Arista 7170 Multi-function Programmable Networking

Much of the world’s data is generated by popular applications, such as music and video streaming, social media, messaging, pattern recognition, peer-to-peer banking, photo sharing, interactive maps and directions, and online retail. The massive data growth spawned by cloud-based applications and new artificial intelligence tools in turn mandate a critical requisite to scale legacy networks to handle highly programmable paths to drive intelligence without compromising performance.

To facilitate modern applications across diverse wireless and wired online connectivity options and ensure access to the data, networks of today require large numbers of additional complex and expensive systems that redirect, filter, monitor and translate requests for information and services from the smartphone, laptop or voice-controlled digital assistant.

Stepping up once again as the trail-blazer, Arista has introduced a new generation of fully-programmable multi-functional platforms with uncompromised performance that enable legacy networks to migrate to a high-powered networking suite that boosts capacity, lowers capital expenditure, and speeds up service deployment.

Arista’s 7170 Series is the first in a new generation of fully programmable leaf and spine platforms that supports a rich set of data plane, and control plane features, facilitating protocols to be moved out of hardware into software without risking performance.
Multi-function Programmable Platform
The Arista 7170 Series, based on the Barefoot Tofino switch chips, are multi-functional programmable 100GbE systems built for the highest performance environments in the largest data centers. With a highly flexible and programmable architecture, the 7170 series with Arista EOS® (Extensible Operating System) enable forwarding profiles to specifically customize the data plane for an array of multiple network roles for cloud, enterprise, service provider and content delivery networks. The forwarding profiles, written in open-source P4 Language, address advanced telemetry, network security, and flexible tunnel termination and address translation, enabling network operators to control granular packet processing to customize specific use-cases.

The 7170 series supports a unique programmable pipeline that allows the parser and pipeline to flexibly add new functionality, and vary look-up table size and operations. This capability provides extraordinary investment protection besides support for highly personalized use cases to solve a range of network challenges, not feasible with traditional switching platforms. Arista EOS' comprehensive set of features, and programmability for cognitive data and control lets the same platform be used in multiple roles, each with its own profile, ensuring consistent management and provisioning.

Here are some solutions that the 7170 Series enable as shown in Figure 1:

- Network Overlay and Virtualization to offload network functions such as traffic segmentation or tunnel encapsulation from virtual servers in a bare metal environment, free up compute, and accelerate applications;
- Flexible routing and segmentation to support a large FIB along with customizations, such as remove/change VLAN tags, modify MPLS labels, and implement custom protocols;
- Network security with firewall functionality, such as learn and track micro flows and sessions, identify anomalies, and take dynamic preventive measures; Large Scale Network Address Translation to conceal internal networks, allow duplicate address spaces, and facilitate IPv4 to IPv6 migration;
- Network and application telemetry for flow-level visibility, custom application counters, configurable thresholds and alarms, timestamping, and end-to-end latency; and,
- Large-scale ACL for stateful ACLs, rule-based policies, hierarchical ACLs and conditional filtering for enhanced security and application performance.

Network services such as firewalling, load-balancing and NAT have so far been delivered in physical network appliances, router service modules, and increasingly in software instances to scale performance, reliability and ensure flexibility with programmable architectures.
Network appliances and routers, however, have limited throughput, high latency, and complicated support and licensing models. In sharp contrast, the Arista 7170 Series replaces expensive, legacy networking appliances and routers by moving critical services into the network.

The multi-functional platform increases server performance by off-loading vSwitch networking functions, such as tunnel termination, security policy enforcement and address translation on to the leaf switch, embedding the network services back into the leaf and spine of the network avoiding additional devices, hairpin traffic through appliances, and new operational models. The compute pool accomplishes more workload while lowering equipment, power and cooling costs.

New protocols and functions can be enabled through custom profiles developed in collaboration with Arista’s engineering team. Profiles can be installed as RPMs and initiated by a single command. The combination of Arista 7170 series and EOS with an open, programmable SDK creates a fungible networking data plane, eliminating costly and disruptive hardware upgrades.

**Arista 7170 Series**

The 7170 Series is available in various models to suit modern datacenter network requirements. The 7170 Series, as in Figure 2, are optimized for deployment, as either the leaf or spine tier, consuming less than 5W per 100G port, with a choice of airflow directions.

![Figure 2: 7170 Series](image)

Each 7170 model is a fixed configuration system designed around a single high performance system-on-chip packet processor, as shown in Figure 3, that has up to 256 x 25G SerDes, allowing for up to 64 x 100G, or up to 256 x 25G interfaces. The packet processor pipeline comprises a programmable parser, fully programmable match-action engines, and both TCAM and SRAM memory.

![Figure 3: 7170 Series System-on-Chip Packet Processor](image)
The Table 1 below provides system details on the 7170 Series.

<table>
<thead>
<tr>
<th>Table 1: 7170 Series – System Specifications</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>7170 Models</strong></td>
</tr>
<tr>
<td>Switch Height (RU)</td>
</tr>
<tr>
<td>10G SFP+</td>
</tr>
<tr>
<td>25G SFP</td>
</tr>
<tr>
<td>40G QSFP+</td>
</tr>
<tr>
<td>100G QSFP</td>
</tr>
<tr>
<td>Maximum Density 10GbE ports</td>
</tr>
<tr>
<td>Maximum Density 25GbE ports</td>
</tr>
<tr>
<td>Maximum Density 40GbE ports</td>
</tr>
<tr>
<td>Maximum Density 100GbE ports</td>
</tr>
<tr>
<td>Maximum HW System Throughput (Tbps)</td>
</tr>
<tr>
<td>Maximum Forwarding Rate (Bpps)</td>
</tr>
<tr>
<td>Latency</td>
</tr>
<tr>
<td>Packet Buffer Memory</td>
</tr>
</tbody>
</table>

**Arista 7170 Series Deployment Scenarios**

The Arista 7170 Series, in combination with other fixed and modular Arista 7000 Series switches delivers the ability to build large-scale leaf and spine data center designs.

The flexible profiles coupled with the programmable pipeline ensure a range of deployment scenarios. A few of the typical use cases as in Figure 4 are as follows:

- Smart top-of-rack supporting multi-tenant networks in server racks with both 10GbE and 25GbE systems, the 7170 series offer high performance tunneling, NAT, Stateful firewalling, QoS, and packet timestamping.

- The emergence of 50GbE and 100GbE host connectivity for the largest scale of high performance compute (HPC) is driving a requirement for dense 100G in fixed form factors. In these areas, the 7170 series deliver cost-effective advanced scaling features, ensuring long term investment protection.

- DDOS mitigation enables tracking connections to identify DDOS traffic and react by modifying forwarding rules.

- Network function virtualization (NFV) typically require tight integration for service provider services. The programmable architecture and flexible resources of the 7170 series can be leveraged to provide service-chaining with segment routing and scale to thousands of tunnels.

- The 7170 series also provide solutions for large-scale tunnel termination in a multi-tenant design, NAT64 for IPv4 to IPv6 translation to facilitate migration to IPv6 only networks.
Arista 7170 System Architecture

Arista 7170 Series is designed with a high performance x86 CPU connected to internal flash, bootflash, power supplies, fans, management I/O and peripherals as presented below in Figure 5. The x86 CPU is further connected over PCIe to the Switch on Chip that runs all the data plane forwarding with direct connections to all the front panel ports. The 7170-64C comes with a choice of 16GB or 64GB of system memory, while the 7170-32C comes with 16GB or 32GB of system memory for the control plane.

Arista 7170-64C

The 7170-64C is a 2U 64-port QSFP100 switch with up to 4.8 billion packets per second and 12.8Tbps of throughput.

All ports are 100G and 40G capable, accept a choice of optics or copper cables, and can be configured for 10/25/40/50/100G. An additional 2 SFP+ ports allow for individual 10G dedicated connections. As shown below in Figure 6, additional front panel ports include a 100/1000 Ethernet interface for out-of-band management, a serial RJ45 connection and a USB port.
The 7170-64C is available as either front-to-rear or rear-to-front cooled, and has 4 hot-swap fans and dual AC or DC power supplies. Figure 7 below illustrates the 7170-64C system architecture.

Arista 7170-32C

The Arista 7170-32C is a 1RU high performance system with 32 ports of QSFP100 for up to 2.4 billion packets per second and 6.4Tbps of throughput. As with the 7170-64C, all ports support a combination of optics and cables for interface speeds from 10G to 100G. The 7170-32C have two SFP+ ports for 1/10G dedicated connections on the front. At the rear is a 100/1000 Ethernet interface for out-of-band management, a serial RJ45 connection and a USB port.

The 7170-32C models, seen above in Figure 8, are available as either front-to-rear or rear-to-front cooled, and have 4 hot-swap fans and dual AC or DC power supplies. Figure 9 seen below presents the 7170-32C systems architecture.
**DataCenter Grade Availability and Redundancy**

The Arista 7170 switches are designed for continuous operations with system-wide monitoring of hardware and software components, simple serviceability, and provisioning that prevent single points of failure. Key high availability features include:

- 1+1 hot-swappable power supplies along with four hot-swap fans provide dynamic temperature control combined with N+1 redundancy as illustrated in Table 2,
- Color coded PSU’s and fans that deliver platinum-level power efficiency,
- Both rear-to-front and front-to-rear airflow options,
- Live software patching,
- Self-healing software with Stateful Fault Repair (SFR),
- Changing profiles that do not require a system reload, and
- Smart System Upgrade (SSU) and Accelerated Software Update (ASU)

### Table 2: 7170 Series - Power and Fan Redundancy

<table>
<thead>
<tr>
<th>7170 Models</th>
<th>7170-64C</th>
<th>7170-32C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power Supplies (Redundancy)</td>
<td>2 hot swappable (1+1)</td>
<td></td>
</tr>
<tr>
<td>Fans (Redundant)</td>
<td>4 hot swappable (N+1)</td>
<td></td>
</tr>
<tr>
<td>Airflow</td>
<td>Front to rear and rear to font</td>
<td></td>
</tr>
</tbody>
</table>

**Scaling Control Plane Services**

The central CPU complex on the 7170 Series switches, as presented below in Table 3, is used for all control-plane and management functions while data-plane forwarding logic occurs at the programmable packet processor. 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 system memory and fast multi-core CPUs support an efficient control plane with headroom for running third-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 7170 Series also offer USB2.0 interfaces that can be used for various functions, including the transfer of images or logs.

### Table 3: 7170 Series CPU Complex

<table>
<thead>
<tr>
<th>7170 Models</th>
<th>7170-64C</th>
<th>7170-32C</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU</td>
<td>Multi-Core x86</td>
<td>Multi-Core x86</td>
</tr>
<tr>
<td>System Memory</td>
<td>64GB or 16GB</td>
<td>32GB or 16GB</td>
</tr>
<tr>
<td>SSD</td>
<td>120GB or 30GB</td>
<td>60GB or 30GB</td>
</tr>
</tbody>
</table>

**Programmable and Flexible Packet Processor**

The 7170 Series is built using a single System-on-Chip (SoC) packet processor that provides both the ingress and egress forwarding for all packets that arrive on or are destined to ports located on the packet processor. All stages of the packet forwarding are performed entirely in the hardware/data plane. The 7170 architecture enables a programmable parser and pipeline along with flexible resource allocation through a series of profiles. Below in Figure 10 is the block diagram showing key components of the packet processor and the functionality provided by each component. There are four identical pipelines, allowing the system to scale to 64 x 100G or 256x 25G with a series of control and configuration interfaces to the system CPU.
Each pipeline is shared by a number of ports, up to 16 x 100G or 64 x 25G. Packets arriving at the ingress MAC are processed by the corresponding ingress match-action pipeline and then enqueued in a common shared memory block which connects the ingress and egress ports. On being scheduled for transmission, packets are passed through the egress match-action pipeline to the transmit MAC.

The following section, as illustrated in Figure 11, provides a detailed look at packet processing through the pipeline.

The parser analyzes incoming packets and maps them to corresponding set of fields called Packet Header Vectors (PHVs). PHVs are then passed through a sequence of 12 match-action units.

The MAUs contain Arista-defined tables that are used to make forwarding and packet rewrite decisions. The series of match-action logic, order, operation, and size and structure of the memory tables are defined in the data plane profile. As PHVs pass through match-action units, keys are extracted from the set of packet fields. Each match-action unit has dedicated SRAM and TCAM banks, which hold up to 12 look-up tables. Table hits can trigger counters, policers, PHV writes and memory updates. As PHVs are passed through the match-action units, the pipeline state from one table can be used as the key to another. Finally, the de-parser reassembles the modified packets for transmission.
7170-64C
The 7170-64C series uses all four pipelines for 64 x 100G ports, and each pipeline allows up to 12 stages on ingress and egress. The match-action engines are used for performing L2 and L3 operations, and any remaining unused match-action stages can be leveraged for custom packet processing. Figure 12 below illustrates 7170-64C Series Packet Processor Architecture.

7170-32C
The 7170-32C also has a four pipeline architecture, with two pipelines assigned for the 32 x 100G ports. The match-action engines are used for performing L2 and L3 operations. Unlike the 7170-64C, two additional pipelines are available for extra processing as needed, and can be used to execute custom code in 7170-32C. Packets are circulated to the extra stages as needed, before being scheduled to the egress match-action pipeline. Figure 13 below shows the 7170-32C Series Packet Processor Architecture.

Shared Buffer Architecture
The 7170 Series incorporates an advanced traffic manager with 22MB of packet buffer that is fully shared across all ports, making it an excellent choice for scalable data centers and intensive workloads. Unlike legacy architectures where the total buffer is statically allocated to a port or group of ports, or the packet buffer memory are multiple slices, the 7170 Series packet buffer is dynamically allocated across all ports with the ability to adjust in real time to the demands of bursty applications, mixed interface speeds and congestion.

QOS Architecture
The 7170 Series supports a sophisticated and advanced traffic manager that supports traffic-shaping along with WRR and SP scheduling. The architecture supports up to seven transmit queues per port, shared by unicast and multicast traffic. The Traffic Class (TC), traffic type and egress port select a transmit queue. The 7170 Series supports three trust modes including CoS trusted, DSCP trusted and untrusted along with CoS and DSCP rewrite capability.
Shared Buffer Architecture
The 7170 Series incorporates an advanced traffic manager with 22MB of packet buffer that is fully shared across all ports, making it an excellent choice for scalable data centers and intensive workloads. Unlike legacy architectures where the total buffer is statically allocated to a port or group of ports, or the packet buffer memory are multiple slices, the 7170 Series packet buffer is dynamically allocated across all ports with the ability to adjust in real time to the demands of bursty applications, mixed interface speeds and congestion.

QOS Architecture
The 7170 Series supports a sophisticated and advanced traffic manager that supports traffic-shaping along with WRR and SP scheduling. The architecture supports up to seven transmit queues per port, shared by unicast and multicast traffic. The Traffic Class (TC), traffic type and egress port select a transmit queue. The 7170 Series supports three trust modes including CoS trusted, DSCP trusted and untrusted along with CoS and DSCP rewrite capability.

Arista AlgoMatch™
AlgoMatch is an Arista innovation combining software and hardware to implement Access Control Lists for policy control and network telemetry that are more scalable and require lower power use. AlgoMatch utilizes an efficient packet matching algorithm that enables flow matching for access control, policy and visibility that is more flexible than traditional technology, allowing multiple actions to be performed on a single packet or flow, and user defined filters for packet classification and custom actions. AlgoMatch implements Access Control Lists without hardware TCAMs.

Indeed, Algomatch delivers the efficiency and flexibility required for next generation architectures in data centers to address the demands of increased scale, whether it be dual stack implementations, or flow analysis at Layer-4 and above.

Flexible Switch Forwarding Profiles
As with all products in the Arista portfolio, a single common EOS image supports the 7170 series, ensuring a common control plane for feature consistency across multiple system architectures to deliver higher quality and simplified network operations.

To leverage the programmable capability of the 7170 Series pipeline, the 7170 series introduces a series of forwarding profiles that are packaged with EOS. Profiles are used to define the data plane. Each profile is optimized to maximize the set of forwarding services and scale for a specific use case. As new profiles are introduced, they can be added either through EOS upgrades or by loading a new RPM directly to the system. While profiles are defined in P4, and used to program the data plane, EOS is still the operating system for control and management planes, with a set of open programmable interfaces and SDK.

Toggling between profiles is hitless, with just a brief pause in packet processing without rebooting or restarting the system.

Changing Profiles
Changing profiles is a simple operation, and requires the use of just a single command as shown below in Figure 15. The “default” profile is replaced by the “baremetal” profile as the active profile although both remain available.
Adding Profiles

Additional profiles can be added by installing RPMs as shown below in Figure 16. The file is first copied to the system, and then installed, making it available as a profile.

```
add29(config)#sh platform barefoot profile
Profile Chips
---------------------- -----
baremetal BfN0
default
---------------------- -----
```

```
add29(config)#bash
[admin@add29 ~]$ ls /mnt/flash/p4-ar-switch-nat.1686.rpm
[admin@add29 ~]$ sudo rpm -i /mnt/flash/p4-ar-switch-nat.1686.rpm
[admin@add29 ~]$ exit
```

```
add29(config)#sh platform barefoot profile
Profile Chips
---------------------- -----
baremetal BfN0
default Nat <<---- NAT profile added
---------------------- -----
```

Figure 15: 7170 Series Profile Addition

Telemetry Services

Arista 7170 series provide advanced application-level performance monitoring, parsing of specific fields in packet headers, which then can be combined with flow and packet level statistics for further analysis to reflect the network behavior in real time. Additional capabilities such as, accurate packet time-stamping, packet payload size and queue occupancy help identify sources of congestion from microbursts or misbehaving applications, and give visibility to decisions made along the network path. Offloading network functions to Arista 7170 series switches further provides the added advantage to monitor underlay and overlay networks from a single network element.

Multi-Tenant Bare Metal Hosting

Bare-metal hosting is a public cloud service that lets customers rent hardware resources from a service provider. Flexibility is a key benefit that bare-metal cloud provides to customize a bare-metal system to meet the unique requirements without interference from neighboring virtual machines (VMs). Bare metal environments are ideal for big data applications or high-transaction workloads that do not tolerate latency.

In the bare-metal environment, additional segmentation, security, encapsulation/tunnel and overlay services are enabled in either dedicated hosts or appliances with all traffic directed through these systems, adding to the latency and lowering performance. The Arista 7170 bare metal profile, however, offloads the virtualized services into the network, ensuring consistent performance and comprehensive services. Figure 17 presents the Multi-tenant Bare Metal Profile.

![Multi-tenant Bare Metal Profile](image-url)
Example of Cost Advantages

SmartNIC Offload
A typical rack design consisting of servers with 100G capable SmartNICs connecting to a traditional, fixed pipeline top-of-rack switch with 32 ports of 100G is approximately 30% costlier than one with regular 100G NICs connecting to an Arista 7170 series top-of-rack switch. If the comparison is made with 25G capable SmartNICs, the cost difference increases to 45%. The cost difference varies between 40%-50% when using a 64-port 100G fixed pipeline top-of-rack switch with SmartNICs compared to a 64-port Arista 7170 series switch. Cost savings can also be attributed to centralizing the cost of memory and equipment needed to support a programmable networking dataplane per rack using Arista 7170 series, rather than on every server using SmartNICs.

Routers and Services Module
Many traditional router platforms can deliver additional packet processing through re-circulating a packet through the processor until the set of operations are complete. This is termed a run-to-completion architecture as compared to the pipeline architecture of the 7170 Series. The router throughput linearly reduces with the additional processing steps, and the ability to add new services is limited by the processor specifications and the system vendor, affecting the flexibility.

An alternate approach is to deliver advanced services using “Services Modules” that contain one or more dedicated NPUs or FPGAs that are designed to handle the acceleration of intensive packet processing services that cannot be supported by the regular run-to-completion packet processors. However, complex chipsets combined with high power consumption and an inability to provide throughput aligned to data center scale make service modules expensive. A typical service module has less than 100Gbps of total aggregate throughput, despite being a dedicated line card with no physical interfaces. In contrast, the 7170 Series provides over 20X greater throughput while supporting and accelerating rich services, such as network and application telemetry, large scale network address translation, tunnel encapsulation, and secure traffic-filtering. Additionally, the 7170 is optimized for the performance requirements of cloud scale data centers with a wide range of physical interfaces in dense form factors and power consumption of less than 5W per 100G port. When comparing the overall cost of a services module, the additional feature licenses and higher operational costs also need to be considered. At best, the service modules offer one tenth the performance for two to three times the cost, resulting in approximately a 200%-300% cost advantage of the 7170 Series.

Arista EOS: A platform for scale, stability and extensibility
At the core of the Arista 7170 Series is Arista EOS®. Built from the ground-up using innovations in core technologies since our founding in 2004, EOS contains more than eight 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: non-stop availability, and high feature velocity coupled with high quality software. Drawing on our engineers' extensive over 30-year experience in building networking products and state-of-the-art open systems technology and distributed systems, Arista started with 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 that separates state information from packet forwarding, protocol processing and application logic. In EOS, system state and data is stored and maintained in a highly efficient, centralized System Database (SysDB). The data stored in SysDB is accessed using an automated publish/subscribe/notify model. This architecturally distinct design principle provides self-healing resiliency in Arista’s software, easier software maintenance and module independence, higher software quality overall, and faster time-to-market for new features.
Arista EOS, as seen in Figure 17, contrasts with the legacy approach to building network operating systems developed in the 1980s 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 like SysDB. 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.

Arista took to heart the lessons of the open source world and built EOS on top of an unmodified Linux kernel maintaining 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 architectural approach lets EOS natively support, for example, Docker Containers that simplify the development and deployment of applications on Arista switches. Arista EOS represents a simple but powerful architectural approach that provides a robust platform to swiftly deliver significant new features.

EOS is extensible at every level, with open APIs at the management plane, control-plane, and data-plane, coupled with services-level extensibility, application-level extensibility, and 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) along with OpenConfig and EOS SDK provide well-documented and widely-used programmatic access to configuration, management and monitoring that can stream real-time network telemetry for a superior alternative to traditional polling mechanisms.
Conclusion
Arista’s open and extensible EOS combined with the flexibility and multi-function programmability offered by Arista 7170 series, is engineered to address real-world use cases like scaling, network encapsulation and support for new protocols in the cloud, eliminating the need for costly, dedicated application-specific hardware. The bottom line is Arista’s 7170 Series’ fully-programmable platform lets customers migrate to a powerful networking suite that can be customized to fit multiple customer bases - public cloud, enterprise, service and content providers.

References:
https://searchsdn.techtarget.com/definition/virtual-network-functions
http://www.mellanox.com/related-docs/prod_adapter_cards/PB_BlueField_Smart_NIC.pdf
https://www.sigarch.org/the-new-life-of-smartnics/
https://www.usenix.org/conference/nsdi18/presentation/firestone
https://www.netronome.com/blog/what-makes-a-nic-a-smartnic-and-why-is-it-needed/
https://www.cavium.com/pdfFiles/ProductBrief_LiquidIO_II_10GBASET_CN2358002-00A.pdf