



[Home](#) [Archives](#) [Profile](#) [Subscribe](#)

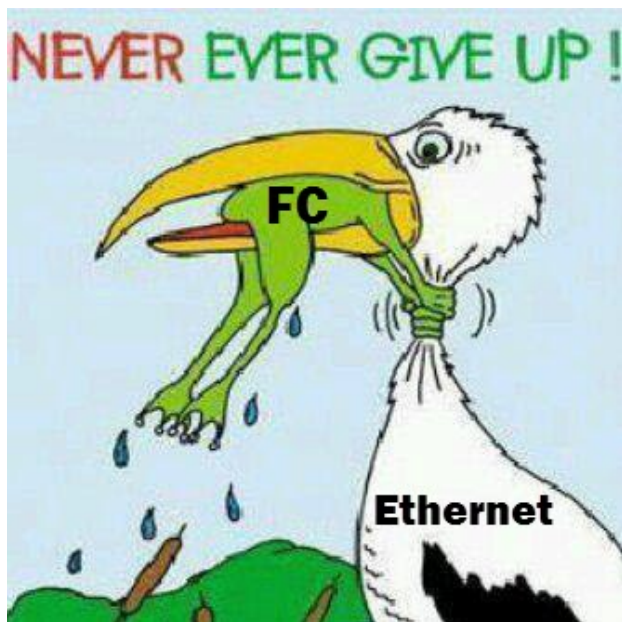
[« Join me at EMC World 2015!](#) | [Main](#) | [Slow Drains are impacting your SAN!](#)

[»](#)

05/19/2015

Fibre Channel is better than Ethernet... ...for Enterprise Storage.

Although this blog post isn't an April fool's joke, I do greatly appreciate them when they are done well. As a result I'm slightly in awe of the amount of thought Tony Bourke has put into his last three ([2013](#), [2014](#) and [2015](#)). Truth be told, I'm actually kind of honored that he sees me as the poster child for FC/FCoE (at least based on his tweets from April 1st). To save you a bit of reading, his posts all rest upon the idea that FC is yesterday's technology, has one foot in the grave, and is desperately trying to prevent itself from being swallowed by the stork by any means necessary.



While Tony's posts were all done in good fun and have given me new respect for the term getting [Bourked](#), I will say that the idea of "FC is dying" is becoming a common theme in the industry; here are a few examples:



[Erik Smith](#)

[ARCHIVES](#)

[September 2022](#)

[August 2020](#)

[April 2019](#)

[December 2017](#)

[November 2017](#)

[July 2017](#)

[April 2017](#)

[September 2016](#)

[July 2016](#)

[February 2016](#)

SEARCH

[CATEGORIES](#)

[Subscribe to this blog's feed](#)

DISCLAIMER

This is not an official EMC blog. The opinions expressed here are my personal opinions. Content published here is not read or approved in advance by EMC and does not necessarily reflect the views and opinions of EMC nor does it constitute any official communication of EMC.

[“Fibre Channel Really Is Dead”](#)



I had many fundamental issues with this article and I plan to do another post shortly that explains why RoCE, and even the new darling of the Ethernet tribe “Routable RoCE”, won’t kill off FC anytime soon.

[“Hyperconvergence and Death of the Storage Array - Interop 2015”](#)



While I do not want to downplay the importance of Server SAN, and I agree that its use in hyperscale environments clearly demonstrates its ability to scale, I think it’s a bit of a stretch to say it will displace all array based storage (and hence the use of FC) in the timeline Wikibon has been promoting. The fundamental problem I have with their line of reasoning is this; they’re comparing the current state of a known entity (e.g., Enterprise Storage Arrays) with an approach (i.e., Server SAN) that is theoretically possible but has yet to reach the peak of its [hype cycle](#) and as a result, is still benefiting from the “Peak of Inflated Expectations”. I say this somewhat tongue in cheek, but feel passionately that anyone who thinks Server SAN is a fully baked technology is fully baked. In addition, the wikibon analysis seems to assume that the storage array vendors will [“go gently into that good night”](#) and not introduce any new array based technologies to differentiate themselves; I believe this is a fundamental flaw in most people’s logic when talking about this topic. As a matter of fact I was quite taken with the [DSSD](#) demo that [Chad Sakac](#) and [Bill Moore](#) put on at EMC World.



It may be early days for the DSSD folks, but I really like where they seem to be headed. The important point is this; they're already providing something in an array form factor that cannot currently be implemented using the Server SAN approach. How does this relate to FC? Well there's talk of running NVMe over FC (i.e., FC-NVME) and what nobody seems to be realizing is that many of the features that the NVMe over fabrics group have expressed an interest in (e.g., a "Name Server", reliable transport) are already fully supported by FC and have been for 15+ years. From where I'm standing it seems like FC has a leg up on the competing transports here, especially if you don't want to place your bets on iSNS.

BrassTacks - "[An Introduction to Virtual Storage Networks](#)"



Yeah, about that... I guess I'm guilty of promoting the idea that the future of storage connectivity is all Ethernet too, and I still think it probably is, but only for platform 3 Applications. Why not platform 2 or 2.5? Read on...

Birds of a feather flock together

At EMC World 2015, during the "FC SAN, Should I stay or should I go?" Birds of a Feather (BoF) session on May 6th, I had the opportunity to speak with a group of

400+ EMC Customers. To start the session off, I pointed at the title of the presentation (FC SAN, Should I stay or should I go?) and asked:

“How many of you are here because you’re asking yourself this question?”

Answer: 90+% of the attendees

I then asked **“How many of you use FC as your primary means for accessing external storage”**

Answer: 90+% of the attendees.

Next **“How many of you use IP Based storage?”**

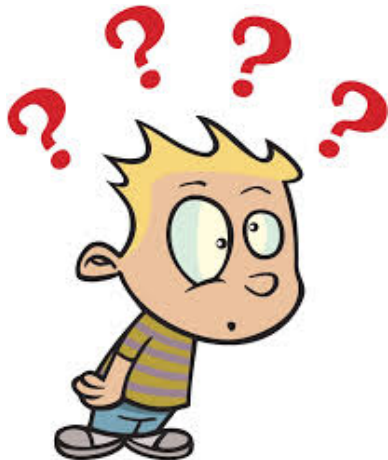
Answer: about 30%

And then **“How many of you see moving to an IP based storage protocol (iSCSI or NAS) for your primary means of accessing your external storage capacity in the next 18-24 months?”**

Answer: 3 people!

Finally **“How many of you see moving to an IP based storage protocol (iSCSI or NAS) for your primary means of accessing your external storage capacity in the long term?”**

Answer: 5 people!



I found this answer perplexing because we’ve been hearing from some big Enterprise accounts as well as many in the networking industry about “everyone’s plans” to move to a converged LAN and IP SAN. In fact, given the current amount of hype around this topic, I assumed that there would be a significant amount of actual storage administrators in the BoF that would be in the middle of planning this move and had structured the BoF presentation accordingly. However, during the BoF, instead of the active “give and take” kind of conversation about IP Storage Networks that I was hoping for, I got the same reaction as I had at previous EMC World BoF Sessions when I talked about IP Storage; [crickets and tumbleweeds...](#)

Nevertheless, since I had prepared to talk about IP Storage Networks, I spent the better part of the next 20 minutes trying to figure out what it would take to move them towards a converged IP network. At every opportunity my audience was giving me kindly worded feedback which I blew past because I was waiting to hear one of the “Actual meaning in American” phrases.

ENGLISH STATEMENT MADE BY JAPANESE PERSON	ACTUAL MEANING IN AMERICAN
I see.	No.
Ah.	No.
Ah-hah.	No.
Yes.	No.
That is difficult.	That is completely impossible.
That is very interesting.	That is the stupidest thing I ever heard.
We still study your proposal.	We will feed your proposal to a goat.

Once I finally got the hint that they were not interested in IP SAN, thankfully it only took 20 minutes and a redirect from the Director of the Connectrix BU, (I can be stubborn), what started to emerge was a very interesting picture that was, in hindsight, consistent with the responses I've received over the past few years especially when talking to these customers about Infrastructure automation. The key questions and answers that tied all of this together for me were:

"How many of you are running OpenStack or some other Infrastructure Automation tool in production"

Answer: 2 people!

"How many of you are evaluating OpenStack or some other Infrastructure Automation tool"

Answer: Less than 10% (~30 people)

When I combined this with:

1. Something Randy Bias indicated during one of the keynotes (I didn't get the exact quote, but it was something to the effect of), "Everyone who has tried to use OpenStack to support an existing platform 2 App has failed"; and
2. The general reaction we got from customers last year when we were demonstrating the VSN-CS, which went something like "Very interesting but I wouldn't know what to do with it. NOW can you tell me where the free beer is?"

I realized that we are dealing with (at least) 2 different very broad classes of users.



The first class of users consists of those that are being forced to evolve and provide services to their customers in an on demand fashion. A good example is the EMC Rubicon team that presented at the devops conference at EMC World. The kind of use case they need to support isn't just do more with less, it's do everything with no human involvement, in zero time, for as little as possible. Yeah, I know, as technologists we're being force fed the idea that everyone involved in IT falls into this category and will eventually need to be like "Amazon" or "VMware vCloud Air". And if you aren't working on this right now you're a loser and you'll be out of a job by next year. And if your product team doesn't have a long term plan to ship your products as collections of container based micro services that are all automatically installed and updated by Puppet and validated continuously by Jenkins which is necessary because you'll be releasing / deploying new versions of your product every millisecond onto your customer's OpenStack Zebra instances which doesn't need to be reliable because fault tolerance is built into your application because well applications are just a bunch of cattle anyway, right?????



AND..... BREATHE.....



....Not that there's anything wrong with this new approach, again, I think it represents the future for some types of applications. But right now, this type of automation (especially infrastructure automation) is in its very early stages, still really more of a framework than something that you can take out of the box and setup by yourself (without a ton of knowledge AND effort). Actually, strictly from an end user point of view, OpenStack specifically kind of reminds me of a very early, sparsely documented Linux distribution from 15+ years ago. That having been said, I think EVENTUALLY it (or something like it) will be just as important to infrastructure as Linux is to compute.

In any case, this new approach has a ton of potential, but it's unstable and making infrastructure stable is really hard to do especially if you need to do so

in an automated fashion. The example I think of is; how to automate the creation of a completely redundant path to your storage array? (Just noodle on that one for a few minutes, if you figure it out let me know, we're hiring). All of this means that running platform 2 Applications in this kind of environment would be challenging because they need stability, because for the most part they rely on the underlying infrastructure to provide redundancy.

This leads me to the second group of people. Perhaps unsurprisingly they're the people who need to support these second platform applications and who I am now affectionately referring to as the "Fibre Channel Tree Huggers".



These FC Tree Huggers are people who value "Stability", "High Availability", and "Predictable performance" and also know that if something happens to one of the applications under their control, it could have significant ramifications not only financially but legally. Keep in mind, some of these people are responsible for maintaining infrastructures that could have life and death consequences were an outage to occur. People of this class are also responsible for the vast amount of infrastructure that keeps our society, financial institutions and even our national defense operational. Is it any wonder that they value stability over automatability? If I had that kind of responsibility, I'd probably be clinging to the FC Tree too, because in my (and more importantly their) opinion, physical FC is a better transport for storage than Ethernet.

Keep in mind I've spent the majority of the previous 7 years working with FCoE (wrote two books on the topic) and have spent the past 3 years focused on IP SANs and Virtual Networks. So I could probably ramble on endlessly about why I feel this way, but I'll boil it down to the FC advantages that seem to resonate most with the FC Tree Huggers (my "peeps").

Top 8 reasons to hug the FC Tree:

1. Isolation - from LAN traffic for performance, predictability, fault isolation or manageability.



This is the most frequently used justification (by far) when we speak to customers who don't want to converge their LAN and SAN. While there are a lot of different dynamics at play behind this seemingly straight-forward sounding reason, they boil down to a general distrust between the Network and Storage guys. I know this reason is an [old-saw](#); we've been hearing about this concern since before FCoE was even released. And, I think many blew this concern off as something that the management layer could somehow organizationally resolve (at least eventually). And, although I am aware of instances where management has been able to get the Network and Storage guys to cooperate, the majority of these teams don't, and this is the root of many of the other issues related to reason of isolation.

A great example where the lack of trust between Network and Storage teams poses a tangible problem is in the area of QOS guarantees and performance monitoring. For the sake of this example, let's assume that the network team has agreed to dedicate bandwidth (call it 40%) to storage traffic across a typical three-layer network topology or let's be generous and assume it's a leaf /spine topology. When users start complaining about bad performance and the Compute guys start pointing at the storage guys as the root cause, how are the storage guys going to be able to troubleshoot the SAN portion of the converged network? According to a very reliable source, it's technically possible to configure RBAC to give the storage team access to only their ports and even to give them read only access, but what about visibility to the shared links (e.g., Leaf to Spine)? How are they going to clear counters, or drop a suspected problem port in the course of troubleshooting? Again, in the vast majority of cases, there is no way the storage guys would even have visibility to the network.

Another interesting concern about converged networks is the [VLAN hopping issue](#) we discovered in the lab.

None of the above is a problem if the Storage team has dedicated network resources (either FC or IP based would be fine).

2. Port fencing



If a port misbehaves either at the physical layer (FC-0, FC-1), the transport protocol layer (FC-2), or the storage protocol layer (FC-4), a FC switch can take the port offline automatically. Yes, an Ethernet switch can do this too (e.g., BPDU guard), but generally speaking not for a transport protocol layer or storage protocol layer violation.

So why does this matter? Well, it's not terribly common, but one example that I've personally worked on a couple of times in an HBA that malfunctions in a way that causes a denial of service attack against the storage port they are logged into. When this happens, an FC switch may be able to detect the condition and shut the port down before it can impact the other devices that may happen to be using the same storage port.

3. Path redundancy based on an [air-gap](#): Provides redundancy and safer upgrades.

Bridge collapses aren't funny but the following picture of the I-5 bridge collapse in Washington State illustrates why many people who are very concerned about catastrophic failures prefer an air gap design.



In the case of a converged SAN, again the concern about the interaction between the Network and Storage teams comes into play. The storage teams are concerned that their paths won't actually be redundant and that upgrades will not be staged (one leg at a time).

As I mentioned in my original [air-gap blog post](#) on the topic of air gaps, there is a technical solution to this problem but again it requires coordination between the Network and Storage teams.

4. Slow drain detection and remediation



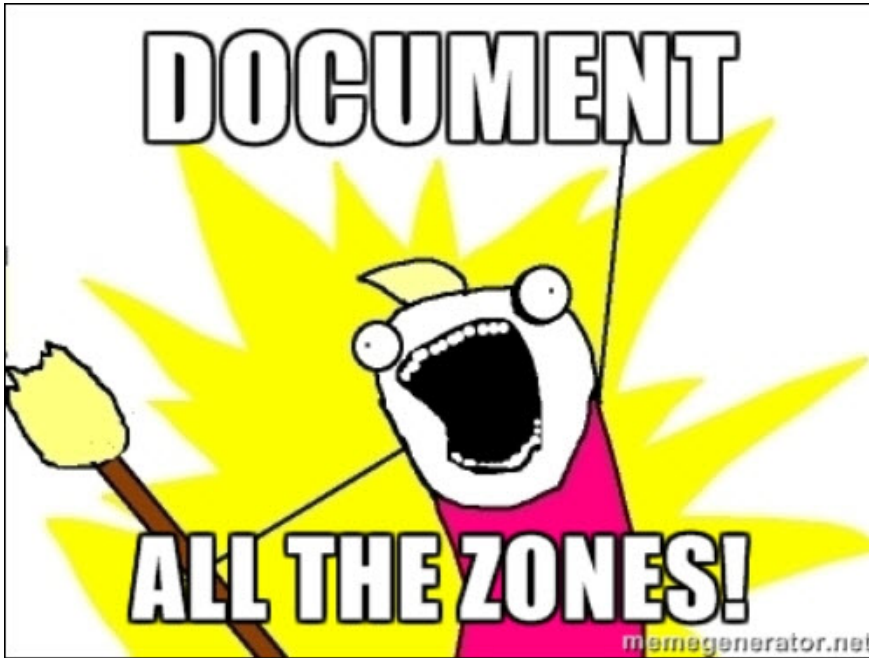
Slow drain devices and congestion spreading are facts of life for lossless networks like FC or DCB Ethernet (used to transport FCoE, Lossless iSCSI and RoCE). For an in depth description of the problem, refer to the [EMC Networked Storage concepts and protocols techbook](#) and look at the Congestion and Backpressure section. The good news is with FC, both switch vendors have been working on solutions to these problems for years; not so with Ethernet. Brocade especially has kicked things up a notch with FOS 7.4 and their [slow drain device quarantining](#) feature.

5. Mature, large scale, centralized name services

OK, not that kind of centralized name service... But while were on it, George, please finish writing books 6 and 7 and PLEASE stop killing all of my favorite characters!

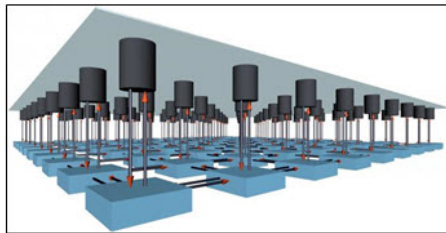
In addition to enabling the “self-documenting” and “network centric management model” discussed below, the distributed FC Name Service provides a very simple way for users to select end devices and add them to a software defined network (more commonly referred to as a FC zone). [Zoning](#) is both a blessing and a curse to FC. It's a blessing because it gives the users the isolation they need, but it's a curse because it takes a bit of skill to properly administer them and this administration is difficult to automate. You can do something like this with Virtual Networks, which is exactly why [we're digging into them](#).

6. FC is self-documenting



When I first pitched a detailed overview of [TDZ](#) at EMC World 2011, many of the concerns I received were in regards to a perception that the users would lose the ability to define the zone names. Actually this wasn't just a perception, that was pretty much my intention from the [get-go](#). In any case, as I talked about this concern with end users, many of them explained that their zone names all follow a specific name format (unique to their environment) that includes bits of information such as, the application, the hostname and the storage port interface; and that this information was of critical importance during troubleshooting. The idea is that an end user calls the storage admin with a concern about a particular application or host and because of the zone naming convention, it's very easy for the storage admin to use the zone name to locate the relevant ports and their WWPNs. Once the WWPNs have been identified, the storage admin can verify that the appropriate ports are logged into the fabric and then drill into the specific physical interfaces involved to look for errors. The same cannot be done with IP Storage.

7. FC is network centric versus end-node centric

O
R

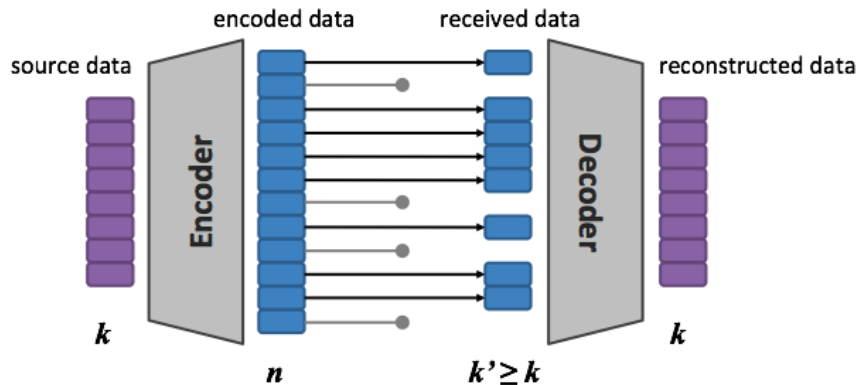
I wrote a [detailed blog post](#) on this topic, so I won't repeat all of those details here. However, the key takeaways are:

FC and FCoE are Network-centric. Both protocols rely on fact that the network will control what each end device has access to. A Network-centric approach is probably better suited for large organizations that need centralized control of access to storage resources.

iSCSI is End-Node-centric. iSCSI relies on the fact that the network will allow communication between the iSCSI Initiator and whatever iSCSI Target the Server Admin points the Initiator at. Since control is managed at each individual end

point, the end devices have evolved so that they will only discover what they are told to discover. The bottom line is iSCSI is probably better suited to smaller organizations that do not need centralized control of access to storage resources.

8. Forward Error Correction (FEC) at 16G



Forward Error Correction is required on 16G FC links but not on 10GbE. So why does this matter? Well, SCSI-FCP (SCSI - Fibre Channel Protocol) was designed to run over a practically lossless transport and as a result, there's no retransmission capability built into FC. This is usually fine because historically frame loss due to drop or corruption is very rare. The problem is, as link speeds increase and as cabling infrastructures age, we're noticing that bit errors due to either dirty fiber terminations or exceeding maximum distance supported for a given fiber type at a given speed are starting to increase.

With FC, when a bit error occurs, if it happens to land in the middle of a frame, the frame will be discarded by the receiving interface. This can have a tremendous impact on the SCSI protocol, sometimes resulting in the need for the SCSI timeout to expire (30-60 seconds by default) before the IO can be retried.

FEC helps prevent this by correcting bit errors and this prevents the frame from being discarded by the receiving interface. This means that the SCSI timeout scenario is much less likely to come into play when you're using 16G FC.

Since FEC is not supported on 10/40GbE, you are more exposed to these kinds of problems when using FCoE at 10GbE. iSCSI / NAS both use TCP, so this problem wouldn't be a much of an issue in an IP SAN.

It's worth mentioning that according to a friend who's familiar with IEEE and the standardization of 25GbE and beyond, it appears as if 25GbE SR will be supporting FEC, so this won't be an issue for Ethernet once you move to 25GbE.

Conclusion

So which is better, FC or Ethernet? It depends on what you need...

If you're working in a traditional enterprise data center and you need stability more than automatability, then embrace your inner FC tree hugger and continue to support the connectivity requirements of your existing platform 2 Applications using the best transport available, FC. FC and the companies that provide FC solutions are not going away any time soon and based on the amount of innovation we've seen recently from our switch partners, you can bet they're not going to stop innovating anytime soon either. As a result, expect to see

32G/128G FC shortly and new protocols being added such as FC-NVME as it makes sense to do so.

If you're working in an environment where automation is king and the use of commodity homologous HW is a foregone conclusion, then you're probably already using either a Server SAN approach or an IP SAN of some kind. However, if you find that you need to use NVME at some point in the future, I would urge you to at least consider FC and see what it can offer you once the protocol has made it through the standardization process.

Thanks for reading!

Posted at 03:00 PM | [Permalink](#)

 Reblog (0)

Comments

 You can follow this conversation by subscribing to the [comment feed](#) for this post.



Hi Erik - thanks for the very long post. I completely agree with you that it is very early days for Server SAN and that is a big point of why Wikibon strongly believes in the potential of that new architecture. If you've seen David Floyer's Flash as Memory Extension (FaME) - it is likely that EMC's DSSD will fit under this heading and FaME fits under the Server SAN umbrella - it's about new ways of combining and scaling storage and compute which is different from the traditional FC or iSCSI SAN (or NAS) which separated compute workloads from storage. FC has been a great tool and is not dying any time soon (we expect that by 2020 it will have dipped but even then won't be gone) - IT is never good at stopping anything, it's always additive.

Cheers,
Stu

Posted by: [Stu Miniman](#) | [06/03/2015 at 07:32 AM](#)



Hi Stu, with regards to the length of the post, despite literally hearing your voice in my head saying "keep it to 1000 words or less", I find it hard to strike a balance between "telling the whole story" and "keeping it short". As a result, I always err on the side of providing more information so that my readers can understand exactly what my facts and assumptions are. This tends to make my posts "very long". On the upside, know that I spend days agonizing over my ability to backup every word I write, so hopefully the quality of the post will make up for the excessive quantity.

Thanks for the pointer to David's FaME post, it looks really interesting and I'll spend some time getting familiar with the concept.

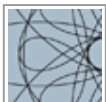
With regards to FaME's applicability to DSSD. Let me state up front that I have no idea what DSSD is planning in the short or long term (perhaps you have some insight that I don't) but since they are currently packaged as an array, I'm confused as to how they would

fit under the Server SAN umbrella regardless of their choice of connectivity protocol?

In any case, my blog post was intended to call attention to the fact that the enterprise customers that I spoke with have little interest in moving away from FC and have a list of (mostly) technical reasons for this position. I also hoped to encourage those who rely on FC for stability to embrace their inner FC Tree Hugger and stay on FC unless they have a compelling reason (e.g., automation) to move to IP Storage or Server SAN.

Regards, Erik

Posted by: [Erik Smith](#) | [06/03/2015 at 09:50 AM](#)



My I give you a lossles highspeed Fist bump?
Would you mind terribly if I just attach this whole blog to my next FC SAN RFP?

Posted by: Gary Olson | [06/08/2015 at 04:03 PM](#)



Hey Gary, absolutely! Please feel free to use it to advance the cause..

FC Tree Huggers of the world, unite! :)

Posted by: [Erik Smith](#) | [06/08/2015 at 08:49 PM](#)



Erik,

you made good points, but still just need to look at the numbers compare the growth in FC attached storage to Cloud, Hyper-converged, NAS, Object, Hadoop, .. and all the rest of the non FC storage

if FC is so great why don't Amazon, Google, Azure use it ? why do we see Nutanix and other options (including 2 competing EMC hyper-converged solution) growing so fast ? why isn't Oracle Exadata or Teradata built on FC ?

I think the traditional (FC SAN) storage camp will be challenged to provide the same cost effectiveness and usability (and performance) provided by all those IP based solutions, they will find ways to provide high-availability and resiliency without LUNs, WWNs and Zones, having 40Gb or 100Gb is better than having 16Gb with congestion control, and SDN will provide pretty nice isolation and QoS.

talking to FC SAN users in an EMC show, seems a little biased POV IMO, they are not likely to adopt the new technologies, but other groups in the organisation will build their fast growing VDI, BigData, .. solutions without turning to the SAN guys for help, or consume pre-integrated "SANless" appliances, some Application guys may even take out the credit card and use AWS services for non sensitive data without the need to wait for IT to build it all, especially when they can get it at fraction of the cost.

BTW i have a blog post on that topic:

<http://sdsblog.com/2014/11/18/why-the-heck-do-we-need-san/>

Regards, Yaron

Posted by: [Yaron Haviv](#) | [06/13/2015 at 07:18 PM](#)



Hey Erik, great post, as usual. Another thing to consider in Ethernet based storage, iSCSI and NAS in particular is that while TCP will recover the corrupted frame, in the case of high throughput streams, this will invariably pile up additional losses and delays as well. These will eventually affect the SCSI operations at some level. I am reminded of a paper Mikkell Hagen wrote at UNH regarding TCP resends in iSCSI networks, where his experiments showed a lot of BW being used, but a full third of it was resends from dropped packets. (M. Hagen and E. Varki, "iSCSI on a converged data center network," in 22nd International Conference on Parallel and Distributed Computing and Communication Systems (PDCCS), pp. 97-102, ISCA, 2009.) - Sorry can't find a link to full text.

His dissertation discussed the issue as well, but not as specifically as in the targeted paper:

<https://www.iol.unh.edu/sites/default/files/knowledgebase/dcb/ml/dissertation.pdf>

Regardless, FEC would be a welcome addition to the Ethernet toolbox, that's for sure.

Posted by: Bob Smith | [06/17/2015 at 11:44 AM](#)



Hi Yaron, thanks for taking the time to comment. I actually agree with most of your points as they apply to non-traditional (e.g., Platform 3) applications. I'll stand by the statements I made in the post as they apply to traditional apps for the reasons I provided.

Regards, Erik

Posted by: [Erik Smith](#) | [06/17/2015 at 03:03 PM](#)



Hi Bob, thanks for reading and commenting! I'm familiar with some of Mikkell's excellent work and enjoyed his dissertation.

It sounds like you're describing the difference between throughput and goodput and the approaches we can use to maximize the efficient use of the network.

In any case, great point about the impact of loss on TCP and how it can have secondary effects.

Posted by: [Erik Smith](#) | [06/17/2015 at 03:16 PM](#)

The comments to this entry are closed.