



NFS over virtio-vsock

Host/guest file sharing for virtual machines

Stefan Hajnoczi <stefanha@redhat.com>

Connectathon MMXVI

Agenda

- Host/guest file sharing
- What is host/guest communication?
- Overview of virtio-vsock
 - AF_VSOCK address family
- Adding a new transport to NFS
 - Extending Linux NFS and nfs-utils
- Status of NFS over virtio-vsock



About me

Work on KVM in Red Hat's Virtualization team

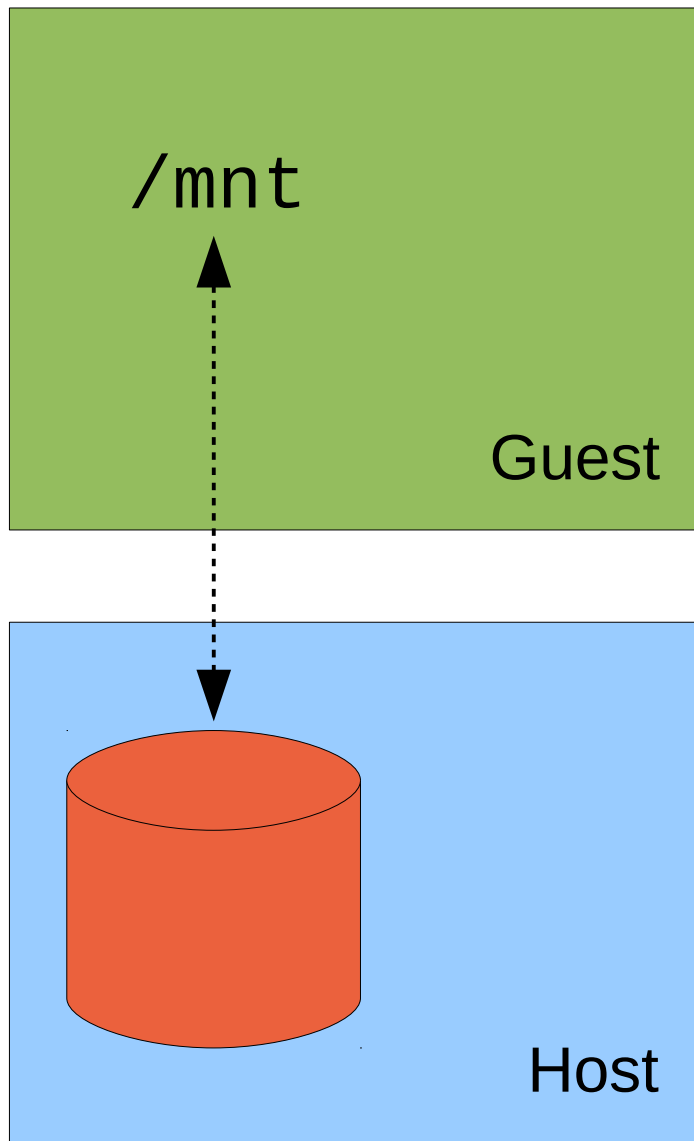


Open source contributor

- QEMU machine emulator and virtualizer (used by KVM and Xen hypervisors)
- Linux (mostly virtualization-related topics)



Host/guest file sharing



- Range of use cases:
 - **NAS as a Service** in cloud or hosting environments
 - Providing files to virtual machine during **installation**
 - File system from host for **quick development/test cycles**

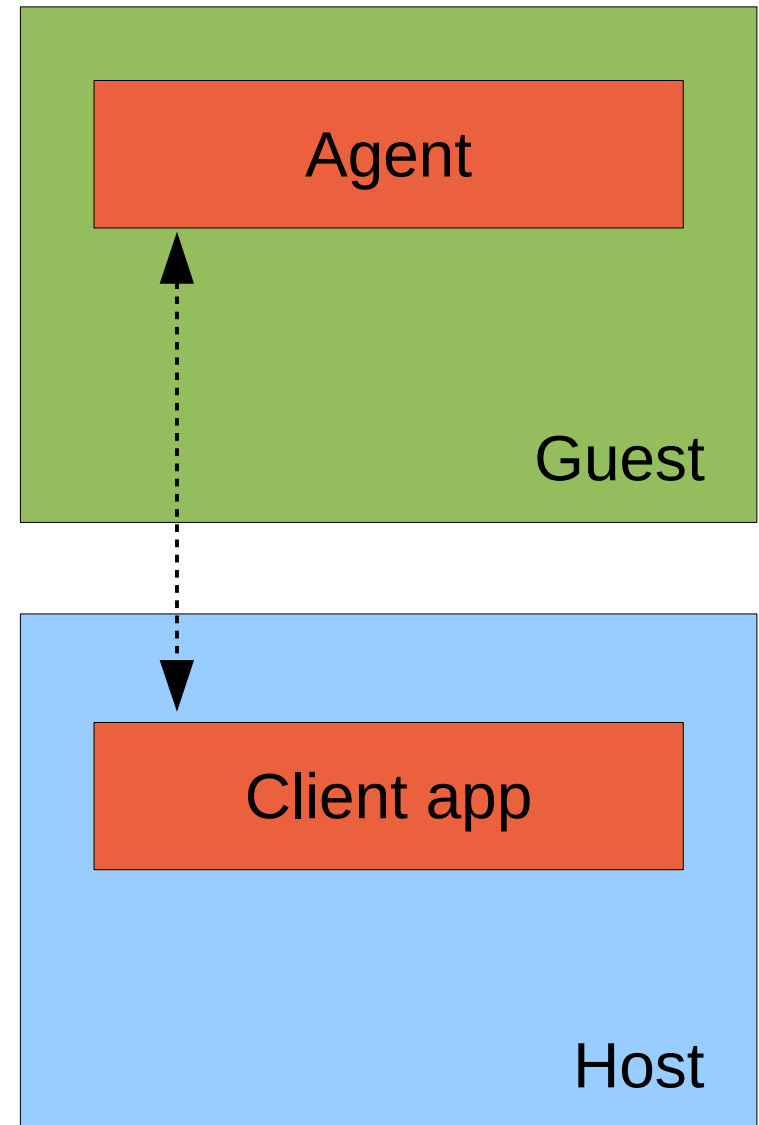
Guest - virtual machine
Host - hypervisor,
physical machine



Host/guest communication use cases

Communications channel between virtual machine and hypervisor.

- qemu-guest-agent
 - Backups, suspend, etc
- SPICE vdaagent
 - Clipboard sharing, etc
- Custom agents
- Host services (file sharing)



Possible solution: Ethernet

Pro: TCP/IP and NIC support already exists

Con:

- Adding & configuring guest interfaces is invasive
- Prone to break due to config changes inside guest
- Creates network interfaces on host that must be managed

No other hypervisor uses Ethernet for host/guest communication...they hit the same problems.



Solutions in KVM

- Virtio-serial
 - In use today
 - Only supports 1:1 communication, not N:1 client/server
 - Does not support Sockets API (inconvenient)
- Virtio-vsock
 - In development
 - Supports Sockets API
 - Shares AF_VSOCK address family with VMware VSOCKets



AF_VSOCK in Linux

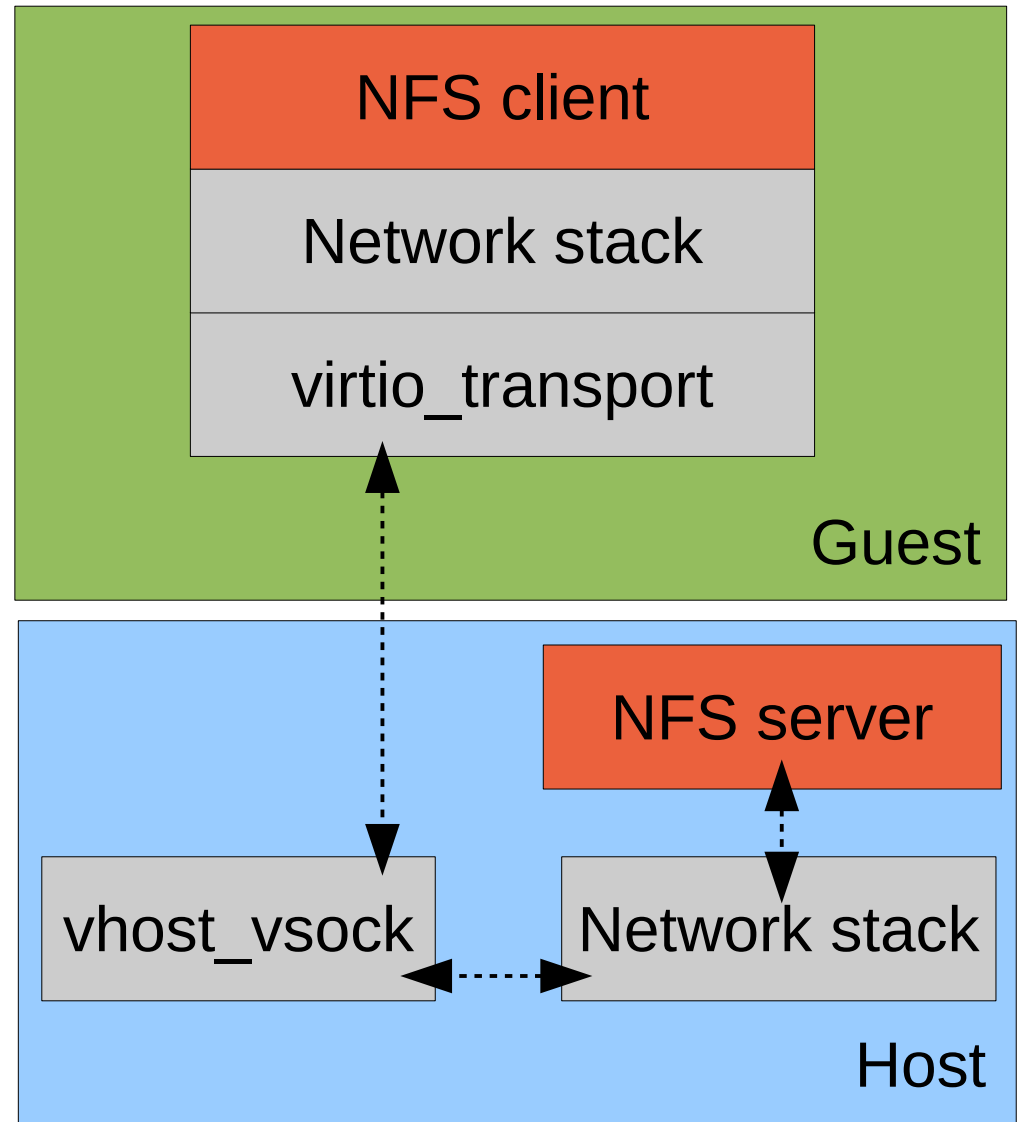
- New socket address family for host/guest communication
- Can support datagram and stream semantics
 - Virtio-vsock only implements stream for now
- Addresses are <u32 cid, u32 port>
 - Each guest has unique cid
 - Host has well-known cid
- Contributed to Linux by VMware in 2013
- Gerd Hoffmann and Asias He prototyped a virtio transport for vsock



vhost-vsock architecture

Uses vhost driver framework to integrate with host network stack

Both guest and host applications use sockets API



Modifications to NFS stack

- RFC 5531 RPC: Remote Procedure Call
 - Extensions to describe AF_VSOCK transport
- Core NFS and NFSD mostly unaffected
 - AF_VSOCK stream semantics match TCP (reliable, ordered)
 - Target NFSv4.1+ so no datagram transport or port mapper usage
- Client and server configuration
 - Passing AF_VSOCK host addresses
 - Syntax for /etc/exports



RFC 5531 RPC: Remote Procedure Call

- AF_VSOCK transport similar to TCP
 - Record Marking required due to stream semantics
- ONC RPC Netid
 - “vsock”
- ONC RPC Uaddr format
 - “vsock:cid.port”, e.g. “vsock:2.2049”
 - The “vsock:” prefix seems useful to distinguish between other transports when netid isn't available

Not yet proposed to IANA. Feedback appreciated!



nfs-utils command-line interface

Proposed command-line syntax:

- Server

```
nfsd -no-tcp -no-udp \  
-nfs-version 4.1  
-vsock 2049
```

- Client

```
mount.nfs 2:/export /mnt \  
-o proto=vsock
```



Configuration in `/etc/exports`

- **Security:** Existing exports must not be visible over `AF_VSOCK` automatically
 - * matches only IPv4/IPv6 clients!
- Wildcard `AF_VSOCK` exports
`/srv/publicfs vsock:*(ro)`
- Single guest address
`/srv/vm001 vsock:5(rw)`



Linux userspace ABI changes

- New transport name “vsock”
- Client addresses represented as “vsock:5”
- `/proc/net/rpc/auth.unix.ip` cache
 - No longer IP-only
 - Outdated `rpc.mountd` will reject vsock clients
 - Backwards compatible



Status of patches

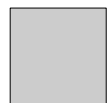
Linux NFS
client

Nfs-ganesha
(Thanks to
Matthew Benjamin)

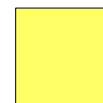
Linux NFS
server

nfs-utils

Can be tested without
nfsd support using port
forwarding to TCP nfsd.
See “nc-vsock” tool.



Not yet implemented



Patches not yet merged



Patches in development



Patches merged



Questions?

Email: stefanha@redhat.com

IRC: stefanha on #qemu irc.oftc.net

Blog: <http://blog.vmsplice.net/>

Specification: <http://goo.gl/mi6LCR>

Code:

- <https://github.com/stefanha/nfs-utils> vsock
- <https://github.com/stefanha/linux> vsock-nfs
- <https://github.com/stefanha/qemu> vsock

Slides available on my website: <http://vmsplice.net/>

