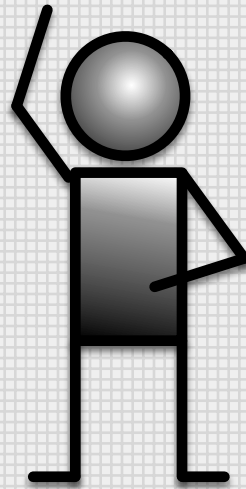


Hello  
Welcome to



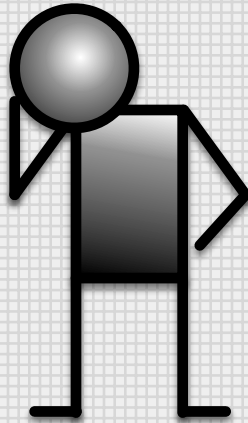
RSIP Alternatives for Cray Compute Node  
External Network Access

Jeremey Enos

National Center for Supercomputing Applications

Problem:

With 27k compute nodes, RSIP doesn't scale to meet applications needs for network communications outside of the CRAY HSN



Looked at a half-dozen different ways\* to provide external connectivity to the compute nodes.

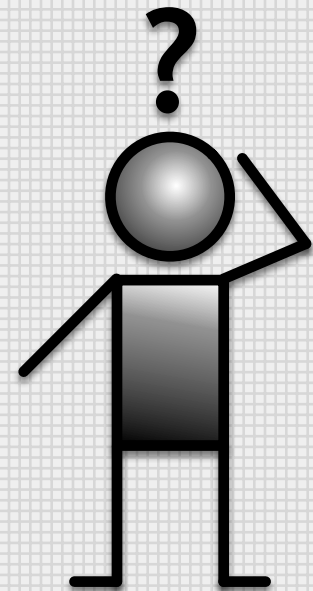
How to choose the right one?

Port Availability/Scalability

Performance

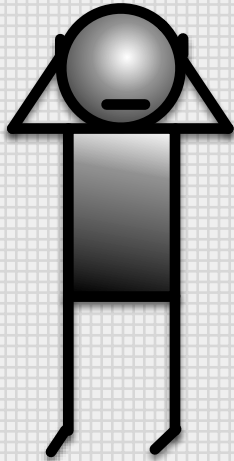
Security

Resiliency



\*Read the paper if you want more information!

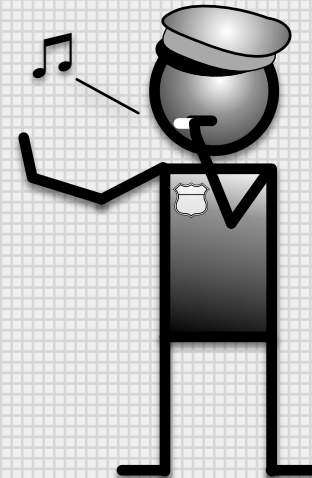
# NAT to the Rescue?



NAT is a Frienemy technology

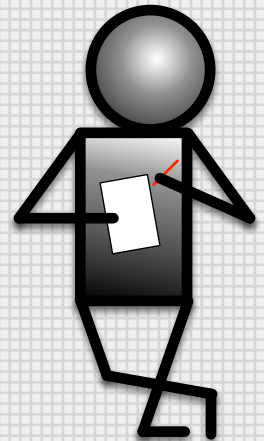
Provides an 'easy out' to many common networking problems

But at the expense of breaking the 'end-to-end' connectivity paradigm of the Internet



# Our solution

1. Replace RSIP with NAT on the service nodes using  
iptables (provides firewall and NAT services)  
contrackd (userspace daemon provides xml state info from iptables)
2. Run a custom script 'the\_keymaster.sh' on the boot node to make initial compute node to NAT GW assignments.
3. GW\_Tracker process running on the boot node monitors the availability and state of the NAT gateways.
  - + remaps compute nodes to remaining NAT GW's if a NAT GW dies
  - + remaps compute nodes to new NAT GW if a NAT GW is running out of resources
4. Use 'Firewall Builder' to keep configurations in sync on the NAT GW's



# The Win

We gain many more ports per compute - limited only by the NAT IP address pool

We utilize all available ports and IP's by dynamic remapping so we are not wasting resources

We gain reliability through dynamic re-mapping if a gateway node dies at the cost of session state

We gain security through implementing firewall rules

It is still to be determined if we gain performance. We can add more NAT GW's to deliver more performance as needed. Stay tuned.

