

Meeting Minutes from Joint Techs 2-Feb-09 in College Station

Meeting Minutes from Joint Techs 2-Feb-09

Attendees:

Carla Hunt, MCNC (Chair)
Tom Throckmorton, MCNC
Greg Henkle
John ?
Eric Boyd, Internet2
Jennifer Schopf
Brian Tierney, ESnet
Ken Lindahl, UC Berkeley
Russ Hobby, Internet2
Someone from APAN
Grant Miller
Azher Mughai, Cal Tech
Mike Van Norman, UCLA
Sandor Rozsa, CERN
Matt Zekauskus, Internet2
Rich Carlson, Internet2
Jeff Boote, Internet2
Aaron Brown, Internet2 (scribe)
Jason Zurawski, Internet2
Chris Robb, Internet2
Emily Eisbruch, Internet2

Discussion

Before getting into the presentations, Carla discussed modifying the working group structure. Originally, we had split into a group of task forces that were each tasked with looking into best-common practices. However, it was decided that it would be better to have the group work on all of the topics than to break up into the task forces. She also asked whether meetings should happen on a monthly or biweekly basis, but no consensus was reached.

Measurement Lab:

Richard Carlson gave a presentation about measurement lab which is a consortium of parties interested in monitoring commodity networks. The model is more like IETF where it's a group of individuals, instead of a corporate project. Measurement Lab started six months ago at a workshop in CA. The topic of the workshop was the Network Neutrality issue. The idea that came from it was that if consumers were given accurate information about how an ISP is shaping their traffic, they will make better decisions in their ISP selection. The problem statement that was developed for this group shares significant overlap with problems encountered in R+E networks. The biggest issue is that diagnosing performance problems in these environments is difficult. The problem could be host configuration, application choice, or infrastructure anywhere in the network between the end hosts. As to application choice, Rich gave the example of SCP which has internal buffers that limit its performance even if the host and the network are properly configured and working.

As Rich pointed out, there are solutions to these problems, but it's difficult to get them out to researchers, let alone getting them out to broadband and commodity users. The difficulty does not stop there. Even if the user knows of a tool's existence, some of the tools are difficult to use and can require specialized expertise to understand the results. If Measurement Lab waited until the tools were usable by the average person, it would be years off before they'd be able to make anything available. The idea was to make some tools available, and work to improve them over time. Meanwhile, if users are testing, the system saves the test results, so the tests may be useful, if not immediately for the user. Joe asked if that data was anonymized. Rich said that it wasn't, but that a user had to click a start on a page that makes it apparent that their IP will be made available.

The initial implementation plan is to deploy three servers, Dell 2950s, at 12 of Google's POPs. These machines will have a gigabit connection to the outside world. The tools deployed are NDT, NPAD and Glastnost, a Bit-Torrent Degradation Detector. Currently, there are three servers deployed at one POP, Mountain View, CA. The next one to come online will be in Dallas with the rest coming online later. All of the POPs are in the US except one which is in the EU.

The hope is to bring various developers on-board. They'd like to bring on tool developers who can provide new tools and improve the existing ones to make it easy for end users to use, and provide more in-depth information about the network. They'd also like analysis engine developers so that they can provide analyzers for the data being collected so that this information can be better used. If anyone is interested in helping the project, they can click the "Get Involved" link and the consortium will get back to them on how to proceed.

Tom Throckmorton asked which virtual machine was being used on the machines. Rich said that they were using Planetlab software which uses VServer. The folks at Princeton are maintaining the machines. They can create "slices" on the machines, and then give Rich a login. Carla asked if these slices have the tools on them. Rich said that it starts as a blank linux machine with web100, and then he and Matt Matthis make sure all the packages are installed. They are working on scripts to make it easy to duplicate the tools on each machine, and said that having RPMs for the tools would help. Matt Zekauskus asked whether there was one slice per machine. Rich said that each physical node has three slices, each running their own server, and sharing the GigE link. Matt was worried that the servers might conflict and produce the bad measurement performance numbers seen on Planet Lab. Rich said that to keep that from being the case, they were limiting the number of tests, and even if that did occur, NDT and NPAD can tell if the server is the bottleneck, and flag the data appropriately.

John asked what the plan was for expanding the resources. Rich said that there were two ways to go about it. First, Google has agreed to put up 36 machines in 12 POPs where Google already has a presence. These machines will have a 1 Gigabit Ethernet interface to the outside. The other option is to have other groups donate resources. The Measurement Lab folks will make available a set of requirements on machines and network connectivity, and, if the donated machines meet those requirements and the institution is okay with Measurement Lab running those machines, Measurement Lab will bring them online.

Tom Throckmorton asked about the traffic graph Rich showed, and what the latent interest was, and what the scalability options were. Rich said the graphs were made with MRTG on the switch feeding the cluster. The missing pieces on the graph were where the servers crashed. After restarting, its taking two minutes to fill up the NDT queues. He noticed that people were becoming frustrated and leaving because of how long the queue was. Since these slots weren't reaped, it became a self-perpetuating cycle where a user would get queued, leave, and cause future users to be further back. So there are things they can do in the software itself. Also there are other options that they may use to load balance. They may also start running multiple tests simultaneously since most users will likely be running broadband, and the servers are running GigE.

Joe asked whether there were usage limits, or if he could setup crontab to check his connection to twelve places on the Internet every 5 minutes. Rich said he could, but he'd be competing with a large number of other users, which creates a natural limit. Rich hadn't been thinking about making this available, but there have been some requests to allow a policy based limits for NDT. Though, whether these limits get enabled would need to be discussed by the entire Measurement Lab group. Jeff said that the plan was to make some of the NDT tests available in BWCTL, which would let one use the BWCTL policy limits to limit users.

Matt asked whether or not a campus would be able to use Measurement Lab to test their commodity links. Rich said that the servers are just machine available on the public Internet so it shouldn't be an issue.

Eric asked if a CIO would be able to say "i'd like a Measurement Lab on my campus, for the campus". Rich said that the tools are all available, and could be stood up. The easiest way is the Performance Node ISO, but a suite of tools might be provided to easily allow standing up a custom Measurement Lab node.

Eric asked what the plan was to make the data available using perfSONAR. Rich said that they were going to work with the perfSONAR community to develop the right plans. When there are the mechanisms available to store the generated data in an MA, they'll be able to make that data available.

Carla asked whether they'd had any insights behind the choice of tools. Rich said that that the tools select were what the researchers promoting Measurement Lab decided should be on there. Jeff asked if there was a process for getting new tools on there. Rich said that its available on the webpage under Getting Involved for Reseachers. What they were looking to avoid was twelve tools that measured throughput. He thought OWAMP might be a good candidate since there's no way to measure one-way latency. He said the community decides which tools are on there, so an argument needs to be made as to why it's useful. It would also depend on how well supported the tool is and how easily it is for a commodity end user to use it. Since Internet2 is supporting OWAMP, and there's a Java applet for it, it might be a reasonable tool to include.

Carla asked how the tools were being registered in the Lookup Service. Rich said that he was using a script written by Aaron Brown to register them. The script would periodically poll the service to see if it was running, and refresh the LS registration. Aaron said that he has an RPM, and that the software might be made available during the next formal release of the perfSONAR-PS software. Jeff said that they were still looking at the best way to do the registration, whether to have a daemon register on behalf of the service or having the service register itself. Rich pointed out an issue with the current NPAD/NDT registration where the web frontend is being registered, but not the test frontend, which makes it difficult to write automated tools.

Multi-Vendor 10 Gigabit Testing:

Matt Zekauskus and Tom Throk Morton talked about some testing that they've been doing for the last year and a half with some folks in EU. They've been testing how well 1G and higher speed circuits work between differing vendor hardware over long distance. They've been testing an Alcatel in England, an HDXc and an OME in MANLAN, and a CoreDirector in New York.

For the 1G testing, using Smartbits hardware, they saw loss at various packet sizes, though not the smallest nor the biggest. The cause was a Ciena GFP processing issue which has been fixed.

For the 10G testing, they attached PCs with 10G Myricom cards to the CoreDirector and the Alcatel. When doing this testing, they saw highly asymmetric bandwidth. The cause was that Ciena didn't support PAUSE frames, which is an Ethernet flow control feature. Ciena will be adding support this feature which should mitigate these issues.

Greg asked if this meant that there's not been much 10GigE interoperability testing. Matt said that there wasn't much testing being done. Most of the testing between networks has been single vendor equipment.

Measurement Update:

Jeff gave an update of what they've been working on, and what their priorities are. They've released some software recently. The perfSONAR-PS Lookup Service has been deployed at a number of locations. BWCTL and OWAMP saw updated releases recently. perfSONAR-BUOY had an release for the MDM appliance, but they're working on making some nicer packaging for a wider release. In general, the future plan is to increase the usability, performance and stability of the current tools.

The plan for the Performance Node is to do one more release of the Knoppix-based Toolkit with some feature updates and update kernel to support more hardware. The release after that will be using a Fedora Core based LiveCD. The reason for the change is to make it easier to share the development work, making it easier for community involvement. The other plan is to make it easier to support working with a group of nodes. Currently, the administrative GUIs for the Performance Node are for a single node, but since people are going to want to deploy several hosts, the plan is to make it easy to do that.

Another focus will be on GUIs for the data already out there. Currently, the focus has been on getting the middleware working and deployed, and the GUIs have reflected that focus. Now that there are enough deployments and data, they are going to start writing GUIs for end users.

DCN integration is important for Internet2. The current DCN software suite supports a Notification Broker so that interested parties can be informed when events are occurring on the hardware. The plan is to have the measurement software listen to these events, and monitor the circuits when they come up. Once that is working, they are going to provide an analysis GUI to show this information.

There was a discussion about getting end sites using the tools, and how the Performance Node might fit into that. There could be an appliance, where individuals buy hardware containing the Performance Node which they can plug in. Jeff said this wouldn't make sense for MCNC who are already deploying the tools, but it might make sense for other groups who don't have time to deploy their own. Tom commented that that described their k-12/higher education customers. Some will be able to setup the tools on their own, and others will need to be walked through completely.

Dan asked what Jeff meant by "appliance"? Jeff said it'd be something they build and possibly manage. This is the model the EU folks went, and it seems useful to certain folks. The question is what people want?

Tom said that MCNC was providing the whole spectrum, 1. source code, 2. packages, 3. a VMWare image, 4. an ISO.

Dan said that he's had difficulty getting stuff out to the campus edges. NIH had a serious problem a year ago, and said they were going to put in measurement gear, but they haven't done anything. If they could order something and drop it on their edge, they'd do that. Dan figured the appliance being administered by an outside organization would probably be a non-starter for them. Jeff said that people put remotely controlled video conferencing devices in their networks regularly. Tom said that the big difference is that the vendor is willing to bear liability for the device.

John said that something like an Ethernet Termination Device might be a reasonable compromise. It's a box that sits at the edge of a carrier network. It doesn't look like a server, though someone from Internet2 might have access to it, but it doesn't get you deep into campus. Jeff thought that having boxes at the demarcation points might be good enough. Dan said he'd kicked around the idea of deploying something like that at the edges for customers' networks so that it would be easier to figure out whose fault an issue is.

There seemed to be a consensus that an appliance would be good as it'd give them something they could point their customers at. The only big issue is support. Jeff was wary of becoming another Arbor Networks. This, however, is an ongoing conversation.