Backing Up Data On the Fortress Server

Less complicated than changing the transmission on your automobile, and less painful than a punch in the face!

Brian M. Napoletano and Luis J. Villanueva-Rivera

Human and Environment Modeling and Analysis Laboratory

February 17, 2010

BEFORE MAKING BACKUPS

Step One: Request an account with RCAC

Visit http://www.rcac.purdue.edu/userinfo/accountrequest.cfm. You are requesting an account on their DXUL/Fortress system. Unless you're used to a different Linux shell, request the *bash* shell. Also include Bryan's career account name and "cc" him the email. Your career account will be used to create your account on the server, so once it is created you can access it with the same username and password you use to access any other resource in the ONEPURDUE domain.

The full path to your user space will be `/archive/fortress/home/username` where `username` is your ONEPURDUE account, but you will probably not ever need to use the full path.

Step Two: Decide what to backup

Think about which types files you want to back up and how frequently you will want to back it up. All data that you consider critical to your research should be backed up, but if it you're not changing the data regularly, then you don't need to back it up periodically (e.g. once you make a backup copy of your base maps, you probably don't need to back them up again). On the other hand, e-mails, field data, results from analyses, data used in publications, drafts of publications, etc. should generally be backed up regularly. Some files, such as temporary files, derived files that are easy to generate, or data that is available on public servers should probably not be backed up if they're going to consume all your available disk space. It is much easier to install new copies of commercially available software like ArcGIS and Microsoft Office than to restore them from a backup, so do not bother backing these up.

When thinking about what data to back up and how often to do it, we recommend the following principle: **do not ask *if* a hard drive will fail, ask *when*.** If you think that losing it might cause a problem, back it up.

Step Three: Document your backup

A backup is useless and wasted effort if recovery of the files in the backup is not possible. Use these tips to make this painful process easier:

- Write a text file with the contents of the backup. A text file should include what the files are, the format, where the metadata is, and how it was processed. Put this text file in the folder that you will backup.
- The filename of the backup file should be descriptive and have the date when the archive

was created. `Backup002.zip` will not mean anything to you after 3 months. `Workstation_backup_01feb2010.zip` and `Shapefiles_Indiana_USGS_Dec2009.zip` are good examples.

- Do not mix files from different projects in a single backup file. You will curse yourself for doing this the first time you have to search through your archives for a particular dataset.
- If you plan to back up several different data sets on Fortress, you should probably first come up with a logical storage hierarchy so that you know where things are when you need them. You can create directories in your designated archive space to help keep things organized.
- Data used in a publication should be kept in a backup file of its own.

MANUAL BACKUPS

Step One: Create the backup file

For the purposes of this tutorial, we are going to assume that you want to back up an entire folder (in this case: `LaSelva` under `My Documents`). Right-click the folder and choose "Send to" and "Compressed (zipped) folder":

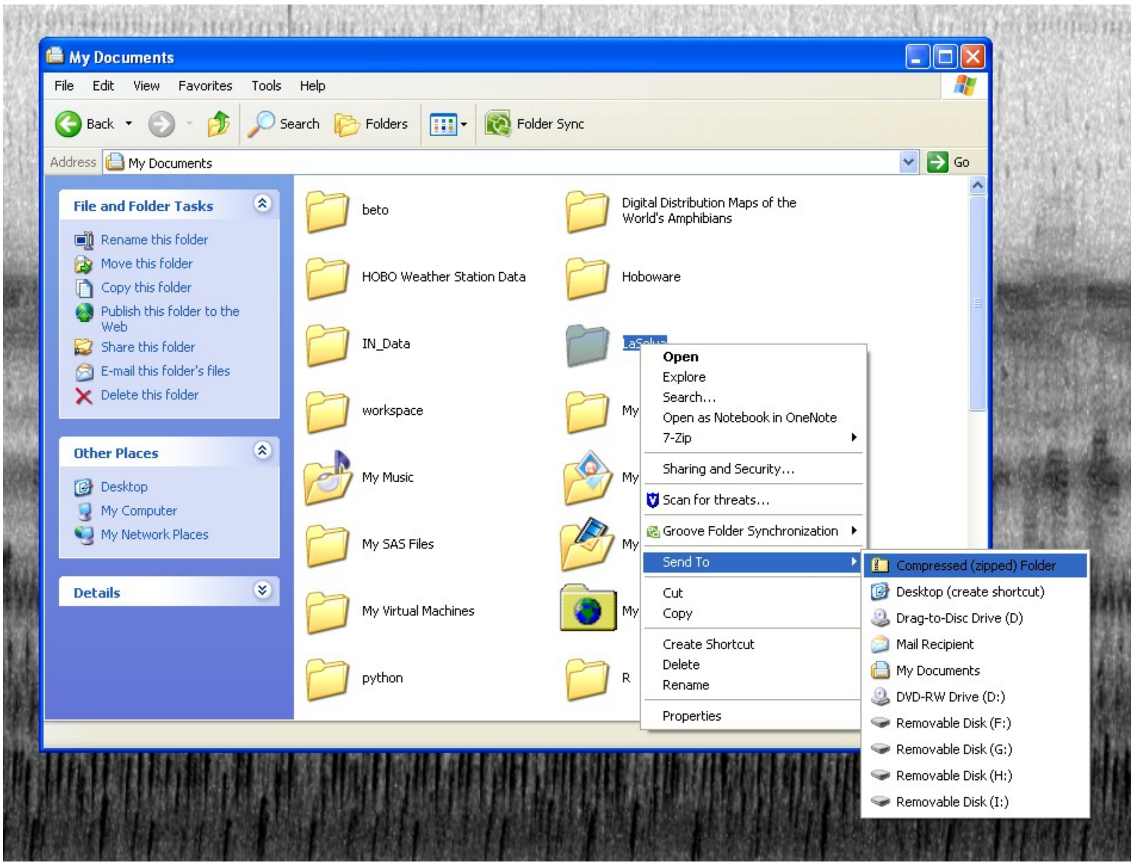


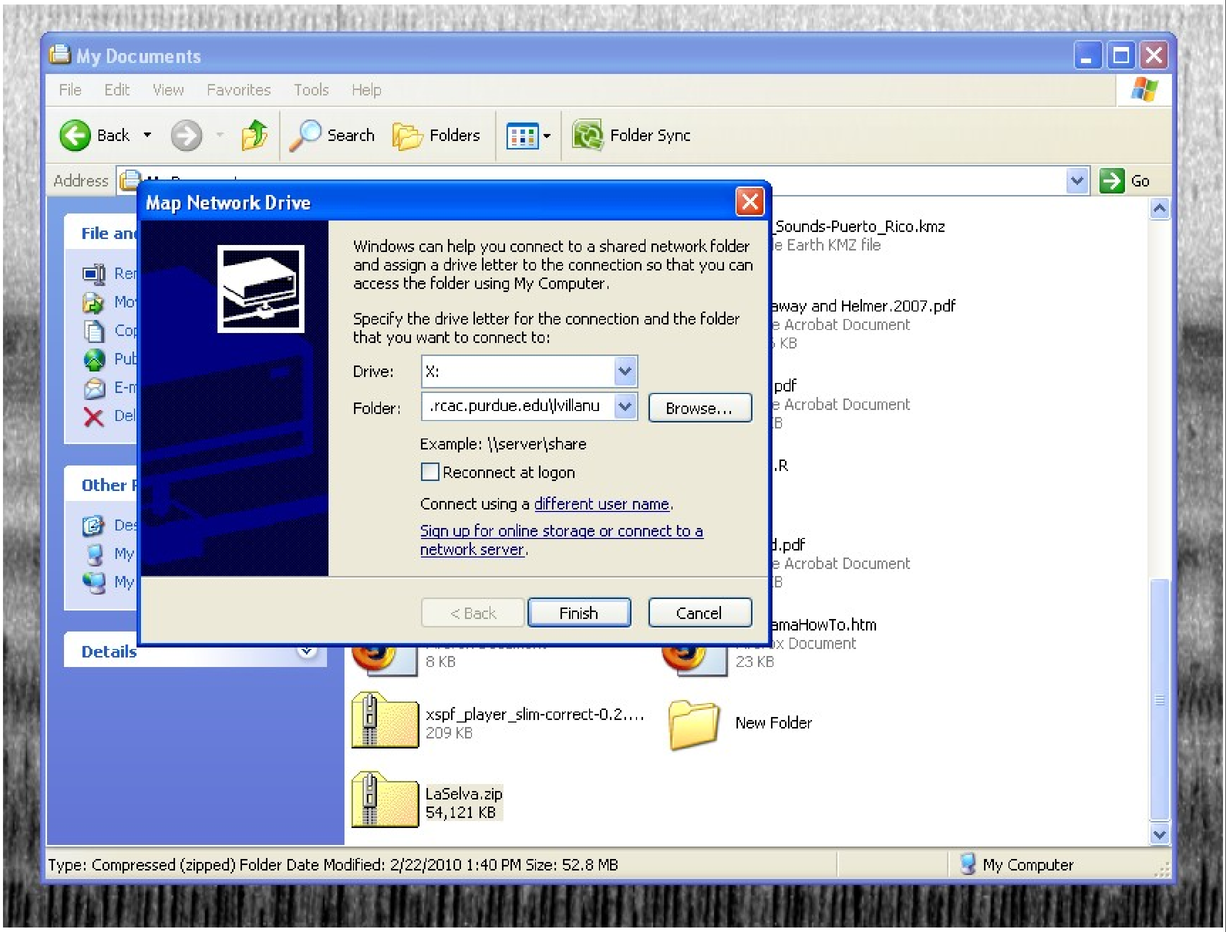
Illustration 1: Sending a folder to a zipped archive

You will end up with a zip file with the name of the folder, in this case: `LaSelva.zip`.

Step Two: Connect to Fortress

Map your folder in Fortress in Windows Explorer by selecting “Map Network Drive.” The settings you will need are:

- Drive: Any one that is available in your computer.
- Folder: `\\fortress.rcac.purdue.edu\username` (`username` is your ONEPURDUE username)
- Do not select the “Reconnect at logon” option



Map Network Drive

Windows can help you connect to a shared network folder and assign a drive letter to the connection so that you can access the folder using My Computer.

Specify the drive letter for the connection and the folder that you want to connect to:

Drive: X:

Folder: .rcac.purdue.edu\villanu

Example: \\server\share

☐ Reconnect at logon

Connect using a different user name.

Sign up for online storage or connect to a network server.

Type: Compressed (zipped) Folder Date Modified: 2/22/2010 1:40 PM Size: 52.8 MB

Illustration 2: Mapping a network drive in Windows Explorer

Click “Finish”. A window will open showing the contents of your folder in the Fortress archive.

Step Three: Copy the files

Using either copy-and-paste or drag-and-drop, copy the archive `.zip` file to your folder in Fortress. Then delete the zipped file from your computer to keep it from taking up disk space.

AUTOMATIC BACKUPS

Step One: Create the appropriate directories on your hard drive

For the purposes of this tutorial, we are going to assume that you want to back up the entire workspace (i.e. `C:\workspace`) in a single archive. As we mentioned earlier, this is usually a bad idea, but we're just doing it now for the sake of illustration. Create a directory named “bin” and a directory named “bak” on the hard drive (i.e. `C:\bin` and `C:\bak`).

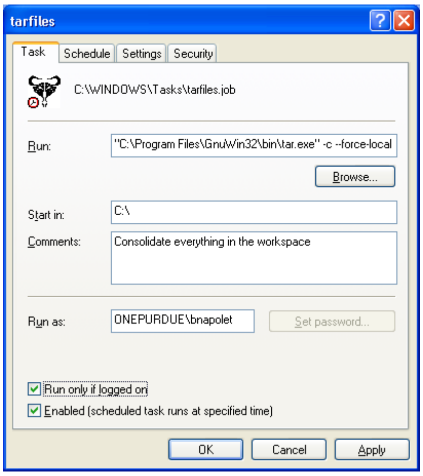
Step Two: Download and install the necessary software

You will need to download and install three separate software packages.

1. PuTTY
   http://www.chiark.greenend.org.uk/~sgtatham/putty/
2. Tar for Windows
   http://gnuwin32.sourceforge.net/packages/gtar.htm
3. Gzip for Windows
   http://gnuwin32.sourceforge.net/packages/gzip.htm

PuTTY will arrive as a single .zip archive. Use whatever archive manager you have installed to extract the contents of the PuTTY archive into the `C:\bin` directory that you created earlier. When you download both Tar and Gzip, be sure to select the complete package (i.e. download the “Setup” option). Once these files have been downloaded, execute them (i.e. “double-click”) and they will install all the libraries and such in `C:\Program Files\GnuWin32\bin\`. Remember this directory, because you will be entering it several times.

Step Three: Use the task scheduler to configure a process to automatically consolidate and compress your workspace



C:\WINDOWS\Tasks\tarfiles.job

Run: "C:\Program Files\GnuWin32\bin\tar.exe" -c --force-local

Start in: C:\

Comments: Consolidate everything in the workspace

Run as: ONEPURDUE\bnapolet

☑ Run only if logged on

☑ Enabled (scheduled task runs at specified time)

Illustration 3: Task Scheduler Interface

Open the Control Panel and select the Task Scheduler. Or, if you don't know how to get to the Control Panel, open the “Start” menu, select “All Programs,” “Accessories,” “System Tools,” and “Scheduled Tasks.” If the Task Scheduler automatically opens a wizard to schedule a new task, press the “Cancel” button. We're going to be hard core and create the tasks ourselves. You should now be looking at the “Scheduled Tasks” window, which should be a mostly empty window with an icon depicting a clock stuck to a notepad that reads “Add Scheduled Task.” Do not double-click this icon! It will send a horrible virus to everyone on your mailing list and then erase everything on your hard drive. Okay, that's a lie, but don't double-click that icon anyway. It just opens the wizard, and I just told you that we're not using the wizard. Instead, right-click anywhere in the open area and select “New” and then “Scheduled Task.” Once the new icon appears, give it a useful name like “tarfiles.” Then, right-click on the icon and select “Properties.” A window similar to the one depicted in Illustration 3 should appear. Select the text box after “Run:” and enter the following text, verbatim (including the double quotes):

```
“C:\Program Files\GnuWin32\bin\tar.exe” -c --force-local --file=”C:\bak\workspace.tar” C:\workspace
```

The first line of this command tells Windows where to find the file to be executed (`tar.exe`). Because Windows is insane, it stores its executable files in a directory named “Program Files,” which means that you have to enclose any calls to files in there in quotes. The `-c` switch tells `tar.exe` to create a new, empty archive, and the `--force-local` switch tells it that this archive will be located on the local hard drive. The `--file=` switch tells `tar.exe` the path to and name of the new archive that it's creating, and then `tar.exe` takes everything it finds in `C:\workspace` (including `C:\workspace` itself) and puts it in this new archive. If you want, you can enter something to the effect of what I just told you in the “Comments” section, so that you can figure out what this task is a year or two from now when you stumble across it. Also, select the “Run only if logged on” option. Since your connection to the Fortress server uses your career account's password, the backup sequence will only work when you are actually logged on. Your workspace can still be locked, but you have to be logged on.

Next, select the “Schedule” tab, and schedule the task to run weekly (you pick the time, but make it a time when you're not planning on using your computer).

The tar utility consolidates your workspace into a single file. Next, you need to compress this file using the gzip utility. To do this, create another new scheduled task. Rename this one something like “zipfiles” and open the “Properties” dialog once again. In the “Run:” box, enter the following text, verbatim:

```
“C:\Program Files\GnuWin32\bin\gzip.exe” C:\bak\workspace.tar
```

This instructs Windows to run `gzip.exe`, which then takes `workspace.tar`, compresses it, and creates a new, smaller file named `workspace.tar.gz`. Once again, you can enter some useful comments to help you remember what this task does, and then schedule it to run weekly. You'll need to schedule this task to run *after* the task that consolidates everything into a tar file is finished, so leave enough time for the tar process to complete. An hour or two should be sufficient, so schedule this task to run two hours after the tar process you scheduled.

Step Four: Mount the Fortress server as a network drive

You can mount the Fortress server just like you would any other shared network drive. Open the “My Computer” Explorer window, and then select “Map Network Drive...” from the “Tools” menu. Assign whatever drive letter you prefer to the network drive, but this tutorial is going to assume that you use the X drive (`X:`). Enter the following address in the “Folder” text box:

```
\\fortress.rcac.purdue.edu\username
```

Replace “`username`” with your Purdue Career Account ID, which should also be your ID on Fortress. Make sure you tick the query box that reads “Reconnect at logon.” This will instruct Windows to automatically map the network share every time you log on to your machine, so you don't need to remember to do it manually. If you forgot what the dialog box for mapping network drives looks like, check out Illustration 2 again.

Step Five: Have your backup file delivered to the server

You have now configured Windows to automatically backup all the data in your “`workspace`” directory once per week. Now all you need to do is to have Windows send your backup off to the server, and you'll be able to maintain a reliable archive of all your work. Return to the "Task Scheduler" window that you used in Step Four, and add another new task. Name it something like "sendfiles," and open the task for editing again. Enter the following command in the "Run:" text box, verbatim:

```
move /Y C:\bak\workspace.tar.gz X:\archives\
```

Once again, notice that I assumed that the server was mapped to the X drive. If you are using a different drive letter, then replace `X` with that letter. If you want to keep a copy of the backup file on your local machine, you can use the "copy" command instead of the "move" command. The syntax is the same, so you would simply enter the following:

```
copy /Y C:\bak\workspace.tar.gz X:\archives\
```

The copy command will create a duplicate of the original file in the second location while the move command will remove the file from its original location. The `/Y` switch prevents the computer from asking you whether you want to overwrite a file that already exists (because, presumably, in this case you do). Once again, enter a useful description in the "Comment" section and then schedule the task to run an hour or two after the task that compresses the archive.

In addition to the ability to recover from something like a catastrophic disk failure, another advantage to backing data up is the ability to go back in time to undo a major mistake. This ability is most effective when you have a logical timeline of changes to chose from. Therefore, when you copy your backup file to the server, it may be useful to add a date stamp to the filename. Note, though, that this means you will be adding an entirely new file to the server every time the task runs, so you should only do this with data that you are actively changing. To add a date stamp to the filename on the server, use the following command instead of the previous one:

```
move  /Y  C:\bak\workspace.tar.gz  X:\archives\workspace-%DATE:~10%%DATE:~4,2%  \
%DATE:~7,2%.tar.gz
```

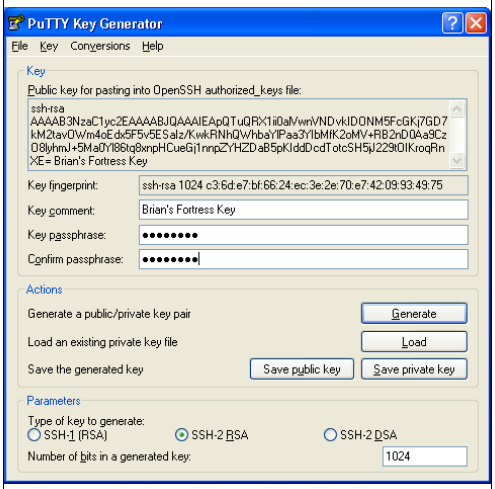
Note that the whitespace and the slash (\) are typographical constructs to indicate that I had to break this line to fit it on the page. When you enter it, do so without the whitespace or the slash (\). The sequence of strange characters that you added to the target filename instructs Windows to include the year, month, and date in the filename(e.g. `workspace-20100215.tar.gz` for 15 February 2010). Note that there should be no whitespaces in the filename.

Congratulations! You have now configured your computer to automatically back up your data on the Fortress server!

Alternative Strategy: Use PSCP with a public/private keypair

Instead of mounting the Fortress server as a network share, you can use a keypair in conjunction with the PSCP application to send your backup file via Secure SHell (SSH). If you decide to use SSH, note that you will still need to create the tasks to consolidate and compress your files in the Task Scheduler (Step Four above).Alternative Step One: Generate the public/private keypair

The public/private keypair allows you to log in to the Fortress server remotely without entering a password. In other words, you need it if you want to automatically backup your files. To generate the keypair, execute the "PUTTYGEN.EXE" file that was extracted into your `C:\bin` directory. Illustration 4 depicts the friendly interface that the PuTTY Key Generator uses to guide you through the key generation process. The Key Comment is a string of text for you to use to help you identify your keys. Give your keypair a meaningful comment (e.g. "Sarah's Fortress Key"), and select a passphrase that you will remember. Once the key is generated, be sure to press both the “Save public key” and the “Save private key” buttons. Save the public key as `C:\bin\pubkey.txt`, and save the private key as `C:\bin\privatekey.ppk`. Make sure you save them in `C:\bin`, lest you run into trouble later. The default key settings of SSH-2 RSA and 1024 bits are sufficient, although you can always increase the number of bits if you would like a key that is more difficult to replicate. Once you have successfully generated your new keypair, you need to share your public key with the server. Do not close this window after you've saved the keys, because you'll need to use it in the next step.



Public key for pasting into OpenSSH authorized_keys file:

```
ssh-rsa
AAAAB3NzaC1yc2EAAAABJQAAAIEApQTuQRX1ii0aMwnVNDvkIDONM5FcGKj7GD7
kM2tavOWm4oEdx5F5v5ESalz/KwkRNhQWhbaYlPaa3YIbMfK2oMV+RB2nD0Aa9Cz
O8lyhmJ+5Ma0Yl86tq8xnpHCueGj1nnpZYHZDaB5pKIddDcdTotcSH5jJ229t0IKroqRn
XE= Brian's Fortress Key
```

Key fingerprint: ssh-rsa 1024 c3:6d:e7:bf:66:24:ec:3e:2e:70:e7:42:09:93:49:75

Key comment: Brian's Fortress Key

Key passphrase: ••••••••

Confirm passphrase: ••••••••

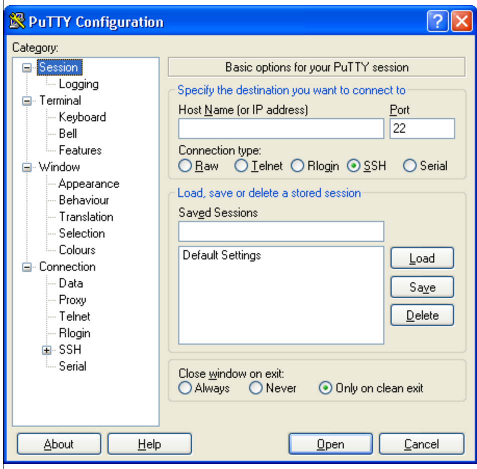
Type of key to generate: ○ SSH-1 (RSA) ● SSH-2 RSA ○ SSH-2 DSA

Number of bits in a generated key: 1024

Illustration 4: The PuTTY Key Generator

Alternative Step Two: Place your public key on the server

The public/private keypair works by matching a public key up with a private key. The names imply the functions. You share the public key with whomever you want to recognize you, and you keep the private key to yourself. This means that you need to place a copy of your public key on the Fortress server so that it will recognize your private key when you try to contact it. To start a new interactive session on the server, double-click the file named “PUTTY.EXE” in the `C:\bin` window that you should still have open from the previous step. This will open a friendly GUI window that will allow you to tell PuTTY what server you want to contact (Illustration 5). In the text box under “Host Name (or IP address),” enter either the name of the host (`fortress.rcac.purdue.edu`) or the IP address (`128.211.158.12`). The “Port” text box should have the number twenty-two in it, and “SSH” should be selected as the connection type. As you are just logging on to provide your key, don't bother with any of the other options, and just press the “Open” button. A terminal window should appear, and you should be prompted to enter your username and then your password. Assuming you entered everything correctly, you should now be looking at a Linux shell interface. If your prompt looks like mine (Illustration 6), then you should see “`frsh>`” followed by a blinking cursor.



Host Name (or IP address):

Port: 22

Connection type: ○ Raw ○ Telnet ○ Rlogin ● SSH ○ Serial

Saved Sessions:

Default Settings

Close window on exit: ○ Always ○ Never ● Only on clean exit

Illustration 5: The configuration window for a PuTTY session

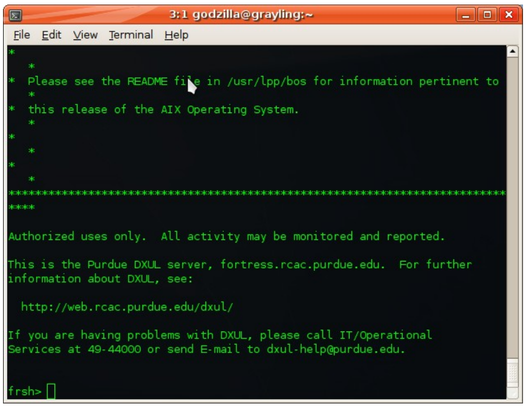
Once you're interacting with the server, the first thing to do is to make sure a directory named “.ssh” exists in your home directory. To find out, enter the following command at the prompt:

```
ls -a
```

This lists all the files and directories, including the hidden ones. If the command does not return a directory named “.ssh,” then you will need to create it with the following command:

```
mkdir .ssh
```

You are going to place the public key that you created previously on the server. To do this, copy all the text inside the text box beneath the words “Public key for pasting into OpenSSH authorized_keys file:” in the PuTTY Key Generator GUI. Now enter the following command in the shell interface (the black window with the “`frsh>`” prompt):



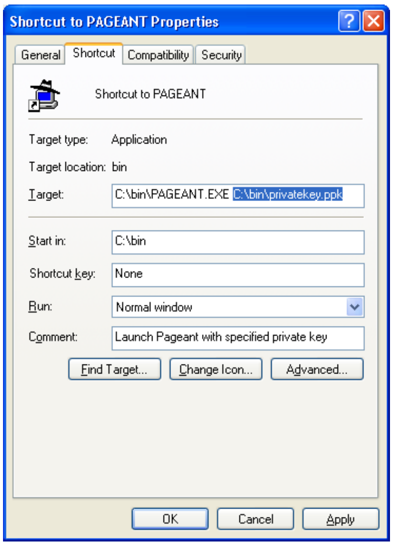
```
*
   *
*  Please see the README file in /usr/lpp/bos for information pertinent to
   *
*  this release of the AIX Operating System.
   *
*
   *
*
   *
***************************************************************************
****

Authorized uses only.  All activity may be monitored and reported.

This is the Purdue DXUL server, fortress.rcac.purdue.edu.  For further
information about DXUL, see:

  http://web.rcac.purdue.edu/dxul/

If you are having problems with DXUL, please call IT/Operational
Services at 49-44000 or send E-mail to dxul-help@purdue.edu.


frsh> 
```

Illustration 6: An example of the shell interface on the server

```
echo "[public key]" > .ssh/authorized_keys
```

Replace `[public key]` with the text you just copied by right-clicking anywhere inside the shell interface. You are telling the the shell to print the string of characters that represents your public key inside a file named “authorized_keys” located in the “.ssh” directory. This is where the server will look for your public key when you try to log on to it. Once you've placed your public key on the server, you can log out by typing “`exit`” at the command prompt and pressing the “Enter” key. Now that the server knows who you are, you're ready to configure Pageant to run automatically. You can safely close the “PuTTY Key Generator” window now.

Alternative Step Three: Configure Pageant to automatically launch with your specified keypair whenever you log on



Target type: Application
Target location: bin
Target: C:\bin\PAGEANT.EXE C:\bin\privatekey.ppk
Start in: C:\bin
Shortcut key: None
Run: Normal window
Comment: Launch Pageant with specified private key

Illustration 7: Properties for the shortcut to Pageant

To have Pageant automatically load your key, first open `C:\bin` in a Windows Explorer window (e.g. Illustration 1). Then, right-click on the “Start” button, and select “Explore” from the menu that appears. This will launch another Windows Explorer window with your Start Menu selected. From there, go into “Programs,” and then “Startup.” Windows will automatically run anything in this folder whenever you log on to your machine. Now, right-click the “PAGEANT.EXE” icon inside the “`C:\bin`” window, and drag it into the “Startup” window. Release the right mouse button, and Windows will ask you what you want to do with the file. Select “Create shortcut here.” A new icon that says “Shortcut to PAGEANT” should appear inside your “Startup” window. Congratulations! You're halfway through this step. Now, right-click on this new “Shortcut to PAGEANT” icon, and select “Properties.” A screen similar to the one shown in Illustration 7 should appear. Select the text box where the screen reads “Target:” and press the “End” key to move to the end of the line. Enter a single space, and then enter the full path to the private key you created in the previous step (if you're following these instructions verbatim, you should enter “`C:\bin\privatekey.ppk`”). This tells windows to launch Pageant when you log on, and the path you just entered tells Pageant where to find the key that you created so that it can load it for you. Now, whenever you log in, you will be asked for the passphrase that you specified when you generated your key. Log out and log back in now to make sure this works.

Alternative Step Four: Have your data copied to the server

Before beginning this step, you will still need to complete step four in the original sequence (i.e. you still need to create the backup file). Once you have the backup file, you can use the “pscp” application to send a copy to the server. PSCP is simply a Windows implementation of the Secure Copy Protocol, which allows you to safely move files across a network using the Secure Shell

(SSH) protocols. Once again, right click in the “Scheduled Tasks” window, and create a new scheduled task. Name this task something like “copyfiles.” Now right-click on the new task and select “Properties.” Enter the following command in the text box next to “Run:”:

```
C:\bin\pscp.exe          -batch          C:\bak\workspace.tar.gz          \
username@fortress.rcac.purdue.edu:~/workspace.tar.gz
```

Once again, note that the slash (\) in this case denotes the fact that I had to break the line to fit it on this page. When you enter the command, enter it without the slash (\), but keep the whitespace this time. The command `pscp.exe` calls PSCP, and the `-batch` switch instructs PSCP to run silently (i.e. it tells it not to stop to ask you for input, since you're not going to be there to give it). PSCP will then send the file named “`workspace.tar.gz`” to your home directory on the Fortress server. If you want the backup file on the server to have the date in the filename, replace the second instance of `workspace.tar.gz` with the following, as in the previous Step Six above:

```
workspace-%DATE:~10%%DATE:~4,2%%DATE:~7,2%.tar.gz
```

Enter some useful information in the “Comments” section, and then schedule this task to run an hour or two after the task that compresses the file (once again, you need to give the previous task some time to finish). Congratulations! You have now configured your computer to automatically back up your data on the Fortress server without mounting your storage space as a network share.

Alternative Step Five (Optional): Delete the local copy of the backup file

If you don't want the backup files sitting around on your computer taking up space, then you can schedule an additional task to delete the local copies after they've been sent to the server. To do this, you simply use the Task Scheduler to schedule a task with the following command in the “Run:” text box:

```
delete C:\bak\workspace.tar.gz
```

Make sure you schedule this task to run several hours *after* the task that sends the backup file to the server. If you schedule the “delete” task to run too soon, it will erase your archive before it is sent to the server.

OTHER CONSIDERATIONS

File System

If you are using Windows XP or any of its successors, your hard drive should be formatted with NTFS (Networking Technologies File System). Some drives, however, may still be using the FAT 32 (File Allocation Tables) system, which cannot handle files larger than 2 GB. If you plan to create files larger than 2 GB, then you need to use NTFS.