K
IPv6 LPM Routes Unicast (Prefix Length <= /64)	192K	12K	8K	4K	--
IPv6 LPM Routes Unicast (any prefix length)	40K	2K	4K	6K	8K

In ALPM mode all UFT banks are dedicated to expansion of the LPM table. This reduces the host routes to 16K and MAC addresses to 32K. In ALPM mode expansion of the L3 Host and MAC tables are not possible.

The UFT and ALPM flexibility allows customers to standardize datacenter switching with the 7050X3 Series, deployed across multiple use cases, each most efficiently using all the switch resources available.

Dynamic Fully Shared Buffer Architecture

In cut-through mode, the Arista 7050X3 switches forward packets with a consistent low latency of 800 nanoseconds. Upon congestion, the packets are buffered in an intelligent fully shared packet memory that has a total size of 32MB for superior burst absorption. Unlike other architectures that have fixed per-port packet memory, the 7050X3 Series use dynamic thresholds to allocate packet memory based on traffic class, queue depth and quality of service policy, ensuring a fair allocation to all ports of both lossy and lossless classes. Buffer utilization, occupancy and thresholds are all visible with Arista LANZ and can be exported to monitoring tools to identify hotspots and measure latency at the device and end to end.

The packet buffer is designed to handle both lossy and lossless traffic classes. For traffic classes requiring lossless frame delivery, some fixed buffer amount is set aside to absorb any in-flight packets that arrive after flow control such as PFC/PAUSE is issued. The lossless buffer is a shared pool across all ingress ports with a defined minimum and maximum buffer space for each port. This allows conservative allocation of buffer space across all ports without the possibility of any one port overflowing the buffer. The reserved pool in the diagram shows the total pool allocated for the lossless traffic classes. The remaining buffer is aggregated into a shared buffer pool for the lossy traffic classes. Another important aspect of buffer management is the cell size. The cell size of 208B on the 7050X3 systems ensures the buffering is highly granular, maximizing performance by minimizing unusable buffer.

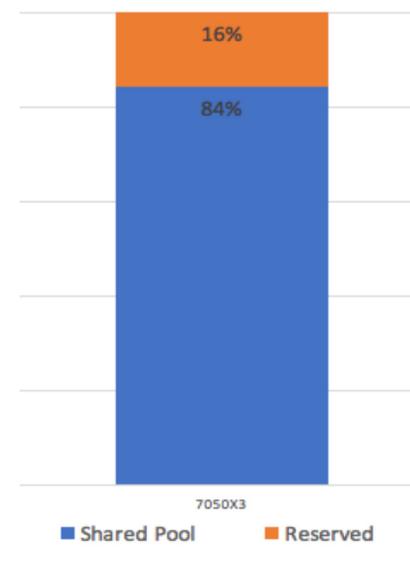


Figure 11: 7050X3 Buffer Allocation

Automated Network Load Balancing

The 7050X3 Series provides optimized network load balancing for traffic in layer 3 ECMP and Layer 2 MLAG environments that improves overall network performance. Enhancements to the 7050X3 load balancing hash algorithms consider the real-time load on links and dynamically assigns new flows to best link. In addition when imbalances are detected, active flows are automatically rebalanced to reduce the probability of link congestion and achieve the maximum throughput.

Cut-Through Switching & Low Latency

The 7050X3 series architecture supports consistent low latency across all ports in the switch. The packet processor supports both cut-through switching and store-and-forward switching. Cut-through switching is supported between any two ports of same speed or from higher speed port to lower speed port. The table below shows the support for cut-through switching between different speed configurations.

Table 7: 7050X3 Cut-through forwarding speeds

Similar Port Speeds	Extended Port Speeds
10GbE to 10GbE	40GbE to 10GbE
25GbE to 25GbE	50GbE to 10GbE
40GbE to 40GbE	50GbE to 25GbE
50GbE to 50GbE	100GbE to 40GbE
100GbE to 100GbE	100GbE to 50GbE
	100GbE to 25GbE

Similar speed and extended speed cut-through forwarding enables the 7050X3 Series to provide low latency for a broad set of customer deployments. Store and forward mode is enabled for low speed to high speed forwarding at packet boundaries, with no impact on cut-through behavior for other flows.

Figure 12 shows two-tier scenarios for typical HFT and HPC use cases. In these leaf spine designs, server to server latency across the two-tiers can be as low as 2us for 100G and 3us for 25G end points.

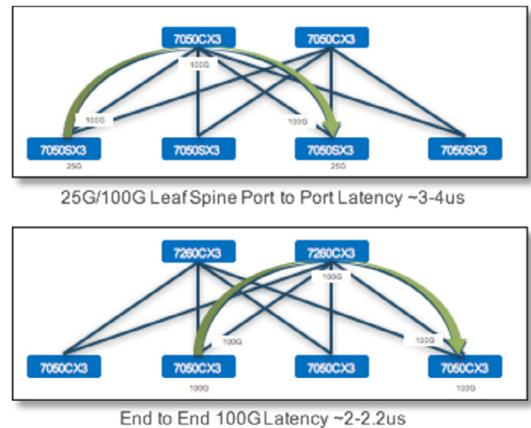


Figure 12:
7050X3 server to server latency

Networks Address Translation (NAT)

Network address translation (NAT) is often leveraged inside a network to mask an internal network or to overcome a limited IP allocation that may be too small for the number of hosts you wish to use. This is common in enterprises, financial trading venues and market exchanges where co-location participants are limited to a small range of IP addresses allocated by the location, exchange or venue. For example, if the venue gives the trading entity 30 addresses and there are 60 servers, a way around this is needed either by obtaining another 30 addresses from the exchange, if available, or using NAT to mask the server addresses. In another example, an exchange may be using a well-known multicast address to publish a feed and need to change the sources or even the group internally without modifying the externally well known S,G information. Typically NAT has been a function of modular routers, firewalls, or software- based routers. All of these devices introduce latency, typically from 10µs up to milliseconds in the worst case.

The Arista 7050X3 offers NAT functionality in a high performance, compact datacenter switch. It implements NAT at the same low latency as standard L2 bridging, L3 routing, or L4 inspection. Delivering address translation in the same device not only reduces latency, it also enables device consolidation, which brings significant reductions in CapEx and OpEx. Multicast NAT is a useful feature for those who are subscribing to or publishing data and can be used to transform traffic so it appears that it came from a single source. It also enables translating multicast groups and destinations to avoid conflicts in the IP infrastructure, or differentiate identical traffic being received at multiple points.

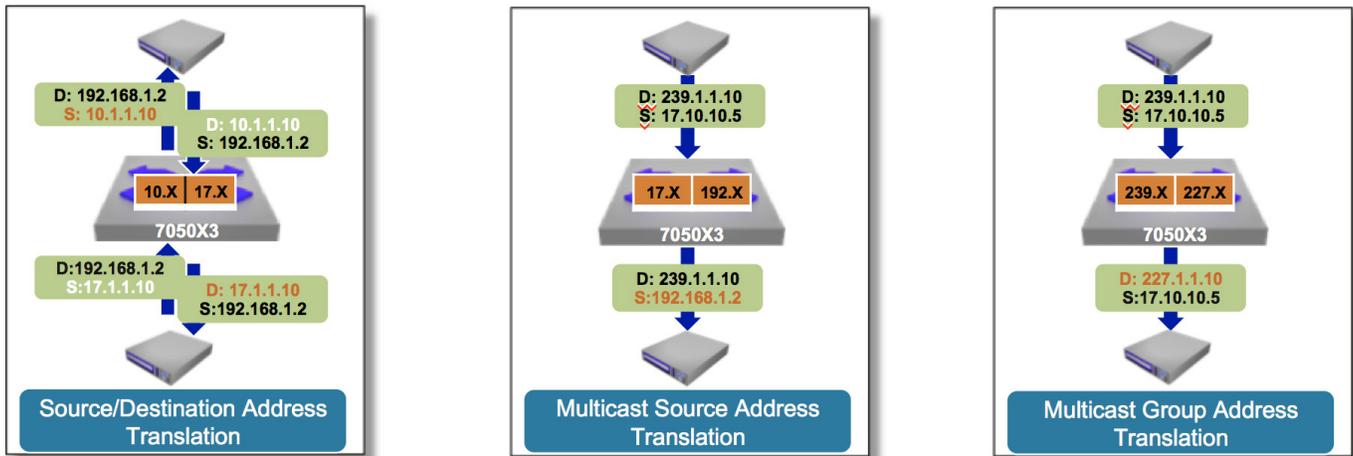


Figure 13: 7050X3 Network Address Translation

802.3by IEEE 25GbE Specification

The 7050X3 Series offers full support for the IEEE 802.3by 25 Gigabit Ethernet standard, ensuring long term investment protection and support for the 25G and 50G Consortium specification for backward compatibility to existing 25G devices.

Unrelenting traffic growth and storage capacity expansion is driving demands for higher performance networks. The introduction of 25GbE provides a 2.5X performance improvement over 10GbE while using the same familiar cabling and designs. Support for 10G/25GbE modes allows for future investment protection with the ability to migrate as needed without expensive network upgrades.

Some of the advantages for migrating servers and storage to 25GbE include:

- Maximize the switch and server throughput and efficiency, by using all available bandwidth on high performance systems
- Reduce capital expense by lowering the number of cables and switch ports compared to using multiple 10G ports to increase bandwidth
- Lower operational expenses by reducing power and cooling compared to a 40G alternative
- Lowest cost per bit of performance since 25GbE provides 2.5X higher throughput using the same switch technology and cabling

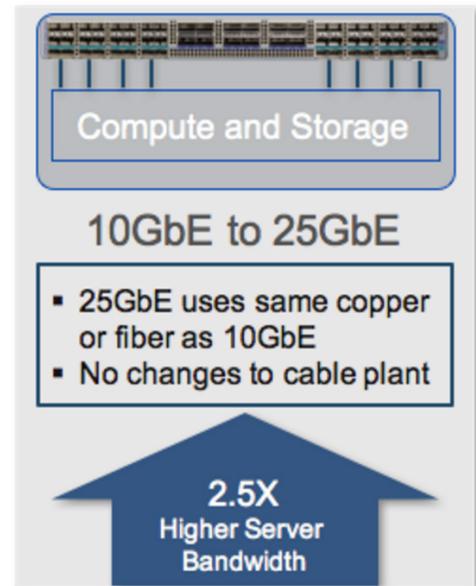


Figure 14: 10GbE to 25GbE Migration Advantages

The 7050X3 Series are fully compliant to both the 25G Consortium specification and the IEEE 802.3by 25GbE standard and provide an easy migration path for many 10G networks. The 25GbE interfaces are backward compatible with a wide range of 10G SFP+ optics and cables, allowing for dual speed support with each port independent of the others. Changing interface speeds from 10G to 25G is performed hitlessly without disrupting traffic on other ports, allowing for migrations to take place over time. In addition, all 7050X3 Series 100GbE ports allow for 5 interface speeds, including 10G, 25G and 50G with parallel optics and breakout cables, as well as both 40G and 100G.

Arista EOS: A Platform for Scale, Stability and Extensibility

At the core of the Arista 7050X3 Series is Arista EOS® (Extensible Operating System). Built from the ground up using innovations in core technologies since our founding in 2004, EOS contains more than 10 million lines of code and over 1000 man-years of advanced distributed systems software engineering. EOS is built to be open and standards-based, and its modern architecture delivers better reliability and is uniquely programmable at all system levels.

EOS has been built to address two fundamental issues that exist in cloud networks: the need for non-stop availability and the need for high-feature velocity coupled to high quality software. Drawing on our engineers' experience in building networking products for more than 30 years and on state-of-the-art open systems technology and distributed systems, Arista started from a clean sheet of paper to build an operating system suitable for the cloud era.

At its foundation, EOS uses a unique multi-process state-sharing architecture where there is separation of state information from packet forwarding and from protocol processing and application logic. In EOS, system state and data is stored and maintained in a highly efficient, centralized system database. The data stored is accessed using an automated publish/subscribe/notify model. This architecturally distinct design principle supports self-healing resiliency in our software, easier software maintenance and module independence, higher software quality overall, and faster time-to-market for new features that customers require.

Arista EOS contrasts with the legacy approach to building network operating systems developed in the 1980's that relied upon embedding system state held within each independent process, extensive use of inter-process communications (IPC) mechanisms to maintain state across the system, and manual integration of subsystems without an automated structured core. In legacy network operating systems, as dynamic events occur in large networks or in the face of a system process failure and restart, recovery can be difficult if not impossible.

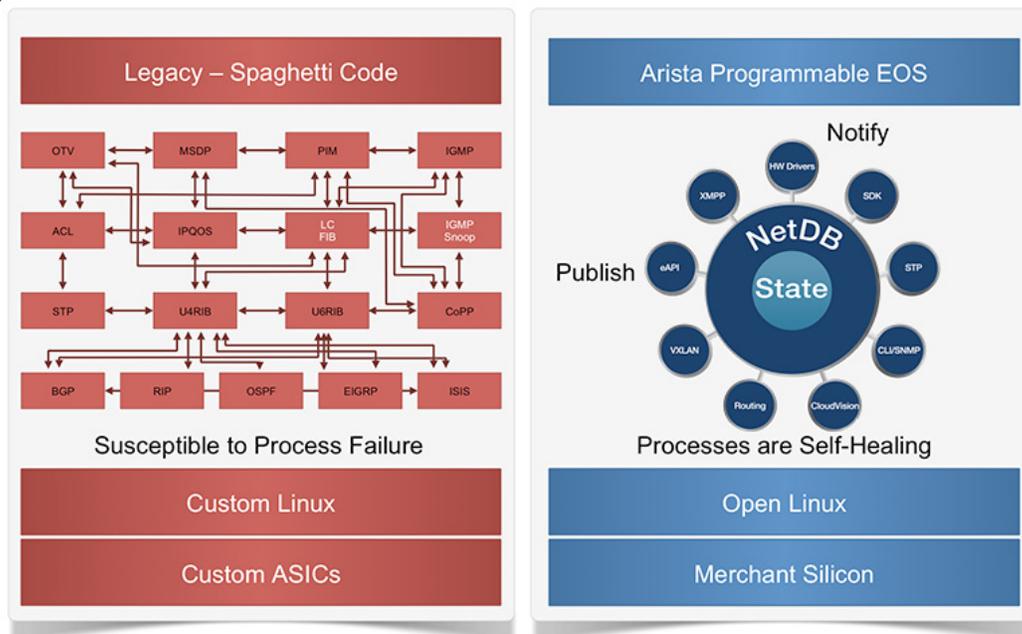


Figure 15: Legacy approaches to network operating systems (left), Arista EOS (right)

Arista took to heart the lessons of the open source world and built EOS on top of an unmodified Linux kernel. We have also maintained full, secured access to the Linux shell and utilities. This allows EOS to preserve the security, feature development and tools of the Linux community on an on-going basis, unlike legacy approaches where the original OS kernel is modified or based on older and less well-maintained versions of Unix. This has made it possible for EOS to natively support things like Docker Containers to simplify the development and deployment of applications on Arista switches. Arista EOS represents a simple but powerful architectural approach that results in a higher quality platform on which Arista is faster to deliver significant new features to customers.

EOS is extensible at every level, with open APIs at every level: management plane, control-plane, data plane, services-level extensibility, application-level extensibility and with access to all Linux operating system facilities including shell-level access. Arista EOS can be extended with unmodified Linux applications and a growing number of open source management tools to meet the needs of network engineering and operations.

Open APIs such as EOS API (eAPI), OpenConfig and EOS SDK provide well-documented and widely used programmatic access to configuration, management and monitoring that can stream real-time network telemetry, providing a superior alternative to traditional polling mechanisms.

Conclusion

The 7050X3 switches support a consistent forwarding architecture and set of EOS features that are already supported on other Arista X-Series systems including Smart System Upgrade, LANZ and Network Telemetry as well as packet timestamping. Maintaining operational and feature consistency while increasing performance and scale lowers the qualification time typically associated with introducing new products, and the 7050X3 systems seamlessly insert into existing networks.

With the increased performance and scale, low latency, higher power efficiency, consistent features and networking innovations, the 7050X3 platforms are ideally suited for the evolution of large Enterprise, big data and machine learning environments, and Service Provider edge networking roles.

Santa Clara—Corporate Headquarters

5453 Great America Parkway,
Santa Clara, CA 95054

Phone: +1-408-547-5500

Fax: +1-408-538-8920

Email: info@arista.com

Ireland—International Headquarters

3130 Atlantic Avenue
Westpark Business Campus
Shannon, Co. Clare
Ireland

Vancouver—R&D Office

9200 Glenlyon Pkwy, Unit 300
Burnaby, British Columbia
Canada V5J 5J8

San Francisco—R&D and Sales Office

1390 Market Street, Suite 800
San Francisco, CA 94102

India—R&D Office

Global Tech Park, Tower A & B, 11th Floor
Marathahalli Outer Ring Road
Devarabeesanahalli Village, Varthur Hobli
Bangalore, India 560103

Singapore—APAC Administrative Office

9 Temasek Boulevard
#29-01, Suntec Tower Two
Singapore 038989

Nashua—R&D Office

10 Tara Boulevard
Nashua, NH 03062



Copyright © 2018 Arista Networks, Inc. All rights reserved. CloudVision, and EOS are registered trademarks and Arista Networks is a trademark of Arista Networks, Inc. All other company names are trademarks of their respective holders. Information in this document is subject to change without notice. Certain features may not yet be available. Arista Networks, Inc. assumes no responsibility for any errors that may appear in this document. Mar 29, 2018 02-0077-01