2014-08-28 CENIC and Research Opportunities

Background material for review prior to the meeting:

>>> To understand the extent of CENIC's role in network access in California, please have a look at the Needs Assessment and Spending Plan for High Speed Broadband in California Libraries (attached as PDF). Cliff recommends the data analysis sections; Steve recommends the Introduction as well.

>>> For an incisive, thoughtful, and accessible overview of the significance of Software Defined Networking, watch a 30 minute presentation by UCB-EECS Professor Scott Shenker, here on YouTube. Prof. Shenker's presentation runs from 2m43s to ~30m, but the questions that follow are quite interesting as well. The video is imperfect; slides are available on-line as a PDF.

Warren Hall access: For those who do not have keycard access to the building, please take the elevator to the second floor (stairwell door requires keycard). Before noon, let the receptionist know you're joining the Reading Group in 200C and s/he will let you in and show you the way. After noon, look for a sign next to the receptionist window to the right as you exit the elevators. We'll post a note with a phone number that you can call or text, and someone will come out to open the locked doors.

From e-mail to the RIT Reading Group listserve:
Our next Research IT Reading Group topic will be: CENIC & Research Opportunities -- From Libraries, K12s, Cultural Institutions to OpenFlow SDN. [CENIC is the Corporation for Education Network Initiatives in California; SDN is Software Defined Networking.]

We'll meet Thursday 28 Aug 2014, from noon - 1pm at 200C Warren Hall (see access instructions below).

Discussion will be facilitated by Cliff Frost, Senior Advisor for Cloud Services at CENIC, and former director of IST - Infrastructure Services.

Cliff writes: CENIC is very interested in helping researchers make the most of the high-speed network and its rich connectivity. I plan to present an overview of CENIC with an emphasis on new developments that may be of interest to researchers. [...] I plan to spend at least half the time discussing Software Defined Networking (SDN) in general and OpenFlow in particular—aiming at knowledgeable IT professionals who are not network geeks (although network geeks are more than welcome!). There is an awful lot of hype and confusion about SDN and I hope to leave people with a reasonable framework for thinking about it.

Attending:

Cliff Frost
Aaron Culich
Chris Hoffman
David Fulmer
David Greenbaum
Erik McCroskey
James McCarthy
John Lowe
Patrick Schmitz
Quinn Dombrowski
Rick Jaffe
Steve Masover

Cliff's slides as PDF: CENIC_CliffFrost_SDNinContextv2.pdf

CENIC & Research Opportunities — Cliff Frost
Framing question: looking at fundamental infrastructure resources, how do we hook up researchers who could benefit from those
Consortium started by UC, CSU, Caltech, Stanford, USC in mid-90's
NSF stopped supporting regional networks, internet went commercial
University costs skyrocketed, lost control over design and policy
Bay Area Regional Research Network — owned by Stanford, they sold it and were happy
CENIC built CalREN — California Research & Education Network; single network across the state
First in the country to happen that way
Immediately paid off; lower costs, focusing on community needs
No separate org; volunteers from campuses built first network
Designed 3-tier network: regular traffic, high performance research network (around the same time, Internet2 being formed), experimental
Internet2 has remained research universities; CENIC expanded
Experimental network was mostly theory at first; now there’s infrastructure, but it’s pretty light
Incorporated as a non-profit in late 1990s
Added community colleges, K12s
Public libraries join this year
Reorganization of how state libraries do businesses internally; this is being reorganized first, legislature put out couple million to cover it, but other people will want it if not used
1200 public libraries in the state, 183 jurisdictions for libraries (largest is LA public library, 73 branches; half are 1-2 branches)
Potential research opportunity, educational opportunity, public service, curriculum
Aspirational goal of getting at least a gigabit in every library in the state; requires fiber to them, can be upgraded more easily
Could be opportunities for working with public libraries that weren’t there before; already deeply involved w/ K-12/community colleges, less so with research university
Scientific and cultural organizations — Exploratorium, SFJazz, more coming (up to 10-15)
SFJazz— something like the symphony, has beautiful new building in downtown SF
Talking to symphony, opera, ballet
Hooking up Balboa Park — already have fiber ring around it
Not really members, don’t have board members, etc. — no org can represent all these organizations
Like the “museum problem” on campus— going after same pot of money, same benefactors
Hard to organize them
Board wanted to add a few, see if they add to the mix
All public ed in state, a lot of private ed on board
What could come out of connecting symphonies, museums, to network and having high bandwidth available between institutions
K-12 is ahead; have strong relationships with symphonies, opera, ballet
SFJazz sponsored by Stanford
Performances, master classes are obvious, must be some researchers who have interest in this, or curriculum development
Not paying — just covering cost of connecting
Couldn’t do it at scale, not paying overhead costs so it wouldn’t scale up
If 3-5 years go by and there’s no interesting activity, won’t expand the program
Experimental network is expensive — use supported by grant funding
$26 million going to last-mile connectivity problems
Lots of schools can’t do online assessment because they’re too remote, never invested in infrastructure on site
Out of 300 schools that need help, not sure who’ll qualify for $26 Million
$26 M won’t be enough; this is a big problem, especially in remote areas
People line up outside libraries — if you want a job, you need to submit resume online
Researchers are few, hard to reach
Collaboration is about relationships; labor intensive, one individual at a time
CENIC wants to help you reach beyond the usual suspects
Can CENIC start small grant program?
- No, have no money of our own
They could give us more for us as a grant program, but no one’s doing that
That’s not our strength anyway
Do get grants from NSF to do things
SDN & OpenFlow
- Do have research network that’s an OpenFlow testbed network
Berkeley & Stanford faculty created OpenFlow to begin with
Disappointed that there’s not more OpenFlow going on between Berkeley and Stanford, but they haven’t needed it (that we know of)
There’s a definition of software-defined networking, involving separation of “control plane” from “data plane”
All the words mean something, you might think you know what it means, used loosely/colloquially
Dictionary definition isn’t that important, want to explain some of other stuff going on
Heart of current notions of SDN: fast, cheap packet-forwarding
Network boxes receive and forward packets
TCAM - content addressable memory, very fast forwarding table
Get a packet, look up something in the packet by looking up the content
TCAM holds fast forwarding table
What fills TCAM in the past is routing protocol
Because of how networking is developed, Develop cheap box, write software for it SDN evolved this way — people noticed we can do things with cheap hardware
No one thought about that abstraction in vacuum
Flavors of SDN: Virtualized data centers, enterprise network management, programmable networking for applications, fast, cheap programmable packet-forwarding (where OpenFlow works best today)
VMs — sharing hardware with other computers, but your own personal machine
Want VMs to be able to jump between pieces of hardware
All works great, but networking gets complicated
Current example: VMware flavors
Puppet & Chef solve a pile of problems, but not the networking part SDN lets you fit the networking part in the category of things Puppet & Chef solve for you
Doing it mediocre — it works, but not well enough
VMware makes it easy — you don’t have to know about what you’re doing (until it stops working, and people pass the responsibility around in a bad way)
Enterprise network management — 10k boxes that are part of the network, have routers, switches, NAT boxes, firewalls, load balancers, all
different generations, manufacturers, software, no idea what’s out there
Makes people very nervous, creates problems that you don’t catch for years
People figure out workarounds to network problems
Holy grail of network stuff: “let’s have a single management view of the whole thing”
Define network ahead of time, impose it on all the boxes
HBOpen View and other such tools have been around for a while; try to impose this on standard networking boxes; also have Nagios, RT, little
pieces of software that try to glue things together and watch over everything
One thing people like about separating forwarding layer from decision-making/control layer: getting rid of some routing retools, imposing definition
on top of boxes, big motivation for SDN
OpenFlow — widely used, Google’s own network (worldwide network) is running SDN, probably OpenFlow or close to it
A reason that works well for them is that they define the whole space
In Google’s network, if a packet doesn’t match the table, they throw it away
They know ahead of time all the valid addresses
Adds simplicity to management, makes things more secure
Orchestration part is easy if you know everything
SDN so application can request services from network — most famous example, Oscars system at LBNL
Program going to do something that needs bandwidth asks for a reservation of bandwidth, network comes back and says yes/no; if no, ask again
Send LHC data, know it’ll go in a reasonable amount of time because there’s a reasonable path
Could also ask for low latency path, predictable getting rid of jitter, can ask for different qualities from the network
Northbound API — control plane, data plane, between the two is southbound API (trying to standardize here)
Northbound API: applications can talk to controllers
Northbound API is a research project, controllers are a research project
Fast, cheap (well, not quite cheap yet), programmable packet forwarding
OpenFlow SDN
- Where the research and education action is
Everyone now has to say they support OpenFlow; question is which version (1.3 - 1.5)
Could throw things away; gets passed up to the control plane (over the network, to the box running controller software; should controller software
be distributed around the boxes?)
Controller software figures out what should happen to the packet, passes it
Many flavors of controllers; OpenDaylight seems hottest now
Each controller written for a specific kind of thing/application
Controller software is half-baked at best; research opportunities in that
Lots of vendors are still doing 1.0, not as usable
Network functions virtualization
— New thing, less baked than SDN (itself not fully baked)
Clearly aimed at Management and Orchestration (MANO); Puppet/Chef kind of thing
Focusing on functions (e.g. firewalls, load balancing)
Also aiming at applications programming the network
Nothing about it that rules out SDN
White paper that open network foundation put out about SDN & OpenFlow — “oh yeah, they’re compatible”
Easy to say, network functions are so ill-defined
National infrastructure?
Internet-scale SDN; real research & education opportunity
GENIE project; networking geeks know; computing infrastructure, more than just network; is using OpenFlow for machines to talk to each other
Packet comes in, send unknown packet to controller, “do you want to pay me for this?” — e.g. data roaming in Mexico
“Want a fat pipe? Do you have a fat wallet?”
Authentication, access, authorization, IDM, etc.
Requesting a gigabit pipe — putting demand on resources at someone else’s institution; no one knows how to do that right now
Conversations about how to get things to talk to each other across the network — research opportunity
How to ask for something even if people are willing to share? How do you know you can route between sites using protocols?
Leads to software-defined exchanges (SDXes); applied for NSF grant to create one in California (but no one knows how it’ll happen)
Ideas is to build infrastructure to allow OpenFlow clouds to talk to one another
Economics of SDN; will get away from paying as much for boxes, software won’t be as important
Off-the-shelf components, boxes that have everything; some flavor of open source software to control them, can build a whole network and
control it with open source controller

Software defined exchanges to provide infrastructure for defining and requesting network resources that cross resources owned by multiple
parties.
The “baked” use cases are networking within a set of resources owned by the entity controlling and orchestrating it.
Less baked: crossing through resources owned by multiple parties/domains.
Aaron: how does OpenFlow / SDN intersect with cloud vendor space
Cliff: don't know of cloud vendors using openflow.

Cliff: CENIC has 40GB connectivity into Google. And similar across other popular vendors. 20% of campus traffic is to Netflix (think dorms).

John: OpenFlow and NetNeutrality?
Cliff: Not baked enough to be on policymakers' radar
Erik: Different layer. Political issue as opposed to OpenFlow at a technology level.
Patrick: But political level is predicated on capability.
Cliff: Economics lifts the question into the realm of political debate / consideration.

Patrick: Impact of VMs in data center, associated challenges. Does growth of Docker and containers -- and how they map -- as a problem
Erik: Docker's implementation of networking is a disaster. If Docker uses IP address of host, network doesn't see anything extra. If Docker gets an IP address it becomes more complicated. Docker doesn't keep an IP address, container can use whatever IP address is available. If application in Docker wants the IP address to follow it, then there's a problem.
AC: Open VSwitch -- thinking through the networking component of containerized software

Cliff: I think of Docker as a way to standardizing one's virtual machine. Mobility between VM providers in the cloud.

Steve: Where would CENIC want to become involved in a research project? How well developed? How well-developed a relationship between researcher and data source (where data is what needs to be transported via a fat pipe).
Cliff: Well-developed is fine. But we're happy to become involved earlier as well.

David Greenbaum: Larry Smarr (UCSD). CIOs across UCs looking at IT support for research across the campuses. Perhaps good for CENIC to become involved in that, to give a similar talk to this.
Cliff: Yes. Getting involved with Stanford as well. Larry is one of "the usual suspects" -- people who are already well involved with CENIC.

Patrick: We need some communication materials that explain to researchers what CENIC is, what Science DMZ is, what it can do for them and what it can't do. This is the only way to scale engagement with researchers, can't bring Erik to every engagement, for example.

David Fulmer: Researchers will become confused over what's meant by "speed" -- the difference, for example between raw speed capability increases (fatter pipe), and increased efficiency using the pipe that's already in place.

Patrick: Would like to distill technology into research use cases, work with Erik et al. to do that.

[...]