Setting Up a Beowulf Cluster

Caleb Kates, HPC Meeting September 18th

Bible Verse of the Day

I am not ashamed of the gospel, because it is the power of God for the salvation of everyone who believes. -

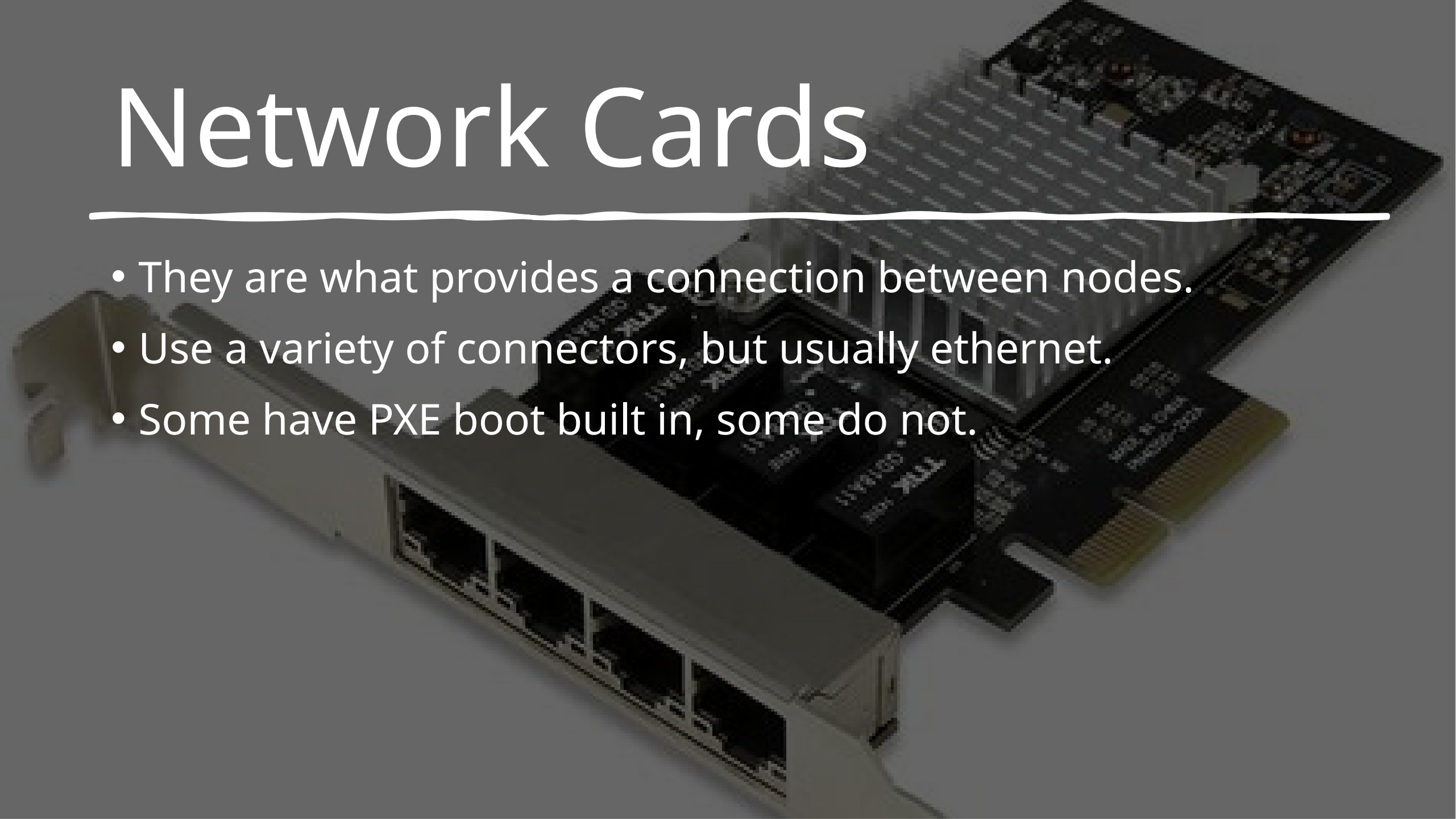
Romans 1:16

Parts involved in a Beowulf Cluster

- PXE
- DHCP (Dynamic Host Configuration Protocol)
- TFTP (Trivial File Transfer Protocol)
- NFS (Network File Sharing)
- MPI (Message Passing Interface)

Network Cards

- They are what provides a connection between nodes.
- Use a variety of connectors, but usually ethernet.
- Some have PXE boot built in, some do not.



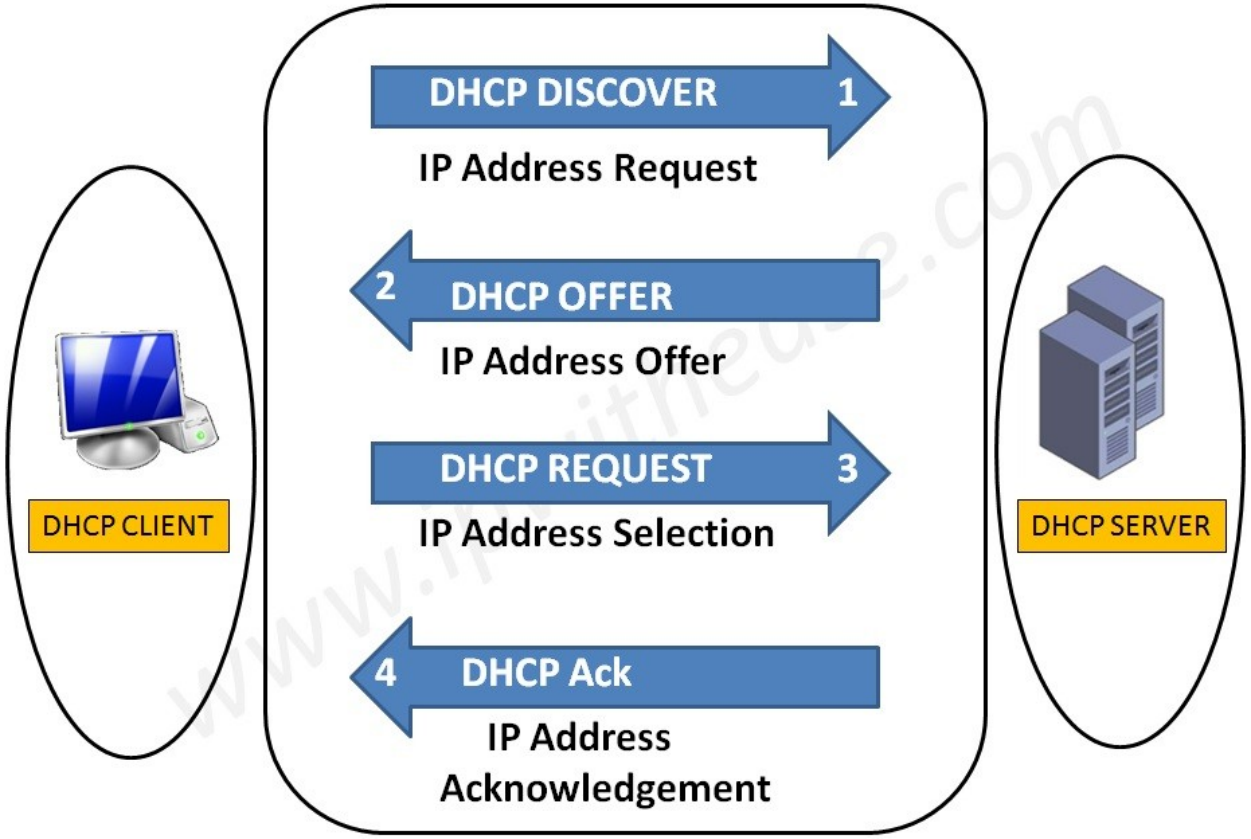
What is PXE?

- “PXE is a standardized client-server-environment that boots software retrieved from a network device, for a client with a PXE-enabled NIC.”
- With this, you can remote boot into a server, remote install an O.S. onto a drive, and even remote boot install an O.S. into RAM!
- Uses TFTP (Trivial File Transfer Protocol) in this case.

Has nothing to do with pirates, but it makes me think of them. Probably because of Peter Pan.

Now, what is DHCP?

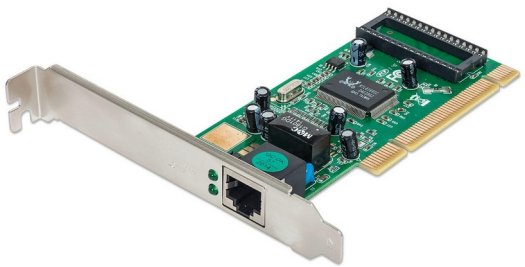
- For this context, it’s a tool for assigning I.P. addresses automatically.
- Used **everywhere.**
- The IPs are normally leased, but you can program a static IP if you want one to stay the same.
- Not all DHCP devices support PXE, but you can fix this with ProxyDHCP.



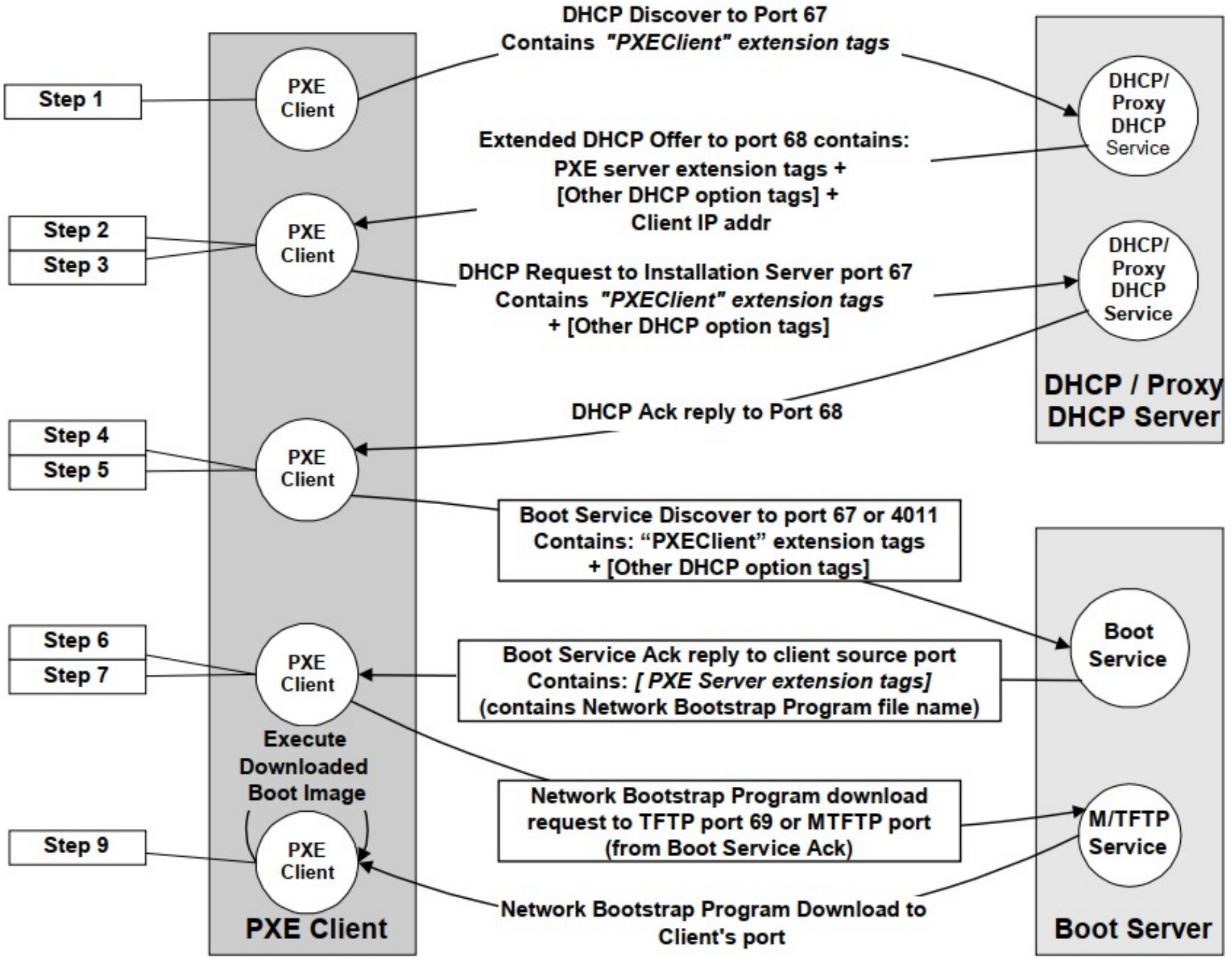
TFTP (Trivial File Transfer Protocol

- A protocol is a set of rules for formatting and processing data
- Usually used for small, “trivial” files.
- Related to the more complex FTP and SFTP protocol
- This includes firmware images (serialized firmware) and config files (“parameters and initial settings.)
- Here, you need it for the NBP file.

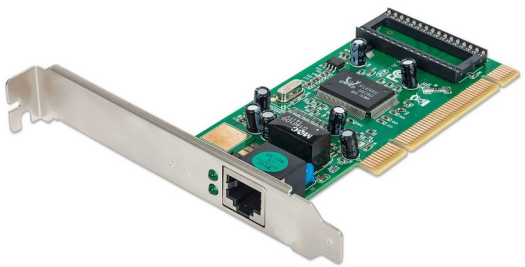
Simple!



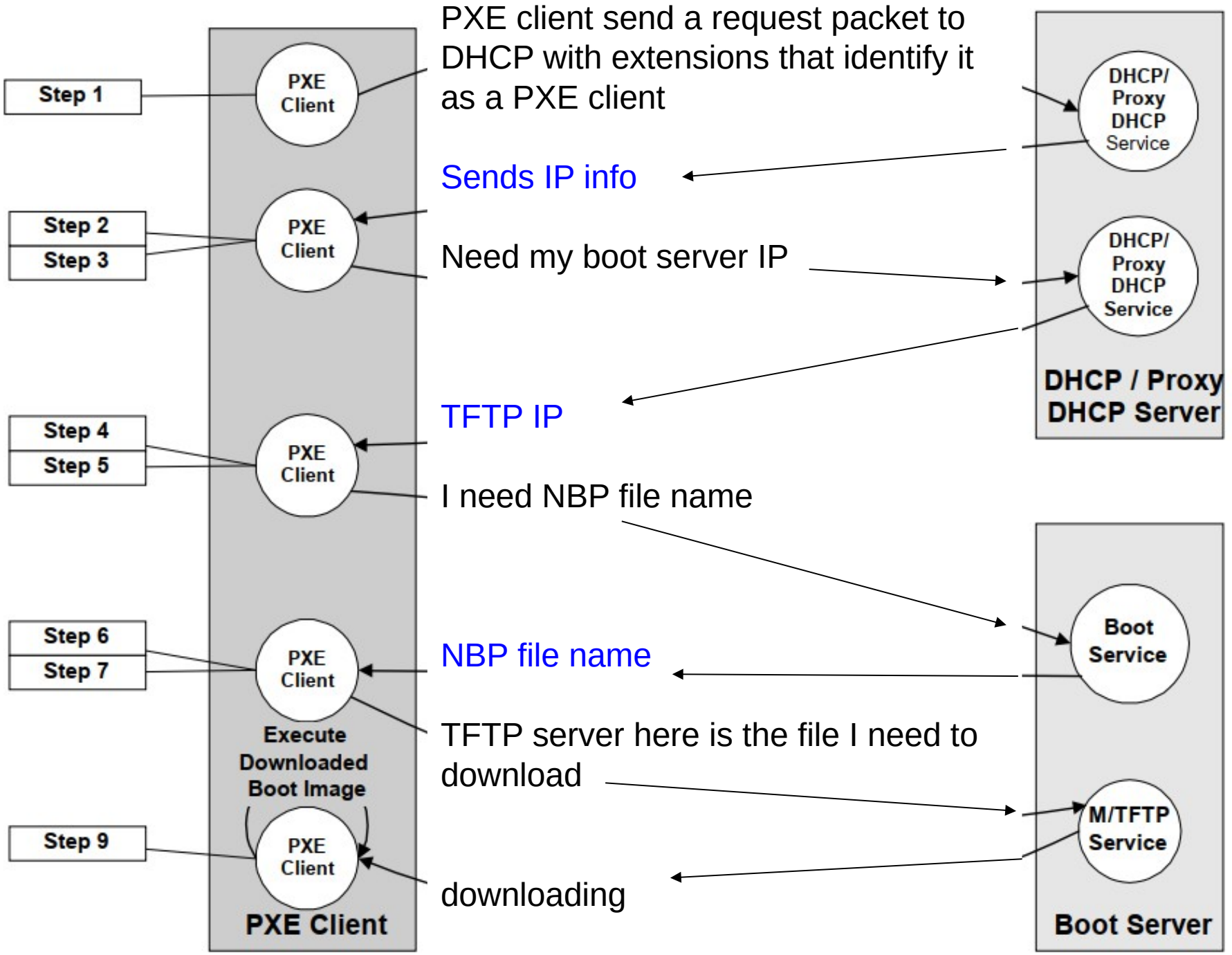
Blatant thievery!



Mr.V's version



Blatant thievery!



The Real Magic: NBP (Network Boot Protocol)

- This is the second of the domino steps to get an O.S. booted via the boot chain process.
- Not limited to O.S.s. It can be as simple as a menu, or as complex as a whole deployment system.
- “Only” used here to boot the O.S., and uses TFTP to do so.

What is chroot?

- Chroot allows us to create a file system inside a file system.
- Crucial to the system we are building.
- Chroot stands for change root, and it a root file system, which will act as our node.
- Each of our slave nodes will boot, and they will have this chroot as their environment.

How to create a chroot

- You use debootstrap on Debian based systems.
- Editing the environment inside the chroot will edit in all the nodes.
- We than make the chroot file system bootable by adding a Linux kernel.
- We then add this chroot file system to be shared via TFTP

Now, O.S. is booted

- The O.S. kernel is booted via TFTP
- But the file system is mounted via NFS

File System

- The file system is provided via NFS
- And to do this, you have to edit the /etc/fstab in order to mount the network file system.
- It runs the O.S. inside of RAM on the node! No hard drive needed.

What is NFS (Network file share)

- NFS is a basic file sharing service.
- That allows us to mount the chroot environments file system on the node as a permanent disk

At this point, it should all be booted, and you should have a working system!

- YAY!
- Now the hard part….installing MPI

What is MPI?

- Refer to the President’s presentation from last week.
- But basically, it is a way of running multi-core operations over a cluster

A bit of MPI install

- I will install MPI by downloading OpenMPI from GitHub into the chroot environment
- This is because the chroot environment is what all the nodes boot off of, and it is a sort of node.
- I will add the ssh key of the nodes once they were booted from this environment
- Then I’ll add a host file that contains the I.P.s of the nodes.
- Then I’ll test by running a hello world program

In conclusion

- PXE with TFTP
- DHCP (Dynamic Host Configuration Protocol)
- NFS (Network File Sharing)
- MPI (Message Passing Interface)

In conclusion

- PXE with TFTP
- DHCP (Dynamic Host Configuration Protocol)
- NFS (Network File Sharing)
- MPI (Message Passing Interface)

Questions?

- I may or may not be able to answer them.
- My liability is limited to an apology and a Lifesaver mint.