A layman's guide to Layer 1 Switching

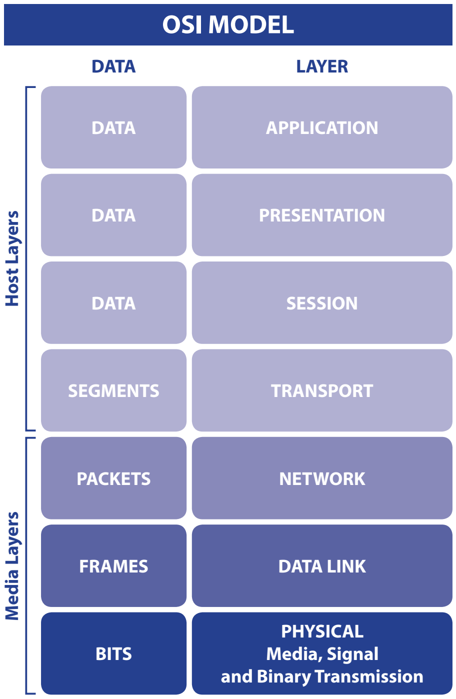
The world of network engineering and platforms is complex and full of acronyms and new vocabulary. This guide serves as an introduction to Layer 1 switching, explaining in layman's terms the OSI model, the functionality of crosspoint switches, the concept of latency, clock and data recovery as well as programmable switches. Great if you are getting started, useful if you are just looking for a quick refresher.

The OSI Model

The Open Systems Interconnection (OSI) model is a conceptual model that characterizes and standardizes the internal functions of a communication system by partitioning it into abstract layers of functionality. It has been ratified in 1984 and has ever since been a key reference as to how network protocols and devices should communicate and interoperate with each other.

The lowest layer of the internal functions of a communication system is known as layer 1, the physical layer.

The physical layer consists of the basic networking hardware technologies which transmit data, moving it across the network interface. All of the other layers of a network perform useful functions to create and / or interpret messages sent, but they must all be transmitted down through a layer 1 device, where they are physically sent out over the network.



The main functions of the physical layer are:

- Encoding and Signalling: The physical layer transforms the data from bits that reside within a computer or other device into signals that can be sent over the network as voltage, light pulses or radio waves that represent ones and zeroes.
- Data Transmission and Reception: After encoding the data appropriately, the physical layer actually transmits the data, and of course, receives it. Note that this applies equally to wired and wireless networks, even if there is no tangible cable in a wireless network.

To summarise, physical layer technologies are the ones that are at the very lowest level and deal with communicating actual ones and zeroes (i.e. sending and receiving) over the network.

These devices have absolutely no knowledge of the contents of a message. They just take input bits and send them as output. Typically, devices like switches and routers operate at higher layers and look at and act upon the contents of the data packets that they receive.

Still a little confused? An analogy may help!

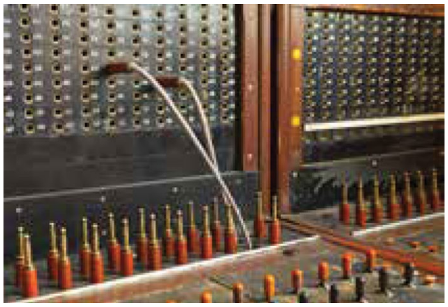
Computer networks are organised a bit like a traditional business with executives at the top of the stack, managers in the middle and workers at the bottom (the factory floor). Work gets done at the bottom but often it needs to percolate up the layers of management for a decision to be made. This process takes time and is not very efficient but is necessary if the decisions can't be made by the workers.

If we think of the network stack then the bottom layer (layer 1 or the physical layer) of the network does the work of sending and receiving bits across the physical network wires without doing a lot of thinking.

In a conventional network switch, each data message is received at layer 1, then passed up the management chain of the network stack (to layers 2 or 3) where a decision is made based on the content of the package. Then, the results of the decision percolate back down the chain to the physical layer to transmit. This all takes time.

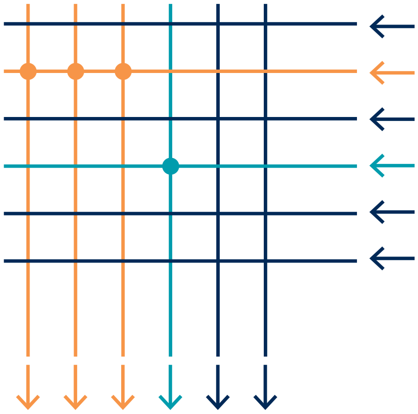
The idea of a physical layer switch is that the switching decisions are made in advance so that the work never gets held up. This is called circuit switching and as a result it is approximately one hundred times faster than the conventional switch because the decisions are never required to go above the physical layer. This is the layer where we talk about high speed or ultra-low latency. The physical layer switch isn't as smart or as flexible as a layer 2 or 3 switch but if latency is important then a physical layer switch should give you a significant speed edge.

What is a Crosspoint Switch?



Now we have gotten our heads around what a layer 1 device does but we still need to define what a crosspoint switch is.

Crosspoint switches originated in telecommunications around 1915 as a way to create a circuit between two phone lines. Originally these were electromechanical devices that used a metal bar to create an interconnect and therefore were known as crossbar switches. Eventually these were replaced by semiconductor technology but the concept remained the same of having a matrix formed from the set of inputs and outputs. (hence also the common name of matrix switches).



In short: crosspoint switches are designed to connect multiple inputs to multiple outputs in a matrix manner.

Modern crosspoint switches are programmable Integrated Circuits (ICs) with the capability to tweak where the signal being communicated goes to. As programmable switches they have become widely used in networks and telecommunications.

What is Latency and why it matters

You'd think that all there is to trading, machine learning or virtual reality these days is being fast. Well back in the day, there wasn't a trader around who knew what the word latency meant. Seriously.

Latency in a network is the measure of delay from the start of packet transmission to the start of packet reception. The latency of a switch is the measure of how long it takes a packet on the wire or fibre to enter and transit the switch onto the output wire.

If network latency is a concern, then the lowest latency switching solution is a physical layer switch with a latency now measured in single digit nanoseconds compared to hundreds of nanoseconds for many layer 2 switches. This is possible since the latency of the crosspoint switch at the heart of these switches are now specified in picoseconds (that is a trillionth of a second or 0.000 000 000 001 seconds).

How are single-digit latency figures possible?

Crosspoint switches do not allow for the connection of multiple inputs to a single output, that is, they do not allow packets from multiple links to be aggregated into a single link. (sometimes called muxing or multiplexing). This means that there is hence no need to buffer any packets in a queue that might arrive at the same time. As we are talking a simple 'one in/one out' approach, packets pass through at close to the speed of light (or 1metre in 3.3 ns)

Another benefit of crosspoints is that they can connect a single input to more than one output at a time without introducing any overhead. This means that a physical layer crosspoint switch can broadcast data to many ports at the same time.

Circuit switching is not for everyone but it turns out that there are a range of applications where low latency is a winner:

- Top of rack patch panel in the data centre or test lab
- Monitoring of links and bypassing of faults
- Data broadcast (i.e. market data broadcast in finance)
- Media conversion (e.g convert from optical fibre to copper).

Clock and Data Recovery (CDR)

Physical layer switches act like a piece of wire but in fact they do a little more than that. They typically have circuitry that regenerates the signal that is passing through. The reason to regenerate is to remove noise from the incoming signal and start again with a fresh clean signal on the output. That is, the individual output bits are replicas of the original input bits where the bits were first generated. This process is called clock and data recovery, CDR or retiming and is a standard part of all network switches.

There is such a big market for CDR circuitry that many manufacturers produce ICs that perform the function with many other bells and whistles.

Arista also includes clock and data recovery - that is, it regenerates a new clean signal. The benefit to users is that it reduces errors in the data and allows the data to be broadcast over longer distances at lower latency.

What are programmable switches?

We noted above that OSI layer 1 switching is all about speed as there is very little intelligence needed from the side of the device – it literally just transmits data as fast as possible.

Programmable switches on the other hand contain FPGAs or Field Programmable Gate Arrays to introduce logic blocks to the equation. FPGAs are basically programmable hardware ie. the functionality of the device can be changed after manufacturing which adds a large amount of flexibility and cost savings. FPGAs hence have the flexibility of software and the performance of hardware. They allow for more 'intelligence' by giving companies the ability to run high-performance, low-latency applications directly on the layer 1 switch. Such applications can be trading algorithms, pre-trade risk checks, analytics functions or securit-related applications, to give only a few examples.

Layer 1 switching technology

Patch panels, taps, signal regenerators, media converters and switches – all these devices introduce latency overheads, possible points of failure and tie up valuable rack space (not to mention cost).

Given that we have been working relentlessly to improve performance and reduce latency in the networking world, it would be remiss not to give a little more information about Arista whilst we have your attention. Just bear with us for another 60 seconds. It will be worth your while ...

Arista Switching Platform

Arista allows users to replace a significant number of these boxes (Patch panels, taps, signal regenerators, media converters, switches) in their server rack with a single intelligent device which has negligible latency.

Arista is unique in combining ultra-low-latency performance of less than four nanoseconds with high levels of monitoring functionality normally only found on much slower devices. It also provides management statistics and network monitoring with no impact on network latency, improves signal quality and saves rack space.

Arista has been designed from the ground up using the latest hardware and software components and incorporating a unique combination of features.

Here are some of the real-world benefits:

- Broadcast data in less than 4 nanoseconds with virtually undetectable jitter.
- Combine layer 1+ switching and FPGA technology to run high performance applications.
- Combine two layers of 48 port devices, broadcasts to over 2,300 in less than 16 nanoseconds.
- Increase network visibility by providing packet statistics and telemetry on every link (via Influx DB).
- Reduce visits to the data centre by providing remote patching capabilities which can be automated via JSON RPC API.
- Rationalise infrastructure costs by performing inline media conversion (fibre to copper) and the ability to leverage passive DAC.
- Simplify DWDM architecture by leveraging tunable DWDM optics.
- Leverage precise, nanosecond resolution timestamping of packets.
- Improve network manageability via connection monitoring and packet statistics.
- Open standards x86 management platform for future extensibility and customization.
- Improve signal integrity on problematic long haul links (Signal regeneration).
- Flexible SFP/SFP+ support allows the use of less expensive modules.

Where network engineers would have been hesitant to introduce an additional network hop for manageability, compliance, monitoring or signal regeneration, they can now do so with virtually no impact on latency.

To characterise it in terms of the OSI model: **It's more like a layer 1+ device than a layer 1!**

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