

Rollback – Myth and Truth

How It Works and How to Avoid Pitfalls

Thorsten Kukuk

Senior Architect SUSE Linux Enterprise Server

kukuk@suse.com



Me :-)

- Master degree in computer science (Dipl.-Inf.)
- Started with Linux 1992
- Working for SUSE since 1999
 - Senior Architect SUSE Linux Enterprise Server
 - Formerly Release Manager of SUSE Linux Enterprise Server
- Open Source development:
 - Glibc
 - NIS
 - Linux-PAM

rm -rf / ?

Rollback, what is needed for it, what I want to talk about today:

- Btrfs in 60 seconds
- Snapshots on SUSE Linux Enterprise Server 11
- Rollback on SUSE Linux Enterprise Server 12
- Grub2 and rollback
- Caveats and risks
- Managing subvolumes

Btrfs in general

- Copy on Write (CoW) general purpose file system
- Provides snapshotting
- Trees for
 - Data
 - Metadata
- Details
 - Every snapshot is again a subvolume of its own
 - Snapshots (as subvolumes) can be mounted and accessed as every other subvolume
 - Snapshots can be created read-only

Btrfs

on SUSE Linux Enterprise Server 11

- Supported as root file system since SUSE Linux Enterprise Server 11 SP2
- Advantage: snapshotting
- Bootloader cannot boot from btrfs
 - /boot on extra partition needed, not on btrfs
- File based rollback / “undochange”
- Integrated in zypper and YaST2
- Snapshots vs. different partitions and subvolumes
- No consistent snapshots cross partition boundaries
- Automatic snapshots every hour for / root filesystem

Snapper Functionality

- Manage snapshots
 - Automatically create snapshots
- Display differences between snapshots
- Rollback (“undochange”)
- Integrated with systems management stacks (YaST, zypp, yum, ...)
- User Interfaces
 - CLI
 - DBUS
 - Graphical (YaST)

Btrfs integration with ...

- Installer

- Btrfs as root fs
- Recommendation for subvolume layout
- Check size of root partition
- Automated snapper configuration

- Partitioner

- Create Btrfs
- Create subvolumes

File Based Rollback

How it works

- single files are copied from the snapshot to the “root” subvolume – using clone ioctl

Benefits

- Subvolumes are treated as read-only
- Subvolumes can be used for Backup
- Supports Pick and Choose

Disadvantages

- rollback not “atomic”
- /boot partition (kernel!) excluded

→ Implemented in Snapper as “undochange”



But ...

... what happens, if the kernel does not boot?

Btrfs

on SUSE Linux Enterprise Server 12

- Default root filesystem for root partition
- Advantage: rollback including kernel
- Bootloader can boot from btrfs
- Rollback per subvolume
- Rollback integrated into bootloader
- No consistent snapshots cross partition boundaries
- No automatic snapshots per cron for root filesystem

Snapshot/Rollback – Overview

Past (SLES 11)

- “snapper undochange”
- Selective Rollback for Package updates
Administrative changes
- No rollback of Kernel / initrd
Bootloader
System data, e.g. /var/log

Present (SLES 12)

- “snapper rollback”
- Full Rollback for Package updates
Administrative changes
- **Kernel / initrd (initramfs)**
- No rollback of Bootloader
- Customer data: “/home”, if on own partition (default)
- System data, e.g. /var/log

High Demand



Rollback per Subvolume

How it works

- Instead of the original subvolume, the snapshot is mounted with the options “subvol=<name>”
 - Remember: snapshots are subvolumes
- “*btrfs subvolume set-default ...*” for permanent assignments

Benefits

- “atomic” operation
- Very fast

Disadvantages

- Additional complexity
 - Require explicit mounting of subvolumes
- No “rollback” per single file

→ Implemented in Snapper as “rollback”

Full System Rollback

Reboot *Later* Mode

- Administrator is in a current read-write filesystem, but wants to rollback
- “snapper list” to view and select a snapshot
- Call

“snapper rollback <number>”
this will

- Create a new read-only snapshot of the currently running system
- Create a new read-write snapshot of the snapshot <number>, linearly after the just recently created read-only snapshot
- “setdefault” to the new read-write snapshot

and reboot

Reboot Now Mode

- Boot into an existing read-only snapshot
- Text console and some services should work
 - Most data is in writeable subvolumes
- To continue to work in this snapshot, the admin should call

“snapper rollback”

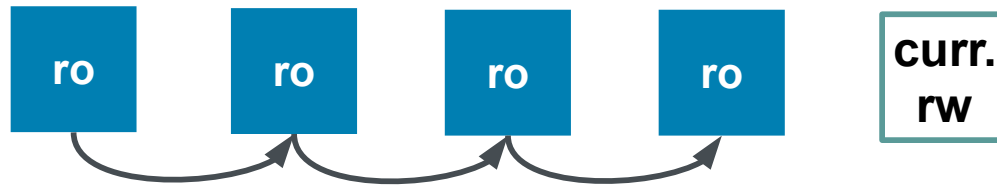
- this will
 - Create a new read-only snapshot of the old read-write one, linearly after the last existing one
 - Create a new read-write snapshot of the snapshot you are currently in in read-only mode, linearly after the just recently created read-only snapshot
 - “setdefault” to the new read-write snapshot

and reboot

User View on Snapshot History

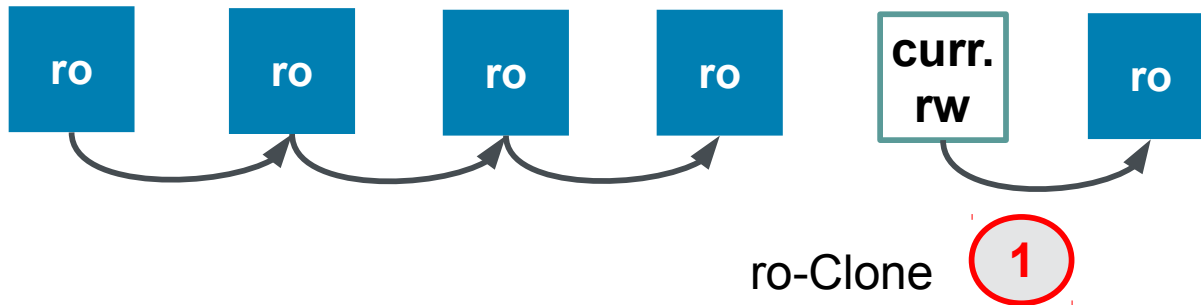
Snapshot / Rollback

User View on Snapshot History (1)

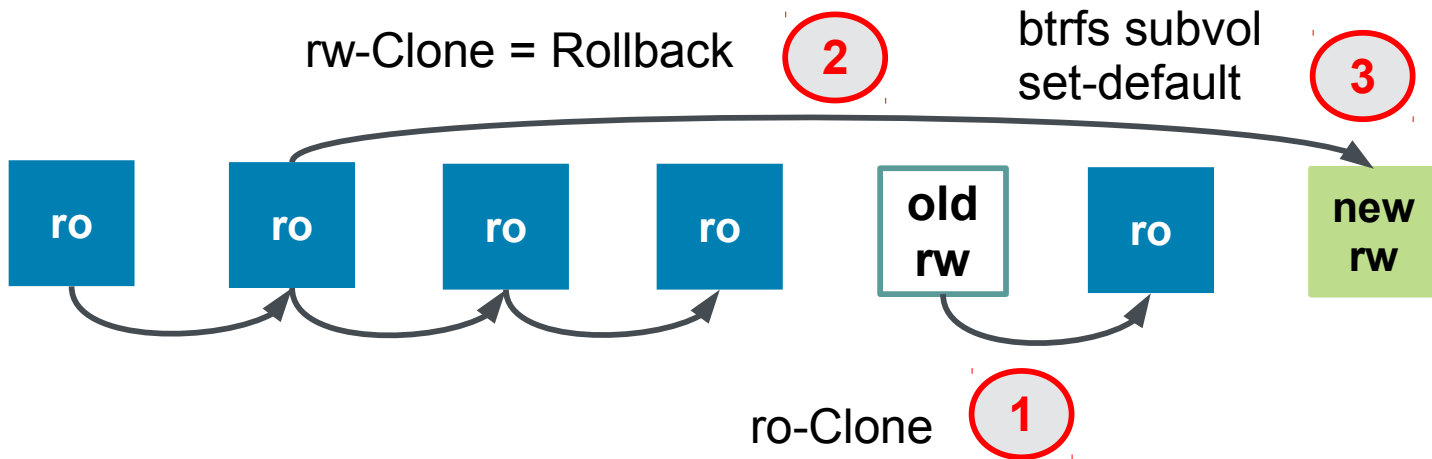


Snapshot / Rollback

User View on Snapshot History (2)

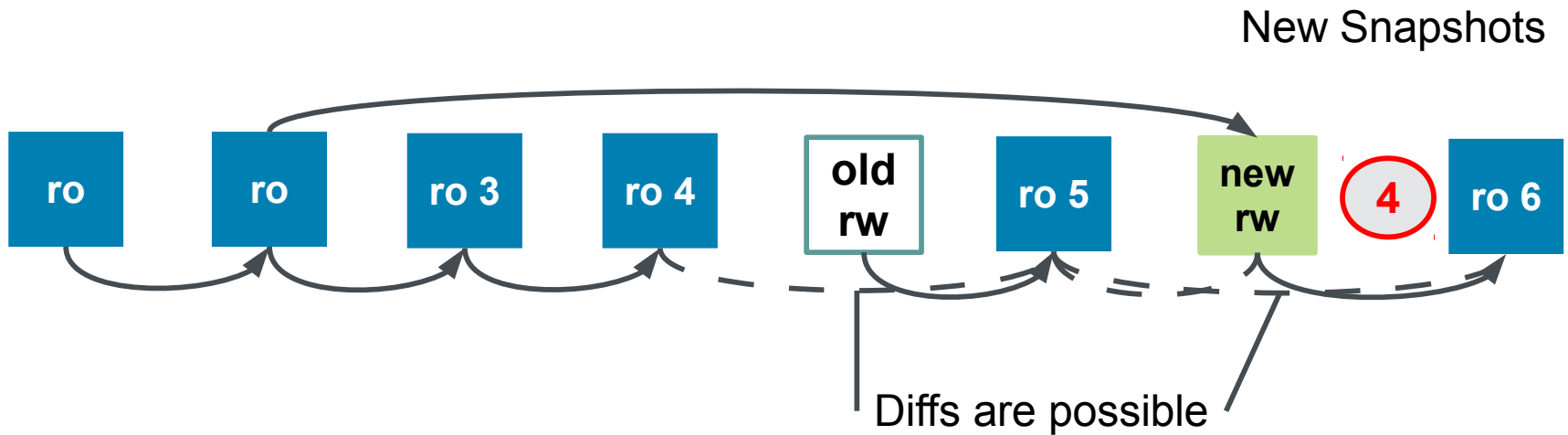


User View on Snapshot History (3)



Snapshot / Rollback

User View on Snapshot History (4)



Snapshot / Rollback

User View on Snapshot History (5)

Condensed view

What happens,
if we rollback again?



Caveat: this does not reflect 1:1 what happens technically.



User View On grub2 Interface

Snapper Headers

- **Type:** [Pre | Post | Single]
- **#:** Nr of snapshot
- **Pre #:** if type is “Post” the matching Pre nr.
- **Date:** timestamp
- **Cleanup:** cleanup algorithm for this snapshot
- **Description:** A fitting description of the snapshot (free text)
- **Userdata:** key=value pairs to record all sorts of useful information about the snapshot in an easily parsable format

Important Snapshots

Snapshots are marked as important (“*”), if a package affecting the boot process is updated:

- kernel
- dracut
- glibc
- systemd
- udev

You can configure that in `/etc/snapper/zypp-plugin.conf`





SLES12

Advanced options for SLES12

Start bootloader from a read-only snapshot



SLES 12 (3.12.28-4,2014-10-17T09:07,pre,zypp{zypper})

SLES 12 (3.12.28-4,2014-10-17T08:59,post)

*SLES 12 (3.12.28-4,2014-10-17T08:59,post)

*SLES 12 (3.12.28-4,2014-10-17T08:59,pre,zypp{y2base})

SLES 12 (3.12.28-4,2014-10-17T08:53,pre,yast sw_single)

SLES 12 (3.12.28-4,2014-10-17T08:46,post)

SLES 12 (3.12.28-4,2014-10-17T08:46,pre,yast users)

```

g227:~ # snapper list
Type | # | Pre # | Date | User | Cleanup | Description | Userdata
-----|-----|-----|-----|-----|-----|-----|-----
single | 0 | | | root | | current | |
pre | 3 | | Fri Oct 17 10:46:22 2014 | root | number | yast users | |
post | 4 | 3 | Fri Oct 17 10:46:49 2014 | root | number | | |
pre | 5 | | Fri Oct 17 10:53:14 2014 | root | number | yast sw_single | |
post | 6 | | Fri Oct 17 10:59:12 2014 | root | number | zypp(y2base) | important=yes
post | 7 | 6 | Fri Oct 17 10:59:16 2014 | root | number | | important=yes
post | 8 | 5 | Fri Oct 17 10:59:19 2014 | root | number | | |
pre | 9 | | Fri Oct 17 11:07:17 2014 | root | number | zypp(zypper) | important=no
post | 10 | 9 | Fri Oct 17 11:07:18 2014 | root | number | | important=no
g227:~ # _

```



Bootable snapshot #4

SLES12

Advanced options for SLES12

```
Starting System Logging Service...
[ OK ] Started Permit User Sessions.
[ OK ] Started /etc/init.d/boot.local Compatibility.
[ OK ] Started Name Service Cache Daemon.
[ OK ] Reached target User and Group Name Lookups.
Starting Login Service...
[ OK ] Reached target Host and Network Name Lookups.
Starting Wait for Plymouth Boot Screen to Quit...
Starting Terminate Plymouth Boot Screen...
[ OK ] Started Wait for Plymouth Boot Screen to Quit.
Starting Getty on tty1...
[ OK ] Started Getty on tty1.
[ OK ] Reached target Login Prompts.
Starting /etc/init.d/after.local Compatibility...
[ OK ] Started /etc/init.d/after.local Compatibility.
[ OK ] Started Terminate Plymouth Boot Screen.
[ OK ] Started Login Service.
[ OK ] Started wicked DHCPv6 supplicant service.
[ OK ] Started wicked AutoIPv4 supplicant service.
[ OK ] Started Update cron periods from /etc/sysconfig/btrfsmaintenance.
[ OK ] Started wicked DHCPv4 supplicant service.
Starting wicked network management service daemon...
[ OK ] Started wicked network management service daemon.
Starting wicked network nanny service...
[ OK ] Started wicked network nanny service.
Starting wicked managed network interfaces...
[ OK ] Started System Logging Service.
[ OK ] Started Load kdump kernel on startup.

Welcome to SUSE Linux Enterprise Server 12 (x86_64) - Kernel 3.12.28-4-default (tty1).

linux-ake5 login: root
Password:
Last login: Fri Oct 17 11:26:47 on tty1
linux-ake5:~ #
```

Caveats and Risks

Snapshotting “/” – Challenges

- Consistent system / “atomic”
 - We cannot do that cross partition boundaries
- Kernel and initrd / initramfs = “/boot”
 - /boot not as extra partition
- Different stages of bootloader needs to match
 - Exclude /boot/grub2/<grub2 arch> from snapshot
 - Grub2 configuration is part of the snapshot
 - new grub2 needs to be able to read old configs

Data and Rollback

I made a rollback, but ...

... what happens with my new data???

System Integrity and Compliance

Don't allow to roll back certain log-files, databases etc.

Solution: subvolumes instead of directories for

- /boot/grub2/*
- /opt
- /srv
- /tmp
- /usr/local
- /var/crash
- /var/lib/{mailman,named,pgsql} (No mysql!)
- /var/log
- /var/opt
- /var/spool
- /var/tmp

What Can Be Broken After a Rollback?

- /home/<user> exists
 - but no entry in /etc/passwd
- /opt can contain Add-Ons like OES,
 - but dependencies are no longer fulfilled
- /srv can contain web applications
 - but wrong php/ruby on rails version installed
- Database was not in a subvolume or extra partition
 - all data after creating the snapshot for rollback are lost

What Can Be Broken After a Rollback?

- Create a snapshot and reboot immediately
 - Still high risk of data loss!
 - Still time between taking snapshot and reboot
 - Enough time for processes to modify data
- Copy modified data from snapshot of old root subvolume into new root subvolume

Managing Subvolumes

Handling of Subvolumes

Subvolumes are automatically mounted with their parent volume

- SLES 11: Only the root subvolume in /etc/fstab
- SLES 12: All subvolumes are listed in /etc/fstab

Why?

- After rollback, old subvolumes are not part of the new parent subvolume!

How to Create/Delete a Subvolume

- Create a subvolume
 - Move old directory (/path/name) away
 - *btrfs subvolume create /path/name*
 - Move old data into new subvolume
 - Add new subvolume to /etc/fstab

- Delete a subvolume
 - Create temporary directory /path/name.tmp
 - Copy data into this directory
 - *btrfs subvolume delete /path/name*
 - Remove /path/name from /etc/fstab
 - *mv /path/name.tmp /path/name*

Automatic Cleanup of Snapshots

Old snapshots are automatically deleted

- Timeline Snapshots
 - Disabled by default for root partition
 - First snapshot of the last 10 days/months/years are kept
- Installation Snapshots/Administration Snapshots
 - New snapshot when calling YaST or zypper
 - Last 10 important snapshots are kept
 - Last 10 “regular” snapshots are kept

Questions?

Thank you.



Missing Answer

rm -rf /

Is it now safe?

No!

Why not?

It will not stop on subvolumes.



Unpublished Work of SUSE LLC. All Rights Reserved.

This work is an unpublished work and contains confidential, proprietary and trade secret information of SUSE LLC. Access to this work is restricted to SUSE employees who have a need to know to perform tasks within the scope of their assignments. No part of this work may be practiced, performed, copied, distributed, revised, modified, translated, abridged, condensed, expanded, collected, or adapted without the prior written consent of SUSE. Any use or exploitation of this work without authorization could subject the perpetrator to criminal and civil liability.

General Disclaimer

This document is not to be construed as a promise by any participating company to develop, deliver, or market a product. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. SUSE makes no representations or warranties with respect to the contents of this document, and specifically disclaims any express or implied warranties of merchantability or fitness for any particular purpose. The development, release, and timing of features or functionality described for SUSE products remains at the sole discretion of SUSE. Further, SUSE reserves the right to revise this document and to make changes to its content, at any time, without obligation to notify any person or entity of such revisions or changes. All SUSE marks referenced in this presentation are trademarks or registered trademarks of Novell, Inc. in the United States and other countries. All third-party trademarks are the property of their respective owners.

