Campus Architectures and Stackable Switches

History of Stackable Switches

Traditional wiring closet switch architectures are built using either fixed configuration standalone switches, stackable switches or modular chassis systems. Each approach has its own benefits and operational trade-offs. The objective of this paper is to analyze traditional stackable switch architectures and to discuss key considerations when designing next generation cognitive campus deployments in the enterprise.

Normally a 24 or 48 port fixed configuration switch can operate as a standalone switch or as a member in a stack of other switches. Typically one switch within the stack is elected as the primary and serves as the master controller for the other switches in the stack. Individual switches within a stack are usually connected through a proprietary link. Stacked architectures were originally developed in the mid 1990s to address the following needs:

- Simplified Management: Providing a single administrative management point by making the stack of switches appear as a single entity.
- Incremental Expansion: Providing a "pay as you grow" deployment model where additional switches may be added to the stack as more capacity is needed.
- Aggregated Connectivity: Offering a means to connect a cluster of Layer-2 switches to the Distribution switch through an aggregated uplink without the need for traditional loop avoidance protocols.

These high level objectives are still important, but traditional stack architectures also include several disadvantages for the next generation campus with wired, wireless and IOT device proliferation; for example:

- Merging the control plane of the stack members is perceived as a simplification but in reality significant complexity is added to the software running the stack itself. Since a stack is inherently fate-sharing, software issues often impact the entire stack.
- The proprietary nature of stacking software and interfaces results in not only limited interoperability between different vendors but also between different product families and models from the same vendor.
- The stacking ports used to interconnect members usually require a proprietary ring topology using dedicated interfaces, inhibiting the flexibility of deployment and apportionment of limited bandwidth.
- Stack port bandwidth varies between implementations and there are often hidden performance limitations. Understanding stack bandwidth is critical to ensure the stack ring does not become a choke point. To confuse matters, how vendors report or calculate bandwidth varies. For example, a Full Duplex bidirectional Gigabit Ethernet link is always referred to as 1G, not 1G in each direction x 2, or 2G! With some stack port bandwidth calculations, bandwidth is counted separately for each direction, and administrators may be misled into believing they have twice as much bandwidth as they actually have.
- Maintainability considerations with a traditional stacked architecture must be considered such as removing a stack member or a faulty stack member may cause the switch stack to divide or partition into two or more stacks, each partition is a mirror of the other.
- Members within a stack are distance and topology limited due to the proprietary stacking cables or the timing requirements of the centralised control plane.

Introducing Arista's Cognitive Management Approach

In a large enterprise, automated network day 1 and day 2 management are two important considerations. Manually managing individual devices - whether in the Campus or Datacenter - is error prone, time consuming and operationally does not scale. For many administrators the primary driver for using stackable switches is ease of management. However, while reducing the number of touch points is beneficial, it does not address the need to provide an end-to-end operational paradigm for day 1 and day 2 management.

Arista's CloudVision offers a unique approach to building the state-driven cognitive management plane, providing many benefits beyond single point management. CloudVision provides end-to-end single point management from the Datacenter to the Campus.

Day 1 Auto Commissioning & Configuration

Beginning from the core premise that all devices should be zero-touch provisioned into the network, CloudVision manages configuration deployment automatically using a hierarchical scheme to ensure consistent deployment throughout the infrastructure. Devices can be arbitrarily grouped into containers by function, location or other common attributes regardless of their vintage or product family without dependency on proprietary protocols.

This approach means it is possible to move from operations that apply only to stacks of switches, to floors, buildings or entire campuses as best fits your organisation's needs.

Day 2 Change Management, Compliance and Visibility

Post deployment, the dynamic nature of modern environments guarantees a need for a robust change control process. Using the same organisationally focused paradigm described above CloudVision's workflows enable the planning, automation and roll back of configuration and image changes without CLI intervention. This reduces deployment times from days to minutes and avoids failures introduced by human error or unexpected interdependent control plane complexities.

A deployed infrastructure also needs to remain in compliance to ensure security and relevant regulations are upheld. CloudVision continually monitors for unexpected or unauthorized changes to the expected steady state environment. Furthermore, CloudVision provides alerts for potential defects or newly discovered security advisories (e.g. PSIRTs).

Infrastructure-wide visibility is also critical to maintaining a high level of service. CloudVision acts as the engine to collect, collate and visualize thousands of parameters in real time from Arista's streaming telemetry integrated into every device. Continuous monitoring of both connected devices and the health of the underlying infrastructure ensures administrators have relevant information at their fingertips to investigate anomalies anywhere in the domain.

2020 Campus Expansion

In the 1990's some administrators preferred stackable architectures because they provided an easy way to add capacity. In practice adding capacity with traditional stacking architectures imposes several limitations. The added switch needs to not only be from the same vendor as the other members of the stack but the same model, software version, and license pack as well. Stacks preclude the Administrator from taking advantage of newer switching platforms which may introduce new capabilities that are important to the organisation. Furthermore, introducing new units to an existing stack often requires significant manual preparation of both the new device, the stack itself and the need to carefully follow a deployment plan to avoid issues of incompatibility and control plane synchronisation. Due to the complexity involved, expansion often requires scheduling maintenance for all switches and in some cases the entire stack may need to be reset. A stack investment is not well preserved when the only expansion option is deploying an obsolete switch, and capacity is not easily added if manual preparation and downtime needs to be scheduled.

Alternatively in the 2020 era, the Arista Extensible Operating System, EOS, provides a self-healing active-active MLAG architecture. The architecture addresses the concerns of spanning tree loops, high availability and traffic load balancing without the limitations of proprietary stack architectures. Switches can be added to an Arista CMP, Cognitive Management Plane, without disruption to the other switches and independent of the switching platform. The Arista architecture is based on "always on" principles from cloud designs where hitless expansion is critical - with no downtime. By avoiding the complexity of fate-sharing combined control planes, Arista's solution allows independent deployment of any number of devices of any product family within the wiring closet. It is possible to deploy mixed copper, fiber, POE, mGig, low cost and advanced products as suits the specific use case together without compatibility issues.

2020 Campus Connectivity

Stackables Today

As mentioned, switches in a stacked wiring closet architecture are commonly connected with proprietary stack ports in a ring topology. The stack ports provide an inexpensive means to connect 2 or more switches, however the stack bus and its protocols are vendor proprietary and the switches often need to be physically close together to meet the timing requirements of the combined control plane. Below is an example of a common stack architecture that provides 240G of bidirectional connectivity.

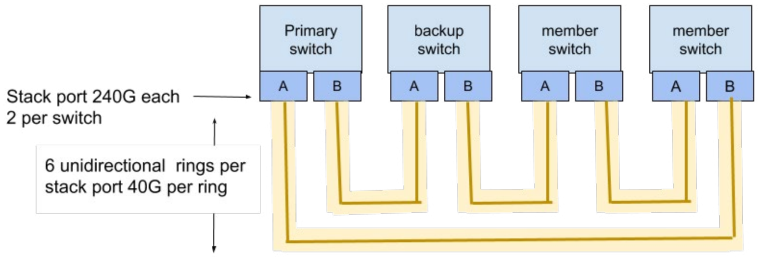

Figure 1: Common Stack Connectivity

Arista's Solution

Arista's EOS provides rich Layer-2 and 3 capabilities. The Arista wiring closet can connect to the Spline/Distribution layer or directly to the Core using either Layer-2 or Layer-3 uplinks. Layer-2 is more common in legacy networks and small environments and doesn't require a VXLAN overlay network to provide Layer-2 connectivity to other wiring closets. Many organisations are migrating away from Layer-2 uplinks to Layer-3. Factors driving this change include, improved policy deployment and automation, smaller fault domains, better segmentation for security and easier troubleshooting as paths are ore deterministic. Both Layer-2 and Layer-3 access are supported and discussed below in Connecting the Spine.

MLAG

Layer-2 networks based on the Spanning Tree Protocol (STP) can be difficult to maintain. Negative experiences with STP are common and include misconfigurations and large scale outages. Arista's solution to this problem was initially developed for the Datacenter, where loops and broadcast storms were extremely disruptive.

Multichassis LAG (MLAG) eliminates the need for STP and provides flow based load balancing and multi-homing while maintaining control plane independence. MLAG is extremely well suited for the wiring closet as it allows administrators to build large, horizontally scalable networks with efficient traffic forwarding.

Connecting Switches Within the Wiring Closet

Arista recommends the use of standard high speed Ethernet interfaces that can be deployed either as uplinks or interconnections as appropriate for the use case. For example, the Arista wiring closet architecture when constructed with ten 48 port switches can provide two 100G uplinks and connectivity to over 500 clients - more than sufficient for most IDF closet PoE requirements.

Below shows some examples of how switches can be connected to form an Arista wiring closet with dual load balanced uplinks traversing an IDF riser.

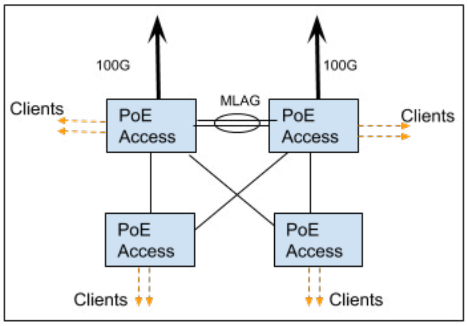

Figure 2: Arista Wiring Closet - 4 Switches for 184 access ports

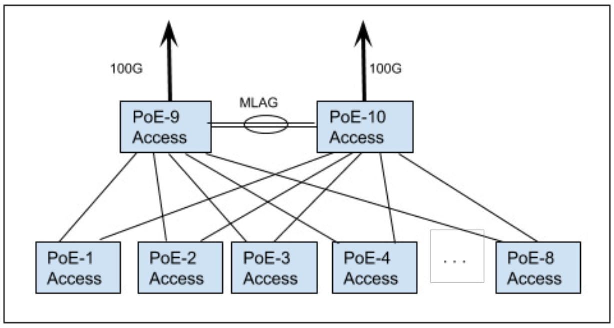

Figure 3: Arista Wiring Closet with Arista 720XP-48ZC2 Switches - 10 Switches with 100G Breakout Cable for 500+ Access Ports

Unlike legacy stack topologies, additional switches can be added without disruption. And since the connecting links are standard Ethernet, the switches do not have to be colocated. Additionally, the switches added may be any current or future platform which is possible through the use of standards-based connectivity to the Arista MLAG pair.

Connecting to the Spline - Layer-2 Uplinks

The Layer-2 uplinks leverage MLAG technology to provide flow level load balancing and redundancy to the Spline/Distribution layer. Gateways for the access VLANs could reside on the Spline/Distribution or Core layers. The below diagram illustrates connecting five 48-port switches. Configuration with 2 Layer-2 uplinks as well as 4 Layer-2 uplinks are illustrated.

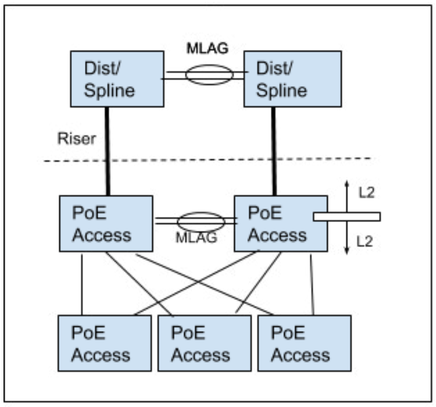

Figure 4: Arista Wiring Closet - L2 Connectivity to Distribution/Spline 2 Uplink

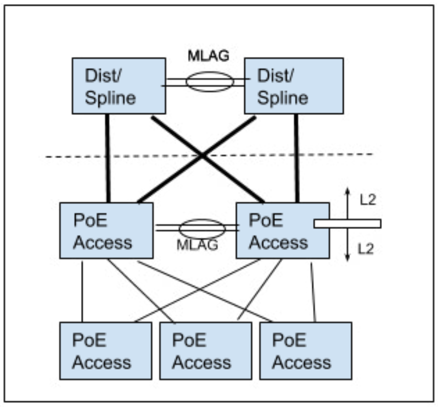

Figure 5: Arista Wiring Closet - L2 Connectivity to Distribution/Spline 4 Uplinks (Bow Tie)

Connecting to the Spline Layer-3 Uplinks

The Layer-3 uplinks leverage Equal Cost Multi-Path (ECMP) for load balancing and high redundancy. Connectivity to a dual set of Spline/Distribution switches is the most commonly deployed topology. However some organizations, particularly organizations requiring high levels of availability or bandwidth, are deploying n-way designs. Below illustrates a typical 2-way campus Spline architecture as well as the 4-way alternative.

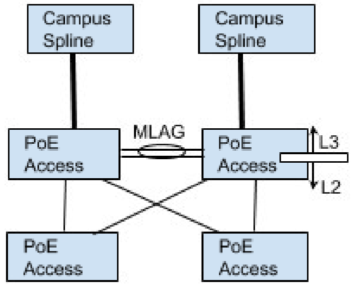

Figure 6: 2-way Campus Spline Architecture - 4 Access Switches

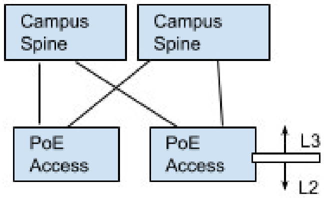

Figure 7: 2-way Campus Spline Architecture - 2 Access Switches

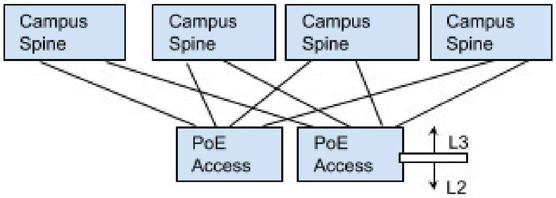

Figure 8: 4-way Campus Spline Architecture - 2 Access Switches

VXLANs for Segmentation and Extending L2 Adjacency over L3 Network

Layer-3 uplinks from the wiring closet may break some applications that require clients to have L2 adjacency across multiple wiring closets. Furthermore, as the number of connected endpoints continues to grow, there is an increased demand for segmentation as well as abstraction of user networks from infrastructure.

VXLAN is a network overlay technology that provides dynamic provisioning of network segments in the form of Software Defined tunnels overlaid on a normal Layer-2 / 3 topology. VXLAN provide Layer-2 adjacency of clients that are separated by Layer-3 infrastructure.

Separating user traffic from the protocols that define the underlying topology removes the need to modify the network to accommodate user adds, moves and changes. This simplifies the design and provides for the additional security by not exposing the network topology, network protocols or devices to malicious users. In addition, VXLAN overlays provide Layer-2 adjacency to remote locations that are connected over standard Layer-3 routing.

All Arista platforms, whether they be for WAN, Datacenter or Campus, support VXLAN technology. As a true multi-vendor open standard, it ensures organisations can have a single, continuous end-to-end approach to segmentation and Layer-2 extension.

Maintainability and Troubleshooting Considerations

Maintainability and troubleshooting are important considerations in a campus architecture. Traditionally, removing a powered on stack member causes the switch stack to divide or partition into two or more stacks, each partition is a mirror of the other. This is because each member of the stack has an independent copy of the entire stack, including themselves. This condition is referred to as "split brain" and since each of the divided halves contain the same configuration, IP address conflicts will occur and must be manually resolved. Of course, the stack can become divided for many reasons including a malfunctioning switch or faulty cable. A faulty ring inside a stack port can flap from connected to disconnect. When this occurs the available bandwidth needs to be recalculated and state tables updated. Continual flapping and recalculations can have severe effects and can also be very difficult to troubleshoot.

Arista Cognitive Campus Portfolio

Arista's portfolio consists of a broad variety of platforms designed to meet different use cases. As all Arista platforms run a common, single image operating system (EOS) and consistent features, there is no restriction as to which platforms can be deployed in campus environments.

For typical campus deployments, the 7XX series products provide a broad set of common capabilities, including support for emerging Multi-Gig (mGig) and 803.bt Type 3 Power Over Ethernet (POE) together with Wifi Edge and Splines.

Table 1: Ports on 720XP Series

Models	Downlink copper ports	Uplink & High Speed ports
720XP-48ZC2	40 x 100M/1G/2.5G + 8 x 1/2.5/5G	4 x 1/10/25G SFP + 2 x 40/100G QSFP
720XP-24ZY4	16 x 100M/1/2.5G + 8 x 100M/1/2.5/5G	4 x 1/10/25G SFP
720XP-48Y6	40 x 10M/100M/1G + 8 x 100M/1/2.5G	6 x 1/10/25G SFP
720XP-24Y6	16 x 10M/100M/1G + 8 x 100M/1/2.5G	6 x 1/10/25G SFP

For additional uplink capacity the 100G uplinks can be divided into 4 25G uplinks with a breakout cable. Note that any port on an Arista switch can be used for either host connectivity or uplink to another switch, operating in layer-2 or layer-3 mode.

The Arista 7050 and 7300 family of switches are ideal for power users, traditional 3-tier or collapsed 2-tier topologies and those requiring higher densities of optical connectivity from 1G to 100G. All products support rich functionality for Layer-2, Layer-3 and Overlay networking using common, open standard features that are consistent with the entire Arista Networks portfolio.

Arista is rooted in leveraging standards based solutions whenever possible. While proprietary stack architectures are familiar and low cost, they have many limitations and are operationally more expensive. They are proprietary, difficult to troubleshoot, can become a choke-point and can become divided; the state of the art Arista PoE switches addresses these limitations. With Arista EOS and industry standard protocols, the Arista campus architecture has been proven in the Datacenter by the largest Cloud Providers and Enterprise Architects. Starting in the mid 1990s stack architectures have provided several benefits when compared to traditional methods of connecting fixed configuration switches. However, when examined closer they do not meet the needs of emerging campus networks. Arista chose a different approach that address the shortcomings of traditional stack architectures associated with management, expansion, connectivity and maintainability. Modern campus management is increasingly critical for consistent day-1 provisioning and day-2 operational management across the entire enterprise. Arista's CloudVision provides single point administrative control and oversight, the ability to consistently apply configuration at any level of defined hierarchy, proactive compliance management of software defects and security threats, while offering unprecedented client visibility with telemetry tracking to guard against malware.

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