ProCurve Networking

ProCurve Switch 5400 and 3500 Power Over Ethernet

Introduction

This paper explores Power over Ethernet (PoE) as a standard method of supplying power to networked devices over Cat 5 Ethernet cabling, much the same way traditional wired telephones received their operating power from the telephone line. The IEEE 802.3af-2003 standard specifies how power should be distributed over standard Ethernet cabling.

We will examine how the ProCurve 5400 and 3500 switches operate in order to allow customers to more fully understand the behaviors they see from the switch, and allow them to fully utilize available power without over-subscribing the power supplies.

A short glossary is given at the end of this paper for those few PoE terms used in this paper

PoE States and their Meanings

- **Disabled**: The switch has disabled the port, either due to configuration or due to lack of available power. (Note: This only affects the ports ability to deliver PoE power and does not affect its ability to detect link and switch packets.)
- **Searching**: The switch is waiting for a valid PD to be attached to the port. This is the normal state for the port to be in if there is not a valid PD attached to it.
- **Delivering**: The switch has detected a valid PD on a port and is delivering PoE power.
- **Fault**: The switch has detected an overload or short circuit condition on the port. Note: All faults of this nature are transitory. After the fault is detected, the port returns to the searching state and the switch can re-power the port when it detects a valid PD.
- **Other Fault**: The switch has detected a hardware fault that will prevent the port from detecting and delivering PoE power. If a switch or module has one or more ports that remain in this state following a reset or power cycle, please contact ProCurve support.

PoE LEDs

The PoE status LED is a cumulative LED for all the ports and slots on the switch. It allows the user to see the PoE summary of the switch at a glance without putting all the LEDs into PoE mode. The LED displays the highest level of failure or alert of all the ports and slots. For example, if one of the ports is denied power and the port LED is flashing an alert in PoE mode, then the PoE status LED will also be flashing an alert. If one of the ports is faulted because of an internal hardware problem, then the PoE status LED will also be flashing a slow-flash amber.

PoE status LED description:

PoE status = green: There is at least one PoE-capable port in the system, and everything is OK.

PoE status = slow flash amber: One or more ports have an internal hardware fault.

PoE status = fast flash amber: One or more ports are denied power or detecting an external PD fault.

PoE status = off: There are no PoE-capable ports in the system or the PoE sub-system is not fully initialized yet.

To determine which port or ports is the source of the fault or warning, press and release the Mode button until the LED mode indication reads PoE.

Note: If there are multiple PoE statuses that need to be displayed, they are shown in the following order: slow-flash amber, fast-flash amber, on, off.

When the Port LEDs are in PoE mode, the Port Link LED is no longer an indication of port link. It is used in conjunction with the mode LED to fully indicate the port’s PoE state. The table below shows all the possible states that the Port LEDs can be in.

Link LED	Mode LED	Port State
On	Off	Searching
On	On	Delivering

Off	Off	Disabled by configuration
Fast-Blink Amber	Off	Disabled due to lack of power or PD Fault
Slow-Blink Amber	Off	Other Fault

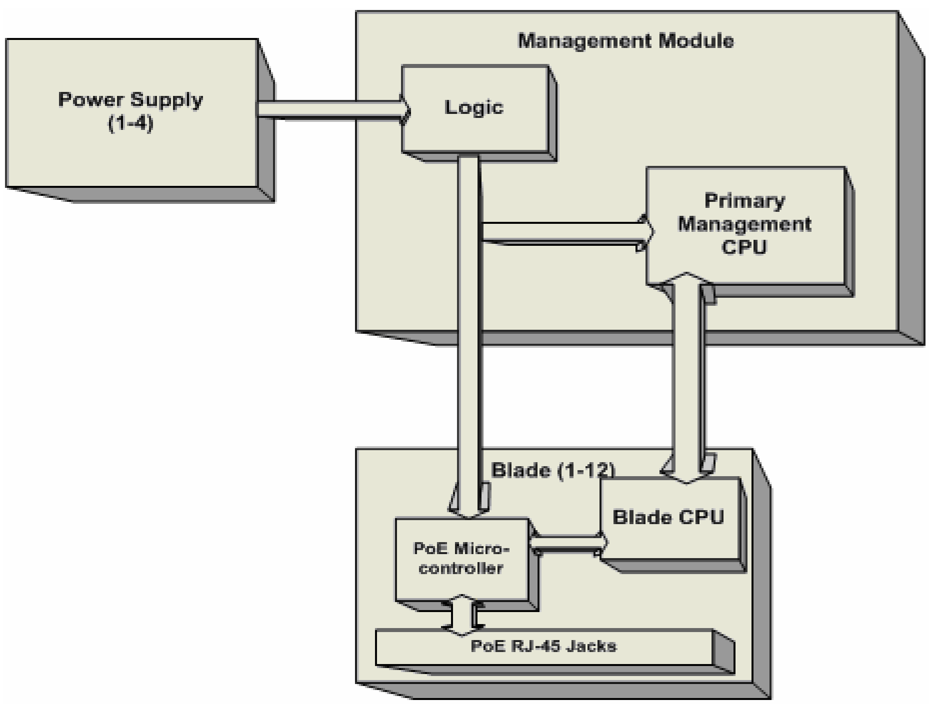
Building Blocks

PoE RJ45 Jacks – These are the connectors on the front of the blade or switch, and contain ASICs to detect PDs and inject power onto the Ethernet cable.

PoE Micro-controller – There is one micro-controller per blade or port group (on the ProCurve 3500 series, ports 1-24 and 25-48 constitute two separate port groups). They are responsible for maintaining the power budget and allocating power according to port priority for each blade or port group. Throughout this document, blade, slot, and port group will be used interchangeably.

Power Supplies (Internal and External) – These are the source of the power being allocated and transmitted by the switch.

Switch PoE system SW – This is divided into two layers. The first resides on the primary management CPU and the second resides on the blade or port group CPU. The primary management CPU is responsible for managing all the power supplies and determining how much power is available for the system as a whole to use, allocating power between all the blades according to need and priority. The blade CPU is responsible for monitoring the power usage requirements and reporting physical state of all the ports on the blade, so that the primary management CPU can dynamically distribute the power across the entire system appropriately.



Dynamic Power Balancing

This section explains how power is balanced between the blades within the system, allowing the maximum number of ports to be powered without exceeding the capacity of the power supplies.

Dynamic power balancing is the process of distributing the available system power between all the blades and ports. All ports have a designated power priority within the system based on their 802.3af priority configuration and port number. As specified by 802.3af, there are three priority levels that a port can have (from highest to lowest): Critical, High and Low. Within each priority level, ports with a lower port number are given precedence over ports with a higher port number.

For example, if ports A3 and F21 are both configured as high priority and there is only enough power left in the system to power one of the ports, port A3 will receive the power. All slots will receive an initial power allocation of 22 W. This is the minimum power necessary to allow a blade to begin detecting PDs, plus a 5 W reserve to allow for power load fluctuations. The remaining power is divided up between all the blades based on priority.

The power priority sequence is as follows:

1. Allocate power for all critical priority ports, starting with Slot A and proceeding to Slots B, C, D, etc. until all critical priority ports have received their required power allocation.
2. After all critical ports have received power, power is allocated to high priority ports, starting with Slot A and moving on to the last slot.
3. After all high priority ports have received power, power is allocated to low priority ports, starting with Slot A and moving on to the last slot.
4. Any remaining power is equally divided between all powered blades.
5. If at any time during the allocation process, a blade has less than 17 W of PoE power available, then all remaining ports on that blade will remain un-powered.

After the switch reaches a steady state, if a new PD is connected, the port will receive power from the power allocation pool. If there is not enough power available, power will be taken away from lower priority ports to allow the new PD on a higher priority port to come up. A typical sequence of events would look like this:

1. The switch has one Low Power Supply (273 W total available power) and is sourcing ~260 W across 4 blades with all ports configured at low priority (default). Since there is less than 17W remaining, there is not enough power left to power up another port.
2. A PD is plugged into port A5. Since there is insufficient power left to allocate, the new PD will be denied power and the port LED will blink amber. A message will be logged in the event log.
3. The switch will rebalance power, taking some power away from Slot D since it is a lower priority slot. The highest number powered port on Slot D will be denied power. The port LED will blink amber and a message will be logged in the event log. Port A5 will be powered up and the port LED will return to green. The mode indication will light up to indicate that the port is sourcing power.

Note: It is possible that more than one port on Slot D would have to be denied power to return enough power to the allocation pool to total more than 17 W and to allow port A5 to be powered up. The number of ports denied power depends on how much power lower priority ports are drawing. Once A5 is powered up, it is possible that one or more of the ports that were powered down would be powered up again depending on the actual power draw of the new PD plugged into port A5. This is normal behavior during a rebalance. To avoid such a situation, it is recommended that sufficient additional power be provided in the system to ensure that adding a port will not cause lower priority ports to be denied power.

Power Usage and Allocation

This section explains how the blades use their allocated power, and how adding powered devices to one blade can affect others.

Once each blade receives its initial power allocation, it can begin to detect attached PDs. PDs are detected sequentially starting from port 1 to the last port on each blade. There must be a minimum of 17W available before applying power to a new device to guarantee there will be enough power available to power the new PD. Once a PD is detected and begins drawing power, the blade will continuously monitor actual usage to the nearest watt, and return any unused power to the slots allocation pool so that it can be used by other devices.

For example, if a slot has an available power allocation of 23 watts and has 4 Class 0 IP phones connected, which draw 3 W each on hook and 5 W ringing at maximum volume, the switch will begin the detection process because there is at least 17 W of power available. Since the actual usage of the first phone only draws 3 W of power, there will be 20 W remaining.

After detecting the second phone, the switch is providing 6 W of power and has 17 W remaining, which is sufficient to detect the third phone. However, the switch cannot detect the fourth phone because there is only 14 W available. The fourth phone will be denied power. A log event will be generated to notify the user that a device was denied power.

In contrast, if power allocation was based on the PD's Classification (as in some third-party switches), then there would only have been sufficient power for a single phone because the maximum 15.4 W would need to be reserved at the PSE jack for a Class 0 device. See the table below for the maximum power draw by PoE Classification. Had CDP (Cisco Discovery Protocol) been used to negotiate the maximum power draw for devices, a similar but smaller effect would be seen, since the CDP negotiation would allocate 5 W per phone. In that case, only two phones could be powered (keeping in mind that 17 W are required to begin the detection process).

Class	Usage	Minimum Power Levels at Output of PSE	Range of Maximum Power Used by the PD
0	Default	15.4 watts	0.44 to 12.95 watts
1	Optional	4.0 watts	0.44 to 3.84 watts
2	Optional	7.0 watts	3.84 to 6.49 watts
3	Optional	15.4 watts	6.49 to 12.95 watts
4	Reserved for future use	Treat as Class 0	Reserved for future use

Note: The difference in power levels provided by the PSE (switch) and available at the PD is to account for line loss. A certain amount of power is lost in the cable between the switch and the powered device (typically less than 16% loss), which can be influenced by cable length, quality, and other factors.

If there is any power available in the global allocation pool, the switch will rebalance the allocations across slots to allow the maximum number of ports to receive power. While this approach provides an optimal power allocation scheme, there is some downside. If the switch is operating at its maximum available power allocation when all the PDs are in their quiescent state, then when multiple PDs become active at the same time (i.e., phones are picked up to make calls, phones start ringing, video cameras start panning, etc.), the dynamic power fluctuation could cause the total power usage to increase above the guard band reserve and the total power available.

This would cause one or more of the lowest priority PDs to be temporarily denied power and shut down. After some of this activity had subsided, these low priority PDs would then come back online. For this reason it is recommended to budget the power supplies in the chassis appropriately.

To be completely safe, it is best to perform an analysis of the number and types of loads that will be attached, along with the nominal and maximum power draw, so that the switch can be provisioned with enough power to meet the maximum anticipated demand.

Power Threshold

The power threshold is a marker that can be used to manage the power utilization of the switch and generate a notification when additional power may be required. It can also be used to manage the PoE power redundancy as explained in the Redundancy section below.

When the power on a blade crosses over or under the threshold value, a message gets logged to the event log and an SNMP trap is sent. For example: A blade is allocated 100 W of power and is currently sourcing 45 W to various loads. If the power threshold is set to 50% and a new PoE load is connected that draws 8 W, then an event is logged, since the new power draw of 53 W exceeds the threshold.

Depending on the PoE power usage throughout the rest of the system, the power may be rebalanced and the allocation to that blade could increase. If additional power is allocated and 53 W is no longer above the power threshold, then another event will be logged stating that the blade has gone below its threshold.

By default, the power threshold is set to 80% and will trigger an SNMP trap and a log message whenever power usage exceeds or drops below that threshold. You can use this to be notified whenever the PoE loads attached to the system are drawing close to the maximum power available, and then adjust either the switch where the loads are connected, or increase the number of power supplies in the system to meet the new power demand.

A recommended default power threshold value is 80% because this allows sufficient dynamic headroom for many mixed types of PoE deployments. However, if you know that most of your devices have a small power usage fluctuation, you can increase the power threshold to 90% or more. If your devices have a large power usage fluctuation (i.e., lots of color display IP phone that light up when off the hook) then you can lower the threshold as needed.

Power Supply Usage

ProCurve Rapid Power Down

The system is designed to be able to use all of the available PoE power from all the installed power supplies and apply it to connected loads. In the event of a power supply failure, the system immediately notifies all components, including the PoE RJ45 jacks. This allows the switch to rapidly turn off the power to lower priority loads without overloading and damaging the remaining power supplies.

How is failover power calculated?

Failover power is the amount of power that is pre-allocated by the system in the event of any single power supply failure. This failover power is calculated based on the number and type of power supplies installed in the switch. The algorithm adds up the total power installed in the switch and then subtracts the power of the largest supply. In a switch with two High Power Supplies (900 W PoE power), this means there is a total of 1800 watts of PoE power with 900 watts of failover PoE power.

What happens when you mix supplies?

When there is a power supply failure in a system with non-uniform power supplies (i.e., one Low Power Supply and one High Power Supply), you may see some of the ports that were originally powered off due to the supply failure come back on. Here is an example:

A ProCurve 5406 switch has one High Power Supply (900 W PoE) and one Low Power Supply (273 W PoE) installed. This configuration provides 1173 W of total PoE power, but only 273 W of failover PoE power. If either power supply fails, the switch will drop down to a power level that can be supported with 273 W.

If it was the 273 W supply that failed, then within one second the switch will increase power back up to 900 W of PoE power. This would result in some of the low priority devices cycling power and then coming back on line if the total power usage exceeded 273 W. All ports that can be supported by the new power level will be back online within approximately five seconds.

Redundancy Usage (PoE Usage and Power Mains)

Note: The discussion below about power redundancy will assume that all power supplies in a switch are uniform. Utilizing non-uniform power supplies greatly lessens the amount of redundant power and is not recommended in a redundant configuration.

Using the power threshold marker, it is easy to maintain a fully redundant PoE power configuration. To maintain full redundancy, you should install an even number of power supplies within the ProCurve 5400 family switch or ProCurve 3500 family switch with optional ProCurve 620 EPS/RPS, and then configure the power threshold to 50% (or less) and maintain the system power utilization below this level.

The power cords from half of the supplies should go to one set of power mains, and the other half to another set, to insure continued operation in the event of one of the power mains failure. This will maintain power to the switch in the event of a power main failure. This is also advantageous if the switch is also on an UPS backup for one of the mains. In order to maintain non-PoE power redundancy, you should also insure that a sufficient number of internal power supplies go to each set of power mains (i.e., for the ProCurve 5406 with two total supplies, one to each and for the ProCurve 5412 with four total supplies, two to each).

Alternatively, you can create an N+1 redundant PoE power configuration, if the maximum PoE power usage is less than or equal to the power provided by two supplies. Just add a third power supply and set the power threshold to 66% to achieve N+1 redundancy. The power cords from an N+1 redundant configuration should also be distributed across multiple power mains, to improve system availability. This would only be applicable to the ProCurve 5412 chassis, since in the ProCurve 5406 or 3500 switches, the maximum number of supplies is two.

Note: Although the ProCurve 5412 chassis requires two power supplies to power up all 12 slots, should one supply fail leaving one active power supply in the chassis, the top six slots will remain powered and fully functional. This includes full PoE functionality on the top six slots.

For more information on how to determine which supplies and how many supplies are necessary for your installation, please refer to the ProCurve Switch 5400zl/3500yl Series Ordering Guide.

Remotely Managing PoE

All PoE settings can be remotely managed via SNMP, and most can be set through the Web UI for the switch. Using remote management, you can easily power cycle any or all PoE device in the network by disabling and then re-enabling power to the specified port(s).

You can also create scripts to monitor power usage on all the blades or any given port within a switch by doing periodic SNMP queries of the MIB objects. Some of the values available include Port Power Usage, Port Current Usage, Port Power Classification and Port Power Priority. See Appendix B for a complete list of available MIB objects.

Appendix A

PoE Log Messages and their Meanings

I 03/17/06 09:30:57 ports: port B1 PD detected.

Indicates that an 802.3af compliant load has been connected to the switch – no customer action is needed.

I 03/17/06 09:30:57 ports: port B1 applying power to PD.

Indicates that we are supplying power to a connected load – no customer action is needed.

I 03/17/06 09:30:57 ports: port B1 PD over current indication.

Indicates that an attached load has attempted to draw too much current from the switch – the customer should check the powered device attached to the indicated port for malfunction.

I 03/17/06 09:30:57 ports: port B1 PD removed.

Indicates that a connected load has been removed – no customer action is needed.

W 03/17/06 09:30:57 ports: port B1 power denied due to insufficient power allocation.

Indicates that an 802.3af compliant load was detected but could not be powered up because the switch has exceeded its power allocation – the customer should add additional power to the switch with an additional or larger power supply, or remove other powered devices that are not needed.

W 03/17/06 09:30:57 ports: port B1 PD invalid signature indication.

Indicates that a non 802.3af compliant device was connected. This is a normal indication for non-PoE devices like other switches and PCs – no customer action is needed.

W 03/17/06 09:30:57 ports: port B1 other fault indication.

Indicates that the switch has detected a problem with the port that does not allow it to detect or deliver power – the customer should try rebooting the switch or hot-swapping the module. If the error persists, the switch or module should be returned.

I 03/17/06 09:27:08 chassis: external ext power supply 1 connected, supplying 385 W of 385 W max.

Indicates that an EPS is connected and supplying power to the switch to use for PoE loads – no customer action is needed.

W 03/17/06 09:27:08 chassis: ext power supply 1: -50V fault failures: 1.

Indicates that the attached EPS has had a 50V supply fault and can no longer supply PoE power to the switch – the customer should check the EPS to insure that it is properly plugged in and receiving power. If it is still indicating a fault, then it should be returned.

W 03/17/06 09:27:08 chassis: ext power supply 1: -50V fault OK failures: 1.

Indicates that the attached EPS has had recovered from a 50V supply fault and can now supply PoE power to the switch – no customer action is needed.

W 03/17/06 09:27:08 chassis: ext power supply 1: temp fault failures: 1.

Indicates that the attached EPS has had a thermal fault. A thermal fault can either be caused by a fan failure in the EPS or a temperature sensor detecting an ambient temperature that is too high. The switch will continue to draw PoE power from it – the customer should check the ambient temperature where the supply is located and ensure that it is within the operating range for the power supply. If the ambient temperature is within the operating range, check the fans and make sure they are all spinning. If not, try power cycling the EPS and if they still do not operate, return the EPS.

W 03/17/06 09:27:08 chassis: ext power supply 1: temp fault OK failures: 1.

Indicates that the attached EPS has recovered from a thermal fault. The switch will continue to draw PoE power from it – no customer action is needed.

W 03/17/06 09:27:08 chassis: EPS not supported by switch code. Please update.

Indicates that the attached EPS is not supported by the current running FW. The switch will not draw any power from it – the customer should update the switch's F/W to a version that supports the EPS.

I 03/17/06 09:27:08 chassis: external ext power supply 1 disconnected.

Indicates that an EPS has been disconnected. All power that it was supplying will be immediately disconnected – no customer action is needed.

W 03/17/06 09:27:08 chassis: POE usage has exceeded threshold of xx%.

Indicates that the POE power usage has exceeded the user-defined threshold – If the customer is using this as a gauge to determine additional power requirements or to maintain redundancy, they should add additional power with a new power supply or remove loads until the switch is again below the threshold.

W 03/17/06 09:27:08 chassis: POE usage is below threshold of xx%.

Indicates that the POE power usage has fallen below the user-defined threshold – no customer action is needed.

Appendix B

Power Over Ethernet MIB Objects

Note: For the examples below, slot number is the numerical equivalent of the Slot Letter, i.e., A=1, B=2, etc. and port number is the ordinal physical port number, i.e. A1 = 1, A2 = 2, B1 = 25, etc.

"pethPsePortAdminEnable"	"1.3.6.1.2.1.105.1.1.1.3.<slot number>.<port number>"
"pethPsePortPowerPairsControlAbility"	"1.3.6.1.2.1.105.1.1.1.4.<slot number>.<port number>"
"pethPsePortPowerPairs"	"1.3.6.1.2.1.105.1.1.1.5.<slot number>.<port number>"
"pethPsePortDetectionStatus"	"1.3.6.1.2.1.105.1.1.1.6.<slot number>.<port number>"
"pethPsePortPowerPriority"	"1.3.6.1.2.1.105.1.1.1.7.<slot number>.<port number>"
"pethPsePortMPSAbsentCounter"	"1.3.6.1.2.1.105.1.1.1.8.<slot number>.<port number>"
"pethPsePortType"	"1.3.6.1.2.1.105.1.1.1.9.<slot number>.<port number>"
"pethPsePortPowerClassifications"	"1.3.6.1.2.1.105.1.1.1.10.<slot number>.<port number>"
"pethPsePortInvalidSignatureCounter"	"1.3.6.1.2.1.105.1.1.1.11.<slot number>.<port number>"

"pethPsePortPowerDeniedCounter"	"1.3.6.1.2.1.105.1.1.1.12.<slot number>.<port number>"
"pethPsePortOverLoadCounter"	"1.3.6.1.2.1.105.1.1.1.13.<slot number>.<port number>"
"pethPsePortShortCounter"	"1.3.6.1.2.1.105.1.1.1.14.<slot number>.<port number>"
"pethMainPsePower"	"1.3.6.1.2.1.105.1.3.1.1.2.<slot number>"
"pethMainPseOperStatus"	"1.3.6.1.2.1.105.1.3.1.1.3.<slot number>"
"pethMainPseConsumptionPower"	"1.3.6.1.2.1.105.1.3.1.1.4.<slot number>"
"pethMainPseUsageThreshold"	"1.3.6.1.2.1.105.1.3.1.1.5.<slot number>"
"pethNotificationControlEnable"	"1.3.6.1.2.1.105.1.4.1.1.2.<slot number>"
"hpicfPoePethPsePortCurrent"	"1.3.6.1.4.1.11.2.14.11.1.9.1.1.1.1.<slot number>.<port number>"
"hpicfPoePethPsePortVoltage"	"1.3.6.1.4.1.11.2.14.11.1.9.1.1.1.2.<slot number>.<port number>"
"hpicfPoePethPsePortPower"	"1.3.6.1.4.1.11.2.14.11.1.9.1.1.1.3.<slot number>.<port number>"
"hpicfPoeAllowPreStdDetect"	"1.3.6.1.4.1.11.2.14.11.1.9.1.3.0"

Glossary of Terms

IEEE 802.3af -2003: Governing standard for Power Over Ethernet (PoE).

CDPv2 - Cisco Discovery Protocol version 2 is a proprietary layer 2 network protocol that is used to share information between directly connected Cisco devices, including information about PoE power usage. The industry standard approach to this capability is provided by LLDP-MED (TIA/ANSI-1057), an extension of the LLDP (Link Layer Discovery Protocol) standard. ProCurve switches allocate PoE power by measuring actual power usage on every port, to the nearest watt, and reserve an additional per-slot guard band to account for dynamic fluctuation.

PoE – Power Over Ethernet: A means of supplying power to networked devices using standard Ethernet cabling.

PD – Powered Device: Any networked device that receives power from its Ethernet cable

PSE – Power Sourcing Equipment: A switch or other network device that can supply power to PDs.

Mid-span: A device that connects between a network device (switch or hub) and the PD and injects power onto the Ethernet cable.

ProCurve Switch zl 1500 W Power Supply – High Power Supply (900 W of PoE power available)

ProCurve Switch zl 875 W Power Supply – Low Power Supply (273 W of PoE power available)

ProCurve 620 External RPS/EPS Power Supply – EPS (385 W or PoE power for each of two connected ProCurve Switch 3500s)

To find out more about
ProCurve Networking
products and solutions,
visit our web site at

www.hp.com/go/procurve

4AA0-9614ENW 11/2006