Minutes - 802.1 Interim Meeting – York, 26th-29th September 2006

Attendees

Maksim	Azarov
Hugh	Barrass
Davide	Bergamasco
Jan	Bialkowski
Rob	Boatright
Matthew	Bocci
Jean-Michel	Bonnamy
Mike	Borza
Paul	Bottorff
Rudolf	Brandner
Robert	Brunner
Paul	Congdon
Uri	Cummings
Kevin	Daines
Arjan	de Heer
Thomas	Dineen
Linda	Dunbar
David	Elie-Dit-Cosaque
Lars	Ellegard
Felix Feifei	Feng
Norm	Finn
John	Fuller
Chris	Gallon
Geoffrey	Garner
Anoop	Ghanwani
Franz	Goetz
Mark	Gravel
Robert M.	Grow
Steve	Haddock
Takafumi	Hamano
Asif	Hazarika
Guy	Hutchison
Guillermo	Ibanez
Romain	Insler
Tony	Jeffree
Michael	Johas Teener
Keti	Kilcrease
Yongbum	Kim
Raghu	Kondapalli
Bruce	Kwan
Anthony	Magee
David	Martin
Alan	McGuire
Menucher	Menuchery
John	Messenger
Gusat	Mitch
Dinesh	Mohan
David	Olsen
Bo	Osterhammel
Glenn	Parsons
Neil	Peers
Haim	Porat
Max	Pritikin
Martin	Rhodes
Robert	Roden
Josef	Roese
Allyn	Romanow
Dan	Romascanu
Eric	Ryu
Ali	Sajassi
Joseph	Salowey
Panagiotis	Saltsidis
Mick	Seaman
Koichiro	Seto
Gopi	Sirineni
John	Sisto
Nurit	Sprecher
Kevin B	Stanton
Bob	Sultan
Richard	Sun
Pat	Thaler
Oliver	Thorp
Maarten	Vissers
Manoj	Wadekar
Brian	Weis
Bert	Wijnen
Hideo	Yosimi

802.1 Opening session

Usual set of opening slides from Tony Jeffree as Chair – slide set is on the website in the “minutes” folder (http://www.ieee802.org/1/files/public/minutes/). As required by IEEE policy, the patent policy slides were presented to the group.

Future interim meetings:

- January meeting 22 Jan 2007 in Monterey, hosted by Broadcom (MJT), following an 802.3 interim the previous week at the same location.
- May 28-31 in Geneva: parallel meetings of 802.1 and 802.3. Then a joint workshop with IEEE and ITU-T.
- There is an SG15 meeting the following week, also in Geneva (June 4th).
- September possibly in Seoul, Korea.

There is now an official IEEE OID arc. We used to use an ISO one. We think we will move to use of iso(1) iso-identified-organization(3) ieee(111) for new OID allocations, but not for existing allocations.

Link aggregation: currently an 802.3 standard; we want to move it into 802.1. There was no real progress on this at the last 802.3 meeting, though Pat and Bob Grow said it had been mentioned, and no major objections.

ITU-T Liaison

There was no incoming liaison.

Mick - Agenda

Added item- discussion of coordinated management approach

Mick - discussion of PARs. How to manage so much work in the group. Going to impose restrictions on what enters, must be highest quality. Right now, we do a standard every 9 months. New groups take 3-4 years to produce a standard. Otherwise would have to fracture into more dot-groups. More on this later, with a proposal.

Interworking and AV tracks:

802.1AB revision

Paul Congdon described the issues which are driving a revision of this standard. The main problem is the use of LLDP in scenarios where it is assumed that the partners are physically connected. This was a misuse of LLDP, but a protocol is needed for such applications. The proposal is to add additional LLDP destination addresses and rules for the propagation of LLDP frames with these addresses. There are also a couple of bugs. The revision may also define new TLVs for AVB (SRP), congestion management. Other ideas include new discovery information, rapid exchange, etc.

802.1Qat AV Bridging Stream Reservation Protocol (SRP)

This is the main protocol in the AVB effort, aimed at automatic configuration of traffic management parameters in AVB-aware bridges. The PAR for this project was discussed at the July meeting and has since been approved by the standards board. In accordance with the new naming convention for amendments to existing standards, this project has been renamed as 802.1Qat, showing that it is an amendment to 802.1Q. The "at" letters are simply part of a sequence "a-z", "aa-az", "ba-bz", etc.

AV Queuing for Time Sensitive streams

The PAR and 5 for this were reviewed in the joint session of AVB and interworking. This PAR will be pre-circulated and submitted for approval at the November meeting.

AV Bridging status update

Michael Johas Teener (Broadcom) updated the group on progress in the audio-video bridging future work. They intend to work together with the ITU-T group as well as IEEE 1588. They are concerned about the interaction with 802.11 in the future. ITU-T has consented G.8261 which is a framework and problem statement in the circuit emulation space. It doesn't say much but is intended as a root for specific approaches to be developed. Additional drafts called G.pacmod and G.pacclk are being developed. These are specifically targeting circuit emulation over Ethernet.

802.1aq Shortest path bridging

There was no work on this project at this meeting.
Suping presented a couple of proposals on shortest path bridging.

Transmitting loopback and Linktrace messages: the status variables were accidentally omitted and will be restored.

802.1ag Connectivity Fault Management

Draft 7.0 has been out for working group ballot, closing a few days before the meeting. 91% qualified voters responded (69 people). The ballot did not pass (58% Yes, 42% No; to pass it needs 75%).

The meeting considered the editor's proposed dispositions on the comments, which he had prepared and posted prior to the meeting. Preparing dispositions on 353 comments in a few days is a mammoth achievement. We did not have time to review every comment disposition in the meeting, so discussion centred on what the editor considered the main issues.

MIB problems

Linktrace message is different from Y.1731. Easily resolvable.

Bad frame format with the extra bytes (e.g., in CCM)

Figure 19-2: MEP architecture. If a MEP has a loss of connectivity, it declares its (internal) interface down (ifOperDown) and does not pass data. That's how it tells the rest of the stack that it has lost connectivity.

There is no per-VLAN macOperational parameter.

Some remaining issues:

Inconsistent use of the term "Service Instance" between this standard and 802.1ad.

Overloading of the term "Maintenance Domain", one of which will need a new name:

- A set of DSAPs that may be configured to provide service instances.
- A row in the Maintenance Domain MIB table, controlling access to a certain number of MSa and supplying a "Maintenance Domain Name" as part of the MAID.

There are many changed resolutions during the ballot comment resolution. A new disposition of comments will be issued today. Then we will run a recirculation ballot prior to November, with the hope that the document may be able to go out to sponsor ballot following the November meeting.

802.1ah Provider Core Bridges

Steve Haddock ran the first part of the ballot resolution process. There are still several difficult issues and time over-ran.

802.1ap Bridge MIB

Glenn Parsons presented progress on this area. The anticipated order of amendments are .1ad, .1ak, .1ag, .1ah, .1ap, .1aj, .1aq. There's lots of progress towards getting a full IEEE-controlled bridge MIB.

The existing Bridge MIB will be re-indexed under the new IEEE arc, and included in 802.1ap.

802.1ak Multiple Registration Protocol

Comment resolution on the recent ballot was performed at this meeting. The comments were not too major and the standard should be finished soon, probably following one more recirculation ballot.

802.1aj Two-port MAC Relay

Mick Seaman presented his new paper on MAC Status Propagation. It is reasonably complicated and **requires study**, because it is likely to end up as part of TPMR and maybe elsewhere.

Data-driven and data-dependent fault management

Mick has developed a PAR and 5 criteria for this work.

There are new CFM opcodes for these functions. There are security concerns. PBB is secured by traffic segregation by VLAN and perimeter security based on MACSEC etc. An attacker has to be able to access the MIPs or MEPs and maybe there are some safety guarantees provided by that.

Linda Dunbar (Huawei) has written an outline draft of this project.

Security Working Group track

802.1af Mick Seaman
Clause 7 – Wake on LAN, newer description in the latest draft.
p.49 Fig 7-14 how to set up VLAN. How to integrate MACsec and key management with VLAN bridge, would be same for router. Integrated system view of key agreement and all related processes.

802.1AR Mike Borza
Discussed definitions of secure device identifier. More descriptive secure device identifiers.

Thursday PM

Security Working Group

802.1AR Mike Borza – discussed what is a device ID
ECC, Elliptic curve issue. Should we do ECC or RSA for keys? ECC is more efficient, has a smaller key size and better processing performance. Unknown IPR conditions. RSA is less efficient but has known favorable IPR conditions.

802.1af Mick Seaman
Went over changes from last version.
Mick has proposed some changes to the .1X state machine to resolve issues with devices that implement both an authenticator and supplicant role.

Annex Z1 contains comments from last ballot

EAP discussion
We'll have to specify a mandatory to implement method, one that is compatible with 802.1AR. Probably EAP-TLS.

Key Confirmation and signaling - This will likely be built off of MKA instead of .11i. It will probably use EAPoL frames (new packet type, not within key descriptor). This will be used for distributing the CAK from a pairwise association and for distributing the SAK. This will probably use key transport.

MKA- not discussed in any detail.

Friday AM, Sept 29

Security Working Group

802.1AR
Will Idevid and Ldevid use the same key pair or not? Yes should.
Anything with a Devid has to store root key. About 10k of storage space needed.

Strategy – give a security module that can be used by networking people, let them complete the solution, rather than giving them the complete solution, which they wouldn't be comfortable with. Like TPM
The audience for the spec is similar to the rest of 802.1 group- networking techies and management, not security people.

Scope of document. There is a normative scope. And a scope that explains the use and is tutorial.
Anything that drives the requirements will to go in the body of spec, not the annex.

Max wants to define an abstract API

Comment- on using PKI in .1AR and not in .1af. Because of frequency. In .1AR not in the critical path, but it is in the critical path in .1af.

For XML perf. Describe how being used, analyze perf of each step. Where is the long pole? XML perf per se may be irrelevant.

DevID modules don't intersect.

Friday PM, Sept 29

Security Working Group

802.1af Joe Salowey presentation on choosing the CAK

Policy- whether run MACsec or not, what algorithm. Call it signaling rather than policy.
4 functions

Pre-shared key option- the Pre-shared key is the CAK
There are multiple models for how to get the CAK.

Joe's diagrams show modularity, can replace the PMK part..
Mick will incorporate Joe's slides

Problem with virtual machines
802.1af will punt on this in the spec
Will push out drafts weekly

802.1AR
v7 in October
Task group ballot for November
Will need another TG cause don't have MIB yet
MIB – ldev and Idev credentials
Other status info
Imprintable state of device

Imprinting and provisioning
Not defining an enrollment protocol
Should define how it would work, though

MIB interface can inject idevid, when device is imprintable
Provision as a one time thing, of idevid
Ldevid – can trigger enrollment, and inject a new ldevid
Here is our opportunity to do enrollment

Service interface vs a MIB
Define operations and data objects
Remotely provision via SNMP
If define as part of MIB, have an enrollment mechanism for AR
Makes the spec complete
We are worried about defining the MIB, not the process, i.e., SNMPv3
Mick thinks we should do this or we are liable for criticism

Requesting a new key, and proof of possession of the new key
Ask device for CSR is a good idea, proof of possession. A MIB object

Max will draft this section and send it to Mike.

Congestion Management track

Thursday 28th September 2006 (9.00am-5.30pm)
Secretary: Manoj Wadekar

Attendees:

1.	Pat Thaler	Broadcom
2.	Manoj Wadekar	Intel
3.	Mitch Gusat	IBM
4.	Mark Gravel	HP
5.	Norman Finn	Cisco
6.	Davide Bergamasco	Cisco
7.	Hugh Barrass	Cisco
8.	Hideo Yoshimi	NEC
9.	Bruce Kwan	Broadcom
10.	Romain Insler	France Telecom
11.	Robert Brunner	Ericsson
12.	Guillemo Ibanet	Univ. Carlos III Madrid
13.	Anoop Ghanwani	Brocade
14.	Uri Cummings	Fulcrum
15.	Menu Menuchehry	Marvell
16.	Anthony Magee	Adva Optical
17.	Takafumi Hamano	NTT
18.	Asif Hazarika	Fujitsu Micro
19.	Bob Grow	Intel
20.	Jan Bialkowski	Brocade

==
===================

Agenda:

1. Review Agenda
2. Simulation Ad-hoc report and results
3. Simulation results: Davide, Bruce, Uri
4. Mitch: Tutorial on LL-FC mechanisms
5. Pat Thaler: Excel based simulation for BCN
6. Review of Objectives
7. Norm Finn: Draft 0 for BCN
8. Paul Congdon: Discussion on "Transmission Selection"

==
===================

Minutes:

2. Manoj Wadekar: CN-SIM AdHoc Report

i. Overview of sim activity since SD Plenary: Weekly calls w/ 15+ members from 10+ companies.

ii. Four distinct sim environments + guidance from the team

1. Q: details on sampling, jitter, time vs. per Byte sampling?
2. Davide: everything is calculated in 64B pages

3. Fairness: Davide - Using RMS as alternative to JFI, which shows (too) good results

   a. Pat: Discussion re. fairness will continue in the future

4. There's now a common understanding of our sims => achieved!

   a. Q: Fullcrum - BCN(0) and Mod1 (proposed by Jain ?) - to be presented by Uri.

----

3. Davide Bergamasco: Simulation results

i. Mitch: Should we be using JFI or focus on RMS? JFI is providing less relevant data. Need to consider Throughput/Latency Fairness index as well.

1. Probably discussion item for next meeting (with Raj in room if possible)
2. Davide: Goal for these FI for simulations was to compare consistency of results.

ii. BCN (0,0): Will broadcast help better to penalize all the contributing flows? This will avoid bipolar behavior (some flows having max rate and some 0 rate etc.)

1. Manoj: Need to compare PAUSE and BCN(0,0) as both operate in same space of the problem. Also need to discuss coexistence, if required.

iii. PAUSE only:

1. PAUSE: B1-B4 do not generate BCNs (to avoid multiple BCN sources)
2. Q: Is sampling changed to account for PAUSE - arrival rate is 0

   a. Sampling mechanism is not changed

   b. To be thought further.. Maybe modify scheme for defining Qdelta when queue is saturated

iv. PAUSE + BCN(0,0):

1. BCN(0,0) drives RPs to 0 quickly - throughput loss on CP link. But PAUSE operates for short duration - good indication that congestion spreading will be limited

2. Discussion: Can BCN(0,0) guarantee "no-drop"? Not really. Number of sources and control loop delay is difficult to tune. Or sampling rate. So, we may need some PAUSE mechanism to achieve "no-drop".

----

4. Bruce Kwan: Simulation Results

   i. Numbers are converging as compared to other simulation environments

----

5. Uri Cummings: Simulation Results

   i. Using no EP latency - loop delay is 6 uS : Difference with other environments

   ii. Work presented to study FI for

      1. Different windows

      2. Different queue sizes - larger queue size makes AFI worse although improves packet-drop

         a. JFI is relatively flat for similar exercise

   iii. Otherwise simulation seem in line with other environments

----

6. Mitch Gussat: Tutorial survey of LL-FC for Data Center Ethernet

   i. Mitch to email reference to a book that has excellent overview of various LL-FC schemes

   ii. Norm: 802.1 does not have mechanism to generate PAUSE. However, 802.1aj has two ports - it has mechanism to send PAUSE.

   iii. Need to discuss which is "primary" congestion - long lived or transient

      1. 1 mS- chronic, nS-uS: Contention, key question is what to call 100s uS congestion? Can be argued to be more prevalent problem.

   iv. Credit/Pause comparison:

      1. Pat - challenge for credits is to have agreement of what "credit" is

         a. FC had used "packet" for credit

         b. IB used "chunk" (say 64B) for credit

         c. PAUSE is architecturally independent

      2. Uri: memory tradeoff with simplicity/interoperability - PAUSE looks more attractive

   v. VL: Pat: IB defined this for multiple usages - Priority, VLAN, multiple spanning tree etc.

   vi. Discussion of further granularity on queues - separate QoS from CM

1. Need to create "good enough" solution, even if it is not "best" and we have "brief throughput loss" during LL-FC duration

2. Mitch: "Good enough" can't be small packet loss. LL-FC needs to covers time till BCN takes over

----

7. Lunch break: 12.20pm - 2.00pm

----

8. Pat Thaler: Presentation of Excel model for BCN

i. Fairness is reasonably good for BCN as compared to other mechanisms seen

ii. Mitch: Fairness should take secondary importance as compared to other factors like work conservation, simplicity etc.

iii. Should consider "time based sampling" vs. "bytes based sampling" for speeding up the convergence

----

9. Open Discussion

i. Pat: Discussion on LLDP

ii. Operates at hop-by-hop

1. LLDP is using 0x0E address for special multicast
   a. Customer bridges don't forward these
   b. Provider bridges forward 6 of these addresses

iii. No ACK, no negotiation, announcement only
iv. Slow protocol
v. We need fast startup - same as AVB group
vi. Carries "Management address" as part of the packet (TLV)

vii. Should participate in discussion triggered by Paul Congdon for .1AB update for supporting AVB and CM needs.

viii. We need people to bring forward proposals for Discovery as well as MIBs.

ix. Common agreement for need of project for "Link level flow control"

x. Need simulations for BCN + PAUSE

1. Will be addressed in CN-SIM

Friday 29th September 2006

----

10. Paul Congdon: Recap of "Proposal to improve expedited forwarding" (from July 2005)

   i. Q: In CM context: There are two groups of traffic "CM compliant" and "CM non-compliant" (or "CM-indifferent"). Both these groups may have high/low traffic. Does the proposal require whole group placed adjacent?

      Answer: No.

   ii. Discussion on Mixed scenario: Does remaining BW go to highest priority or distributed across all the PGs? Ans: Paul intended earlier, however both implementations can be represented by same table

   iii. AVB is interested in traffic shaping in addition to these things. Paul:
      1. Hope is to combine CM and AVB requirements in a single project
      2. However it is not quite clear right now how it can be achieved.
      3. Joint meeting with AVB is next step.

   iv. Pat: More details on Paul's proposal will be useful for discussion with AVB

   v. Bruce: Clear specifications of CM requirements is necessary.

   vi. Lot of discussion about impact of "Minimum BW" for priority group on BCN.
      1. Need to revisit later when simulations advance more.

----

11. Pat Thaler: How does IB do it?

   i. Two tables - Hi and Low priority.

   ii. Hi Table is serviced WRR and if no packets (or credits) in Hi-pri - Low priority table is serviced

----

12. Norm Finn: D0.1 draft for 802.1u discussion

----

13. Future Work Items: Discussion

   i. Simulations for BCN+PAUSE

   ii. Management
      1. List of relevant information
      2. Discovery/Domain Creation protocol

3. MIB objects
4. LLDP Objects (TLVs)

iii. Presentation on:

1. CM Mechanisms
    a. BCN-MAX
    b. PAUSE

    c. BCN Interaction
        i. e.g. what BCN to send when output is PAUSEd
        ii. What BCN to send when input is PAUSEd

    d. To fill in the holes in D0.1 - coordinate directly with Norm

2. Per Priority Pause
    a. List of objectives and constraints

    b. Address, EtherType, Opcode: Same or different than 802.3X?
        i. Coexistence with 802.3X
        ii. Hugh (& Pat)

    c. Discussion on dead/live locks in link level flow control and typical solutions - Mitch

    d. 802.1 architectural positioning of PPP - Request to Norm?

3. Transmission Selection:
    a. Datacenter transmission selection needs

    b. More details on Paul's proposal
        i. How much black box vs. white box
        ii. Scheduling algorithm
        iii. Need for maximum BW%?
        iv. Burstiness?

    c. Possible PAR wording - Pat
    d. Use of DE bit

----

14. Meeting adjourned 12.30pm
