

Gluing a Desktop and a Kernel together

David Zeuthen <davidz@redhat.com>

Red Hat, Inc.

linux.conf.au

Sydney, Australia, January 2007

Kernel.. Desktop.. huh?

- » **Different meaning to different people**
 - » Hackers, “Info Workers”, Home users ... others

Kernel.. Desktop.. huh?

» “Kernel”

» Linux, FreeBSD, Solaris ... others

Kernel.. Desktop.. huh?

- » “Desktop”
 - » GNOME, KDE, XFCE ... others
- » But also
 - » Maemo (Nokia N770 / N800) and OLPC's XO

Desktop/Kernel Integration

- » Today driven primarily by desktop people
 - » GNOME: Project Utopia, KDE: Solid, others
- » **freedesktop.org**
 - » desktop and vendor agnostic
- » Projects
 - » D-Bus, HAL, NM, X.org, Avahi ... others

Examples

Example

- » Attaching a digital camera

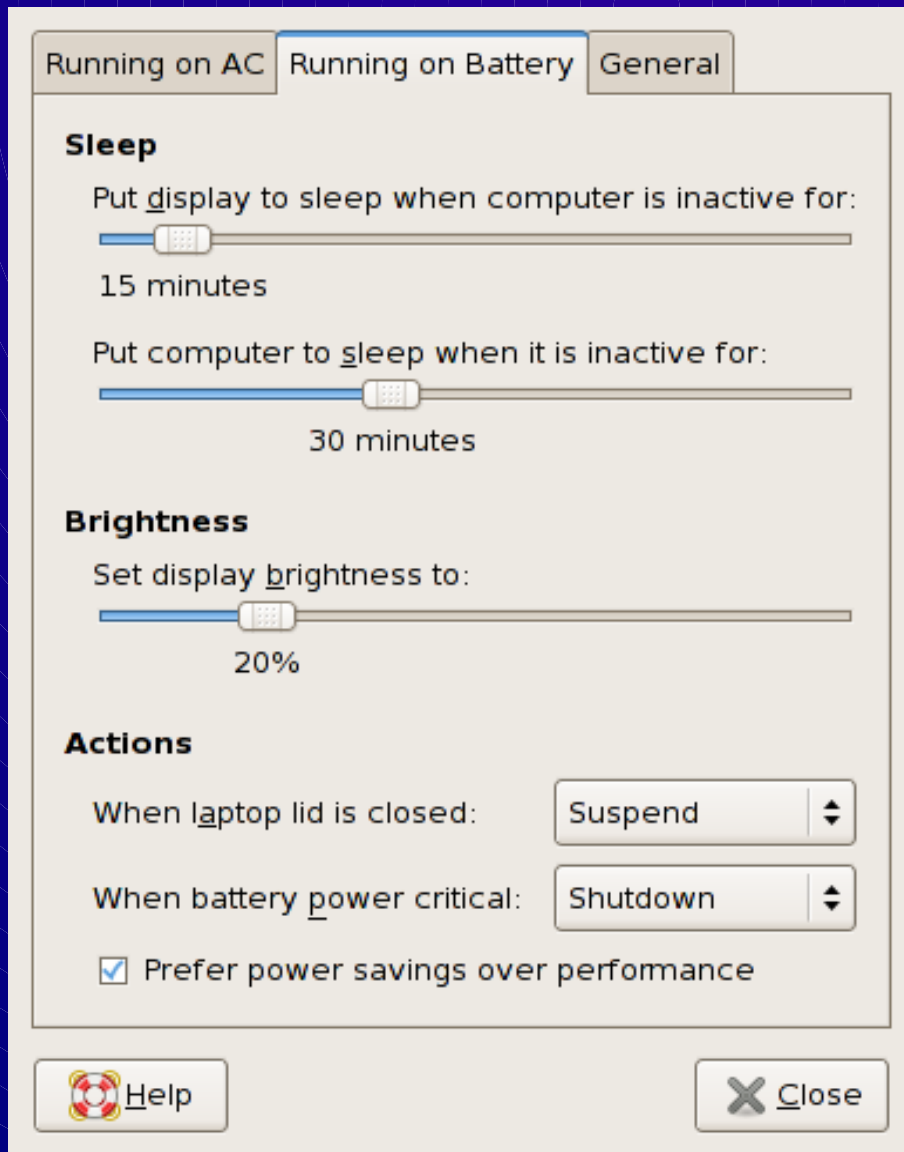


Compare with...

» `$ f-spot --import /media/EOS_DIGITAL`

Example

» Integrated Power Management



Running on AC | **Running on Battery** | General

Sleep

Put display to sleep when computer is inactive for:

15 minutes

Put computer to sleep when it is inactive for:

30 minutes

Brightness

Set display brightness to:

20%

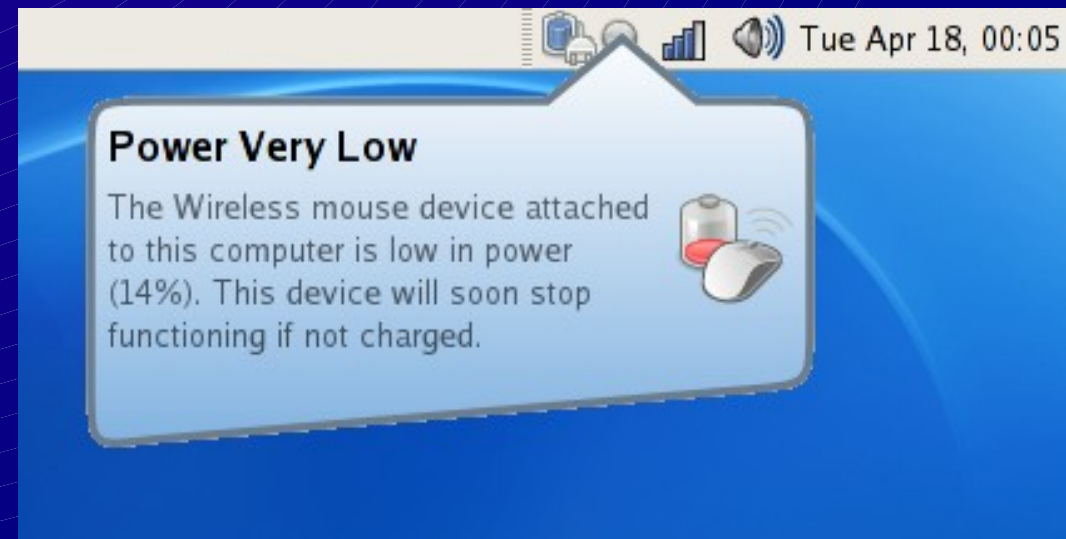
Actions

When laptop lid is closed: Suspend

When battery power critical: Shutdown

Prefer power savings over performance

Help Close



Tue Apr 18, 00:05

Power Very Low

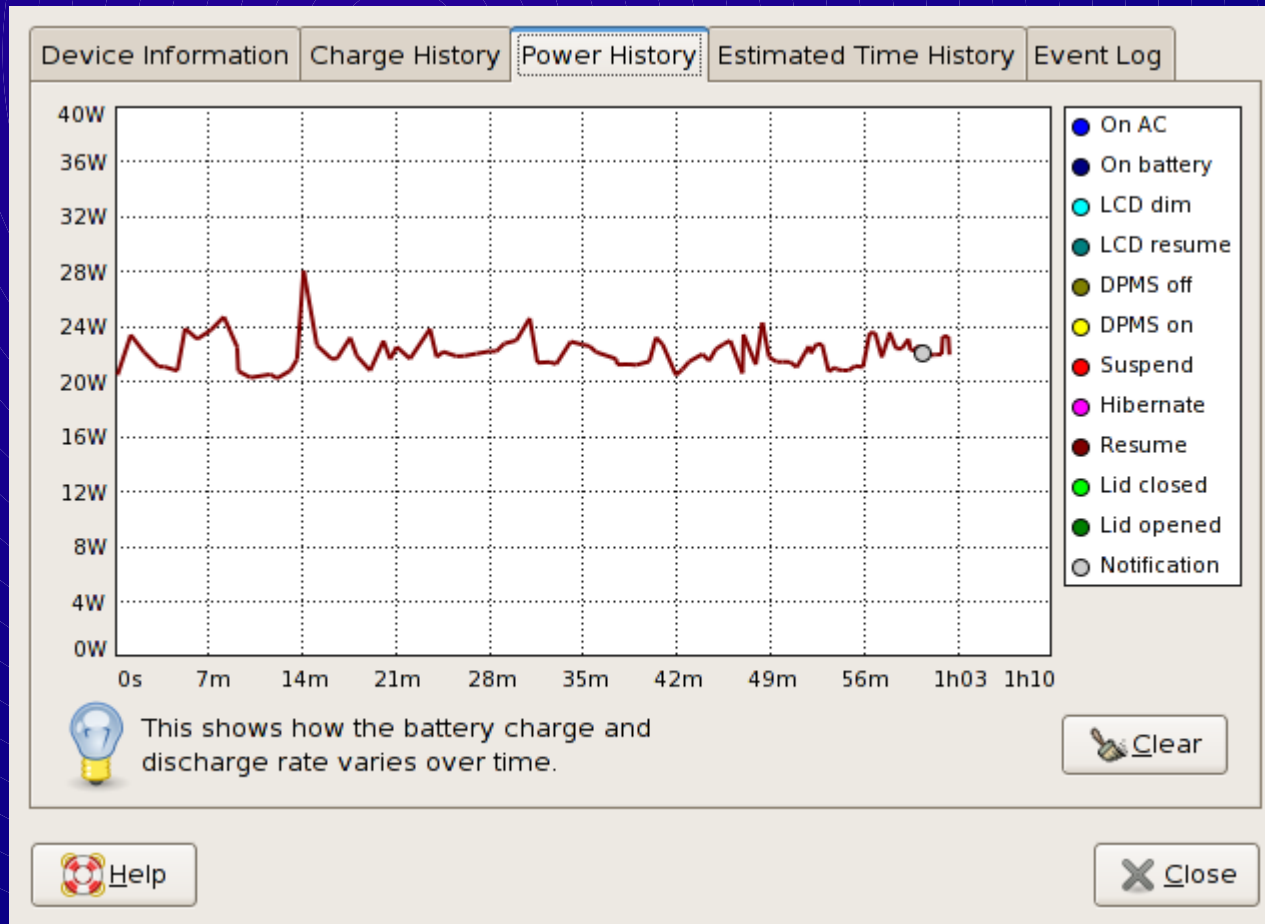
The Wireless mouse device attached to this computer is low in power (14%). This device will soon stop functioning if not charged.

Compare with...

» # emacs /etc/acpi/...

Example

- » Power Management – useful eye-candy

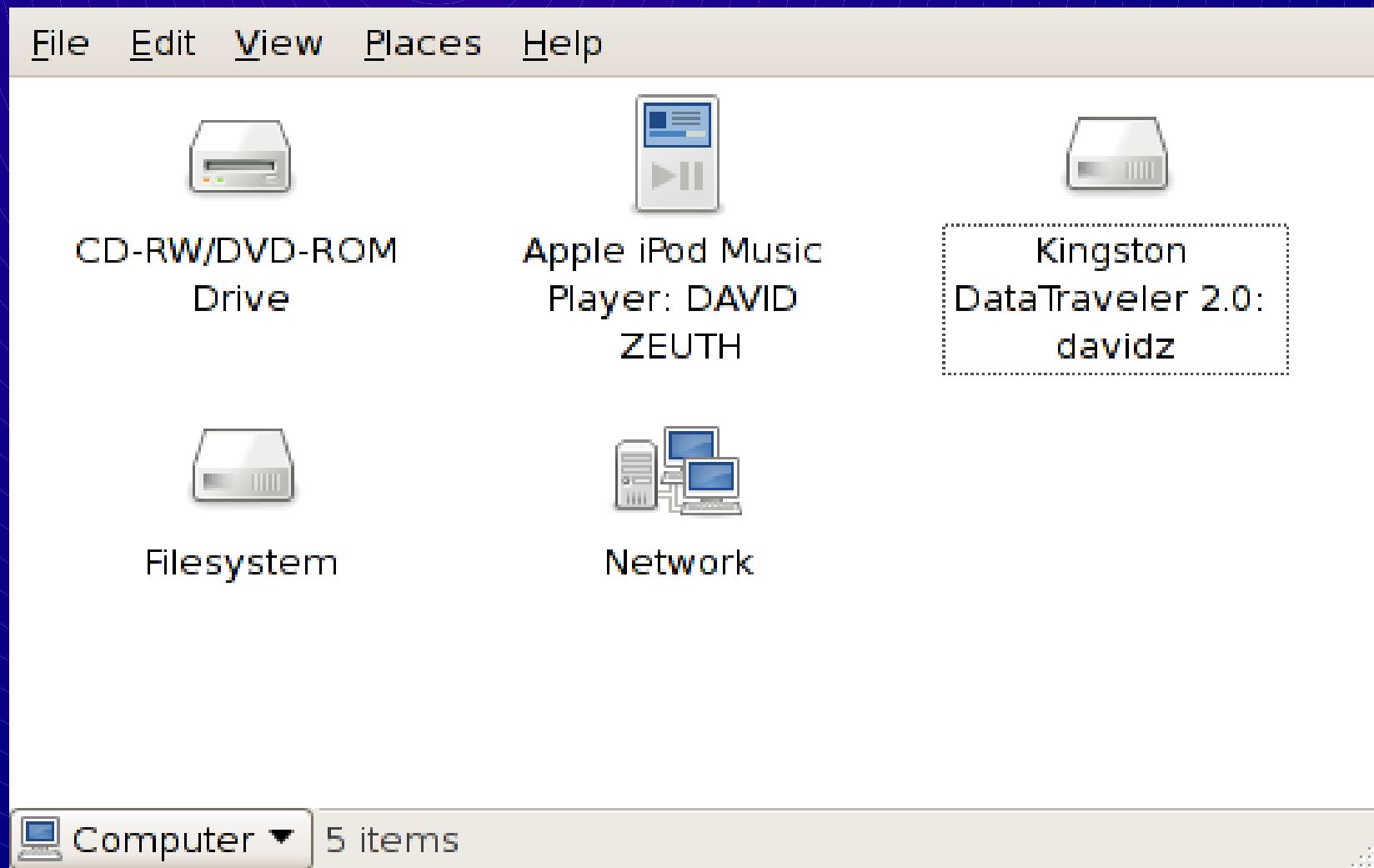


Compare with...

» `$ cat /proc/acpi/battery/BAT0/state`

Example

» Peripherals



Compare with...

» `$ tree /sys/block | less`

Example

» Peripherals cont'd – customization

Basic Emblems Permissions Notes Drive **Volume**

Label: davidz
Size: 481.0 MB
Media: Removable Hard Disk
UUID: 456C-AAA8
File System: vfat (FAT16)

Mount Point: /media/foobar
File System: msdos
Mount Options: rw nosuid nodev noexec uid=500 fmask=0022 dmask=0022 codepage=cp850



▼ **Settings**

Mount Point:

File System:

Mount Options:

Changes in settings will not take effect until the volume is remounted.

 Help  Close

Compare with...

» # emacs /etc/fstab

Wait a minute...

Is this good UI?

Basic Emblems Permissions Notes Drive **Volume**

Label: davidz
Size: 481.0 MB
Media: Removable Hard Disk
UUID: 456C-AAA8
File System: vfat (FAT16)

Mount Point: /media/foobar
File System: msdos
Mount Options: rw nosuid nodev noexec uid=500 fmask=0022 dmask=0022 codepage=cp850



▼ **Settings**

Mount Point:

File System:

Mount Options:

Changes in settings will not take effect until the volume is remounted.

 Help  Close

Uhm.. no..
(but it's fixable)

Example

- » Peripherals cont'd – encrypted thumb drive

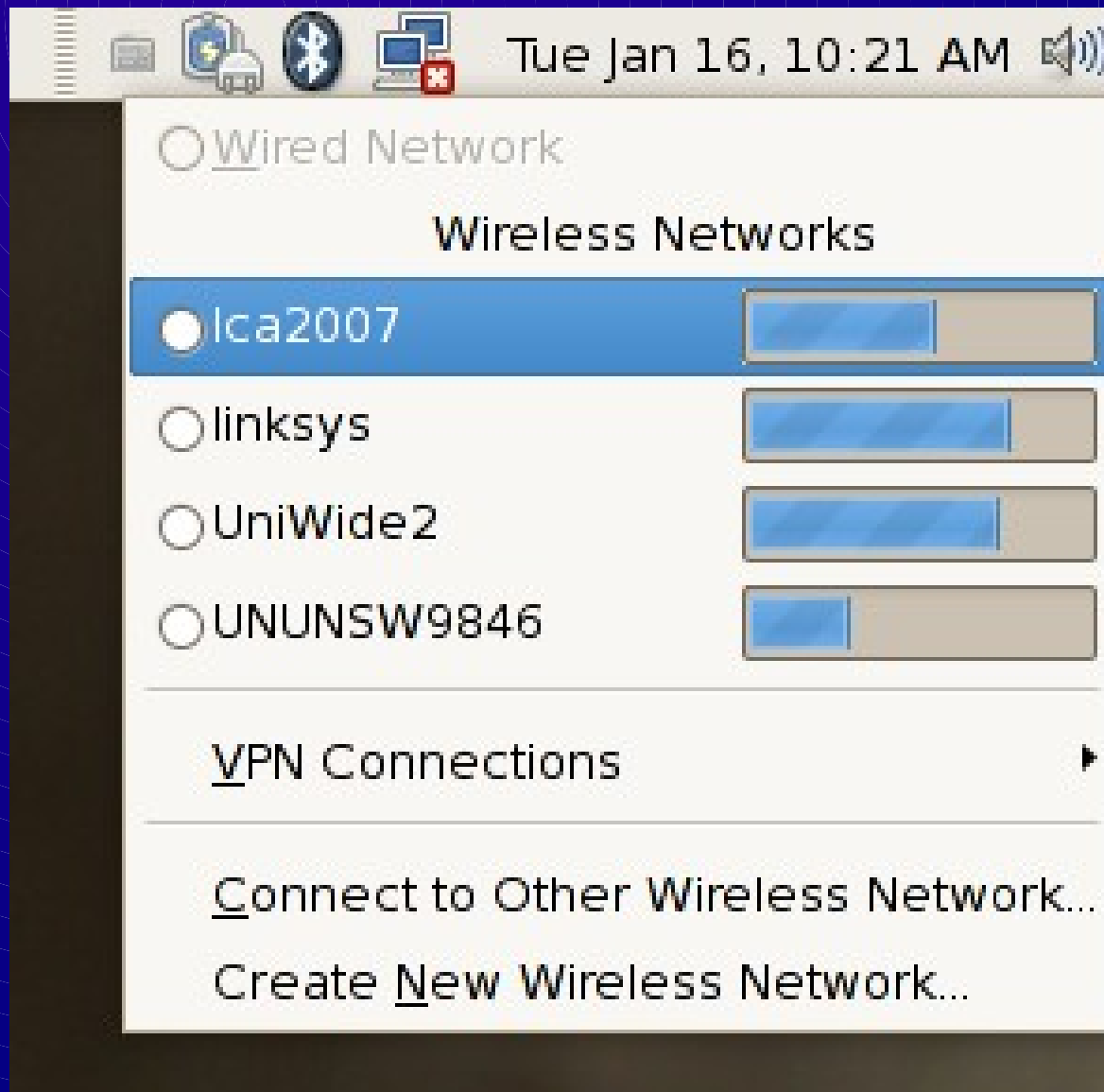


Compare with...

```
» # cryptsetup luksOpen /dev/sda1 my_luks  
Password:  
# mount /dev/mapper/my_luks /media/secret
```

Example

- » Connect to wired and wireless networks



Compare with...

```
» # ifconfig eth0 up  
# iwlist eth0 scan  
...  
# iwconfig eth0 essid lca2007  
# dhclient eth0
```

Example

Create VPN Connection - 2 of 2

Please enter the information provided by your system administrator below. Do not enter your password here as you will be prompted when connecting.

Connection Name

Name used to identify the connection to the private network, e.g. "Campus VPN" or "Corporate Network"

Required Information

Gateway:

Group Name:

Optional Information

Override user name

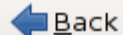
Use domain for authentication

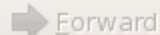
Only use VPN connection for these addresses

example: 172.16.0.0/16 10.11.12.0/24

+ Import Saved Configuration...

 Cancel

 Back

 Forward

» Connect to VPN

» NM supports vpnc,
openvpn, ptp

» plug-in based

Compare with...

- » Compare with:
vpnc -gateway ...

The background features a series of concentric circles in a light blue color, centered on the left side of the frame. The circles radiate outwards, creating a ripple effect. The overall background is a dark blue gradient that transitions to a lighter blue at the bottom.

How does all this work?

How it works

- » (this is on Linux but more about that later!)
- » Linux Kernel
- » udev
- » HAL
- » D-Bus
- » Desktop Session

How it works - kernel

- » Kernel discovers device and publishes it in sysfs

```
/sys/block/sdb/  
|-- dev  
|-- device ->  
..../devices/pci0000:00/0000:00:1d.7/usb5/5-5/5-  
5:1.0/host1/target1:0:0/1:0:0:0  
|-- holders  
|-- queue  
|   |-- <...>  
|-- range  
|-- removable  
|-- size  
|-- slaves  
|-- stat  
|-- subsystem -> .././block  
`-- uevent
```

How it works – udev

- » udev receives uevent from kernel over netlink

```
add@/class/input/devices/input5
ACTION=add
DEVPATH=/class/input/devices/input5
SUBSYSTEM=input
SEQNUM=1166
PHYSDEVPATH=/devices/pci0000:00/0000:00:1d.1/usb2/2-
2/2-2:1.0
...
```

/sbin/hotplug detour

- » Does anyone remember the /sbin/hotplug fork bomb?

How it works - udev

» udev

- » creates / destroys node in /dev
- » run modprobe and other programs
- » passes the event to HAL over a socket

```
KERNEL=="mice", NAME="input/$kernel_name", MODE="0640"  
ACTION=="add", SUBSYSTEM=="pci", MODALIAS=="?*",  
RUN+="/sbin/modprobe $modalias"  
SUBSYSTEM=="scsi_device", ACTION=="add",  
RUN+="/sbin/modprobe sg"
```

```
IMPORT{program}="/sbin/vol_id --export $tempnode" \  
ENV{ID_FS_UUID}=="?*", SYMLINK+="disk/by-  
uuid/$env{ID_FS_UUID}"
```

```
RUN+="socket:/org/freedesktop/hal/udev_event"
```

How it works - HAL

» HAL

- » Receives event from udev
- » Investigates the device
- » Creates representation of device

```
<!-- SanDisk ImageMate II CF Reader -->  
<match key="@storage.physical_device:usb.vendor_id" int="0x0781">  
  <match key="@storage.physical_device:usb.product_id" int="0x0002">  
    <merge key="storage.drive_type" type="string">compact_flash</merge>  
  </match>  
</match>
```


How it works - HAL

- » HAL exports “device objects” in a tree
- » Objects export well defined properties
 - » ... so applications can gather information
- » Objects export well defined interfaces
 - » ... so applications can configure hardware
- » Applications interact with HAL through D-Bus

How it works - HAL

- » Simple device browser (< 1000 lines of Python)

The screenshot shows a graphical user interface for a device browser. On the left, a tree view lists various hardware components. The 'SWAP-99bbcfad-4' device is selected and highlighted. On the right, a detailed view of this device is shown, displaying a list of keys, their types, and their values.

Key	Type	Value
linux.sysfs_path_device	strlist	/sys/block/sda/sd
storage.model	strlist	
volume.block_size	int	512 (0x200)
volume.fstype	strlist	swap
volume.fsusage	strlist	other
volume.fsversion	strlist	2
volume.ignore	bool	true
volume.is_disc	bool	false
volume.is_mounted	bool	false
volume.is_mounted_read_only	bool	false
volume.is_partition	bool	true
volume.label	strlist	SWAP-99bbcfad-4
volume.linux.is_device_mapper	bool	false

Compare with...

» \$ tree /sys | less

blkid | grep /dev/sda2

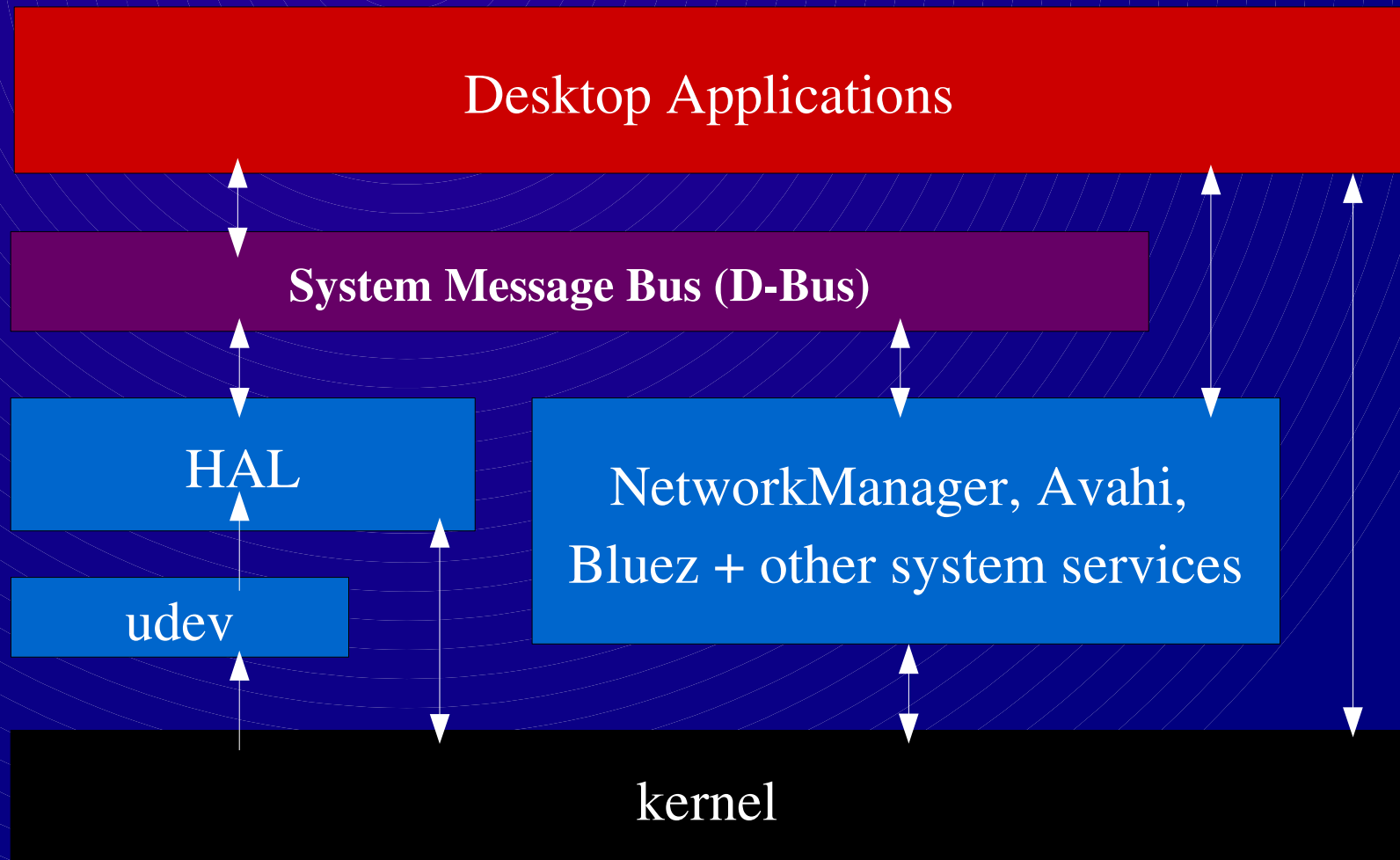
<+ other tools>

How it works - D-Bus

- » IPC (Interprocess Communication)
- » **“Desktop Bus”**
- » **system-wide bus**
 - » enabling desktop apps to communicate with the kernel / core OS (**“Base System”**)
- » **session-wide bus**
 - » Enabling desktop apps to communicate with each other and consume services from each other

How it works - D-Bus

» Typical use of D-Bus in a modern system

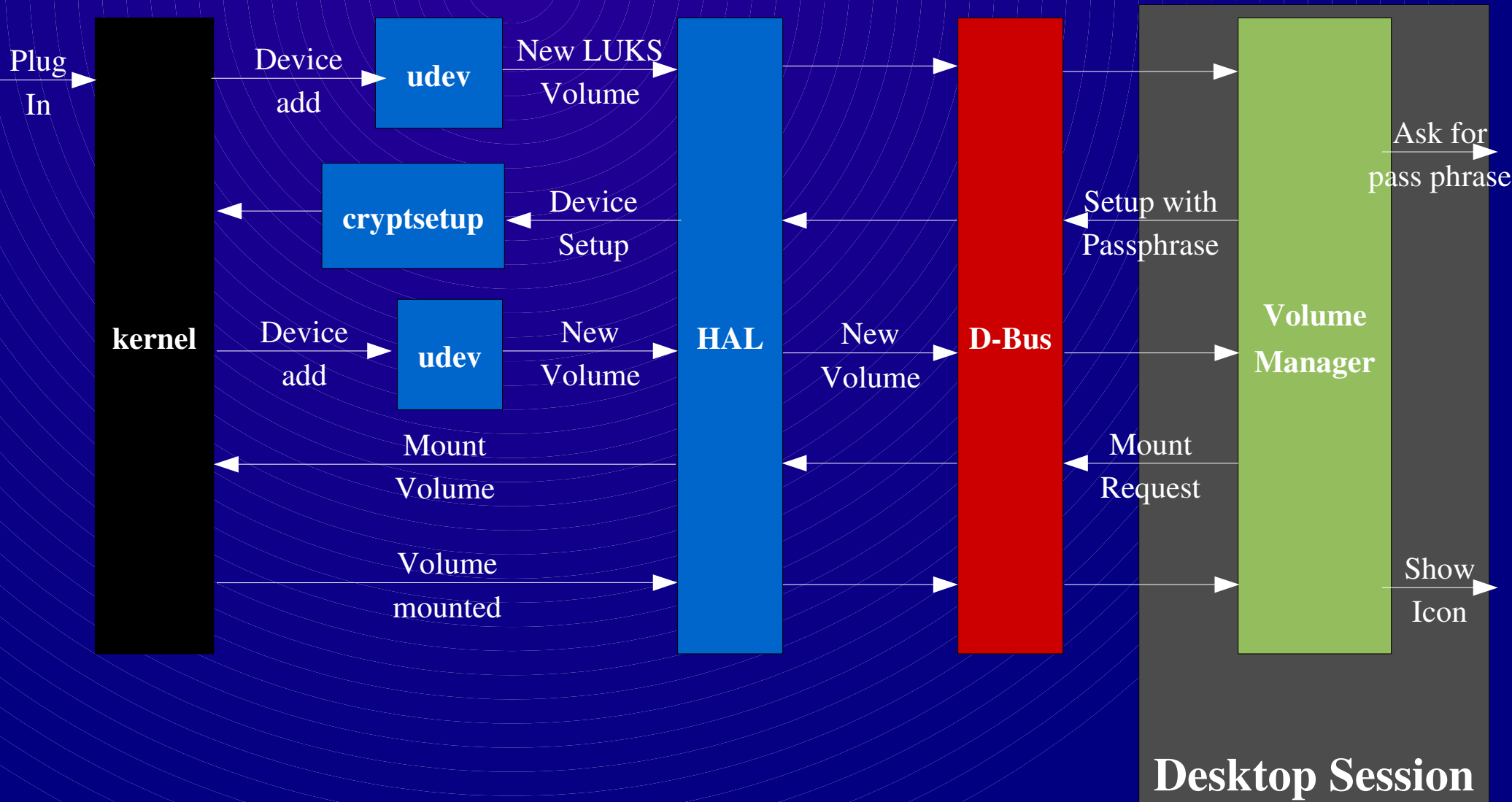


How it works – Desktop Session

- » Policy Manager in desktop session
 - » receives events from HAL over D-Bus
 - » reacts accordingly (draws icons, updates UI)
 - » enforces user policy by invoking methods on HAL
 - » possibly causing HAL to emit new events; and policy manager reacts accordingly

How it works

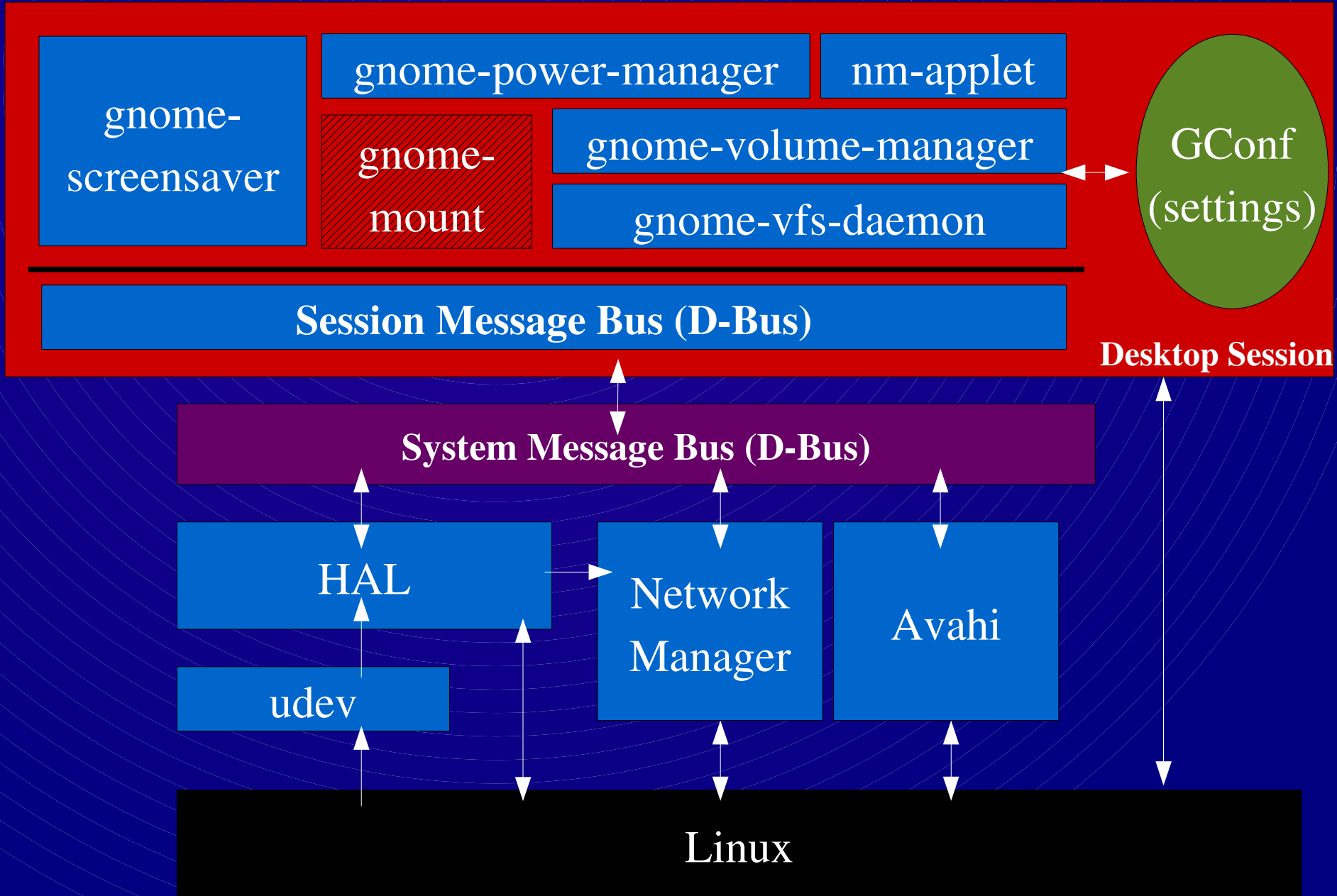
» Thumb drive with LUKS encryption



(sorry about that diagram!)

How it works

» Example: GNOME on Fedora 6 (simplified)



POLICY

MECHANISM

Design Principles

- » Least privilege
- » Clear separation between policy and mechanism
 - » Policy (and settings) originate only from user session
 - » Mechanism originate only from system context
- » Desktop / OS agnostic
- » No polling or useless wakeups (if avoidable)
- » More on this

Least Privilege

- » Users session is unprivileged, however configuring hardware etc. requires privileges

X11 applications should

NEVER

run as uid 0

Least Privilege

- » Modern X11 applications uses GTK+ or Qt and such includes millions of LOC including image loaders and lots of other attack vectors

Least Privilege

- » It breaks accessibility and thus the user experience for impaired users

Least Privilege

- » Section 508 is scheduled for a refresh soon...

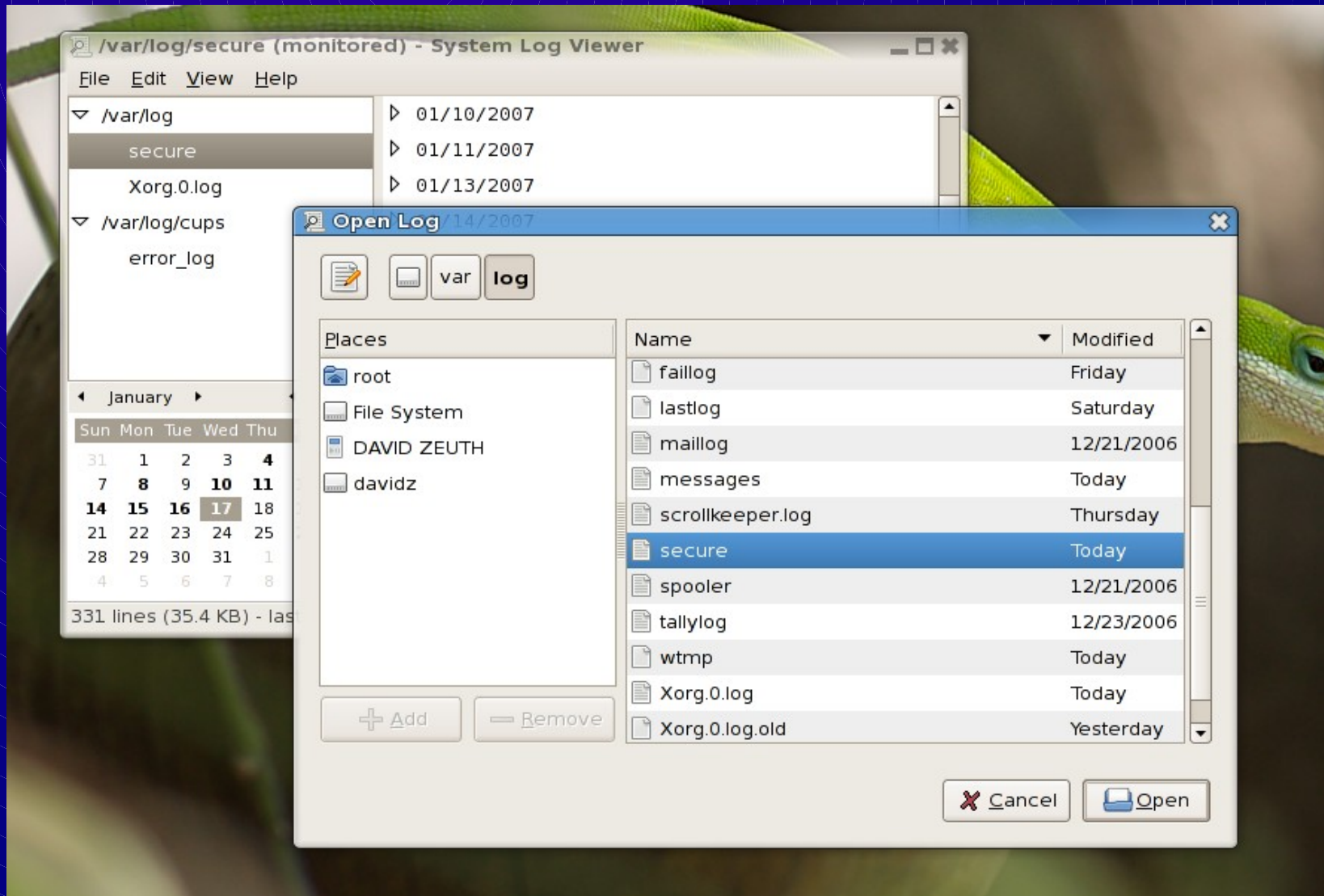
More importantly...

Least Privilege

- » We have a moral applications to make our applications work for impaired users

Least Privilege

» What's wrong with this picture?



**X11 applications should
NEVER
run as uid 0**

</rant>

Least Privilege

- » HAL and other system services provides privileged operations through D-Bus IPC
- » ... and as such does not suffer from the “X11 application running as uid 0” problem

Least Privilege

- » However, access to privileged operations is restricted by either
 - » Session at system console (Red Hat, SUSE)
 - » Group membership (Debian, Ubuntu) e.g. “plugdev”

Least Privilege

- » Problem is not specific to HAL

Privileges – who is allowed to do privileged operations?

- » Today's restriction of privileged operations is basically “All or nothing”

The background of the image is a solid blue color. Overlaid on this background is a pattern of concentric circles. The circles are centered in the upper-left quadrant of the image and radiate outwards, becoming smaller and more densely packed as they move towards the right and bottom edges. The circles are a slightly lighter shade of blue than the background, creating a subtle, hypnotic effect.

All or nothing...

Privileges – who is allowed to do privileged operations?

- » Examples of various privileged operations
 - » Automounting USB thumb drives vs. accessing partitions from internal hard drives
 - » Disk partitioning and file system creation
 - » Installation and upgrades of software
 - » Account creation / deletion / maintenance
 - » Setting the clock / NTP timeserver config / timezone (!)
 - » ... and so on

Privileges – who is allowed to do privileged operations?

- » “Requirements” for a new system
 - » Vendor agnostic
 - » to reduce amount of vendor patches
 - » so upstream projects can start depending on it
 - » Unified file format for describing who can do what and when (!)
 - » Users should be able to obtain privileges by authenticating as either themselves or as uid 0
 - » Probably depends on the privilege in question

12. Drop privilege "hal-mount-fixed"

8. HAL.Mount("/dev/hda3")

Unprivileged Desktop Application
running as uid 500
(e.g. gnome-mount)

4. PermissionDeniedByPolicy
privilege="hal-mount-fixed"

7. Privilege granted

5. PolicyKit.InitiatePrivilegeGrant (
privilege="hal-mount-fixed",
temporaryPrivilege=TRUE,
restrictToDBusConnect=TRUE)

Start here, follow the
numbered arrows

1. HAL.Mount("/dev/hda3")

org.freedesktop.PolicyKit

System Level D-BUS service
(e.g. org.freedesktop.Hal)

2. PolicyKit.IsUserPrivileged (
uid=500,
privilege="hal-mount-fixed")

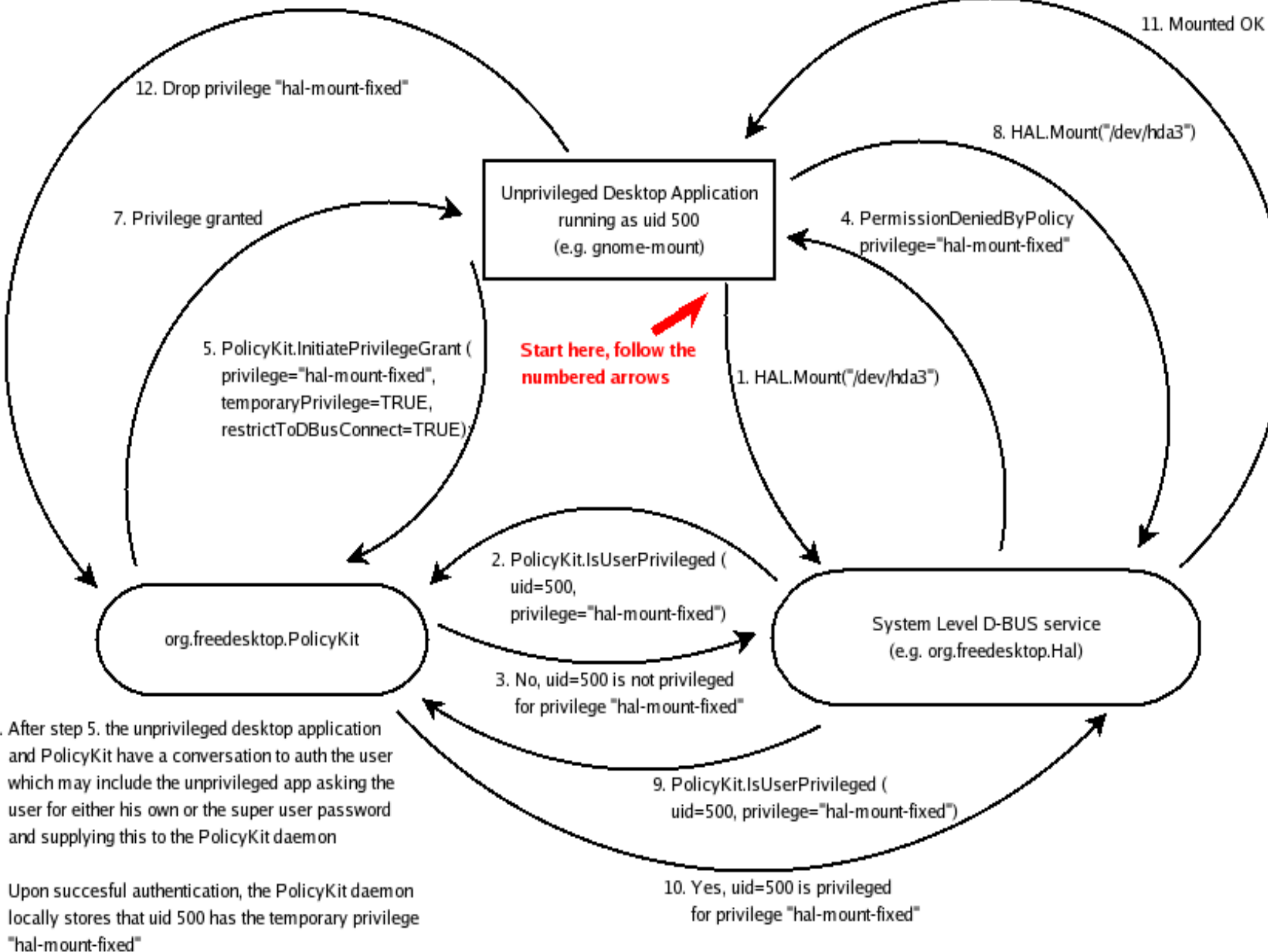
3. No, uid=500 is not privileged
for privilege "hal-mount-fixed"

9. PolicyKit.IsUserPrivileged (
uid=500, privilege="hal-mount-fixed")

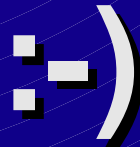
10. Yes, uid=500 is privileged
for privilege "hal-mount-fixed"

6. After step 5. the unprivileged desktop application
and PolicyKit have a conversation to auth the user
which may include the unprivileged app asking the
user for either his own or the super user password
and supplying this to the PolicyKit daemon

Upon succesful authentication, the PolicyKit daemon
locally stores that uid 500 has the temporary privilege
"hal-mount-fixed"



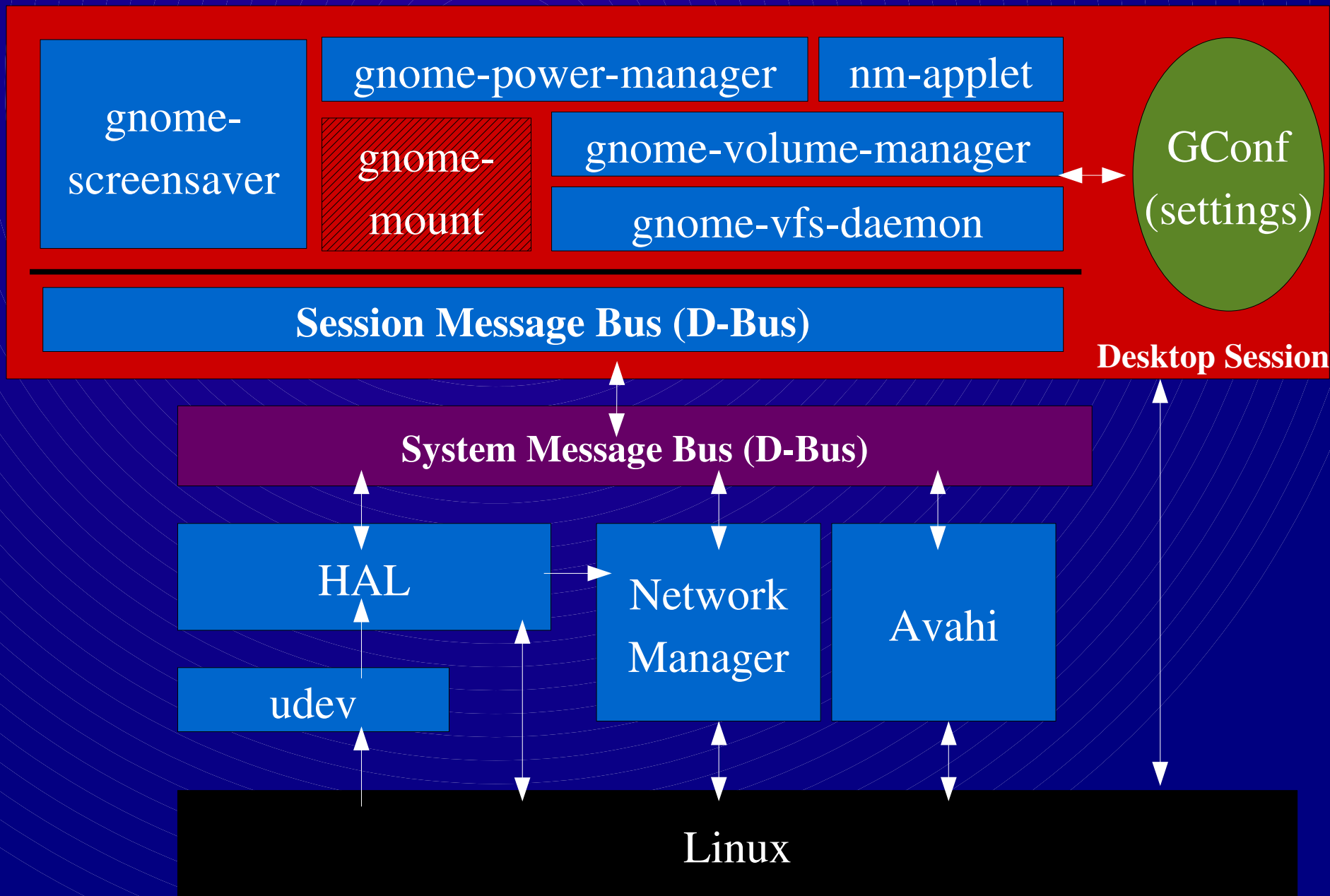
**Did anyone say
“Architecture Astronaut”?**



Policy and Mechanism

- » Policy enforcing daemons in system context...
- » Normally they read settings from /etc
 - » so we get “configuration tools” for rewriting files in /etc
- » Cannot cleanly interact with the desktop
 - » (save for hacks like setting DISPLAY=:0.0 or grepping the process list etc.)
- » They sometimes provide limited interaction with a desktop widget for drawing icons and using the settings from the user

Policy daemons belong in user session



POLICY

MECHANISM

Desktop and OS agnostic

- » Modern Linux: configure it your way!
 - » Good and bad
- » HAL exports only a tiny subset of operations
 - » Driven by what the desktop needs (not much)
 - » Operations exported by HAL are so fundamental that they should be able to be implemented on any UNIX-like system
 - » Wants to keep the HAL interfaces stable (eventually)
 - » Linux kernel developers should be free to change kernel internals (example: sysfs and laptop backlight interfaces)

HAL on Solaris

The screenshot displays a Solaris desktop environment with a blue background and various application icons. A terminal window is open, showing the output of the `uname -a` command. A Device Manager window is also open, displaying a tree view of hardware devices and a detailed view of the selected device's properties.

Terminal Output:

```
$ uname -a
SunOS d-mpk17-229-216 5.11 snv_55 i86pc i386 i86pc
$
```

Device Manager - Selected Device Properties:

Key	Type	Value
info.bus	strlist	usb
info.claimed	bool	true
info.claimed.service	strlist	gnome-volume
info.claimed.uid	int	92521 (0x1696)
info.parent	strlist	/org/freedesktop
info.product	strlist	USB Device In
info.solaris.driver	strlist	scsa2usb
info.udi	strlist	/org/freedesktop
solaris.devfs_path	strlist	/pci@0,0/pci10
usb.device_revision_bcd	int	8243 (0x2033)
usb.product	strlist	Cruzer Micro
usb.product_id	int	28947 (0x7113)
usb.serial	strlist	0002091740
usb.vendor	strlist	SanDisk Corp.
usb.vendor_id	int	1921 (0x781)

The Device Manager window shows a tree view of hardware devices. The selected device is a **CRUZER** USB device, which is a **SCSI Device** connected to a **SCSI Host Adapter**. The device is identified as a **82801DB/DBL/DBM (ICH4/ICH4-L/ICH4-M) USB2 EHC** controller.

HAL on FreeBSD

File Edit View Terminal Tabs Help

```
~ 2:03 shumai marcus %507> uname -a
FreeBSD shumai.marcuscom.com 6.2-PRERELEASE FreeBSD 6.2-PRERELEASE #18: Sat Dec
30 20:35:43 EST 2006 marcus@shumai.marcuscom.com: /usr/obj/usr/src/sys/SHUMAI
i386
~ 20:46 shumai marcus %508> 
```

Device Manager

File View Help

Devices

- UHCI root hub
 - USB Hub Interface
- 82801G (ICH7 Family) USB UHCI #3
 - UHCI root hub
 - USB Hub Interface
- 82801G (ICH7 Family) USB UHCI #4
 - UHCI root hub
 - USB Hub Interface
- 82801G (ICH7 Family) USB2 EHCI Controller
 - EHCI root hub
 - USB Hub Interface
- 82801G (ICH7 Family) PCI Express Port 1
- 82801G (ICH7 Family) High Definition Audio Controller
- 82801G (ICH7 Family) SMBus Controller
- 82801G (ICH7 Family) IDE Controller
 - ATA channel 0
 - IDE Device (Master)
 - Ignored Device
 - SCSI Host Adapter
 - SCSI Device
 - CDRW/DVD TSH492B
 - My Disc
 - ATA channel 1
 - SCSI Host Adapter

Device Advanced

Key	Type	Value
block.device	strlist	/dev/cd0
block.freebsd.cam_path	strlist	2,0,0
block.is_volume	bool	false
block.major	int	0 (0x0)
block.minor	int	138 (0x8a)
block.storage_device	strlist	/org/freedesktop/Hal/devices,
freebsd.driver	strlist	cd
freebsd.unit	int	0 (0x0)
info.addons	strlist	[u'hald-addon-storage']
info.bus	strlist	block
info.capabilities	strlist	[u'block', u'storage', u'storage
info.category	strlist	storage.cdrom
info.parent	strlist	/org/freedesktop/Hal/devices,
info.product	strlist	CDRW/DVD TSH492B
info.udi	strlist	/org/freedesktop/Hal/devices,
info.vendor	strlist	TSSTcorp
storage.automount_enabled_hint	bool	true
storage.bus	strlist	scsi
storage.cdrom.bd	bool	false
storage.cdrom.bdr	bool	false
storage.cdrom.bdre	bool	false
storage.cdrom.cdr	bool	true

**KDE and XFCE uses HAL
too**

HAL is only a mechanism

- » Why not just use the kernel directly?
 - » Our desktops should, as far as possible, be able to easily run on different OS's
 - » Access control is tricky
 - » deny service to inactive sessions (fast-user-switching)
 - » privileges should be finely grained
 - » Information from the kernel is scattered throughout the place (but it's getting better)
 - » Missing abstractions in the kernel but with 'backlight' and 'battery' device classes it's getting better

HAL is only a mechanism

- » Desktop needs information kernel don't know or want to know on device investigation
 - » File system labels / UUID
 - » gphoto2 / SANE and other user space drivers provide XML files so HAL can tag devices accordingly
 - » Integration with other user space driver frameworks such as NUT (for UPS support) underway
 - » Polling mass storage drives (yuck!)
- » Light-weight IPC mechanism needed (D-Bus)
- » No need to run all this code in ring 0

Detour...

- » Is Linux becoming a micro-kernel?

Detour...

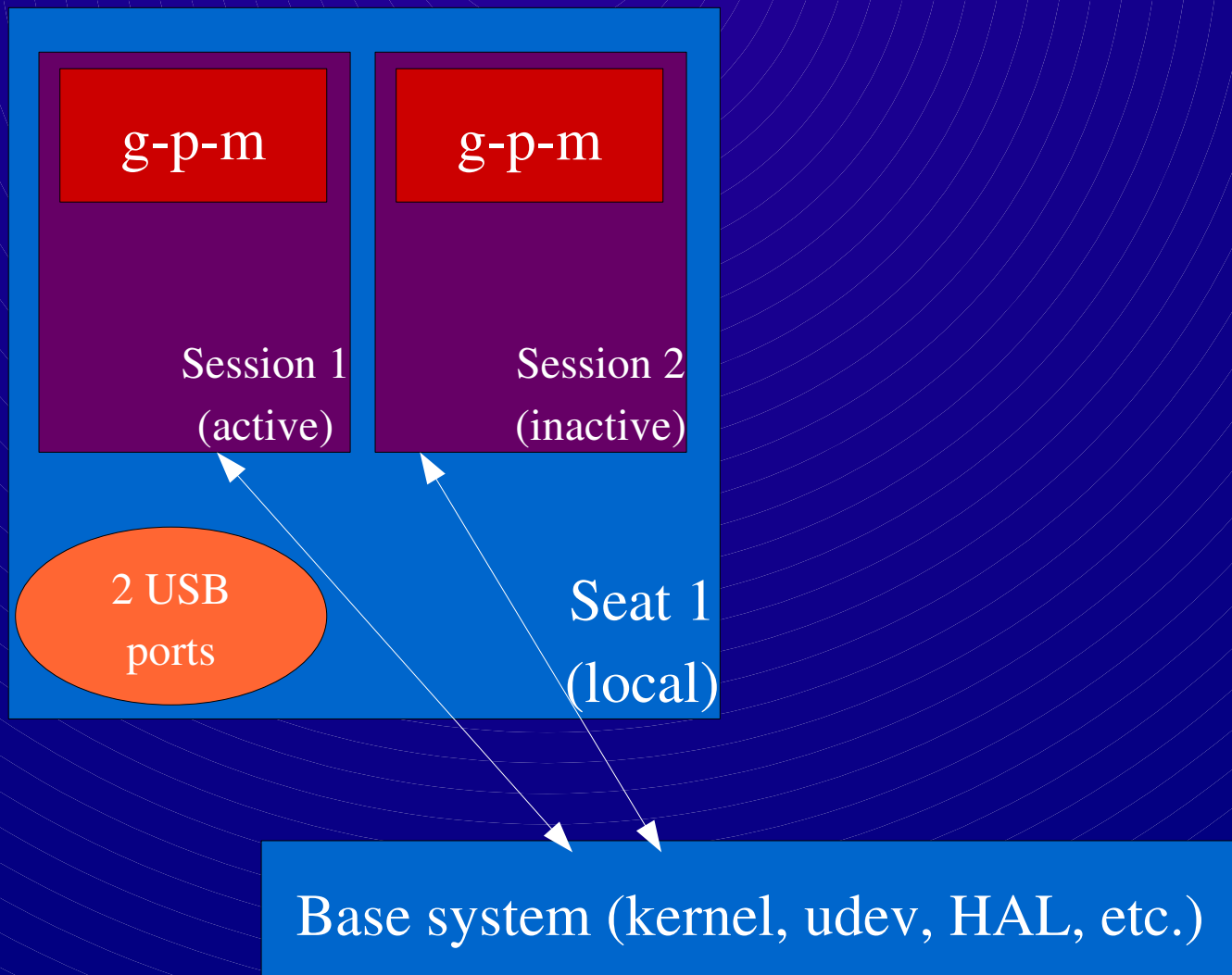
- » Is Linux becoming a micro-kernel?
 - » udevd
 - » dbus-daemon, hald
 - » NetworkManager, dhcdd
 - » restorecond, mcstransd, auditd,
selinuxtroubleshooterd
 - » hcid, sdpc

Future Work

- » Privileges / PolicyKit
 - » Tricky how to model it without making things too complex / over engineered
- » Tracking user sessions / ConsoleKit
 - » A modern replacement for /var/log/wtmp based on D-Bus – designed to be used in fast-user-switching, multi-seat and for thin clients....
 - » Fundamental objects: **Seat** and **Session**. HAL will use ConsoleKit to e.g. deny service to policy daemons in inactive sessions for e.g. Fast-user-switching
 - » See announcement on hal@lists.freedesktop.org

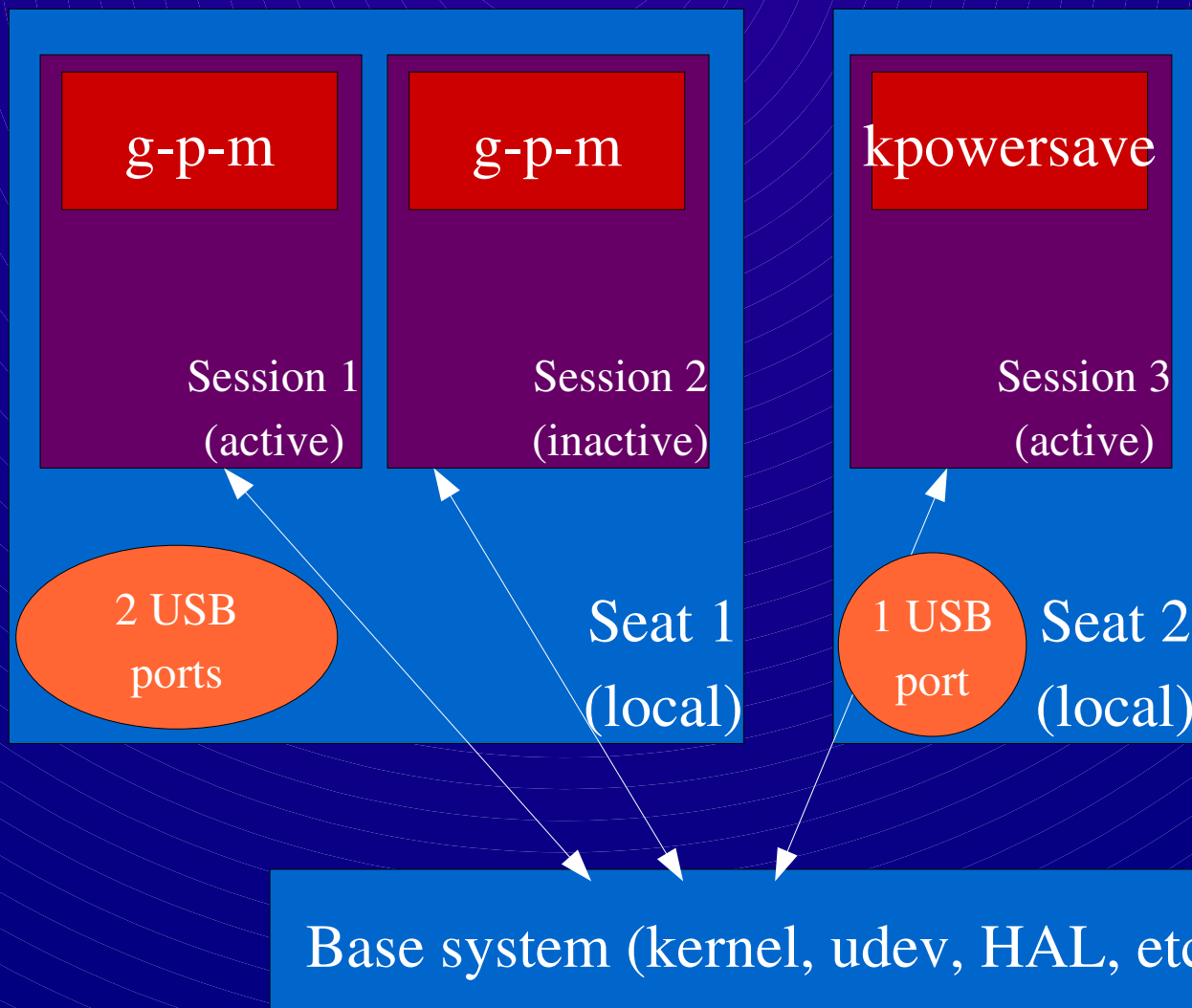
ConsoleKit detour

- » Interesting problems
 - » Synchronization between multiple policy managers, Assigning device ownership... just to name a few



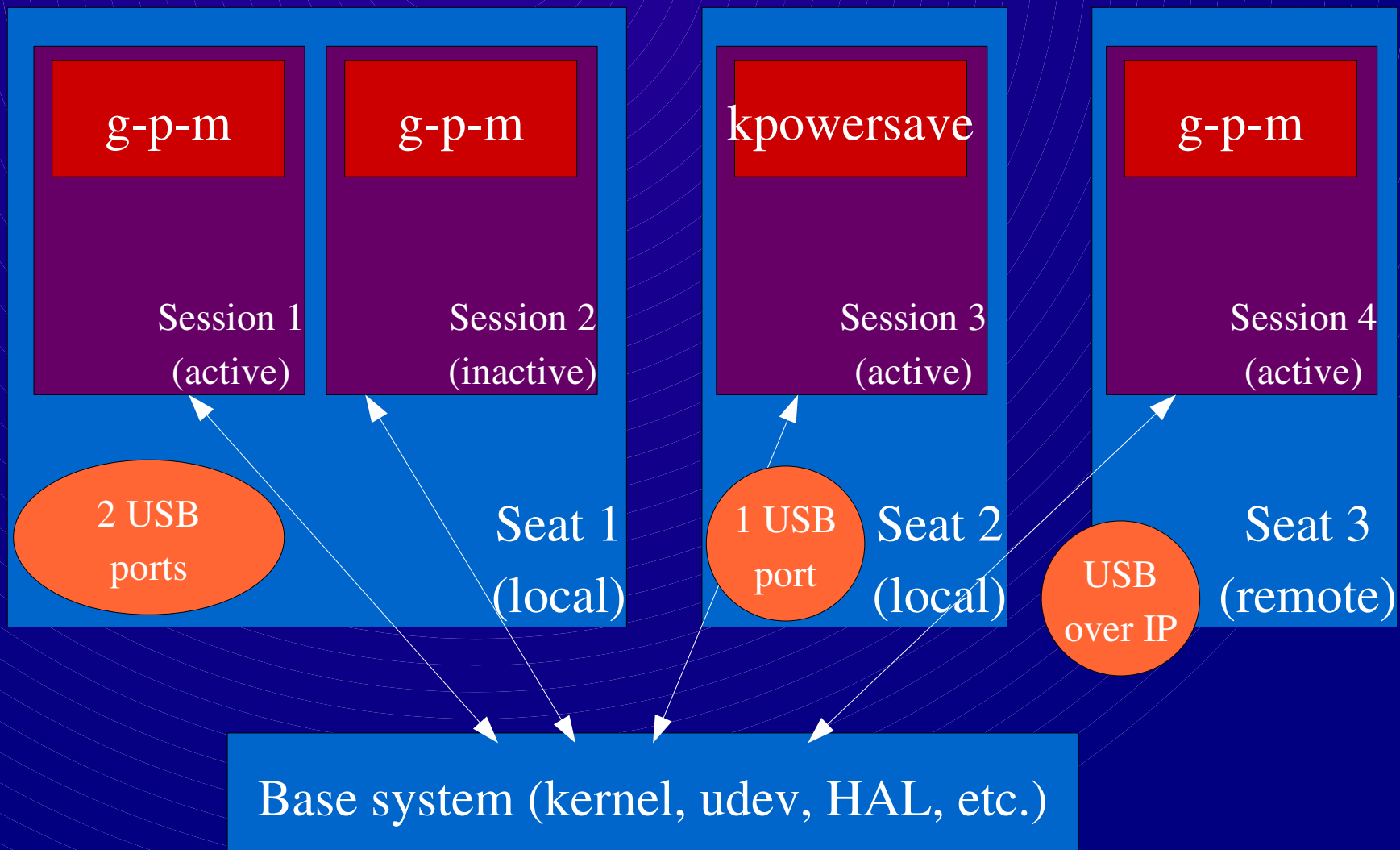
ConsoleKit detour

» Two seats



ConsoleKit detour

» Remote seats...



Future Work

- » Performance / Optimization; HEAD vs. 0.5.8.1
 - » Memory use: Down to 600kb dirty from 2.3MB dirty
 - » Startup: Down to 0.9sec from 2.3sec
- » D-Bus
 - » 1.0 out and ABI is frozen
 - » Activation of services on the system bus targeted for release 1.2. Should make it possible to start privileged services (example: Cups) only when service is needed
- » Other
 - » Work on reusing policy daemons (such as gnome-power-manager) when no one is logged in

Lessons learned so far...

- » Manpower
 - » Though many contributors
- » Bridging communities
 - » No, it's not easier if you work for a vendor!
- » Architecture
 - » Keep it as simple as possible
 - » It's easy to replace a broken implementation; it's much harder to replace a broken architecture.
- » Upstream. Upstream. Always upstream.
- » Desktop integration is just damn hard

Thank you for listening!

Questions?