John Nels Fuller
Interconnect Solutions Architect
48 NE Mulberry Ct.
College Place, WA   99324-2101

JNF

jfuller@compter.org
home: 509-526-4977
fax: 509-529-2073
mobile: 206-409-0338

Ethernet AVB Technology Assessment Report

1 Abstract

Very early in the development of the Energy Efficient Ethernet (EEE) standard by the 802.3 working group it was realized that a project in the 802.1 working group, Audio/Video Bridging (AVB), would be an important technology that the EEE technology would have to work with in order to be successful in the marketplace. The 802.1 working group also realized that AVB needed to work with EEE technology. The two groups kept up informal liaisons to try to address this interoperability requirement. As EEE neared a major milestone it was determined that an engineering resource should be brought in to evaluate how well EEE and AVB would actually work together. This report is the result of that effort.

2 Executive Summary

2.1 Goals

The following goals were defined for this assessment:

- Assess the design and operation of the draft Energy Efficient Ethernet (EEE; IEEE 802.3az) technology as it might affect operation of AVB.
- Assess the design and operation of the AVB technology as it might affect the operation of EEE.
- Derive initial conclusions about the nature of any potential conflicts between EEE and AVB and possible solutions.
- If time allows, assess the effect of Power over Ethernet (PoE) on the operation of EEE, and the potential for energy savings through PoE.
  Note: Time did not allow for this to be pursued in detail, initial thoughts are included in section 3.4.

2.2 Approach

The task leader (John Nels Fuller) has been a member of the 802.1 AVB Task Group since its inception in 2005 and is familiar with that set of standards projects. He undertook to study the 802.3az EEE drafts and ongoing work in order to become familiar with the portion of that technology that relates to the AVB technology. In this -process, he conferred with members of both working groups to clarify and validate his understanding of these technologies.

2.3 Activities

The following activities were performed:

- Attended IEEE 802 March plenary meeting. This included conferring with members of 802.3az and 802.1 AVB task groups.
- Attended 802.1 AVB task group conference calls.
- Attended 802.3az PHY shrink time conference call.
- Attended 803.3az Layer 2 conference call.
- Reviewed documents: 802.3az draft 1.2 and draft 1.3; 802.3az PHY shrink proposal; 802.3az Layer 2 proposal; 802.1AS; 802.1Qat; and 802.1Qav.

2.4 Significant Issues Encountered

The following issues were identified:

- The time for EEE to exit low power idle (LPI) may cause AVB to fail to deliver stream data in a timely manner unless care is taken in the implementation of the interface between the physical layer and the data link layer (i.e. between 802.3 and 802.1).
- The optional additional wait time on exit from LPI that may be negotiated between the two partners of a physical link will need to be restricted while AVB streams are active or it will cause AVB to fail to deliver stream data in a timely manner.
- EEE does not decide when to enter low power idle but merely provides a management interface to assert or de-assert the LPI_REQUEST. There is no guidance from EEE to upper layers about when to request LPI. The appropriate guidance may depend upon the operating environment (home, enterprise, performance venue, etc.).
- A significant AVB standard, 802.1BA, is not making progress because it lacks an editor. This document would be the home for most of the fixes to the above issues.

2.5 Implications

None of the significant issues is fatal to the interoperation of EEE and AVB, provided there is one minor change to EEE (see section 3.3.2), and the AVB document 802.1BA is completed and incorporates the required information (see section 3.3). I do not believe the task groups will find these changes objectionable, but they many require some education about the other's technology in order to accept them.

3 Background to this Report

3.1 Overview of 802.3az Energy Efficient Ethernet (EEE)

This overview is necessarily brief and the focus is on those features that are pertinent to this technology's interoperation with the AVB technology.

EEE adds a power saving mode to the Ethernet physical layer (PHY) technologies called Low Power Idle (LPI). For our purposes, LPI on the transmit path and LPI on the receive path can be viewed as independent and the remote receive path can be viewed as part of the local transmit path and vice versa.

While the transmit path is operating in LPI, or transitioning into or out of LPI, it is unable to carry data for the upper layers of the network stack. If LPI request is asserted and then subsequently upper layer data arrives, then the LPI request must be removed and the transmit path must be given time to exit the LPI before the data can be transmitted. This delay varies with the PHY technology (speed and media type, e.g. 100 Mb/s over twisted pair cable).

The link partners may negotiate an additional delay in either or both directions to allow even more power savings. The delay is limited by the amount of buffer space the transmitting partner is willing to dedicate to the additional delay in that direction, and by the width of the field communicating the delay (sixteen bits allows for 65,535 microseconds). The receiving partner may request this additional delay but must be able to operate correctly if the transmitting partner denies the request. For example, the receiving partner may request an additional 100 microseconds to give it time to bring up its CPU but if the request is denied it will leave its CPU running during LPI.

The following table gives the minimum delay for the twisted pair cable technologies currently supported by EEE:

PHY Technology	**Speed**	**Media**	**Minimum Delay**
100BASE-TX	100 Mb/s	Twisted pair cable	30 µsec
1000BASE-T	1 Gb/s	Twisted pair cable	16.5 µsec
10GBASE-T	10 Gb/s	Twisted pair cable	7.36 µsec

3.2 Overview of Audio Video Bridging (AVB)

AVB is a suite of standards that enable audio and video streams over the local area network with the following quality of service guarantees (in addition to the existing guarantees):

- A maximum end-to-end latency over seven hops for class A traffic of two milliseconds (class B latency is TBD but longer to allow transmission over Wi-Fi).
- Congestion will not cause dropping of stream data packets.

This allows a live video camera and/or microphone to transmit its stream across the local area network to a video display and/or speaker with only 2 ms of buffering required.

The following standards are included in the AVB suite:

- 802.1AS *Timing and Synchronization*;
- 802.1Qat *Stream Reservation Protocol*;
- 802.1Qav *Forwarding and Queuing Enhancements for Time-Sensitive Streams*; and
- 802.1BA *Audio Video Bridging (AVB) Systems*.

Each of these standards is briefly described below, focusing on those features that pertain to interoperability with the EEE technology.

3.2.1 802.1AS Timing and Synchronization

This standard, based on IEEE Std. 1588-2008, synchronizes time of day clocks on local area network nodes to be within 1 microsecond of the grand master’s time of day clock for nodes within seven hops of the grand master.

In order to do this 802.1AS requires the physical layer to provide timestamps for the actual transmission time and reception time of certain packets. Using this timestamp capability, 802.1AS measures the propagation delay on every hop of the network. This measurement is repeated at a TBD rate in the range of 62.5 milliseconds to 1000 milliseconds. This, along with the residence time of packets within a bridge, allows

802.1AS to know the adjustment to add to the time sent by the grand master to get the current time of day at any node.

There are additional adjustments for the variation in frequency of the local node's clock versus the grand master's clock and its nearest neighbor's clock. In order to keep these adjustments valid, periodically the nearest node closer to the grand master sends the time of day. After these adjustments are stabilized this period is TBD but in the range of 7.8 milliseconds to 125 milliseconds.

For both of the above TBD intervals, the shorter interval is known to yield the desired results but the longer interval is the goal and is awaiting verification by simulation.

3.2.2 802.1Qat Stream Reservation Protocol

This standard provides the mechanism by which nodes capable of sourcing an audio or video stream (talkers) announce their offerings and by which nodes capable of consuming streams (listeners) request those offerings. In addition, this mechanism reserves the resources to carry those streams from the talker to the listener(s) of each stream. The mechanism is called Multiple Stream Reservation Protocol (MSRP) and is based on Multiple Registration Protocol (MRP) which is a protocol defined in IEEE Std. 802.1ak-2007. MSRP also interacts with another MRP application called Multiple MAC-Address Registration Protocol (MMRP) if it is present.

The MRP algorithm maintains a distributed database of attribute declarations on the local area network. One of its features is a timer that produces an event about every 10 seconds called LeaveAll. LeaveAll forces all nodes to declare again all their attributes, forcing stale information out of the database. A single MRP data packet may contain many attribute declarations. In a small to medium local area network, a single MRP data packet will contain all the declarations for a particular MRP application (there are separate data packets for MMRP and MSRP).

When a talker declares a stream, it specifies the stream class, either Class A or Class B, and the stream's bandwidth requirements. The bandwidth requirements are specified by two numbers, the maximum packet size (before any network overhead added by layer 2 and below), and the maximum number of packets per class interval. For Class A the class interval is 125 microseconds, for Class B it is 250 microseconds. This information allows 802.1Qav to set up its traffic shaping.

3.2.3 802.1Qav Forwarding and Queuing Enhancements for Time-Sensitive Streams

This standard describes the shaping of stream data for transmission both at the talker and at intermediate bridges. The standard defines a credit-based shaper that ensures that stream data does not consume more bandwidth than is reserved for it. At the talker, shaping is both on a per stream basis and again on a per class basis. At a bridge, shaping is only on a per class basis. The shaper determines when the class queue has a data packet available for transmission, that is, it holds back the availability of a data packet until enough credit is accumulated to send it.

Class A traffic has the highest transmission priority when it is available, followed by class B traffic, and then by all the other levels of priority supported by the bridge or talker. A stream data packet is delayed if a lower priority packet has just started transmission when it becomes available. However, since credits continue to accumulate while that lower priority packet is sent, additional stream packets may become available and freeze out the lower priority queues until the stream queue has caught up.

Streams that generate more than one packet per class interval are not expected in the home environment because up to a 96 Mb/s stream can be supported on class A with a single packet per interval (for class B, up to 48 Mb/s). As a reference, an ATSC high definition video stream with 5-channel sound is about 24 Mb/s. Further, the home environment will mostly be 100 Mb/s Ethernet with only 75 Mb/s available to be reserved. Therefore, streams will probably generate one packet every 125 microseconds (class A) or one packet every 250 microseconds (class B).

3.2.4 802.1BA Audio Video Bridging (AVB) Systems

Work on this standard is in the preliminary stages (they are still looking for an editor). This standard will define profiles for AVB systems in various markets (automotive, consumer, professional A/V, and industrial). Each profile will specify what optional features of various 802 standards must or must not be implemented; what, if any, changes to default parameter values are required; etc. For example, 802.1Qat and 802.1Qav will be required, 802.3X (pause) will be prohibited. Design guidelines may also be included.

3.3 Areas of Concern between AVB and EEE

There are a number of issues to address to allow AVB and EEE to be compatible. The assumption here is that it is desirable to put a link into LPI even for only a very short time. Since AVB usage would provide many such short LPI times, the cumulative total would amount to a worthwhile amount of energy savings. If that is not the case then AVB and EEE become compatible simply by disallowing LPI while streams are active. The following subsections state various problems and their solutions that will allow LPI use even while streams are active on the link.

3.3.1 Delay while exiting LPI

Because AVB is trying to achieve just-in-time delivery of stream data packets, it is concerned with anything that may induce a delay in transmission of a stream data packet. Even without EEE, the timing is tight on 100 Mb/s Ethernet. This is primarily due to the time it takes to transmit a maximum length non-stream packet of 2000 bytes (about 160 microseconds). This happens if a large non-stream packet is chosen for transmission just prior to the availability of a stream packet. Other things that contribute to the delay of forwarding a stream packet are the fixed internal queuing delay of a bridge, and any other stream packets received on other bridge ports for transmission on the same outbound port. The sum of all these sources on a 100 Mb/s link is 255 to 275 microseconds. Two milliseconds (maximum latency over seven hops) divided by seven is about 285 microseconds per hop.

With EEE involved, if we select that non-stream packet for transmission and then find that we are in LPI, the resulting delay (LPI exit plus non-stream packet transmit time) can easily break the latency guarantee for the stream. That is, the LPI exit time (30 microseconds minimum on 100 Mb/s links) plus the other sources of delay (between 255 and 275 microseconds) can be more than the maximum allowable delay per hop (285 microseconds). To avoid this we can initiate exit from LPI when any packet is ready for transmission, but not choose which packet to transmit until the link is fully active as indicated by transition of CARRIER_SENSE to the off state. In this case, any available stream packet is transmitted before any non-stream packet or if no stream packet is available, any non-stream packet can be transmitted as in the case without EEE.

This falls in the gray area between the 802.3 documents and the 802.1 documents as it would be easy for designers to implement using the first method without understanding the need for the second method. The easiest way to address this is to put a description of how it should work into 802.1BA (this is an appropriate place since 802.1BA describes the entire AVB system).

For Ethernet speeds above 100 Mb/s the transmission time for a non-stream packet of maximum length does not dominate the worst-case transmission delay calculation but the above method still leads to the best possible worst-case delay.

3.3.2 Negotiated delay

When there are no reservations for stream traffic over the link, AVB considerations do not limit the negotiated delay after de-asserting LPI. However, if there is at least one reservation for stream traffic over the link then the total negotiated delay (including the minimum delay for the PHY technology) should be limited to a maximum of 160 microseconds (this is an AVB constraint, other considerations may impose even smaller limits). Thanks to the solution in section 3.3.1 the LPI exit delay is never added to a maximum length non-stream packet transmit time at 100 Mb/s so we may let LPI exit delay grow to that same time (even if this isn’t a 100 Mb/s link).

The 802.3az document should specify the ability for upper layers to limit the maximum negotiated delay then the 802.1BA document should utilize the feature to enforce the 160-microsecond limit when any streams are active on the link. Details of this interface must be worked out between the groups and will be included in an upcoming report.

3.3.3 When to assert LPI

The EEE document gives no guidance as to when to use LPI, it merely provides the knobs for others to use. There are many ideas for how to decide when entering a power saving state is appropriate. The more complex ones involve remembering traffic patterns over time in order to predict times of low use. This is probably too costly to implement in consumer devices.

Another idea is to wait for a period of inactivity of at least some defined duration, but inactivity is not a very good predictor, especially with AVB streams which send data every 125 or 250 microseconds.

For the consumer environment, an aggressive approach which asserts LPI whenever there are no packets available for transmission and then lives with the delay when a packet becomes available is simple to implement and probably optimal. This is clearly possible even on the slowest AVB supported technology (100 Mb/s) because the delay to return to an active link can be constrained to be less than 160 microseconds (the time to transmit a maximum length non-stream packet that happens to be available just before a stream packet becomes available). When the link becomes active, it transmits packets continuously until there are no more packets available for transmission (available stream packets before non-stream packets). At that point, the AVB system is again ready to accept a delay of up to 160 microseconds.

As an example, imagine a stream of High Definition Video with 5-channel Audio (an ATSC broadcast stream). This stream generates a packet of about 32 microseconds in length (at 100 Mb/s, or about 400 bytes) every 125 microseconds, leaving about 93 microseconds of unused link time. If the negotiated delay is the 30 microsecond minimum then one packet comes in, waits 30 microseconds for an active link, spends 32 microseconds transmitting then goes back to LPI, then 63 microseconds later another packet comes in and the cycle repeats. If the negotiated delay is 160 microseconds then one packet comes in, waits 160 microseconds for an active link during which time another packet comes in, then both packets are transmitted (64 microseconds) then goes back to LPI, then 26 microseconds later another packet comes in and the cycle repeats.

The appropriate place to describe this mechanism is in the 802.1BA document.

3.3.4 Completing the Specifications

As mentioned previously, the 802.1BA document is in need of an editor and there will be no progress until it finds one. Since there are a number of the above requirements for interoperation with EEE that need to be in that document, this project should consider providing an editor. I estimate this is a 2 to 3 year commitment of one half-time engineer plus travel expenses for six meeting per year.

3.4 Maximizing Power Savings

When there are no streams active over a link there will still be periodic traffic generated by 802.1AS and by 802.1Qat. The shortest period is sending of time of day by 802.1AS (in the range of 128 to 8 times per second). Participation in 802.1AS implies that the device will keep its clock running. Additional power savings are possible if the device stops participating in 802.1AS and turns off its clock. The cost is a long stabilization time (on the order of a second) when the device again begins to participate in 802.1AS. For many applications, this cost is acceptable. For example, a video display may take longer than this to bring up its screen.

If the device is not participating in 802.1AS there is still periodic traffic from 802.1Qat at an interval of about 10 seconds. There is probably no significant power savings to the choice of not participating in this protocol when the device is not active, especially if the device will continue to participate in a higher layer discovery protocol such as UPNP. If the device does not participate in a discovery protocol then it might as well be in a hard off state unless some other device provides a proxy service. Any control of the device's power through PoE mechanisms would have to work in tandem with, or be a part of, the proxy service. I am not aware of any attempt to define such a proxy service, but I believe that it would have to consist of a portion defined in 802.1 (layer 2) and another portion defined in the discovery protocol (layer 3).

3.5 Summary and Conclusions

The key issues and conclusions are summarized below:

- The time for EEE to exit low power idle (LPI) may cause AVB to fail to deliver stream data in a timely manner unless care is taken in the implementation of the interface between the physical layer and the data link layer (i.e. between 802.3 and 802.1). The mechanism described in section 3.3.1 should be described in 802.1BA.
- The optional additional wait time on exit from LPI that may be negotiated between the two partners of a physical link will need to be restricted while AVB streams are active or it will cause AVB to fail to deliver stream data in a timely manner. EEE should provide a management interface to allow the negotiated delay to be limited. 802.1BA should describe the use of that interface to limit the total LPI exit delay to 160 microseconds when there are active streams on the link.
- EEE does not decide when to enter low power idle but merely provides a management interface to assert or de-assert the LPI_REQUEST. There is no guidance from EEE to upper layers about when to request LPI. The appropriate guidance may depend upon the operating environment (home, enterprise, performance venue, etc.). For each operating environment profiled in 802.1BA that document should describe when LPI_REQUEST is asserted and de-asserted;

additionally the Data Center Bridging and Interworking task groups of 802.1 should be consulted to add describe use of LPI in their documents.

- A significant AVB standard, 802.1BA, is not making progress because it lacks an editor. This document would be the home for most of the fixes to the above issues. This project should consider providing an editor.
- None of these issues is fatal to the interoperation of EEE and AVB, provided there is one minor change to EEE (see section 3.3.2), and the AVB document 802.1BA is completed and incorporates the required information (see section 3.3). I do not believe the task groups will find these changes objectionable, but they many require some education about the other’s technology in order to accept them. These education efforts should include:
    - A presentation to EEE about how AVB streams are managed across multiple bridges to achieve the two millisecond maximum latency over seven hops.
    - A presentation to all of 802.1 about EEE focusing on the controls provided for asserting and de-asserting LPI_REQUEST, the LPI exit delay, and the proposed control for limiting the negotiated delay. There should be a specific call for each of 802.1’s task groups to examine EEE to determine if any energy savings are possible in their environments.