

Backing up in the cloud

Mark Janssen – Sig-I/O Automatisering
<http://sig-io.nl>

Focus of the presentation

- Linux (or at least POSIX) systems
- Personal or small-business use
- Significant amount of bytes
- Not afraid of a little DIY

Why make backups ?

Thursday, August 4, 2011

Third Largest Bitcoin Exchange LOSES Wallet.dat

Bitomat.pl, the leading Polish Bitcoin exchange, as well as the third largest Bitcoin exchange internationally, reported several days ago a total loss of their Bitcoin wallet due to an extremely risky server configuration using Amazon's Elastic Compute Cloud service. Bitomat had chosen to use the EC2 default state configuration, which resets itself to "factory default" after each reboot. After changes to increase the available RAM to the virtual machine, the EC2 instance was rebooted and promptly wiped. No off-site backups were taken, so in effect 17,000 BTC (~\$170,000 USD value) have been "deminted" and will forever be unavailable to the network. Bitomat's owner has put the exchange service up for sale in an attempt to recoup the lost funds and repay wallet owners, you can find a copy of his [official statement](#) below:

Why make backups ?

- Harddisks fail (2-4% per year) [1]
- Accidental deletions / corruptions
- Bitrot
- Theft
- Virusses, hackers and worms

[1] <http://www.pcworld.com/article/129558>

What to backup

- Backup everything
 - Takes more time & space
 - Full-restore easy, partial restore more difficult
- Backup exactly what you need
 - Chance to miss important bits
 - Easy to restore
- Make a rough selection
 - Include /etc, /home, /usr/local, /root, /srv, /var, /opt
 - Exclude /proc, /sys, /mnt, /dev, NFS mounts
 - Think about: /usr, /bin, /sbin, /boot

How and where to backup ^{1/2}

- Backup to another local computer
 - Quick
 - Enough space
 - Not disaster-proof
- Backup to removable media (disk/tape/dvd)
 - Cheap
 - Time consuming
 - Doesn't scale
 - Hard to automate

How and where to backup ^{2/2}

- Backup to a third party
 - Limited speed (upstream network)
 - Scales, pay-as-you-go
 - Easy to automate
 - Offsite by definition
 - Do you trust them to keep your data safe

Linux Backup Tools

- tar, cpio + compression
- Rsync
- Amanda
- Bacula
- Rdiff-backup
- Duplicity
- BackupNinja
- Many others

Tar / CPIO / Rsync

- Familiar to any unix user
- Simple archiving tools
- Easy to use

Amanda

- Advanced backup system
- Full backups or incrementals
- Network Backup
- Many clients
- Aimed at larger sites
- Write to disk, tape, cloud (amazon)
- Open Source, Commercial Support, Mature

Bacula

- Client/Server model
- Advanced feature set
- Complex setup
- Write to disk, tape
- Open Source, Commercial support

rdiff-backup

- Python, seems unmaintained (2009/03)
- Requires same rdiff-backup on remote
- Stores latest version and reverse-delta's
- Creates full 'mirror' on remote system
- Easy to read / restore

Duplicity

- Python, replacement for rdiff-backup
- Doesn't need server component (scp/ftp)
- Stores full version + forward delta's
- Stores encrypted archives/images

BackupNinja

- Simple client side interface
- Frontend for many popular backup tools
- Create backups to remote servers
 - rsync
 - rdiff-backup
 - duplicity
 - others

Cloud Backups:

Important things to think about

- Use crypto !
 - You don't know who has access to your data
- Backups take their time
 - Especially over DSL lines
- So do restores
 - Scales with the amount of data
- Know how to access your data
 - When your system is broken

Cloud Backups: Where to store backups

- Exchange backups with a friend
 - Bandwidth sucks on both sides
- Use free cloud storage providers
- Use specialized cloud backup providers
- Use a Virtual Private Server

Free Storage Providers

Service	Capacity	Max File Size	Protocols	Notes
Microsoft Live Skydrive	25 GB	100 MB	Webdav	
Dropbox	2 GB	200 MB		Client needed
Google Docs	1 GB	1 GB		
SugarSync	5 GB			90 day expiration
Gmail FS	7 GB	25 MB	IMAP / Fuse	
Box.net	5 GB	25 MB		
1fichier.com	unlimited	10 GB	FTP / HTTP	60 days expiration
Cloudfsafe.com	2 GB	2 GB	Webdav	SSL

Payed Cloud Storage Providers

- Many many choices, check wikipedia for a incomplete list:
- http://en.wikipedia.org/wiki/Comparison_of_online_
- Things to look for:
 - Maximum filesize
 - Encrypted storage and transfer
 - Possibility to grow in storage capacity
 - Pricing range: \$5 – 40 per month

Virtual Private Servers

- Private instance, full control
- For larger amounts of data usually cheaper than 'cloud-storage'
- More DIY, less guarantees
- Price range: \$10 – \$60 per month
- Storage: 10GB – 200 GB
- Even larger: Colocation (\$50 – \$200 per month)

Pointers and Links

Sig-I/O

- <http://www.sig-io.nl>

Amanda

- <http://www.amanda.org>
- <http://www.zmanda.com>

Rsync

- <http://rsync.samba.org>

Rdiff-Backup / Duplicity

- <http://www.nongnu.org/rdiff-backup>
- <http://duplicity.nongnu.org>

BackupNinja

- <https://labs.rise-up.net/code/projects/backupninja>

Thank You

- Questions ?
- Advertisement: Sig-I/O sells Virtual Private Servers and starting next month we will be offering Duplicity backup space.
- <http://www.sig-io.nl>