

Table of Contents

1. Slide presentation.....pp 2-19
2. Notes on selected slides.....pp 20-25
3. Brief descriptions of various
high level backup programs.....pp 26-31

Backing Up Home Computers

Considerations and Possibilities
from a professional non-expert

by
Jack Warkentin

What is a "backup"?

A *backup* is a copy of some or all of the information on a computer's hard drive (or drives).

The action of creating such a copy is called *backing up*.

Considerations

- Why
- What
- Media
- How
- Types
- Media management
- Number of computers, and their network configuration

Why have backups?

- hard drives fail
- almost always with absolutely no warning
- very rarely is it possible to recover any of the information from a failed drive
- the cost of losing this information can be enormous
- a timely backup can avoid serious/expensive info loss

If still in doubt see <http://www.taobackup.com/>

What all information should be backed up?

- everything that cannot otherwise be cheaply replaced on a drive
- operating systems and other software can usually be reinstalled very cheaply, so not usually backed up
- data of all sorts (company databases, user data, photos, spreadsheets, emails, ...) plus configuration data **must** be backed up.
- in GNU/Linux, for home use, usually this means the /etc and the /home directories, but there could be others

What media can/should be used?

possibilities include

- removable media
 - floppies, flash drives
 - tapes
 - camcorder tapes
 - cdroms/dvds
- other hard discs
 - external, eg USB or NAS
 - on another computer

How can backups be performed?

Possibilities include

- Manual
 - You decide it's time to do a backup, so you make the copy, using cp, cpio, xcdroast, k3b, ...
- Automatic
 - You set up a cron job, or use a special backup program, to perform the backups at certain times
 - The computer must be switched on during those certain times.
- Semi-Manual/Automatic
 - Automatic, except backup program prompts for media changes.
Backups to dvd will probably be manual or semi-manual, unless you have a drive (expensive) capable of holding several disks at the same time.
But there are higher level tools, eg multird, which can prompt when media changes are required.

Types of Backups

- Full
 - every file is completely copied to the backup, typically once per week, in a business shop
- Incremental
 - only files which have changed since the last incremental backup are copied, typically once per day, in a business shop
- Differential
 - only files which have changed since the last full backup are copied, typically once per day, in a business shop
- Continuous/lazy
 - changed/new files backed up periodically throughout day

Media Management

- Removable media need to be inserted, removed, labelled and stored.
 - For automatic backups, insertion must occur before backup, and removal afterwards. Must make sure next backup does not overwrite previous.
- For hard disk media, partitions and mount points must be carefully managed.
- Periodic off-premises storage is highly desirable, eg monthly full dvd backups to safety deposit box.
- Cdroms and dvds deteriorate over time. Don't get caught relying on useless media.

of computers & their network configuration

- Single computer
 - options pretty well limited to cdrom/dvd and external hard drive
- Multiple networked computers
 - client/server configuration
 - easiest configuration for managing backups, since only one machine to be backed up
 - requires server to be always switched on (supports fully automatic backups), or else switched on and off when necessary (not fully automatic)
 - no server
 - difficult to make fully automatic unless switched on all the time

Possibilities

- Tools
 - low level
 - high level
- My personal choices
 - xcdroast, for backups to dvd's
 - unison/unison-gtk for syncing different computers
 - boxbackup for backing up to external hard disks

Tools

- low level
 - copying progs, eg cp (same host), scp, rsync (diff hosts)
 - command line archiving programs, eg tar, cpio, afio, usually combined with compression, eg gzip, bzip2
 - can be automatic or manual
 - cdrom/dvd writers, eg cdrecord, wodim (CLI), **xcdroast**, k3b, brasero, ..., (GUI)
 - manual, unless multidisk writer unit available & CLI tool used
 - media management almost always manual

Tools

- high level
 - abackup, amanda, backintime, backup-manager, backup2l, backupninja, backuppc, bacula, **boxbackup**, cdbackup/cdrestore, cedar-backup2, chiark, dar, deja-dup, dirvish, dump, duplicity, dvbackup, faubackup, flexbackup, hdup, jpilot-backup, keep, luckybackup, mondo, multictd, nilfs-tools, partimage, pdumpfs, pybackpack, rdiff-backup, rsnapshot, slbackup, tob, **unison/unison-gtk**, vbackup
 - full or partial automatic media management frequently available

My Personal Choices

- I currently have a desktop and two laptops.
- So far I have rejected the client server model because I don't want some machine to be switched on all the time; and waiting to switch off until a backup is finished, when I'd rather be in bed, is also not acceptable to me.
- Besides, when a laptop is away from home, the server would be unavailable.
- The laptops are partially synced to the desktop manually, using **unison** and **unison-gtk**.
- The desktop, my main machine, is backed up automatically to a 1 TB USB-attached hard disk, specifically for backups, using **Box Backup**.
- Occasionally I make backups to dvd's using **xcdroast**

unison/unison-gtk

- Licensed under GNU GPL version 3
- Runs on both UNIX (OSX, Linux, Solaris, ...) and MS Windows, even cross platform
- Command line (unison) and GUI (unison-gtk) versions
 - command line version can be fully automated
 - useful for maintaining a full backup to an external hd
 - GUI version excellent for syncing two machines
 - can be set up to view/merge differences between text files
- Big disadvantage
 - only keeps current versions of files, ie cannot recover either deleted files or older versions of existing files from a backup

Box Backup

- Distributed under a BSD license
- "An open source, completely automatic on-line backup system."
- Client/server model
 - server designed to be on separate machine from client, and support multiple clients, but may be on same machine as client
- Continuous/lazy mode - clients wake up every hour (configurable) and copy changes to server.
- Developed and deployed on Linux, Darwin, Windows and OpenBSD
 - primary platforms are OpenBSD and Linux
 - important platforms include Windows (client only)
- See www.boxbackup.org and www.boxbackup.org/trac/wiki

Box backup (continued)

- All backed up data is stored on the server in files on a filesystem -- no tape or archive devices are used.
- The server is trusted only to make files available when they are required -- all data is encrypted.
- Only changes within files are sent to the server, just like rsync.
- Old versions of files on the server are stored as changes from the current version.
- Behaves like tape -- old versions and deleted files are available.
- Designed to be easy and cheap to run a server. Portable implementation, and optional RAID (implemented in userland) for reliability without complex server setup or expensive hardware.

Demonstrations

- unison-gtk
- box backup

Backing Up Home Computers

Considerations and Possibilities
from a professional non-expert

by
Jack Warkentin

2

I consider myself to be a computing professional because I have a PhD in Computer Science (University of Waterloo, 1971), spent 10.5 years as a prof at Lakehead University, Thunder Bay, Ontario, and spent 15.5 years as a software engineer at Nortel Networks in Brampton, Ontario.

However, I have had no professional experience related to backups, so cannot consider myself to be an expert, or professional, with respect to them.

What all information should be backed up?

- everything that cannot otherwise be cheaply replaced on a drive
- operating systems and other software can usually be reinstalled very cheaply, so not usually backed up
- data of all sorts (company databases, user data, photos, spreadsheets, emails, ...) plus configuration data **must** be backed up.
- in GNU/Linux, for home use, usually this means the /etc and the /home directories, but there could be others

6

Many would argue that the operating system and other software should also be backed up. This should be a value judgment, based on how long it would take, availability (or not) of OS and other software media, and the cost of performing the re-installations.

What media can/should be used?

possibilities include

- removable media
 - floppies, flash drives
 - tapes
 - camcorder tapes
 - cdroms/dvds
- other hard discs
 - external, eg USB or NAS
 - on another computer

7

Floppies were viable at one time, but now have far too insufficient capacity. The capacities of flash drives make them of limited use. Tape drives are expensive and so of unlikely use for home computers.

There is an interesting program called dvbackup which can be used to back up to 13 GB camcorder DV cartridges, but again this capacity is quite small by current standards.

That leaves cdroms/dvds and other hard drives as the most likely candidates for backing up home computers.

Tools

- high level
 - abackup, amanda, backintime, backup-manager, backup2l, backupninja, backuppc, bacula, **boxbackup**, cdbackup/cdrestore, cedar-backup2, chiark, dar, deja-dup, dirvish, dump, duplicity, dvbackup, faubackup, flexbackup, hdup, jpilot-backup, keep, luckybackup, mondo, multird, nilfs-tools, partimage, pdumpfs, pybackpack, rdiff-backup, rsnapshot, slbackup, tob, **unison/unison-gtk**, vbackup
 - full or partial automatic media management frequently available

14

Brief descriptions of all of these programs are included in the pdf file containing this presentation.

My Personal Choices

- I currently have a desktop and two laptops.
- So far I have rejected the client server model because I don't want some machine to be switched on all the time; and waiting to switch off until a backup is finished, when I'd rather be in bed, is also not acceptable to me.
- Besides, when a laptop is away from home, the server would be unavailable.
- The laptops are partially synced to the desktop manually, using **unison** and **unison-gtk**.
- The desktop, my main machine, is backed up automatically to a 1 TB USB-attached hard disk, specifically for backups, using **Box Backup**.
- Occasionally I make backups to dvd's using **xcdroast**

15

xcdroast is a GUI program for ripping and writing cd's and dvd's. I like it because it can be set up to make the medium bootable, *if you want it to be*. k3b, for example, didn't allow me to configure this.

unison/unison-gtk

- Licensed under GNU GPL version 3
- Runs on both UNIX (OSX, Linux, Solaris, ...) and MS Windows, even cross platform
- Command line (unison) and GUI (unison-gtk) versions
 - command line version can be fully automated
 - useful for maintaining a full backup to an external hd
 - GUI version excellent for syncing two machines
 - can be set up to view/merge differences between text files
- Big disadvantage
 - only keeps current versions of files, ie cannot recover either deleted files or older versions of existing files from a backup

16

unison is very well documented. There is a config file for selecting which files to sync. I have mine set up so that the differences between text files on the two copies can be viewed, and merged if necessary.

High Level Backup Programs

The programs listed here are available via Debian packages. In most cases the package name is given. In some case, there are several packages involved. With these, a title is selected that forms part of the package names. The descriptions given are taken from the Debian package descriptions.

afbackup - Client-Server Backup System

This is a client-server backup system offering several workstations a centralized backup to a special backup server. Backing up only one computer is easily possible, too.

amanda - Advanced Maryland Automatic Network Disk Archiver

Amanda is a backup system designed to archive many computers on a network to a single large-capacity tape drive. This package is suitable for large amounts of data to backup. For smaller solutions take a look at afbackup, tob, ...

backintime - simple backup/snapshot system

Back In Time is a framework for rsync, diff and cron for the purpose of taking snapshots and backups of specified folders.

backup-manager - command-line backup tool

This is a backup program, designed to help you make daily archives of your file system.

Written in bash and perl, it can make tar, tar.gz, tar.bz2, and zip archives and can be run in a parallel mode with different configuration files. Other archives are possible: MySQL or SVN dumps, incremental backups...

Archives are kept for a given number of days and the upload system can use FTP, SSH or RSYNC to transfer the generated archives to a list of remote hosts.

Automatically burning archives to removable media such as CD or DVD is also possible.

backup2l - low-maintenance backup/restore tool

backup2l [backup-too-l] is a tool for autonomously generating, maintaining and restoring backups on a mountable file system (e. g. hard disk). In a default installation, backups are created regularly by a cron script. backup2l features differential backups at multiple hierarchical levels. An integrated split-and-collect function allows to comfortably transfer all or selected archives to a set of CDs or other removable media.

backupninja - lightweight, extensible meta-backup system

Backupninja lets you drop simple config files in /etc/backup.d to coordinate system backups.

bacula - network backup, recovery and verification

Bacula is a set of programs to manage backup, recovery and verification of computer data across a network of computers of different kinds.

It is efficient and relatively easy to use, while offering many advanced storage management features that make it easy to find and recover lost or damaged files. Due to its modular design, Bacula is scalable from small single computer systems to networks of hundreds of machines.

backuppc - high-performance, enterprise-grade system for backing up PCs

BackupPC is disk based and not tape based. This particularity allows features not found in any other backup solution.

boxbackup - BoxBackup remote backup system

BoxBackup is an automatic on-line backup system. The client watches for changes on the local file system, connects to a BoxBackup server and sends the changes via a secure channel. All data is encrypted before being sent to the server. A command-line tool is provided for restoration of backups including deleted files and old versions.

cdbackup/cdrestore - CD-R(W) backup and restore utility

cdbackup and cdrestore are a pair of utilities designed to facilitate streaming backup to and from CD-R(W) disks. Specifically, they were designed to work with dump/restore, but tar/cpio/whatever you want should work, so long as it writes to stdout for backups and reads from stdin for restores.

cedar-backup2 - local and remote backups to CD or DVD media

Cedar Backup is a software package designed to manage system backups for a pool of local and remote machines. Cedar Backup understands how to back up filesystem data as well as MySQL and PostgreSQL databases and Subversion repositories. It can also be easily extended to support other kinds of data sources.

Cedar Backup is focused around weekly backups to a single CD or DVD disc, with the expectation that the disc will be changed or overwritten at the beginning of each week. If your hardware is new enough, Cedar Backup can write multisession discs, allowing you to add incremental data to a disc on a daily basis.

chiark - backup system for small systems and networks

These are the backup scripts used by chiark.greenend.org.uk and other systems belonging to the Sinister Greenend Organisation. Features:

- * Suitable for single systems and small networks.
- * Reasonably simple; they do what you tell it to.
- * Hard failures when individual systems fail, to encourage fixing !

If you have a larger site you may wish to look at Amanda.

dar - Disk ARchive: Backup directory tree and files

Full featured archiver with support for differential backups, slices, compression, ATTR/ACL support. DAR also supports Pipes for remote operations, including with ssh.

deja-dup - Backup utility

Déjà Dup is a simple backup program. It hides the complexity of doing backups the 'right way' (encrypted, off-site, and regular) and uses duplicity as the backend.

dirvish - Filesystem based backup system using rsync

A utility to maintain multiple backups on online storage, each backup is available as a sort of snapshot directory, where common files are shared between the different backup generations. It uses rsync to do the actual copying.

Backups can be made locally or over the network (using ssh).

dump - 4.4bsd dump and restore for ext2 filesystems

Dump examines files on a filesystem and determines which files need to be backed up. These files are copied to the given disk, tape or other storage medium for safe keeping.

The restore command performs the inverse function of dump. A full backup of a file system may be restored and subsequent incremental backups layered on top of it. Single files and directory subtrees may be restored from full or partial backups.

duplicity - encrypted bandwidth-efficient backup

Duplicity backs directories by producing encrypted tar-format volumes and uploading them to a remote or local file server. Because duplicity uses librsync, the incremental archives are space efficient and only record the parts of files that have changed since the last backup. Because duplicity uses GnuPG to encrypt and/or sign these archives, they will be safe from spying and/or modification by the server.

dvbackup - backup tool using MiniDV camcorders

Dvbackup hides the data it receives on standard input in a perfectly legal DV (digital video) stream. This allows you to use your DV camcorder and your DV cartridges as a potent mass storage system. Obviously, your system and your camcorder have to be connected via IEEE1394 (aka Firewire, iLink).

Current digital camcorders can save approximately 13 GB of data on those tiny DV cartridges at a speed of 3.6 MB/sec.

faubakup - Backup System using a Filesystem for Storage

This Program uses a filesystem on a hard drive for incremental and full backups. All Backups can easily be accessed by standard filesystem tools (ls, find, grep, cp, ...)

Later Backups to the same filesystem will automatically be incremental, as unchanged files are only hard-linked with the existing version of the file.

flexbackup - Flexible backup tool for small to medium sized installations

Flexbackup is a flexible backup tool that works well for small to medium sized tasks for which solutions like amanda are overkill.

It features:

- * Easy configuration
- * Uses afio, dump, GNU tar, cpio, star, pax, lha, or zip archivers
- * Full and numbered levels of incremental backup (similar to "dump")
- * Compression and buffering options for all backup types
- * Handles remote filesystems with ssh/rsh; no special services required
- * Writes to tapes, on-disk archive files, or on-disk directory trees
- * Keeps a table of contents so you know what archives are on each tape
- * Extensive logging options

hdup - Filesystem duplicator and backup

hdup is a backup utility with the following characteristics:

remote backups (move the backup to another machine); encrypted backups; no obscure format for the backups; compression (gzip/bzip or none); simple to use.

jpilot-backup - Backup plugin for J-Pilot

This program enhances jpilot's own backup, making it much more flexible, handling automatic backups, configuring which applications/databases to back up and which to ignore, etc.

keep - backup system for KDE

Keep is an automatic backup program that allows users to set the parameters of the backup, including the frequency and the number of backups.

Homepage: <http://jr.falleri.free.fr/keep>

luckybackup - rsync-based GUI data backup utility

luckyBackup is an application that backs-up and/or synchronizes any directories with the power of rsync. Its main features are: backup, safety, synchronization, exclude/only include options, allows custom rsync options, remote connections, restore and dry-run operations, scheduling, profiles and command line mode.

It is simple to use, fast (transfers over only changes made and not all data), safe (keeps your data safe by checking all declared directories before proceeding in any data manipulation), reliable and fully customizable.

mondo - powerful disaster recovery suite

Mondo is reliable. It backs up your Debian GNU/Linux server or workstation to tape, CD-R, CD-RW, NFS or hard disk partition. In the event of catastrophic data loss, you will be able to restore all of your data [or as much as you want], from bare metal if necessary. Mondo is in use by numerous blue-chip enterprises and large organizations, dozens of smaller companies, and tens of thousands of users.

Mondo is comprehensive. Mondo supports LVM, RAID, ext2, ext3, JFS, XFS, ReiserFS, VFAT, and can support additional file systems easily. It supports adjustments in disk geometry, including migration from non-RAID to RAID. Mondo runs on all major Linux distributions and is getting better all the time. You may even use it to backup non-Linux partitions, such as NTFS.

multicd - Backup your data to CD-R/CD-RW

multicd provides an easy way to backup a large number of files to multiple CDs. Give multicd the files/directories you want backed up and it will create as many CDs as it needs to, prompting the user to put in a new disc whenever needed. It can be configured to run in a multi-threading mode, where it will burn one image to a disc while it is copying files to another image. This feature can be disabled for slower machines.

nilfs-tools - Continuous Snapshotting Log-structured Filesystem

NILFS is a new implementation of a log-structured filesystem. It allows continuous snapshotting, so that users can restore files mistakenly overwritten or destroyed even when it happens after the last backup.

This package contains utilities (mkfs, fsck, etc.) used to work with nilfs.

pdumpfs - a daily backup system similar to Plan9's dumpfs

pdumpfs is a simple daily backup system similar to Plan9's dumpfs which preserves every daily snapshot. pdumpfs is written in Ruby. You can access the past snapshots at any time for retrieving a certain day's file. Let's backup your home directory with pdumpfs!

partimage - backup partitions into a compressed image file

Partition Image is a partition imaging utility. It has support for the following file systems:

- * Ext2/3, the Linux standard
- * ReiserFS, a journalised and powerful file system
- * FAT16/32, DOS and Windows file systems
- * HPFS, IBM OS/2 file system
- * JFS, journalised file system, from IBM, used on AIX
- * XFS, another journalised and efficient file system, from SGI, used on Irix
- * UFS (beta), Unix file system
- * HFS (beta), MacOS File system
- * NTFS (experimental), Windows NT, 2000 and XP

Only used blocks are copied and stored into an image file.

The image file can be compressed in the GZIP/BZIP2 formats to save disk space, and split into multiple files to be copied onto removable media (ZIP for example), burned on a CD-R, etc.

This makes it possible to save a full Linux/Windows system with a single operation. In case of a problem (virus, crash, error, etc.), you just have to restore, and after several minutes, your entire system is restored (boot, files, etc.), and fully working.

This is very useful when installing the same software on many machines: just install one of them, create an image, and restore the image on all other machines.

rdiff-backup - remote incremental backup

rdiff-backup backs up one directory to another, possibly over a network. The target directory ends up a copy of the source directory, but extra reverse diffs are stored in a special subdirectory of that target directory, so you can still recover files lost some time ago. The idea is to combine the best features of a mirror and an incremental backup. rdiff-backup also preserves subdirectories, hard links, dev files, permissions, uid/gid ownership, modification times, extended attributes, acls, and resource forks.

Also, rdiff-backup can operate in a bandwidth efficient manner over a pipe, like rsync. Thus you can use rdiff-backup and ssh to securely back a hard drive up to a remote location, and only the differences will be transmitted. Finally, rdiff-backup is easy to use and settings have sensible defaults.

pybackup - user friendly file backup tool for GNOME

A graphical tool to make file backup simpler. Provides a user-friendly interface to rdiff-backup which allows you to back up your files easily to an archive, to a CD/DVD or to a remote location using SSH.

rsnapshot - local and remote filesystem snapshot utility

rsnapshot is an rsync-based filesystem snapshot utility. It can take incremental backups of local and remote filesystems for any number of machines. rsnapshot makes extensive use of hard links, so disk space is only used when absolutely necessary.

slbackup - Skolelinux Backup system

A backup system designed to back up multiple clients to a server. slbackup uses rdiff-backup as the underlying backup software. Clients and hosts can be any computer reachable via an SSH connection (using SSH keys).

tob - Small yet powerful program for tape oriented backups

tob is a simple yet configurable shell-script which, given a set of `volume definitions', runs tar or afio based backups, either for use with tape drives or floppy disks. You could also mount and unmount partitions for backups in the pre- and post-commands. tob is a general driver for the making and maintaining of backups. It makes full backups, differential backups (of the files which were changed since the last full backup), incremental backups (of files changed since all previous backups), lets you determine the size of the backup before actually making it, maintain listings of made backups, make remote backups and possibly more.

unison/unison-gtk - A file-synchronization tool for Unix and Windows

Unison is a file-synchronization tool for Unix and Windows, written in OCaml. It allows two replicas of a collection of files and directories to be stored on different hosts (or different disks on the same host), modified separately, and then brought up to date by propagating the changes in each replica to the other.

Unison offers several advantages over various synchronization methods such as CVS, Coda, rsync, Intellisync, etc. Unison can run on and synchronize between Windows and many UNIX platforms. Unison requires no root privileges, system access or kernel changes to function. Unison can synchronize changes to files and directories in both directions, on the same machine, or across a network using ssh or a direct socket connection.

Transfers are optimised using a version of the rsync protocol, making it ideal for slower links. Unison has a clear and precise specification, and is resilient to failure due to its careful handling of the replicas and its private structures.

vbackup - A modular backup utility

vbackup consists of a set of scripts that handle the various aspects of a backup. It can be easily extended to support other backup methods too. It can be used for full or incremental backups. It also includes a wizard for basic (quick) configuration.

Currently it supports/backups:

- * XFS using xfsdump
- * Generic filesystem backups using tar
- * PostgreSQL
- * MySQL
- * dpkg and RPM database
- * Partition tables, disk MBRs, MD and LVM information