Can Linux network configuration suck less?

Pavel Šimerda
pavlix@pavlix.net

FOSDEM 2013, Brussels

http://data.pavlix.net/fosdem/2013/

Pavel Šimerda
pavlix@pavlix.net

The story of network management

Pavel Šimerda
pavlix@pavlix.net

Kernel automatic configuration

- Loopback address configuration
- Link-local address configuration (IPv6 only)
- Global address configuration (IPv6 SLAAC only)
- Device route configuration (IPv6 single device only)
- Default route configuration (IPv6 single device only)
- Lifetime-based address removal (IPv6 only)

Pavel Šimerda
pavlix@pavlix.net

Kernel network configuration APIs

- Netlink API for links, addresses and routes
- Provides RDNSS and DNSSL (IPv6 RA only, not cached)
- Sysctl API for various networking settings
- Various legacy APIs
- Quickly access rtnetlink API using *iproute* commands

Pavel Šimerda
pavlix@pavlix.net

A daemon is needed for

- Stateful configuration (not in kernel)
- RDNSS and DNSSL configuration (requires file access)
- Auxiliary scripts (e.g. for setting NTP servers)

Pavel Šimerda
pavlix@pavlix.net

DHCP clients as configuration daemons

- Most DHCP client daemons can do that
- Some of them support multiple interfaces
- Kernel configuration via Netlink or other API
- DNS configuration through /etc/resolv.conf

Pavel Šimerda
pavlix@pavlix.net

So what's the problem?

- Multi-interface support conflicts with per-interface configuration
- Proper policy decisions require additional complexity
- Doesn't integrate well other configuration daemons
- Badly violates the UNIX philosophy (do one thing and do it well)
- **All of this can be solved with a dedicated network configuration daemon**

Pavel Šimerda
pavlix@pavlix.net

Now to the real world

Pavel Šimerda
pavlix@pavlix.net

Simple scripts using *iproute* commands

- Suitable for static configuration or networking tests
- Links, addresses and routes via `ip` command
- Traffic control via `tc` command
- Supports all sorts of virtual network interfaces
- More static configuration can be added (e.g. iptables)

Pavel Šimerda
pavlix@pavlix.net

Scripts with per-interface configuration

- Suitable for static configuration
- Usable for DHCPv4 on a single interface
- Conflicts with...
  - DHCP on multiple interfaces
  - Dualprotocol DHCP
  - IPv6 DNS configuration
  - Combination of DHCP and any VPN software and/or DNSSEC
- This solution is generally limiting and inflexible
- Every now and then you miss a proper policy decision

Pavel Šimerda
pavlix@pavlix.net

Network script implementations

- ifcfg variants (Fedora, openSUSE, Mandriva, ...)
- ifupdown (Debian, Ubuntu, ...)
- ifnet (Gentoo, ...)
- uci network (OpenWRT)
- ...

Pavel Šimerda
pavlix@pavlix.net

Network configuration daemons

- Reads on-disk configuration files at startup
- Provides API for runtime and permanent configuration
- Performs policy decisions (e.g. default output interface)
- Communicates with the kernel (via Netlink and /proc/sys)
- Manages all sorts of network-related daemons
- Broadcasts events via IPC (e.g. D-Bus signals)
- Runs auxiliary scripts (e.g. NTP service configuration)

Pavel Šimerda
pavlix@pavlix.net

Network configuration daemons – use cases

- Multiple managed network interfaces
- Static, dynamic and mixed network configuration
- Policy decisions on connections and routing
- Event-based coordination of networking tools
- API for user configuration tools

Pavel Šimerda
pavlix@pavlix.net

Network configuration daemons – implementations

- NetworkManager (used by most distributions)
- connman (came from Intel's Meego project)
- Wicd (a network daemon written in Python)
- netcfg (Archlinux)
- netifd (new OpenWRT development)
- wicked (is not Wicd)
- . . . ?

Almost everybody wants to have his own. . . ;)

Pavel Šimerda
pavlix@pavlix.net

NetworkManager, distributions' daemon of choice

- Stable branch 0.9.6, development branch 0.9.7
- Long history of laptop and desktop usage
- Gradually expands to the server and virtualization world
- Four regular developers (all Red Hat employees)
- Good number of regular and occasional contributors
- We often contribute to other projects in the networking ecosystem
- Users who avoid using NM still benefit from our work

NetworkManager in distributions

Pavel Šimerda
pavlix@pavlix.net

Fedora and Red Hat

- Packaged by developers themselves
- Best level of integration of other packages
- Issues with systemd/dbus integration
- Supports ifcfg-style configuration
- Cooperation with network-scripts is far from perfect
- Network test week in December 2012 for Fedora 18
- We plan to have another test week for Fedora 19

Pavel Šimerda
pavlix@pavlix.net

Debian and Ubuntu

- Quite a number of contributions
- Supports ifupdown-style configuration (not so well)
- Similar problems with parallel configuration scripts
- I recommend switching to NetworkManager's native configuration

Pavel Šimerda
pavlix@pavlix.net

SUSE

- Occasional contributions
- Separate ifcfg-style plugin (would be nice to merge it)
- Supports DNS setting through SUSE netconfig tool
- Minor integration issues

Pavel Šimerda
pavlix@pavlix.net

Gentoo

- Occasional contributions
- Good level of integration with OpenRC service manager
- Supports ifnet-style configuration (not tested)
- Includes a live git ebuild (since 2013-01-28)
- Great for various tests

Pavel Šimerda
pavlix@pavlix.net

NetworkManager on other distributions

- Native configuration format (keyfile)
- Builds without special configure options (for 0.9.8)
- Any distribution plugin can be explicitly enabled
- Loopback device is handled generically

Pavel Šimerda
pavlix@pavlix.net

NetworkManager features

Pavel Šimerda
pavlix@pavlix.net

Physical connection types

- Wired and wireless Ethernet connections (including 802.1x)
- ADSL connections
- Mobile broadband (including bluetooth DUN)
- Bluetooth PAN
- OLPC mesh
- Wimax connections
- Infiniband

Pavel Šimerda
pavlix@pavlix.net

Virtual connection types

- VLAN interfaces
- Bridges and bonds for wired Ethernet (for 0.9.8)
- Bridges and bonds also for Wifi (in planning)
- Team driver integration (in planning)
- VPN plugin interface (several plugins available)

Pavel Šimerda
pavlix@pavlix.net

Connection sharing

- IPv4 connection sharing over Ethernet
- IPv4 sharing over Ad-hoc Wifi and Hotspot mode
- Considering DHCPv6 PD-based sharing
- IPv6 masquerade seems to be too fragile
- Sharing based on NDP proxy is also a possibility
- In some cases bridging might be a better idea

Pavel Šimerda
pavlix@pavlix.net

Conneciton dependencies and autoconnection

- Bridges and bonds works pretty well now
- Mobile broadband doesn't autoconnect or reconnect
- VPN autoconnect switch is not supported
- Physical connection can autoconnect a single VPN

Pavel Šimerda
pavlix@pavlix.net

Addresses and routes on wired/wireless Ethernet

- Static IPv4/IPv6 configuration is now in the same format (for 0.9.8)
- Dropped support for dhclient 3.x (for 0.9.8)
- Dynamic IPv4 works well even with bridges/bonds (for 0.9.8)
- Basic scenarios of dynamic IPv6 configuration work well (since 0.9.6)
- DHCP for IPv6 is now equivalent to IPv4 (for 0.9.8)
- IPv6 needs to be heavily reworked (in kernel first)

Pavel Šimerda
pavlix@pavlix.net

DNS

- Defaults to direct modification `/etc/resolv.conf`
- Supports `resolvconf` and SUSE `netconfig`
- Supports local resolution through `dnsmasq`
- When using `dnsmasq`, suppots zone-specific servers
- DNSSEC support belongs to long-term plans

Pavel Šimerda
pavlix@pavlix.net

Platform interaction refactoring for testability

- New nm-platform framework for OS interaction (WIP, for 0.9.10)
- Dropped support for libnl 1.x and 2.x (¿= 3.2.7 is required)
- Kernel and libnl workarounds moved into one source file
- Tests for the platform code and kernel/libnl behavior
- Tests for NM behavior through a fake platform (in planning)
- Possible separate library or contribution to libnl

Pavel Šimerda
pavlix@pavlix.net

Integration with initramfs

- Important for NetworkManager-enabled network boot
- NetworkManager doesn't have excessive dependencies
- Parts of NetworkManager are already dynamic modules
- Private bus will avoid D-Bus daemon requirement

Pavel Šimerda
pavlix@pavlix.net

Temporary connections and connection take over

- Separate runtime and persistent configuration
- Save, restore and modification APIs
- Generic device configuration
- Accepting modifications from other tools
- Taking over connections on startup and at runtime
- Full-featured API via private or system bus

Pavel Šimerda
pavlix@pavlix.net

Networking ecosystem

Pavel Šimerda
pavlix@pavlix.net

Kernel and libnl

- Fixing MAC address handling for bridges and bonds
- Improvment of loopback and link-local address handling
- Implementation of IPv4 lifetime for DHCP addresses (for 3.9)
- Rethinking and fixing the whole IPv6 configuration
- Fixing parts of the rtnetlink API
- Redesign of libnl's cache synchronization
- Various wireless driver bug fixes

http://fedoraproject.org/wiki/Networking/Bugs#Kernel

Pavel Šimerda
pavlix@pavlix.net

GNU C Library

- getaddrinfo() may return duplicate or even wrong addresses from /etc/hosts
- getaddrinfo() with NULL servname may return duplicate addresses
- getaddrinfo() with AI_PASSIVE may still return address list not suitable for bind()
- getaddrinfo() with AI_ADDRCONFIG may fail to translate literal addresses
- getaddrinfo() with AI_ADDRCONFIG may fail to resolve /etc/hosts addresses
- getaddrinfo() with AI_ADDRCONFIG may send unwanted AAAA queries
- getaddrinfo() has a bad choice of default flags

http://fedoraproject.org/wiki/Features/FixNetworkNameResolution

Pavel Šimerda
pavlix@pavlix.net

Various other projects

- Bugfixes and improvements in dhclient (not upstreamed)
- Reconsidering of DHCP timing, especially on wireless
- Wimax-tools ported to libnl 3.x which is used by NM
- Network connectivity dependencies in software
- And more...

Pavel Šimerda
pavlix@pavlix.net

Standards

- RFC 6106 (RA DNS): Assumes that IP (esp. multicast) is reliable
- RFC 4861 (NDP): Same as above, when stopping RA daemon
- RFC 3493 (getaddrinfo): Breaks name resolution based on the list of addresses
- POSIX.1-2008: Defines AI_ADDRCONFIG so that it's virtually useless

http://fedoraproject.org/wiki/Networking/Bugs#Networking_related_standards

Pavel Šimerda
pavlix@pavlix.net

For the networking community

Pavel Šimerda
pavlix@pavlix.net

Work with us to make linux networking better

- Discuss at #nm at Freenode or on the mailing list
- File and comment bug reports in our bugzilla
- Improve any projects we rely on and send patches
- Reduce excessive work duplication

Contact me at pavlix@pavlix.net. . .

http://data.pavlix.net/fosdem/2013/

pavlix@pavlix.net
psimerda@redhat.com

http://fedoraproject.org/wiki/User:Pavlix

http://fedoraproject.org/wiki/Networking