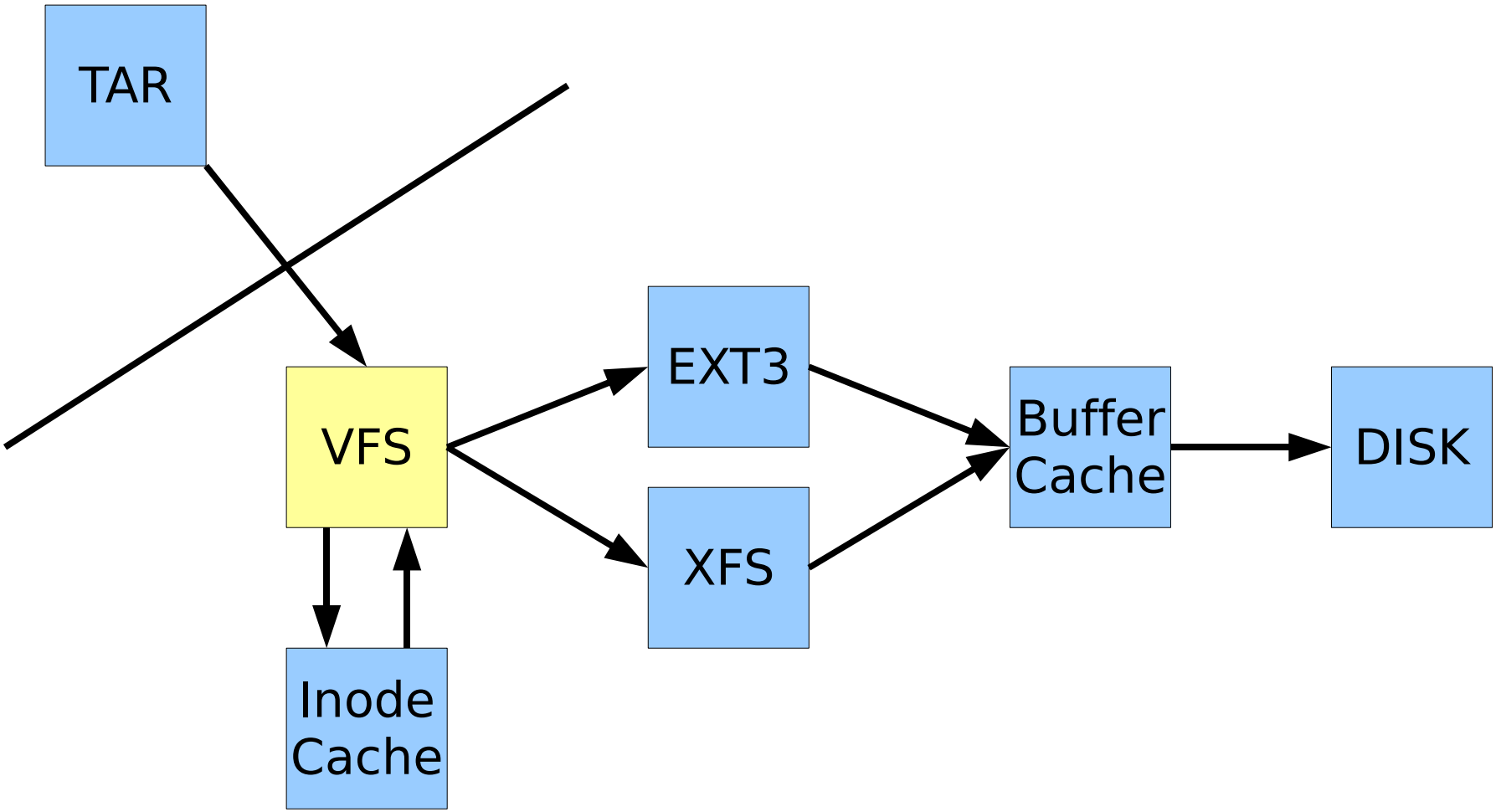
Linux VFS Hacks

Michael Gebetsroither
gebi@grml.org

- Linux VFS
- Shared Subtrees
- Namespaces
- Capabilities

Linux VFS



Shared Subtrees

- mount --bind /from /to
- mount --move /from /to
- mount --make-shared /mnt
- mount --make-slave /mnt
- mount --make-private /mnt
- mount --make-unbindable /mnt

Namespaces

- since 2.4.19
- different set of mountspoints per namespace
- clone(2) CLONE_NEWNS
- unshare(2) CLONE_NEWNS

Namespaces

- new_namespace /bin/zsh
- mount -n --bind /lib32 /lib
- mount -n --bind /bin/tcsh /bin/sh
- <start broken program>

Namespaces

```
int main(int argc, char** argv)
{
    if(syscall(SYS_unshare, CLONE_NEWNS) == -1)
{
        perror("error in unshare");
        return EXIT_FAILURE;
    }
    if (argc > 1)
        return execvp(argv[1], &argv[1]);
    return execv("/bin/sh", NULL);
}
```

Capabilities

- since 2.2
- useable since 2.6.24
- activated per default in Debian 2.6.29 (sid)

Capabilities

- CAP_CHOWN
- CAP_SYS_CHROOT
- CAP_MKNOD
- CAP_NET_ADMIN
- CAP_NET_BIND_SERVICE
- CAP_NET_RAW
- CAP_SYS_BOOT

Capabilities

- getcap /bin/ping
- chmod 755 /bin/ping
- setcap cap_net_raw=ep /bin/ping
- getcap /bin/ping

  /bin/ping = cap_net_raw+ep
- ping localhost (as user)

  ....

Namespaces

- http://lxc.sf.net
- Hostname
- PID
- IPC
- User
- Network
- /proc

Links

http://glandium.org/blog/?p=217

http://etbe.coker.com.au/2008/12/13/per-process-namespace

Shared Subtrees: http://lwn.net/Articles/159077/