

A method to vary the Host interface signaling speeds in a Storage Array driving towards Greener Storage.”

Dr. M. K. Jibbe, Technical Director

NetApp, Wichita, Ks USA

Mahmoud.jibbe@netapp.com

316 636 8810

Arun Rajendran, Software Engineer

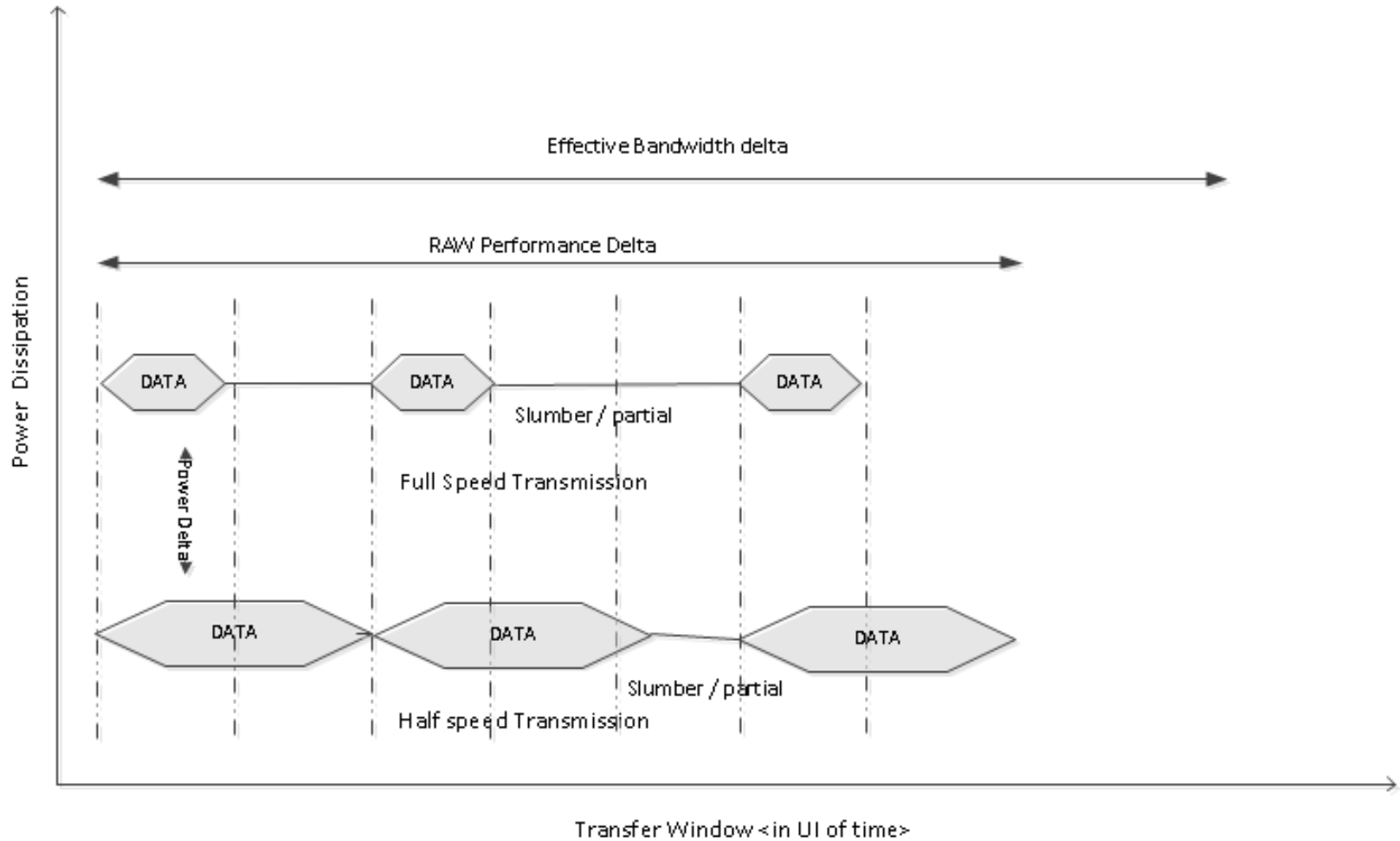
NetApp, Bangalore, India

- ❑ This presentation describes a method which can be used effectively to alter the signaling speeds of a Host Interface based on set performance criterion or user defined application / time of day criterion that are user definable.
- ❑ The end goals are
 - ❑ Considerable power savings by changing the signaling speeds to a lower supported speed. Such savings are confirmed by our background study and analysis.
 - ❑ Reduce the MTBF of components by operating such components at nominal speeds and improving the operable life span of the system
 - ❑ Move towards Greener Storage, low power operation, minimize Heat dissipation and emission reduction

- ❑ Here are some of the Green Storage Cost Cutting, Saving, and Measures
 - ❑ Using Analytics to monitor companies energy fertilize the green to reduce cost and increase sustainability
 - ❑ Cloud is cost effective in providing software services, virtualization, and scalable computing resources
 - ❑ Increase computational power (Reduce Idle time)
 - ❑ Reduce the data center load (Server count and Energy use,
 - ❑ Lengthen the lifespan of PCs /Servers (More Manageable Central Site)
 - ❑ Power down devices based on usage

- ❑ Enterprise Storage Arrays are required to be power efficient.
- ❑ Most of Application Work loads don't saturate the available raw bandwidth over the Storage Interfaces. This is mainly due to be Application specific latencies, protocol snags and other reasons.
- ❑ There is a significant time difference between maximum possible raw data bandwidth and real-time bandwidth realized during a data transfer across a UI between a Storage Host and Target.
 - ❑ A conceptual view of this delta is presented in the subsequent slide
- ❑ Explore further possibilities to dissipate lesser power over a Storage interface beyond ACTIVE/SLUMBER state variations.
 - ❑ Active = Fully operational
 - ❑ Partial = Low Power State
 - ❑ Slumber = Off state / Deep Sleep state
- ❑ A Storage interface signaling at lower Speeds dissipate lower power, this needs to be factored to enable power efficiency.

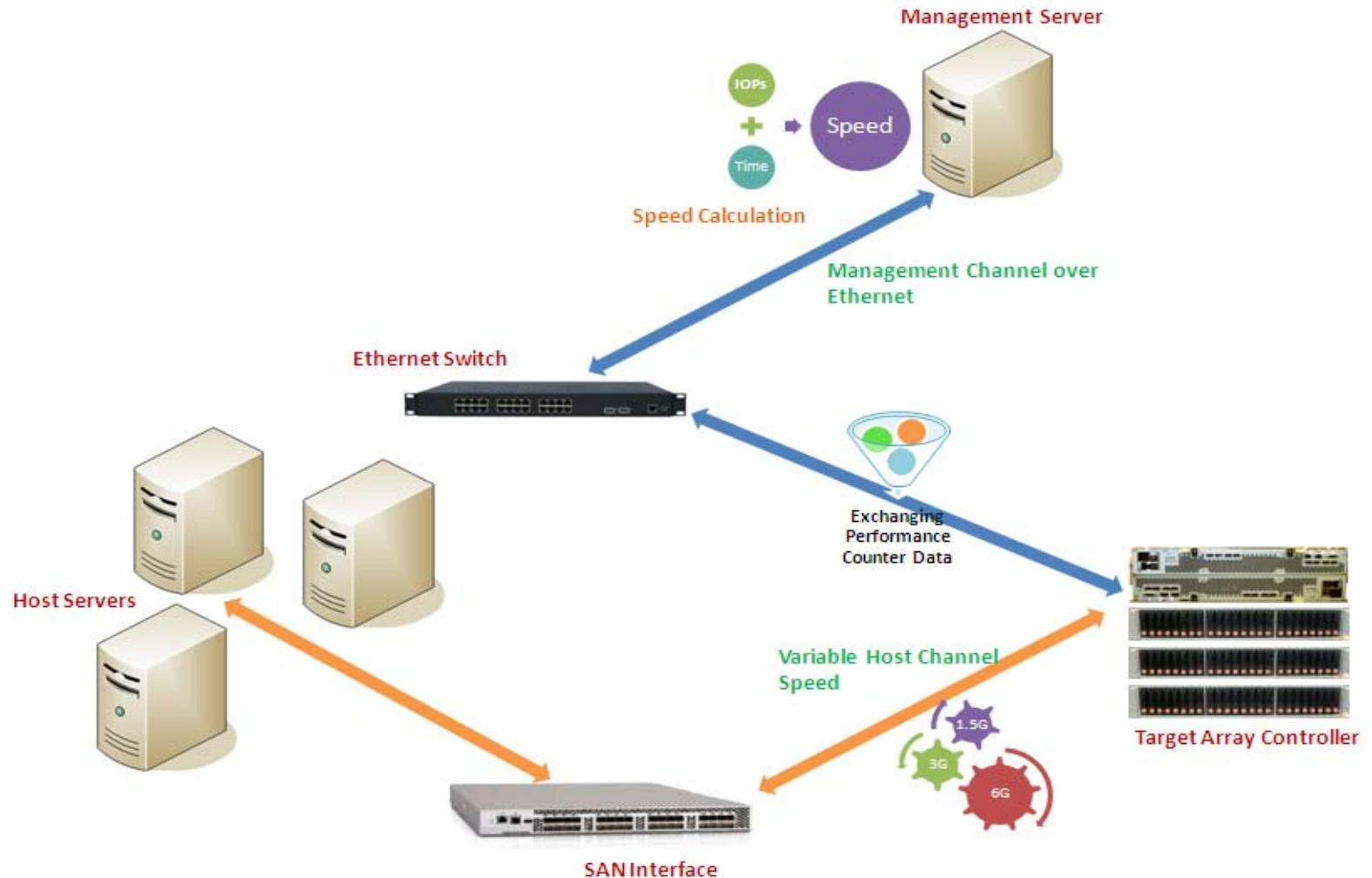
Bandwidth Delta – Conceptual View



Performance / Utilization Based Downscaling

- ❑ Determine the peak performance numbers which are specified in the Product Specifications and arrive at multiple thresholds downwards from such peak.
- ❑ During a real time data transfer the Management application connected to the storage array fetches the Host interface performance counter statistics at regular polling intervals.
- ❑ The decision logic in the management host compares the real time data to the pre-defined performance thresholds during a set monitoring window.
- ❑ Based on the real time performance/utilization trends, the management system can direct the storage array to vary the host interface speeds.
- ❑ The topology elements in this communication flow are mentioned below.

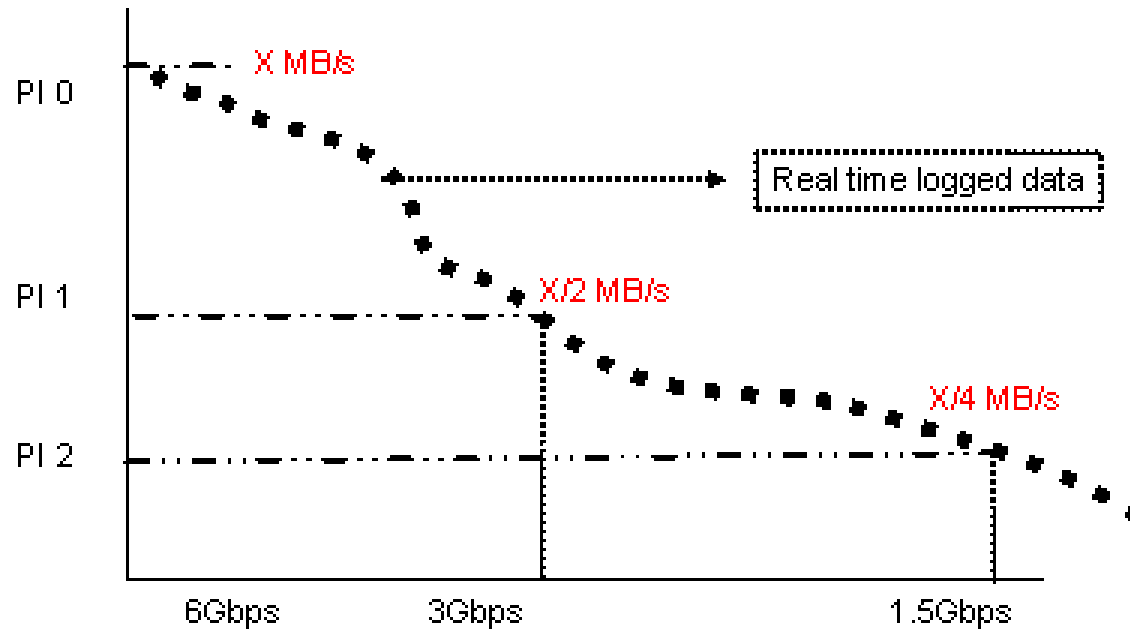
Topology in Details



- ❑ Based on Peak Bandwidth Capability of an Array - RD/WVR / sec. establish the following steps:
 - ❑ Determine the maximum Read / write performance or obtain the aggregate bandwidth per sec from the product specifications
 - ❑ Derive multiple comparison thresholds from this peak value.
- ❑ The number of comparison thresholds defined would be based on the number of variable speeds of a given host interface,
 - ❑ Example: For a SAS2 link based array the number of comparison thresholds would be 3 as the no of speed variants are 3 (1.5, 3, and 6 Gbps.
 - ❑ If X MB/s is the peak bandwidth in the array specifications and the array's host interface speed can be set at three different levels. These performance thresholds/indexes are
 - ❑ $PI(n) = X / 2^n$ where n is the speed variants: (PI(0) X MB/s, PI(1) X/2 MB/s, PI(2) X/4 MB/s are defined)
- ❑ Performance index are defined at the mid point to avoid frequent speed transitions.

Note: This schema can be applied to other specified performance / utilization metrics such as IOPs /sec etc.

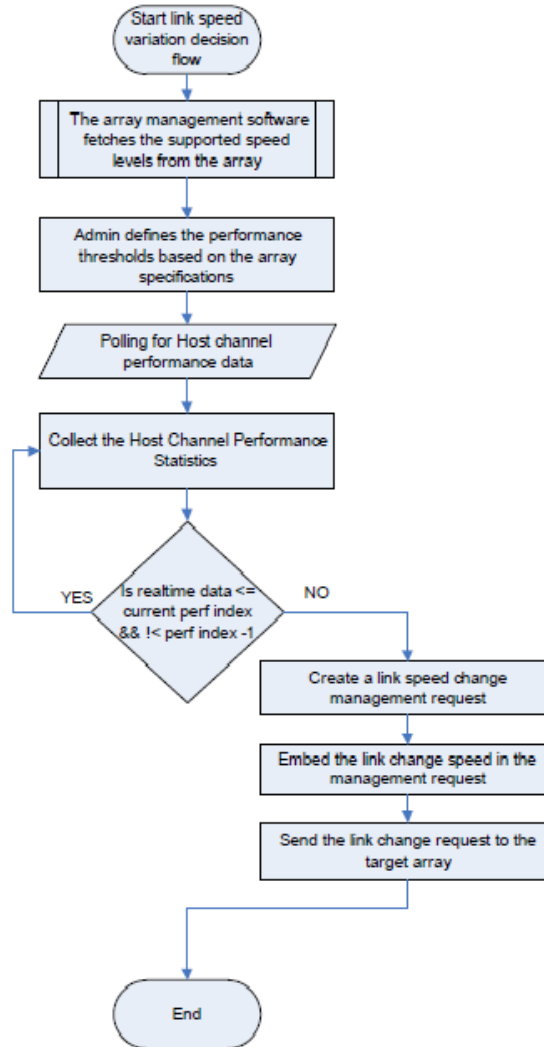
Conceptual View



X axis - Performance Indexes and the **matching thresholds**

Y axis - Host interface speed variations

The Flow



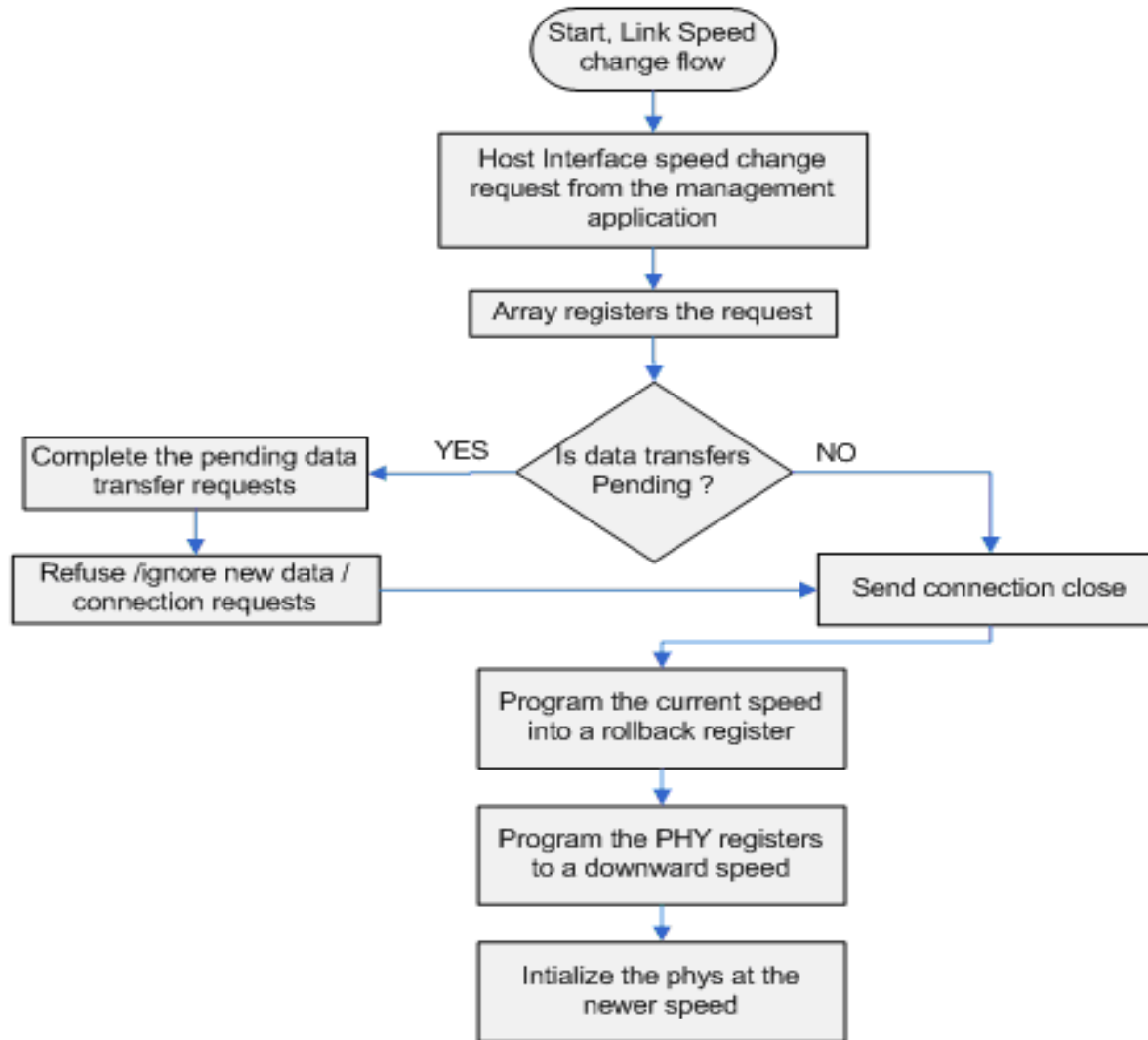
Decision Flow for Performance Data based link Speed Change (Downwards change)

- ❑ The Link Speed Change request shall be honored at the next Open Connection requests from the Specified Host.
- ❑ The link speed change request flow is described in the subsequent slides.

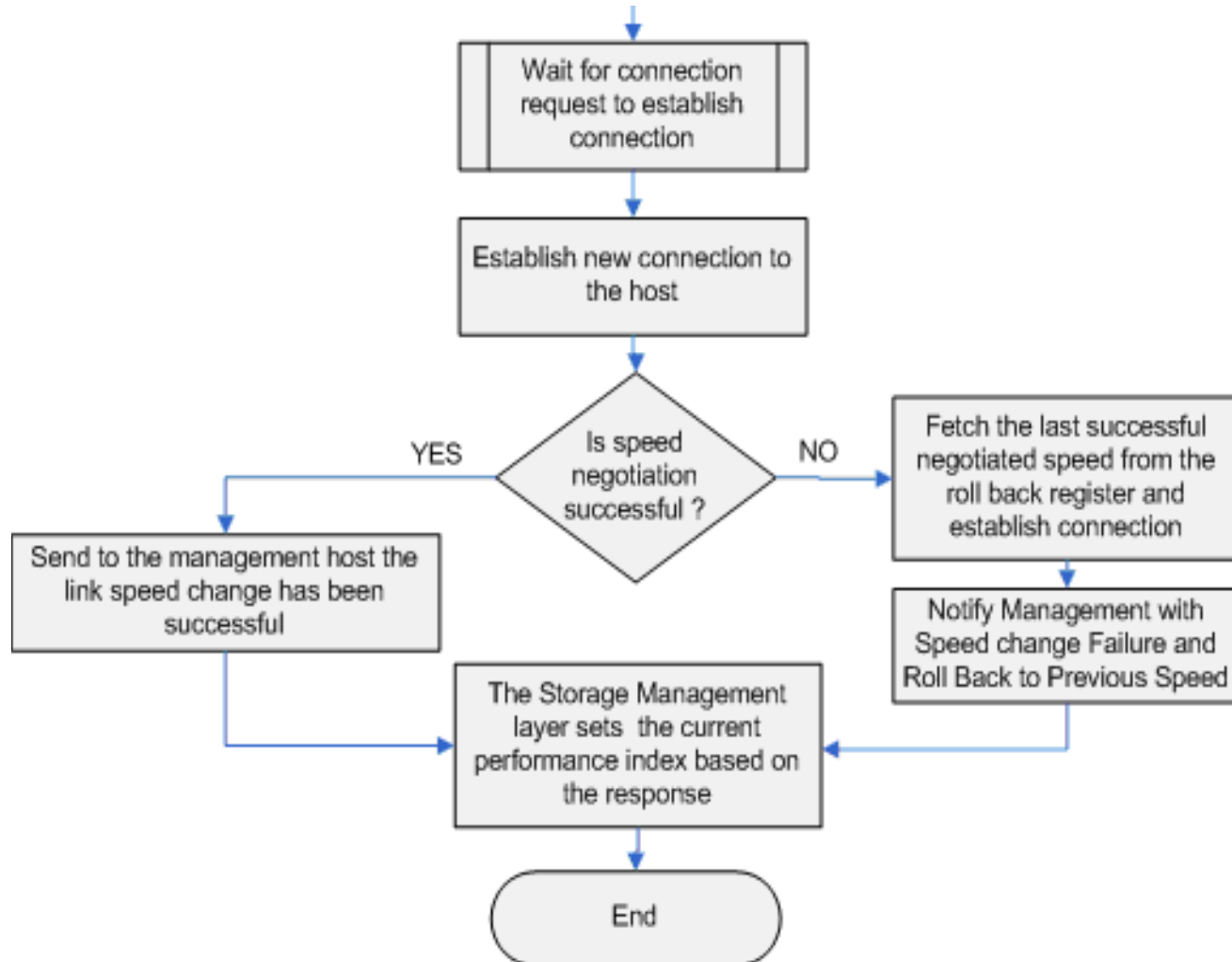
Time of Day Based Downscaling

- ❑ Another approach that we propose is time of day based host interface link speed variation;
 - ❑ Speed variation is directly dependent on the application load pattern
 - ❑ It is known that the load patterns tend to vary a lot across a time of day for e.g. an exchange server can drive data at a much faster (9:00 am – 3:00 pm) rather than the end of day (after 5:00 pm) .
 - ❑ Another example: For overnight backup Windows, Downscaling considers the power savings while running the interfaces at a lesser signaling speed throughout the backup window.
- ❑ Setup a set of time sensitive rules based on which the array can configure to downgrade the link speed
- ❑ The performance based link speed change flow and the time sensitive rules can be combined to reduce energy for the same array
 - ❑ Apply the performance utilization based schema only during the time window specified per rules.

Link Speed Change Flow.



Contd...



Advantages.

- ❑ The proposed method will deliver considerable power savings even when the system is online and processing I/Os.
- ❑ The method also achieve the optimal use of the available raw bandwidth by switching to lesser raw bandwidth if the data rate doesn't utilize the offered initial higher bandwidth rate.
- ❑ The performance thresholds defined are based on the specified performance metrics supported by the array. Such thresholds determines the utilization efficiency of the array capabilities.
- ❑ The link speed variation based power saving can co-exist with the existing protocol specific power saving modes (partial/slumber). A 3Gbps link in partial / slumber mode dissipates lesser power than 6Gbps in partial/slumber mode.
 - ❑ SAS power dissipation is 20% reduction per port between 6 and 3 Gbps
- ❑ The method also addresses exception conditions and allows rollback to the last supported speed if the system can't perform a link speed change / negotiation.

- ❑ The proposed method could create significant overheads if the polling periods for the performance data aren't chosen carefully.
- ❑ Overheads could also be high if there are far too many performance data based on which the host interface speed would be varied, Choose the appropriate Performance metrics.
- ❑ The definition of optimum time intervals for the subsequent interface speed switch might not be possible during Burst IOs or Scattered IOs.
 - ❑ In such condition (Scattered IOs across long duration) , the default Power management modes (like Sleep/Active) is far more beneficial in terms of power savings.
- ❑ The target array might not be able to find adequate window during concurrent data transfers across multiple Hosts. In such cases, target may not switch to a low speed mode.

Thank you

