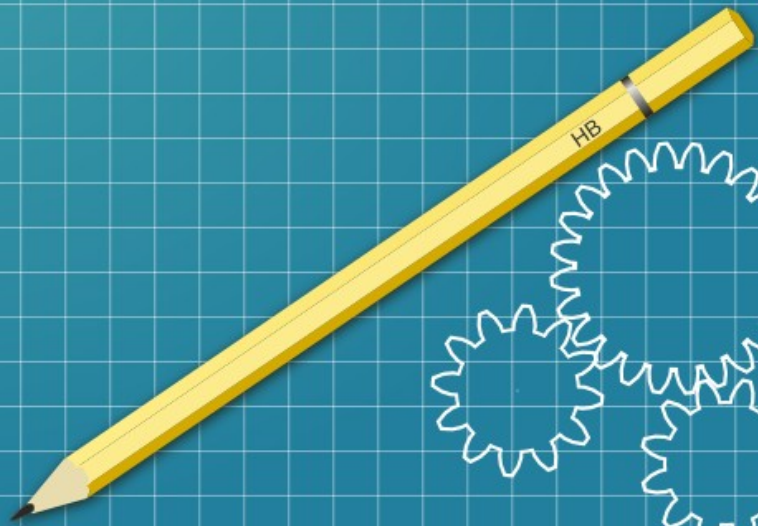
Incremental Backup System

Nick Holland



Backup Systems Suck

- Difficult to set up
  - How does this thing work?
- Difficult to test → uncertain what you actually have
- Difficult to recover data
- Nearly useless for anything but recovery from a disaster
  - Assuming they worked.

How often do you do this?

```
$ ls /etc/pf.conf*
pf.conf
pf.conf-Mar15-2021
pf.conf-old
pf.conf.2022-07-25
```

Isn’t that what your backup system is for?

Your backup solution sucks

Introducing: **Incremental Backup System**

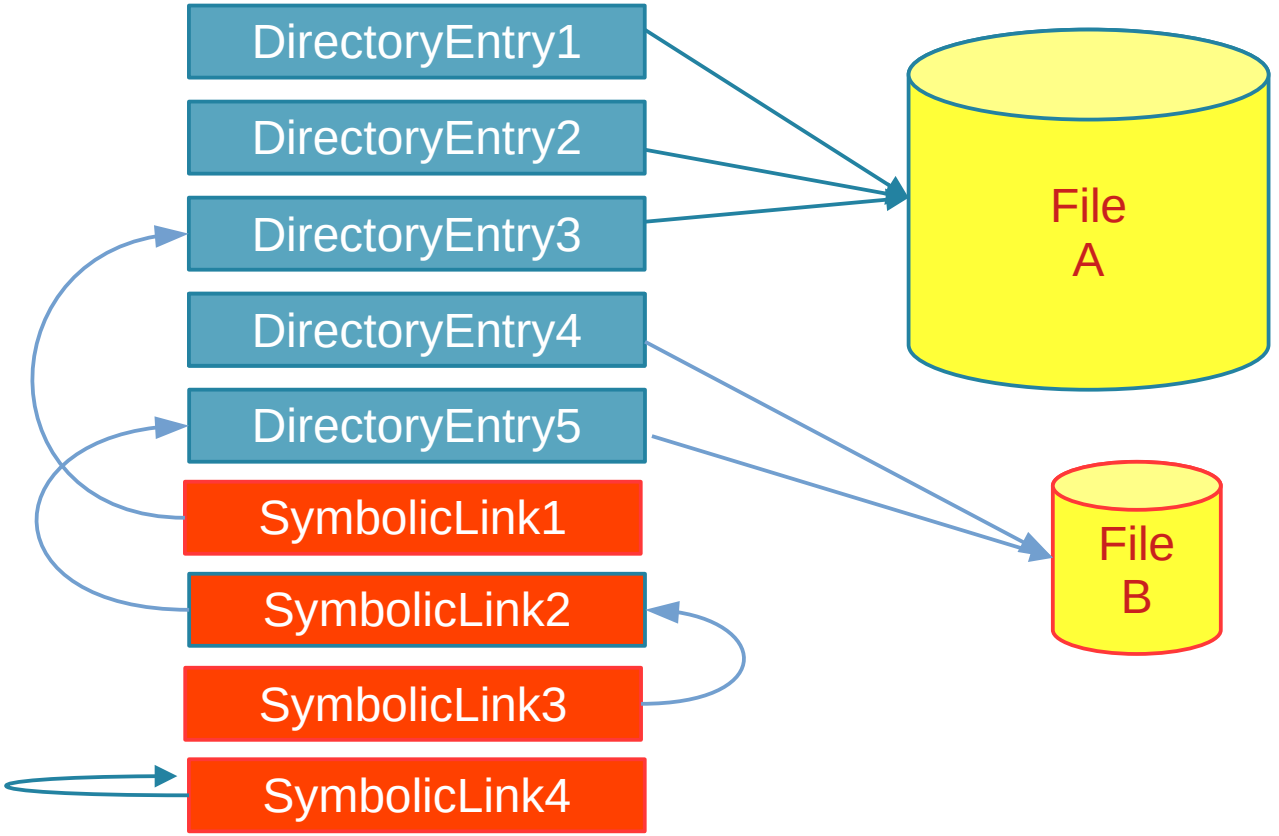
- “Client” software is rsync and SSH
- Backup software is rsync + ksh script.
- Backup hardware is a recycled computer & big disk
- Backups themselves are files in a file system that accurately mimics what was on the source machine
- (Almost) Every backup activity is an “incremental”, but what is on the disk is a “full” backup.

This backup solution sucks less!

Review: hard links vs. symbolic links

- Symbolic link (symlink): points to another directory entry
  - Can cross file systems, can point to directories or files
  - Clearly different from the actual file
  - Yes, like Windows shortcuts
- Hard link: Multiple directory entries (links) pointing at the exact same file on the disk.
  - Can not cross file systems, can only point to files.
  - Each hard link is an “equal” – there is no “master”.
  - As long as ANY link exists, the file exists. When the last link is removed, file is removed from disk.
  - Not at all like Windows shortcuts.

Review: files, links, symlinks



The magic of IBS: `rsync --link-dest`

- Standard rsync: source and destination
  - `rsync -av src:/dir /destdir`
- Three-way rsync: Source, destination, and some other dir
  - `rsync -av **--link-dest /otherdir** src:/dir /destdir`
    - If the file is new/changed, copy from source
    - If the file is unchanged, hard-link from the “link-dest” copy
    - If the file no longer exists, don’t put a copy in the new directory.

Turning `rsync --linkdest` into a backup

- First backup copies everything
- Second backup: New directory
  - Hardlink files that haven't changed from previous BU directory to current BU directory
  - Copy files that have changed
- What moved through the wire was an incremental
- What is now on the disk looks and acts like a full backup!
- All following backups follow same process – hardlink unchanged from previous backup, copy stuff that did change.
- Oldest backup directory is deleted.
- Space freed when last link to an individual file is deleted.

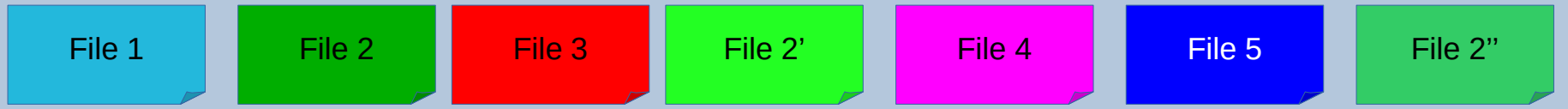
rsync w/links

Backup 1 (full!)
File 1
File 2
File 3

Backup 2
File 1
File 2’ (changed)
(deleted)
File 4 (added)

Backup 3
File 1
File 2’ (no chg)
File 4 (no chg)
File 5 (added)

Backup 4
File 1
File 2’’ (changed)
File 4 (no chg)
File 5 (no chg)



rsync : --filter=”merge <file>”

- Control over what you back up
  - /mnt, /proc, OS files, /dev...
- Filter files provide powerful (but somewhat confusing) control over what goes through rsync
- General gist: you must include everything up to your target, then exclude the stuff you don’t want
- More at: https://holland-consulting.net/tech/rsyncnotes.html

--filter=”merge <file>” is how you select your backup targets

(Tiny) filter-merge file example:

- Task: Backup **`/APP/APP952/BATCH`** and only that.

Incorrect:

```
- /
+ /APP/APP952/BATCH
```

Works:

```
+ /
+ /APP
+ /APP/APP952
+ /APP/APP952/BATCH
- /APP/APP952/*
- /APP/*
- /*
```

- Real world filter files can be simple (+ /, - a few things)
- Real world filter files could be machine generated and complex.

IBS storage structure

- Yesterday's backup:

  `firewall:/etc/pf.conf → bu1:/ibs/firewall/2022-12-19/etc/pf.conf`

- Today's backup:

  `firewall:/etc/pf.conf → bu1:/ibs/firewall/2022-12-20/etc/pf.conf`

- Most Recent Month End:

  `firewall:/etc/pf.conf → bu1:/ibs/firewall/2022-12-01-ME/etc/pf.conf`

- Previous Month End:

  `firewall:/etc/pf.conf → bu1:/ibs/firewall/2022-11-01-ME/etc/pf.conf`

IBS Storage Part 2

- `/v` → home for chunks of storage -- `/v/1, /v/2,` …
- `/ibs` → machine named symlinks to real storage

  `/ibs/firewall → /v/1/firewall`

  `/ibs/fileserver → /v/1/fileserver`

  `/ibs/webserver → /v/2/webserver`
- `/etc/ibs` → Config file, filter files (hard coded)
- `/usr/local/sbin` → scripts

One-line command…

(and 500 lines to make that one line)

- Preflight check: Might this backup work?
- Is this a re-run of today's? Or a first run of the day?
- Is this a month-end? If so, use the -ME rotation
- Identify:
  - Current backup directory (destination)
  - Previous backup directory (--link-dest)
  - Oldest backup in rotation (about to be deleted)
  - Filter file (default or machine specific)
- `rm -r <oldest>` in background; run rsync in background
- Stagger starts, don't run too many simultaneous backups

IBS tips

- Super-fast HW is not a benefit
  - Slow backups mean your systems aren’t bogged down
  - But not too slow – you might need to do a restore or move data!
- Consider encryption for the data store
  - Consider the problems that could create!
- Redundant storage (maybe redundant IBS systems?)
- Backup your backup configuration!! (`/etc/ibs`)
- Restrict access to IBS server
- Potentially good Administrative server/jump box
- rsync delta transfer is probably not your friend.

IBS log files

```
/bu/z-logs $ more node1-2022-12-06
node1 /bu/node1 /bu/node1/2022-12-05 2022-12-06
==== /etc/ibs/node1.bufilter
+ /
+ /usr
- /usr/src
- /usr/obj
- /usr/xenocara
- /usr/xobj
- /usr/share/man
- /dev
- /var/www/ftp/pub
- /bu
Deleting /bu/node1/2022-11-27
receiving incremental file list
./
tmp/
var/backups/
var/cron/
var/cron/log
var/cron/log.0.gz
var/cron/log.1.gz
var/cron/log.2.gz
  ...
  ...
var/spool/smtpd/queue/f0/
var/spool/smtpd/temporary/
var/www/logs/access.log

Number of files: 36,294 (reg: 34,447, dir: 1,780,
link: 63, special: 4)
Number of created files: 34 (reg: 33, dir: 1)
Number of deleted files: 0
Number of regular files transferred: 91
Total file size: 3,695,957,077 bytes
Total transferred file size: 223,648,205 bytes
Literal data: 223,662,041 bytes
Matched data: 0 bytes
File list size: 171,706
File list generation time: 0.001 seconds
File list transfer time: 0.000 seconds
Total bytes sent: 4,168
Total bytes received: 12,755,326

sent 4,168 bytes  received 12,755,326 bytes
16,755.74 bytes/sec
total size is 3,695,957,077  speedup is 289.66
==== BACKUP COMPLETE rc=0 ====
```

Backup Report (`ibs -r`)

From nick <nick@dbu.in.nickh.org>

To nick@dbu.in.nickh.org

Subject **Inhouse backup report (dbu basement)**

Reply Forward Archive Junk Delete Mo

```
Volume      Status              Size Device
softraid0 0 Degraded    4000786726912 sd2     RAID1
          0 Online      4000786726912 0:0.0   noencl <sd0a>
          1 Offline                0 0:1.0   noencl <>
softraid0 1 Online      8001562918912 sd3     CRYPTO
          0 Online      8001562918912 1:0.0   noencl <sd1p>

Disk Space: /v/1 /v/2 /v/3
Mounted on              Size     Used    Avail Capacity
/v/1                    2.1T     1.7T     290G   86%
/v/2                    1.3T     347G     878G   28%
/v/3                    7.2T     6.0T     853G   88%

Last and finished time:
5 Dec 18 05:32 hc1-2022-12-18

System             MostRecent   Oldest   BUs    ME-Recent ME-Oldest  MEs  TotSize   IncSize     vol  rc
================================================================================================================
cvsweb.openbsd.org 2022-12-18 2022-12-10  9    2022-12-01 2022-06-01  7     2201M       56M    /v/1   0
dbu                2022-12-18 2022-12-11  8    2022-12-01 2022-06-01  7     2701M        0M    /v/1   0
dbu1               2022-12-18 2022-12-11  8    2022-12-01 2001-01-01  7     2490M        0M    /v/1   0
fluffy3            2022-12-18 2022-12-02 17    2022-12-01 2022-06-01  7   142600M     4785M    /v/1   0
g2.nickh.org       2022-12-18 2022-12-11  8    2022-12-01 2022-06-01  7     3229M        7M    /v/1   0
gw                 2022-12-18 2022-12-11  8    2022-12-01 2022-06-01  7     2603M        6M    /v/1   0
```

Backup Report

```
Volume      Status               Size Device
softraid0 0 Degraded    4000786726912 sd2     RAID1
          0 Online      4000786726912 0:0.0   noencl <sd0a>
          1 Offline                0 0:1.0   noencl <>
softraid0 1 Online      8001562918912 sd3     CRYPTO
          0 Online      8001562918912 1:0.0   noencl <sd1p>

Disk Space: /v/1 /v/2 /v/3
Mounted on               Size     Used    Avail Capacity
/v/1                     2.1T     1.7T     290G   86%
/v/2                     1.3T     347G     878G   28%
/v/3                     7.2T     6.0T     853G   88%

Last and finished time:
5 Dec 18 05:32 hc1-2022-12-18
```

Backup Report

System	MostRecent	Oldest	BUs	ME-Recent	ME-Oldest	MEs	TotSize	IncSize	vol	rc
cvsweb.openbsd.org	2022-12-18	2022-12-10	9	2022-12-01	2022-06-01	7	2201M	56M	/v/1	0
dbu	2022-12-18	2022-12-11	8	2022-12-01	2022-06-01	7	2701M	0M	/v/1	0
dbu1	2022-12-18	2022-12-11	8	2022-12-01	2001-01-01	7	2490M	0M	/v/1	0
fluffy3	2022-12-18	2022-12-02	17	2022-12-01	2022-06-01	7	142600M	4785M	/v/1	0
g2.nickh.org	2022-12-18	2022-12-11	8	2022-12-01	2022-06-01	7	3229M	7M	/v/1	0
gw	2022-12-18	2022-12-11	8	2022-12-01	2022-06-01	7	2603M	6M	/v/1	0
gw.universalbearin	2022-12-18	2022-12-11	8	2022-12-01	2022-06-01	7	353M	6M	/v/1	0
hc1	2022-12-18	2022-12-10	9	2022-12-01	2022-06-01	7	588204M	139M	/v/1	0
hc1p	2022-12-17	2022-11-29	5	2022-12-01	2022-07-01	5	4985466M	52956M	/v/3	0
man.openbsd.org	2022-12-18	2022-12-08	11	2022-12-01	2022-07-01	6	8650M	297M	/v/1	0
monitor.nickh.org	2022-12-18	2022-12-11	8	2022-12-01	2022-07-01	6	2533M	27M	/v/1	0
node1	2022-12-18	2022-12-11	8	2022-12-01	2022-07-01	6	3394M	171M	/v/1	0
node2	2022-12-18	2022-12-11	8	2022-12-01	2022-07-01	6	2689M	3M	/v/1	0
obsdts.cs.toronto.	2022-12-18	2022-12-11	8	2022-12-01	2022-06-01	7	2602M	3M	/v/1	0
suzy2	2022-11-13	2000-00-07	8	2001-02-01	2000-00-00	3	22196M	13332M	/v/1	0
universalbearing.c	2022-12-18	2022-12-11	8	2022-12-01	2022-07-01	6	1791M	164M	/v/1	0
web.holland-consul	2022-12-18	2022-12-10	9	2022-12-01	2022-06-01	7	309161M	20M	/v/2	0
console	2022-12-18	2022-12-12	7	2022-12-01	2000-00-11	6	3478M	0M	/v/2	0

Customizing to your environment

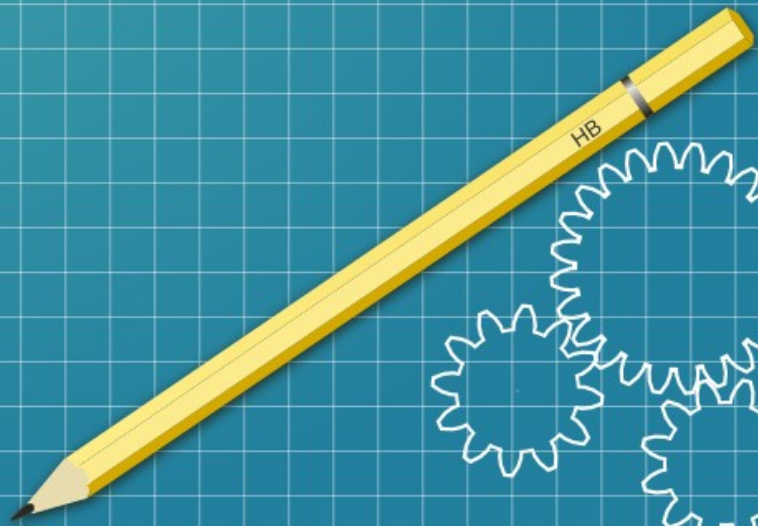
- Machine generated backup filter files
- “on-machine” backups: `rsync --linkdest` makes a great localhost config backup system.
- Force all users & groups to non-root
- Run AV against backups (save putting an AV on every host)
- ZFS snapshots instead of `-link-dest` (& `zfs send`)

I told you that story to tell you this one...

File Alteration Reporting Tool

A security tool that doesn’t stink!

Nick Holland
Putting the SH in IT since 1999



File Alteration Reporting Tool

- Goal: File integrity by spotting unexpected changes to files
- Ultimate goal: Silence except when there's a problem.
- Reality: ultimate goal is futile. For example, `/etc/hosts`
  - COULD be malware redirecting connections
  - COULD be new data added properly by an administrator
  - COULD be new data added incorrectly by an administrator!
- Human review is going to be a requirement. Sorry.
- In real life: found the human review very useful!

How to make a F.A.R.T.

- sha256 on each file? Massive amounts of work
- BUT WAIT!...
- … The IBS log *IS* a list of changed files!!!!
- So … regex-out all the files you expect changes on, what's left is the unexpected.
- `grep -vf <exclusionfile>` is the entire magic. One line!
- ...except for the few hundred lines of script to figure out what the exclusion file should have in it and make it look pretty.

Exclusion Files

- Global exclusions
- Machine specific exclusions
- Special Event exclusions (i.e., OS or application update)

  … all the above combined.
- Exclusion files should support comments and “obvious” syntax
  - `grep -f` doesn’t support comments
  - IBS directory structure complicates things slightly
  - rsync output complicates things

Filter files need pre-processing

- Backup log shows:

```
home/nick/.cache/chromium/Default/Cache/Cache_Data/2013dd9c844666a4_0
```

- Exclusion should **look** like:

```
/home/nick/.cache/  # daily chatter from browser use
man/mandoc.db$ # couple places in OpenBSD, lots of places in my stuff
```

- Exclusion should **act** like:

```
^/home/nick/.cache/
man/mandoc.db$
```

- Tolerate extra white space
- So … exclusion files need to be processed and combined, written to a temp file, and the used as a filter file with grep.

Log files need stripping

- Strip the IBS job info at the top
- Strip the summary at the bottom
- Strip random rsync alerts

Output needs processing

- Limit output to a small number of lines per machine.
  - Typically either very little or a huge flood
  - Floods of output typically due to one thing
    - (but might be hiding something important!)
  - Full output available: `-a`
  - Copy/paste diffs available: `-d`
- Deal with files that are touched/rebuilt, but unchanged
- Make output pretty and readable

Usage Considerations

- Change is cumulative. Today's clean report means nothing if yesterday's report wasn't clean (and inspected)
- Lots of comments in the filter files, change control
- No output isn't the goal. Learning things is
- In corporate environment:
  - TWO people charged with looking over every F.A.R.T. report
    - (maybe on call and previous on-call person?)
    - E-mail sent to team saying what was observed.
    - Goal is to learn something, not to say, "no problem seen".
  - Doesn't have to be done every day, but every day needs to be done

Results seen in real life

- Unexpected (and un-consulted) changes made by admins
- Activity of outside vendors with access
- “Secret” administrative tools left behind by vendors
- Major configuration changes made by other teams
- Bad things...

Quickly went from a compliance check to useful tool!

F.A.R.T. Output

Subject **FART report (dbu basement)**

```
   home/nick/calendar.aniv

======== /bu/z-logs/g2.nickh.org-2022-12-18
   home/nick/calendar.aniv

======== /bu/z-logs/gw-2022-12-18
   etc/pf.conf
   var/backups/etc_pf.conf.backup
   var/backups/etc_pf.conf.current

======== /bu/z-logs/universalbearing.com-2022-12-18
   var/www/data/Switchvox_Backup_F20221218040006-KK2900-72248-350-business.svb

======== /bu/z-logs/web.holland-consulting.net-2022-12-18
   home/nick/.ssh/control/gw.nickh.org:test:22

EOT generated from /bu/z-logs/console-2022-12-18 /bu/z-logs/cvsweb.openbsd.org-2022-12-18 /bu/z-logs/dbu-
2022-12-18 /bu/z-logs/dbu1-2022-12-18 /bu/z-logs/fluffy3-2022-12-18 /bu/z-logs/g2.nickh.org-2022-12-18 /bu/z-
logs/gw-2022-12-18 /bu/z-logs/gw.universalbearing-2022-12-18 /bu/z-logs/hc1-2022-12-18 /bu/z-
```

Summary

- Dirvish (inspiration), rsnapshot (“competition”), Tarsnap (compliment)
- http://holland-consulting.net/scripts/ibs
- Windows through WSL???
- Secure HTML Alteration Reporting Tool. (someday?)
  - Use the F.A.R.T. concepts to populate a set of (static?) web pages allowing a one-pass “drill down” for more info.

Demo / Questions