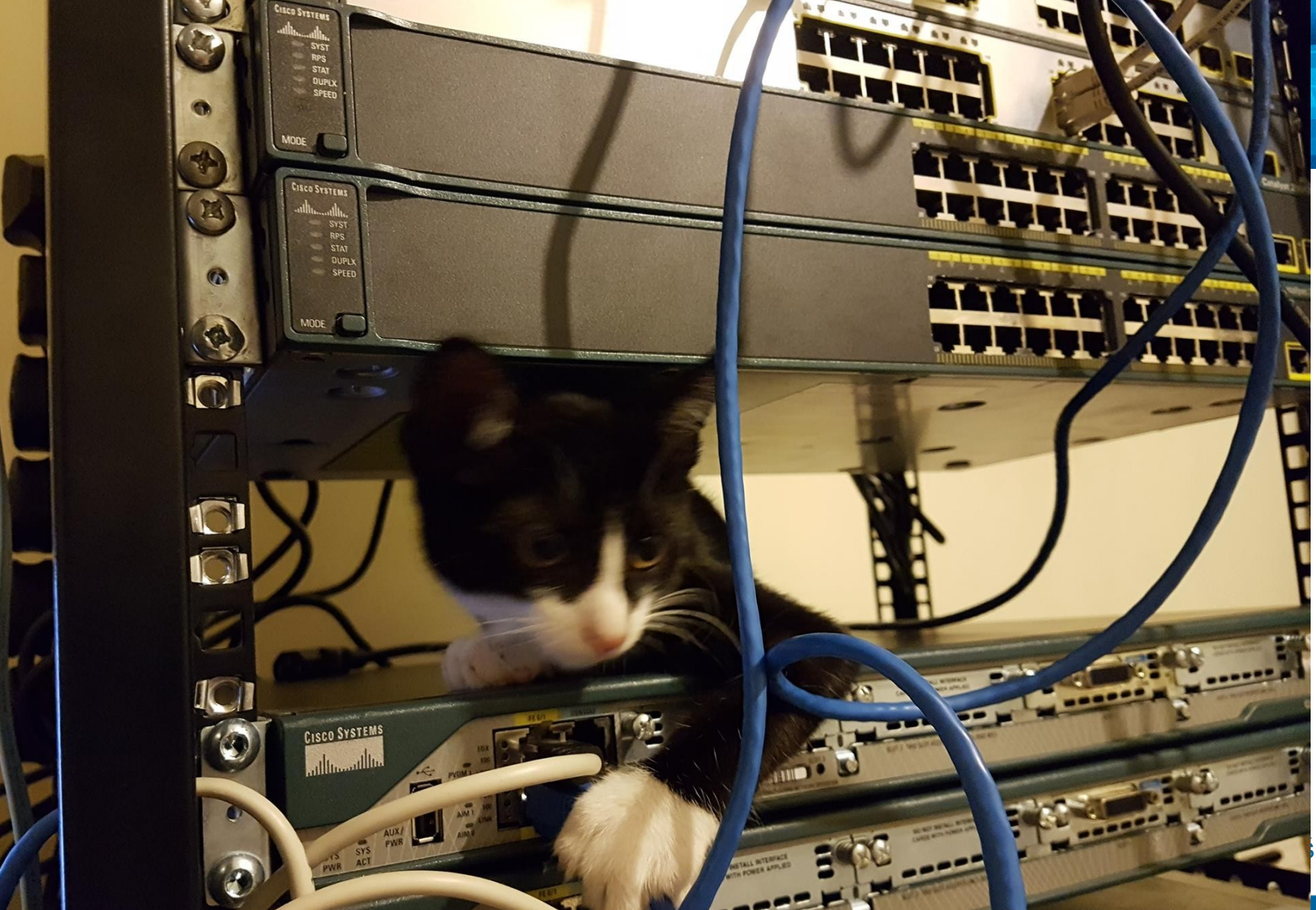
NETCAT



Ručka, CESNET

cesnet

AND SO... NIC

THE PITFALLS OF NOT REINVENTING THE ~~OPENSOURCE~~ WHEEL; CHAPTER 2

Lukáš Ručka
CESNET

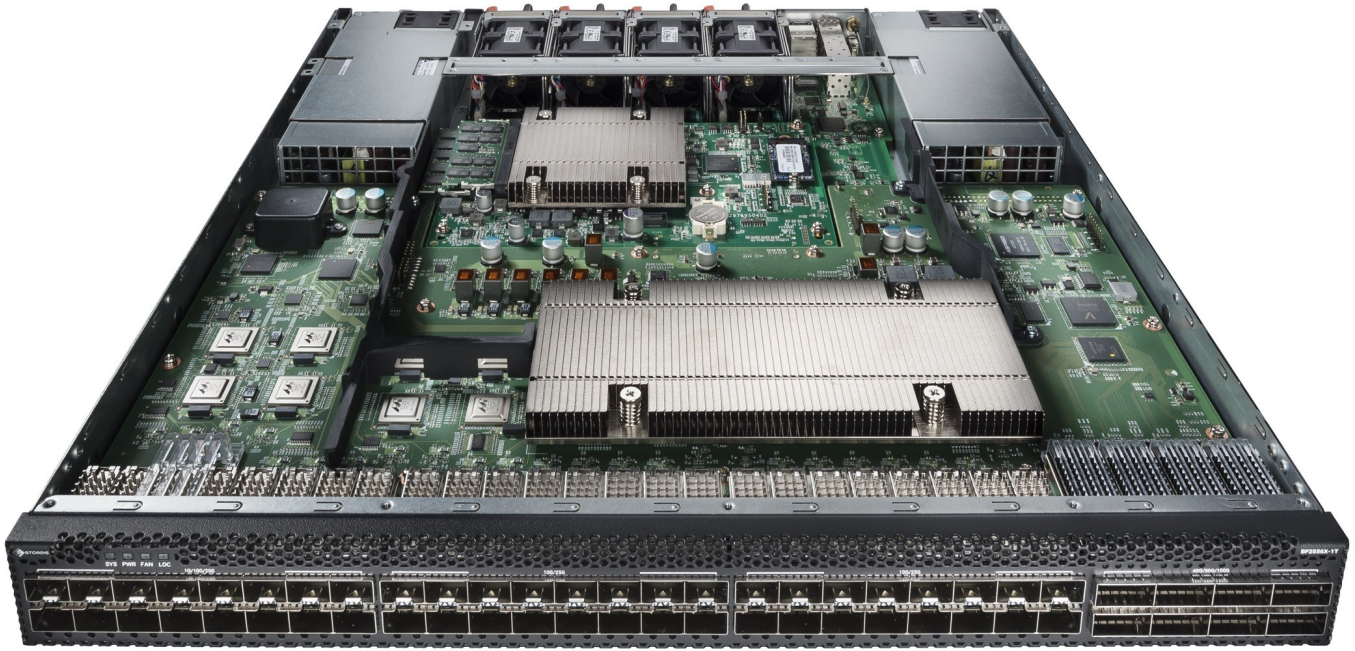
září 2023
Telč

AND SO... NIC

- **40 Gbps uplink (CESNET backbone)**
  - Single pair fiber to Mellanox 56Gbps @A507 rack B
- **10 Gbps uplink**
  - Single pair fiber to DELL 10Gbps @A507 rack B (optical links)
  - Chained to 3x Juniper ex3400 @A507 rack A (144 metallic links, +10 optical links)
- **Tofino – replacing 40 Gbps uplink with 100 Gbps**
  - APSN/Stordis bf2556x-1t (16x 1Gbps[1], 32x 10Gbps[1], 8x 100Gbps[1])
  - Facebook/Accton/Edgecore Wedge100BF-32QS (32x 100Gbps[1])

[1] **Lies**

STORDIS/APS NETWORKS BF2556X-1T



- https://www.opencompute.org/documents/210216-bf2556x-1t-switch-specifications-v2-pdf-1
- 52 pages of technical drawings and I2C register mappings

AND SO... NIC

- **Blackbox / off-the-shelf switch**
  - Hardware with (built-in) closed-source software
- **Whitebox switch**
  - Off-the-shelf switch with open design
  - Semi-customisable software
- **Baremetal**
  - Essentially mid-range server with PCI-express attached switch chip (ASIC)
  - What do you mean – „software“?
  - “““““ Support what“““““?

cesnet

~~TROUBLESHOOTING~~ SOFTWARE

(NETWORK) DESIGN

ASIC – Intel Tofino

- Tofino is essentially CPU – has (instruction) pipelines + memory
- Controlled by switch daemon (switchd) – driver and scheduler
- Application – set of rules applied to packets (match-action paradigm)
- „Multitasking“ – rule chaining

~~SDN~~ (cloud switch) vs standalone

- OpenCompute Stratum – requires SDN control plane
- ~~Azure~~[2] OpenCompute Sonic – standalone switch

[2] Yes, Microsoft’s Linux distribution

MANUFACTURER SUPPORT

- **APSN Sonic for bf2556x-1t**
  - Released as a modification of the 2018 Azure Sonic
    - No support for vlans
- **APSN Sonic platform packages for 2021 Sonic release**
  - Old crashy bf-sde (driver)
  - Requires Linux kernel 3.1X
    - The current Linux kernel stable version is 6.5.3
- **Choice of the day: reinventing the wheel[3]**
  - Low-level Tofino application (switch.p4) is configurable via interactive cli!

[3] ~~Do not~~ try this at ~~home~~ work, running somebody's worked-once software

cesnet

INTERACTIVE CLI
vs
BATCH-COMPATIBLE CLI

Expect, pyexpect

- Common tools, **expect/send/spawn** commands
- Tty-related issues (sudo?)

Tmux

- Terminal multiplexer, client/server architecture, multiple windows in one session
- Nearly every tmux control key sequence has a corresponding command
- Commands to control existing sessions and windows/panes + interactive mode
- Pane preserves context between command invocations
- **send-keys** supports any named key ; however, no **expect** (**save-buffer** hack)

- **tmux new-session -d -s gearboxctl ; tmux new-window -n gearboxcli -t gearboxctl**
  - Create a new tmux session with one interactive shell window and a named window
- **tmux send-keys -t gearboxctl.0 'docker run --rm -i -t ...' Enter || true**
- **tmux send-keys -t gearboxctl:gearboxcli gearboxcli Enter**
  - Beware on exit value for ongoing command (thus || **true**)
- **tmux kill-session -t gearboxctl**
  - Closes the session, signalling all children to end
- **tmux has-session -t gearboxctl:gearboxcli**
  - Check speciffic window exists
- **gearbox-confcheck | while read line ; do sleep 2s ; tmux send-keys -t gearboxctl:gearboxcli "$(echo -n "$line" | tr -d "\r\n")" Enter Enter ; done**

TMUX RESULTS

Results

- Gearbox (first 16 ports) up... (tmux + docker + systemd monster unit)

Caveats

- No persistent config – custom implementation required
- Converting config file to a sequence of tmux commands – handler script
- Has issues with the last line, timing-sensitive
- No restore policy (active / saved config, etc)

cesnet

KERNEL UPGRADE

PLATFORM DRIVER

- **Contexts**
  - Global states consistency
  - ~~Private~~ upper driver data, as somebody just dropped the crucial part of the driver
- **Multi-level multi-host I2C hell**
  - Hardwired dynamic bus numbers
  - No error reporting on failure
- **Kernel cannot kernel**
- **Current best-practices**
  - Use... bios or device tree provided hardware details
  - Use... userspace helpers

EEPROM DRIVER

- **Code management**
  - Kernel is git, APSN support package is just a bunch of files
- **Manually-applied patches**
  - Mixed indents
  - Mixed ordering of functions
- **Probably backported upstream driver**
  - Era uncertain
  - Upstream driver distinct, or not?
  - Upstream completely rewritten

- **git blame**
  - shows which line was affected by which commit
- **git log --reverse --ancestry-path 4bb5..fd15 ./at24.c**
  - shows commits touching file (including merges)
- **git format-patch --stdout -1 021c ./at24.c**
  - exports single commit in patch/mbox format
- **git am patchfile.patch (--abort when glue commit is required)**
  - imports exported commit, including author and comment; can fail (but sets metadata)
- **patch -p 1 patchfile.patch**
  - attempts to weak-apply the commit

ONIE DRIVER

- **ONIE = Open Network Instal Environment**
  - Miniature live Linux environment
  - Similar advancement to PXE boot, as EFI to BIOS
- **Crutial EEPROM driver**
  - Switch SN, model
  - Chasis status
  - IPMI over SMBus?
- **SoNIC watchdog**
  - No platform EEPROM -> reboot upon timeout

cesnet

ST(R)ATUSM

SoNIC

- Build environment consumes 60GB
- Approx 380 cpu-hours, 200GB RAM
- Binary image (Debian + numerous docker service images)
- Switch service crashes, no relevant log
- Dependencies pulled as Github assets, hardwired dependency URL[4]

Takeaways

- Automate interactive apps (console – expect / tmux, graphical - xdotool)
- Preffer glue commits + patch adoption – git can track it
- No hardwiring (IPMI login: `root:0penBMC`)

[4] failed build → Github download limit reached

cesnet

THANK YOU FOR YOUR ATTENTION

QUESTIONS*?

* where/when is my 100G is not a polite question

Switch teardown photo, slide 3

Realwire News: Delivering relevance, releasing influence. In: REALWIRE LIMITED. STORDIS Launches First Optimised Barefoot Tofino Switches with Time Synchronisation and 8-Core Processing Power: STORDIS Launches First Optimised Barefoot Tofino Switches with Time Synchronisation and 8-Core Processing Power [online]. 2019 [cit. 2023-09-14]. Available: https://www.realwire.com/releases/STORDIS-Launches-First-Optimised-Barefoot-Tofino-Switches