



Connecting to NERSC and Transferring Data

February 15, 2013

David Turner
NERSC User Services Group



National Energy Research
Scientific Computing Center



Lawrence Berkeley
National Laboratory



Connecting with SSH

- **Secure Shell (ssh)**
 - Provides *secure channel over insecure network*
 - Encrypts all communications
 - Control
 - Prevents password “sniffing”
 - Data
 - Potential performance penalty usually not a problem
 - Can forward TCP ports and X11 connections
 - Protocol version 2 since 2006



SSH Client/Server Model

- **Server**
 - sshd daemon
- **Clients**
 - ssh, scp, sftp
 - OpenSSH – Linux, Mac
 - PuTTY – Windows
 - Other clients using ssh protocols
 - Subversion (svn)
 - rsync
 - bbcp
 - sshfs



SSH Authentication Models

- **password**
 - Need “keyboard-interactive” method in client to support PAM (pluggable authentication modules) in server
- **key pairs**
 - Password is never transmitted over network
 - Private/public keys
 - Protect private key!



Key Generation

- **Generate keys with “ssh-keygen”**
 - Use “passphrase” to protect (encrypt) keys
 - Private key typically kept on desktop
 - Public key placed in `$HOME/.ssh/authorized_keys` on remote systems
 - Be careful not to introduce newlines into keys!



Key Usage

- **Start *agent* on desktop with “ssh-agent”**
 - Typically a shell or X-server
- **Store private key in memory with “ssh-add”**
 - Prompts for passphrase



Basic Key Example

```
desktop% ssh-keygen -t dsa
Generating public/private dsa key pair.
Enter file in which to save the key (/homes/dpturner/.ssh/
id_dsa): <return>
Enter passphrase (empty for no passphrase): <enter passphrase
here>
Enter same passphrase again: <enter passphrase again>
Your identification has been saved in /homes/dpturner/.ssh/
id_dsa.
Your public key has been saved in /homes/dpturner/.ssh/
id_dsa.pub.
desktop% scp .ssh/id_dsa.pub edison.nersc.gov:~/.ssh/
authorized_keys
Password: <enter password>
desktop% ssh edison.nersc.gov
Enter passphrase for key '/homes/dpturner/.ssh/id_dsa': <enter
passphrase>
edison%
```



Better Key Example

- Assume public key placed in `authorized_keys`

```
desktop% ssh-agent tcsh
```

```
desktop% ssh-add
```

```
Enter passphrase for key '/homes/dpturner/.ssh/id_dsa': <enter  
passphrase>
```

```
desktop% ssh edison.ner.sc.gov
```

```
edison% ssh hopper.ner.sc.gov
```

```
Password: <enter password>
```

```
hopper%
```



Even Better Key Example

- Use “agent forwarding”, with `.ssh/config` file
- Assume agent started on desktop, passphrase entered

```
desktop% cat .ssh/config
Host *
ForwardAgent yes
desktop% ssh edison.nersc.gov
edison% ssh hopper.nersc.gov
hopper% ssh carver.nersc.gov
carver%
```



Useful SSH Client Options

- **-Y**
 - Enable trusted X11 forwarding
- **-A**
 - Enable agent forwarding
- **-l <username>**
 - Username on remote system
 - Can also use “username@hostname”
- **-v/-vv/-vvv**
 - verbose/very verbose/very, very verbose mode
- **Most command-line options have .ssh/config versions**



Password Security

- **Must be changed every 6 months**
- **Must be *at least* 8 characters long**
- **Must contain *at least* one of each of:**
 - upper-case letter
 - lower-case letter
 - numeral
 - “special” character (! @ # \$ % ^ & *)
- **Don’t use common words, names, etc.**
- **Account locked after 5 login failures**
 - Call NERSC Account Support to reset
- **DON’T SHARE PASSWORDS!**



Password Examples

- **Good**
 - j#K01vz\$ewP@!udls
- **Bad**
 - P@ssw0rd
- **My favorite method**
 1. computer security is very important for nersc users
 2. csivifnu
 3. C\$1v1fnu



Data Movement

- **Use NGF to minimize movement and reduce duplication**
 - global home, global scratch, global project/projectb
- **Use cp on locally-mounted file systems**
 - NGF : NGF
 - NGF : Lustre (Cray scratch)
 - Lustre : Lustre
- **Use scp or bbcp for remote file systems**
 - Hopper Lustre : Edison Lustre
 - To/from external locations



Secure Copy

- Uses ssh authentication
- Good for “small” (~100s of MB)
`scp localfile user@host:remotefile`
- Watch out for:
 - “Chatty” dotfiles can cause silent failure
 - Missing “:” results in silent *local* copy
`scp mydata hopper:`
Copies mydata to home directory on Hopper
 - `scp mydata hopper`
Copies mydata to *a file named hopper*



BaBar Copy (bbcp)

- **Developed for BaBar experiment at SLAC**
- **Peer-to-peer model (not client-server)**
 - Must be installed on each end
 - Easy to build and/or install
 - Available on all NERSC systems
 - Can do third-party transfers
- **Uses ssh authentication**
- **Many tuning options**
- **Good for larger files**
- **Somewhat complicated command-line**



bbcp Syntax

```
bbcp [options] source target
bbcp [options] sfile host:tfile
bbcp [options] host:sfile tfile
bbcp [options] host:sfile host:tfile
```

- **Options**

- P *interval*

- Progress messages every *interval* seconds

- V

- Verbose output, include transfer rates

- w *size*

- Set TCP window (buffer) to *size* bytes

- s *streams*

- Use *streams* parallel data streams



Examples of bbcp

```
bbcp -P 5 -V -w 8m -s 8 -T "ssh -x -a -oFallbackToRsh=no %I -l  
%U %H /usr/common/usg/bin/bbcp" bigfile remotesystem:
```

- **More options**

-T

Target system options

-S

Source system options

-Z

Reverse protocol

```
bbcp: Accept timed out on port 5031
```

```
bbcp: Unable to allocate more than 0 of 8 data streams.
```

```
Killed by signal 15.
```



scp vs bbcp

- **From Carver to Hopper on a Thursday night**

```
time scp file.4G hopper:/scratch/scratchdirs/  
dpturner/FileXfer
```

```
time bbcp -T "ssh -x -a -oFallbackToRsh=no %I -l %U  
%H /usr/common/usg/bin/bbcp" file.4G hoppergrid:/  
scratch/scratchdirs/dpturner/FileXfer
```

scp	bbcp
1:09	0:21
1:12	0:24



Hosts

- **Use “grid” name for load-balanced hosts**
 - hoppergrid.nersc.gov
 - carvergrid.nersc.gov
 - edisongrid.nersc.gov
- **Use Data Transfer nodes for wide-area transfers**
 - dtn01.nersc.gov
 - dtn02.nersc.gov
 - dtn03.nersc.gov
 - dtn04.nersc.gov



GlobusOnline

- **Primarily web tool**
 - Based on Globus grid infrastructure
 - CLI and RESTful API
- **Drag and drop file transfer**
- **Tuned for wide area, high-performance**
- **Secure**
- **Reliable**

<http://www.globusonline.org>



More Information

<http://www.nersc.gov/users/getting-help/>

<http://www.nersc.gov/users/data-and-networking/connecting-to-nersc/>

<http://www.nersc.gov/users/accounts/user-accounts/passwords/>

<http://www.nersc.gov/users/data-and-networking/transferring-data/>

<http://www.slac.stanford.edu/~abh/bbcp/>

<http://fasterdata.es.net/fasterdata/data-transfer-tools/>

- **Use Google to locate ssh tutorials!**