IE: The Interactivity Environment

Peter A. Dinda
Department of Computer Science
Northwestern University

Introduction

This document describes the Interactivity Environment (IE) implemented within Northwestern University's Department of Computer Science. It covers the hardware and software configuration of the IE, access policies, and how to use it.

The IE is funded by grants from the National Science Foundation and by investigator funds. The principle investigators for the IE are professors Peter Dinda, Ben Watson, and Brian Dennis, but many people (noted below) have contributed to its construction. ***The main purpose of the IE is to support the research efforts of the principle investigators and their students.*** We have decided to share the IE with the Northwestern CS department where possible, and will provide user accounts, managed by SSG, ***with the understanding that PI research always takes priority.***

The design of the IE has been geared toward research in interactive high performance computing, particularly interactive visualization of large datasets by combining computer graphics and distributed and parallel computing techniques. It also serves other purposes for the PIs, including functioning as a network testbed.

Description

The IE consists of data nodes that serve large volumes of data to other machines, compute nodes that provide computational resources for operating on this data, and visualization nodes that can do final rendering of this data. Final output appears on a CAVE/Active Mural display, visualization workstations, desktops, and high speed wireless devices. Data, compute, and visualization nodes are tightly coupled through a private gigabit network. A secondary 100 megabit network provides support for logins and other utility functions. Some of the compute nodes are extremely tightly coupled using an extremely low latency private barrier synchronization network. The IE connects to the building networks, including wireless, through a firewall. A rudimentary KVM system is implemented in the machine room. The figure summarizes the following discussion.

Components

The core of the IE is divided into eight "IE Components", each of which can be used autonomously. In fact, a component can easily be physically disconnected from the rest of the IE. This makes it easy to divide up the hardware resources to support conflicting experiments. A component consists of a data node and its accompanying disk array, four

compute nodes, and a visualization node. We refer to ie-component-1 through ie-component-8.

Data Nodes

A data node is based on a SuperMicro 370DLE server motherboard and contains two 800 MHz Pentium III Coppermine processors, 1 GB of RAM, and two 30 GB IDE hard drives. Each data node has two 64 bit, 66 MHz PCI slots. The first contains an Adaptec 29160 Ultra3 LVD SCSI controller that connects to an external IDE RAID box containing about 1 TB of raw disk. The RAID boxes are Caen Raptor 8 or Caen Raptor 12s fully populated with drives and 128 MB of cache. These are hardware raid machines with their own serial terminal support, pager support, dual-attach capabilities, and other toys. They appear to the data nodes as very fast giant SCSI hard disks. The second slot contains an Intel PRO/1000 T server adaptor that connects to the private gigabit network. An integrated Intel PRO/100 talks to the 100 mbit network.

Data nodes run Linux. Specifically, the base configuration is a Red Hat Linux 7.2 "Everything" install, supplemented with:

- Custom 2.4.18 kernel including drivers for Intel and Adaptec cards
- GCC 3.0.4 full installation
- Java2 HotSpot JDK 1.4.0
- Perl 5.6.1
- Python 2.2.1
- MPICH 1.2.4
- Matlab 6.1.0.450 Release 12.1 including all NWU site-licensed toolboxes

Network visible services are very stripped down, but some specialized services are enabled:

- SSH for login and remote execution using SSH1 and SSH2
- NIS for maintaining single sign-on
- NFS for maintaining single home directory
- NTP for maintaining clock synchronization
- SNMP for network measurement and administration
- DHCP for dynamic configuration of the networks

The two IDE drives internal to the data node are configured using Linux's software RAID tools. Each mounted volume (/, /home, /boot, swap) is striped (RAID 0) across the two drives to achieve high performance. There is 8 GB of swap space while /home contains about 40 GB of scratch space which can be used by anyone, and that can and will be deleted if needed for other purposes. Please note that the performance of this space comes at the cost of reliability - if either drive dies, data on both drives is lost. ***Use /home only as scratch space, not as reliable storage.***

Each data node's external RAID box is configured as RAID 5 and is quite reliable. The RAID box connected to ie-data-1-1 (alias ie-login) contains the home directory and maintenance directories for the cluster. Each node in the IE mounts these home

directories via NFS under /ie/home. The RAID boxes hooked to other data nodes are used predominantly for experimentation and require special permissions to use.

The eight data nodes are named ie-data-1-1 (alias ie-login) through ie-data-8-1, which resolve via DNS to their 100 mbit interfaces. Their gigabit connections can be accessed via ie-data-1-1g through ie-data-1-8g. Note, however, that the gigabit network is private to the IE, so those interfaces are only visible from within the IE.

Compute Nodes

Compute nodes are identical to data nodes except in the following respects:

- Dual 1 GHz Pentium III coppermine processors
- SuperMicro P3TDLE motherboards (very similar to the 370DLEs)
- Internal hard drives are /dev/hda and /dev/hdc instead of /dev/hda and /dev/hdb.
- No SCSI
- NFS client only

The 32 compute nodes are named ie-compute-1-1 through ie-compute-8-4. The first digit signifies the component, while the second gives the compute node number within the component. Use the names ie-compute-1-1g through ie-compute-8-4g to access the gigabit interfaces.

Visualization Nodes

We have not yet finalized the visualization nodes. They will probably be dual processor P4, Xeon, or Athlon MP machines with high-end consumer graphics cards. They will probably dual-boot Linux in a similar configuration to the above and Windows XP. They will connect to the gigabit and 100 mbit networks as shown in the figure.

The visualization nodes will be named ie-viz-1-1 through ie-viz-8-1. Use ie-viz-1-1g through ie-viz-8-1g for access to the gigabit interfaces.

CAVE and Active Mural

The visualization nodes drive the screens of an immersive, walk-in environment that can also be reconfigured as a giant room-size screen. More details will be provided later.

Wireless Nodes

We have not yet finalized the wireless nodes that will become a part of the IE. Theywill probably be Tablet PCs or Pocket PCs. We do provide both an 802.11a and 802.11b environment for these devices. DNS names for these resources are yet to be determined.

Networks

The gigabit network is built from unmanaged Netgear gigabit switches. The core of each IE component is a Netgear GS508T switch with a 10 gbit/s backplane. The component switches connect via a Netgear GS516T switch with a 20 gbit/s backplane. This switch can be connected to the firewall or to other lab networks such as the Plab, but it usually

remains disconnected. Each component has a free port. The top level switch has seven free ports.

The 100 mbit network is built from unmanaged Netgear and SMC switches. Each IE component contains a Netgear FS108 switch. These switches connect to an SMC EZ Switch 24 which in turn connects to the IE firewall and through it to the outside world. Each Netgear has two free ports which the SMC has 15 free ports.

IP addresses on the 100 mbit and gigabit networks are assigned via DHCP.

There are five custom-built PAPERS network boards. Each board can interconnect four machines or four sub-boards. The connection to a machine is via the parallel port. PAPERS provides 3 microsecond barrier synchronizations.

The building's 802.11b wireless network is being augmented with 802.11a 54 mbit/s access points for IE users.

Using the IE

To use the IE, you must first get an account, the process for which is described below. An IE account provides the following:

- Access to all of the IE resources
- A single sign-on to all of the machines (in Linux)
- A single home directory shared via NFS to all of the machines (in Linux)

To use your account, simply ssh (using ssh2) to any of the machines from a machine which is allowed through the firewall. For obvious reasons, we do not list those machines here, and we may change the firewall rules from time to time. If you enable X11 forwarding in your ssh client, you'll be able to export X Window displays back to your client machine. Once on an IE machine, you can ssh or otherwise use the gigabit network simply by using the "g" suffix on the machine's name.

When you log in, your home directory will be automounted via the gigabit network as /ie/home/you. You have the same home directory on each of the machines. If you really want to be on the same machine as /ie/home/you, you can log into ie-login. However, the gigabit network is actually faster than the RAID Box, so you'll gain nothing in speed by doing this. You may also write into /home on each of the machines. However, as noted above this is intended as scratch space and has no guarantees whatsoever.

You'll find a standard Red Hat 7.2 environment greeting you. Our additions are in /usr/local. The data and compute nodes are identical with respect to the user experience.

Account Policy and Reasonable Use

To acquire an account on the IE, you must convince Peter, Ben, or Brian that you need one. If we decide to give you an account, we'll ask SSG to set it up for you. The account is free provided that you understand that our research and that of our students takes

priority and preempts other users. Accounts are given to individual users. Account sharing is discouraged.
We would like to avoid spending time dealing with quotas, etc. However, if it turns out that a user's disk space, network, or CPU usage is excessive enough to interfere with our research or other research, we will ask that it be reduced, and, if that doesn't help, we'll have SSG enable quotas for that user. Any attempts to hack the system or spy on others will be result in immediate revocation of all privileges.

Mailing List

Users should subscribe to ie-list@cs.northwestern.edu. For heavy use, pseudo reservations, or hardware/software changes, users should send mail to the list.

Hardware Policy

The IE also functions as a hardware pool for research purposes. From time to time, hardware will be reconfigured for experiments (new kernels, for example), disconnected from the network (private testbed use), or temporarily moved to other venues (demos, kernel hacking). All such changes shall be made only by the PIs and their students and will generally be prefaced by email to ie-list@cs.northwestern.edu. When machines are returned to the cluster, they should be cloned back to their original state.

Disk Images, Boot CDs, and Cloning

The RAID array on ie-login (ie-data-1-1) contains a maintenance directory (/mnt/raid/maint) with utilities and disk images that can be used to return data, compute, and visualization nodes to their original states. The machines have custom boot CDs that enable remote access to this data via the gigabit network.

To clone a machine, first boot it from a custom boot CD. Once the penguin appears, type

```
linux single varmem=16000 ramdisk_size=32000
```

This will boot into single user mode with /var on a ramdisk. You may be asked for the root password if the RAID partitions on the internal drives are too damaged. Eventually, you will get to an sh prompt. At the prompt, first bring up the gigabit Ethernet card:

```
ifconfig eth1 up 10.10.10.40
```

Replace 10.10.10.40 with the IP address of the machine's gigabit card. This will automatically load the appropriate kernel modules. Next, mount the maintenance share from ie-login:

```
mount –t nfs 10.10.10.2:/mnt/raid/maint /mnt/cdrom
```

Now you can clone your machine from the maintenance directory:

```
cd /mnt/cdrom/images
```

```
./clone_me_compute
```

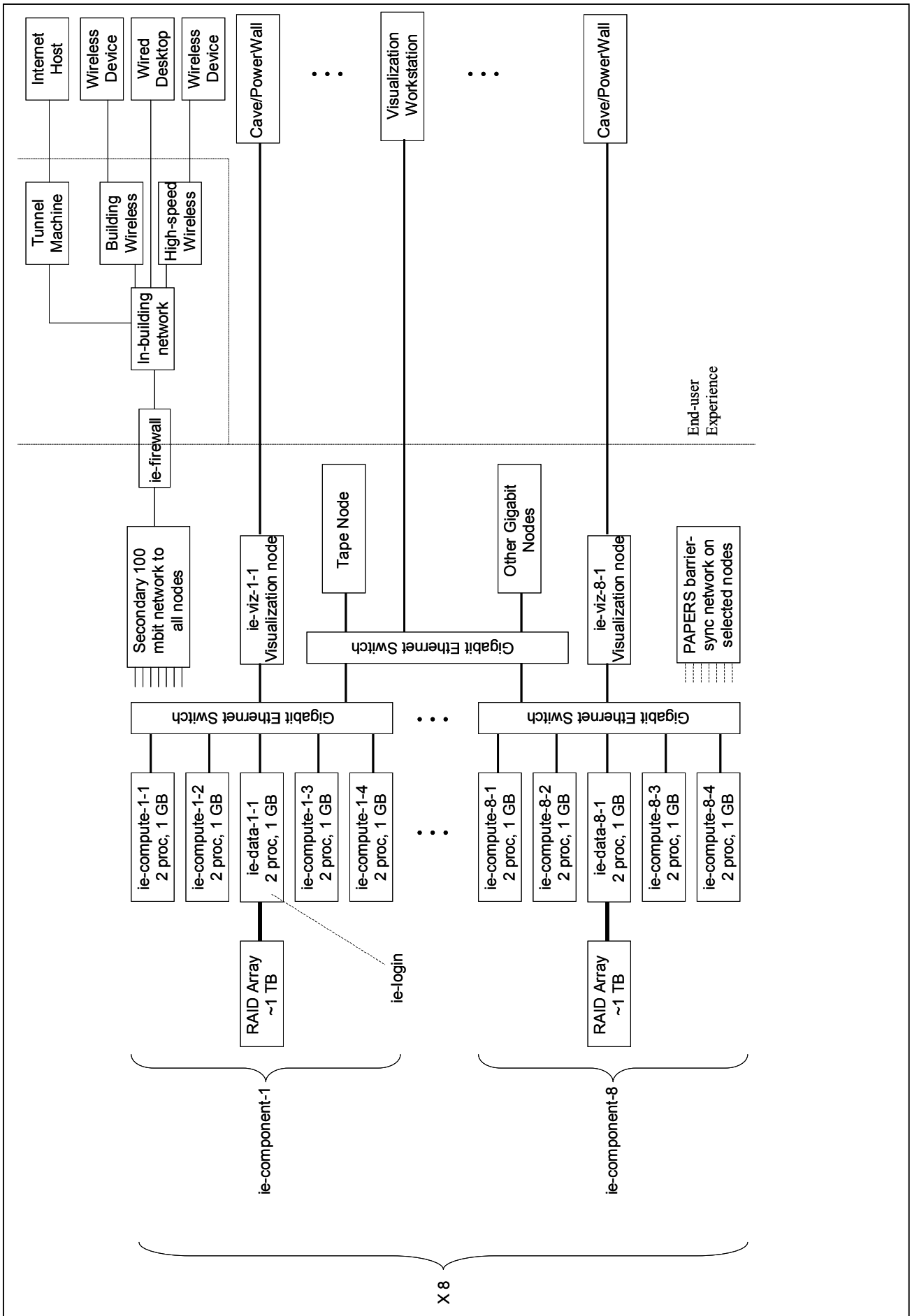
You can choose from among clone_me_data, clone_me_compute, and clone_me_viz. The cloning process will take about an hour. After you get bored with all the blinking lights, go get a cup of coffee.

After cloning, reboot the machine from the hard drive. The first boot will be slow as the NFS mounts will fail. Once you get a login prompt, log in as root and edit /etc/sysconfig/network-scripts/ifcfg-eth1 to set the gigabit card's IP address appropriately. Next, rm /etc/dhcp/* to get rid of cached DHCP information. Reboot and the machine should be back to normal.

Contributors

The following people have contributed to the IE.

Peter Dinda
Ben Watson
Brian Dennis
Jason Skicewicz
Dong Lu
Conrad Albrecht-Buehler
Mike Knop
Scott Hoover
Jack Lange
Ryan Battista
Jim Trieu
Sam Benediktson
Jameel Harris
Anshu Dabriwala
Eric Cheng



The Interactivity Environment