<<DRAFT MINUTES>>

IEEE 802 / IETF Workshop

On Data Center Networks

November 10, 2018

1:30PM – 5:30PM

Attendees (Name and Affiliation):

Ting Ao (ZTE Corporation), Ignas Bagdonas (Equinix), David Black (Dell EMC), Paul Bottroff (Aruba), Bob Briscoe (CableLabs), Stewart Bryant (Huawei), Randy Bush (Arrcus & IIJ), Ayla Chaing (Huawei), Mach Chen (Huawei), Paul Congdon (Huawei), Subir Das (Perspecta Labs), Ian Duncan (Ciena), Donald Eastlake (Huawei), Hesham Elbakoury (Huawei), Janos Farkas (Ericsson), Miao Fuyou (Huawei), Barak Gafni (Mellanox), Steve Haddock (Stephen Haddock Consulting, LLC), Bob Heile (Wi-Sun Alliance), Russ Housley (Vigil Security), Yihong Huang (Huawei), John Kaippallimalil (Huawei), Yong Kim (NIO), Victor Kuarsingh (Oracle), Warren Kumari (Google), Hyeong Ho Lee (Netvision Telecom), Bing Lin (Huawei), Sun Liyang (Huawei), Andy Malis (Huawei), Roger Marks (Huawei), John Messengar (Adva), Greg Mirsky (ZTE Corporation), Tal Mizrahi (Huawei), Al Morton (AT&T Labs), Paul Nikolich (Paul Nikolich), Yoshihiro Ohba (Toshiba Memory Singapore), Glenn Parsons (Ericsson), Keyur Patel (Arrcus), Cristel Pelsser (University of Strasbourg), Tony Przygienda (Juniper), Yinzhen Qu (Huawei), Jon Rosdahl (Qualcomm), Richard Scheffeneger (NetApp), Jeff Tantsura (Apstra), Pascal Thubert (Cisco), Yuji Tochio (Fujitsu), Paul Unbehagen (Extreme Networks), Hao Wang (Fujitsu), Yuehua Wei (ZTE Corporation), Takahiro Yamaura (Toshiba), Bin Yeong Yoon (ETRI), James Young (Commseopa), Yolanda Yu (Huawei), Zheng Zhang (ZTE Corporation), Ning Zong (Huawei)

Minutes – Prepared and submitted by Paul Congdon:

1. The workshop commenced at 1:30PM. The online agenda and description of the workshop is attached below and also available here: https://1.ieee802.org/802-nendica/802-ietf-workshop-data-center-bangkok/
2. The organizer, Paul Congdon, read the IEEE 802 pre-PAR meeting participation slides to the attendees as part of the Workshop opening remarks available here: https://mentor.ieee.org/802.1/dcn/18/1-18-0070-02-ICne-ieee802-ietf-workshop-on-dcns-welcome.pdf
3. Roger Marks presented an update on the Nendica project and Lossless Data Center Network published report. The presentation is available here: https://mentor.ieee.org/802.1/dcn/18/1-18-0068-00-ICne-overview-ieee-802-nendica-report-on-the-lossless-network-for-data-centers.pdf
   a. It was pointed out that this document, while published, is still open for comment and comments will be considered for a future revision. The attendees were requested to read and comment on the paper.

b. It was suggested that a new document that provides technical details from both IEEE 802 and IETF on data center technologies should be created as part of the Nendica activity.   It was pointed out that the 802 Industry Connections activity, of which Nendica is a part of, have rules for operation allow this to easily occur because membership is not required, and all output is open and freely available.

c. It was suggested that the Nendica contributors read the Sigcomm17 'best paper' paper on re-architecting the data center for low latency and low loss (Handley). In response, it was mentioned that this paper is referenced by the Nendica report.

d. There was a comment that some of the approaches described appear to be re-inventing L3 at L2. A clarification from the author was that the paper is trying to purpose NOT propose this and is looking specifically for how the two layers can co-operate with one another.

e. There was support for continued collaboration between the two organizations and a commenter volunteered to actively participate.

4. Paul Congdon presented a technical overview of the new Congestion Isolation project, P802.1Qcz, within IEEE 802.1. The presentation is available here: https://mentor.ieee.org/802.1/dcn/18/1-18-0066-00-ICne-p802-1qcz-technical-overview.pdf

a. There was a comment that the classic approach for isolating an elephant my kill the elephant. The suggestion is essentially that the current approach for congestion control may not be appropriate going forward and something that is more 'scheduled' for large persistent elephant flows may be more appropriate.

b. It was pointed out that identifying flows using a 5-tuple is the right place to start, but the IETF has innovated on flow identification and we may want to consider additional definitions of flows.   There are now many types of flows that contain embedded streams within the traditional flow (e.g. WebRTC). It was pointed out that NVO3 is another example of having flows within flows. This would be an area of future collaboration and documentation.

c. In addition to multiplexed streams in flows, other flow types like MPLS should be considered. What is important, however, is that we are interacting with end-to-end congestion control and that there is some assurance that the source of the flow will reduce its injection rate in response to the isolation action by the data center network. So, how to interact with MPLS would be a good topic of investigation.

d. It was clarified that the CIM signaling of Congestion Isolation is hop-by-hop. A comment was made that perhaps the signaling is not needed and it was voiced that one vendor would prefer to not perform signaling.

e. There was a question if there was an idea of how big the congested flow table needs to be. The presenter responded that simulation has shown it is not huge and only congested flows need to be included. However, it isn't clear exactly what size is needed and specifics could not be provided, but it is believed that existing flow tables for ACLs and OpenFlow are more than enough.

f. A comment was made that notifying an upstream neighbor when a flow is no longer congested may not be necessary. A response commented disagreed and pointed out that end-to-end congestion control and fairness would be impacted if this was not done. The presentation includes an example of a scenario where this is useful.

g. A comment regarding the need for security and authentication of backward congestion messages should be addressed. A second commenter supported this idea and mentioned that research papers discuss this, but no references were provided. Unauthorized CIM messages can impact flow performance by isolating them.

5. Richard Scheffenegger presented a vendor's view of future requirements for converged ethernet and storage appliances. The presentation has been replaced due to embedded copyright statements [per objections noted during the presentation]. A replacement presentation is available at: https://mentor.ieee.org/802.1/dcn/18/1-18-0081-00-ICne-converged-ethernet-and-storage-appliances.pdf.

   a. A clarification question was asked whether traffic from different applications and uses cases is mixed all the time, or if separate classes of service are used. The presenter indicated that traffic is mixed because of the difficulty of configuring QoS and classifiers.

6. Bob Briscoe presented background on congestion control, problems created with classic TCP and a new approach for low latency called L4S. The presentation is available here: https://mentor.ieee.org/802.1/dcn/18/1-18-0073-00-ICne-l4s-ultra-low-queuing-delay-for-all.pdf

   a. It was pointed out that some of the material from the presentation may not represent a position of consensus within the IETF. In particular, the topic of re-ordering and RACK support.

   b. There was a question of how network slicing relates to the L4S work. It was felt that slicing should provide hard walls within the network pipes such that bandwidth and latency control can be applied the same way.

7. Barak Gafni presented Data Center Silicon Considerations. The presentation has been replaced due to embedded copyright statements [per objections noted during the presentation]. A replacement presentation is available at: https://mentor.ieee.org/802.1/dcn/18/1-18-0072-00-ICne-data-center-asic-considerations.pdf

   a. The presenter noted that 3 SDOs are involved in the RoCE standards.

   b. There was a comment that the embedded OAM that was described as a new feature for future ASICs should be aligned with the IETF approach. Currently there are multiple IETF approaches being proposed.

   c. There was a question about how well the ASIC implementations scale at high bandwidth. The presenter was not familiar, but confident it should scale.

   d. There were questions about the appropriate model for the telemetry data. The presenter would like to work with the vendors to develop the appropriate models.

8. Tony Przygienda presented an overview of the RIFT activity within the IETF. The presentation is available here: https://mentor.ieee.org/802.1/dcn/18/1-18-0069-00-ICne-rift-open-standard-zero-opex-ip-fabric-routing-underlay.pdf

9. Randy Bush discussed Network Discovery in the LSVR working group of the IETF and the requirements for a liveness and discovery protocol. The group is proposing LSoE (link state over Ethernet). The presentation is available here: https://mentor.ieee.org/802.1/dcn/18/1-18-0071-02-ICne-ieee802-ietf-workshop-network-discovery.pdf

   a. An objection was made to IPR statements within the presentation. The presenter agreed to remove the statements and re-submit the presentation. The link above is a version of the presentation without the IPR statements.

10. Paul Congdon wrapped up the meeting with a discussion of potential next steps and follow-up. These include the potential for a new Nendica report to further detail and discuss the relationship between IETF and IEEE 802 protocols as they relate to the data center.
11. The meeting concluded at 5:43PM

IEEE 802/IETF Data Center Workshop – Bangkok, 2018-11-10

In November 2018, the IEEE 802 and IETF are meeting on consecutive weeks at the same hotel in Bangkok, Thailand. This fortuitous schedule allows a unique opportunity to assemble key technologists from both organizations to discuss and explore new and developing standardization activities for the Data Center environment. The IEEE 802/IETF Data Center Workshop of 10 November 2018 is intended to act as a catalyst for future collaboration between the two organizations with respect to Data Center technologies, with an emphasis on congestion management, load-balancing, routing, and simplified management.

The workshop is scheduled for Saturday 10 November from 13:30 – 17:30 at the Bangkok Marriott Marquis Queen's Park (Thai Chitlada Room #2, on the 2nd Floor). There is no cost to attend, but registration is required. The **online registration** allows you to also register for the DetNet – TSN joint session the following day.

The following agenda is tentative and will be adjusted as speakers are confirmed. For more information, or to offer your participation, please contact Paul Congdon.

Schedule	Topic	Presenter	Organization	Length
1:30PM - 1:40PM	Opening Remarks	Paul Congdon	Tallac/Huawei	10
1:40PM - 2:10PM	IEEE 802 Network Enhancements for the Next Decade (Nendica): Lossless Data Center Networks Activity and Report	Roger Marks	Chair, IEEE 802 Nendica	30
2:10PM - 2:40PM	P802.1Qcz - Congestion Isolation	Paul Congdon	Tallac/Huawei	30
2:40PM - 3:00PM	Converged Ethernet and Storage Appliances	Richard Scheffenegger	NetApps	20

Schedule	Topic	Presenter	Organization	Length
3:00PM - 3:30PM	L4S: Ultra-Low Queuing Delay for All	Bob Briscoe	CableLabs	30
3:30PM - 3:45PM	Break			15
3:45PM - 4:15PM	Data Center Silicon Considerations	Barak Gafni	Mellanox	30
4:15PM - 4:45PM	RIFT: Open Standard, Zero OPEX, IP Fabric Underlay	Tony Przygienda	Juniper	30
4:45PM - 5:15PM	Network Discovery	Randy Bush	Arrcus & IIJ	30
5:15PM - 5:30PM	Wrap-up	Paul Congdon	Tallac/Huawei	15