HKG18-211: Android Common Kernel and Out of Tree Patchset

Amit Pundir
Session Layout

- Android common kernel
  - Rationale and Brief Introduction
  - Patchset evolution
- linux-v4.14.y vs android-4.14
  - Git diff stats - v4.14..android-4.14
  - Quick intro of out of tree patches/functionality and upstream status
- experimental/android-mainline-tracking
  - Rationale and Brief introduction
  - Git diff stats - linux/master..experimental/android-mainline-tracking
- Android kernel patchset delta and call to action
Android Common Kernel
Android Common Kernel Rationale

- Downstream Long Term Stable (LTS) kernel
- Plus a handful (400+) of out of upstream tree patches
  - Features tailored for AOSP needs e.g. quota2, qtaguid
  - Features rejected by upstream due to implementation concerns e.g. USB gadgets, Paranoid networking
  - Testbed for features to be pushed upstream over time e.g. EAS
  - Features which are available mainline but Android kernel still use in-house implementations e.g. PPPoPNS, PPPoLAC
  - Vendor/OEM features which deemed useful for whole ecosystem e.g. sdcardfs
- Channel for partners to get timely updates including LTS patches, bugfixes and new Android kernel features
Android Common Kernel Tree

Clone this repo:

git clone https://android.googlesource.com/kernel/common

Branches

- master
- android-3.10
- android-3.10-y
- android-3.18
- android-3.18-n-release
- android-3.18-o-mr1
- android-3.18-o-release
- android-4.14
- android-4.4
- android-4.4-eas-test

More »
Android Common Kernel Branch Hierarchy

The branch hierarchy for the android-4.4 kernel uses the following structure:

- Stable Long Term Support (LTS) Kernel (4.4.y)
  - AOSP Common Kernel (android-4.4)
    - AOSP N Kernel (android-4.4-n regular merges from LTS (not shown))
      - AOSP Release Kernel (android-4.4-n-release) security patches
        - AOSP Release Kernel (android-4.4-n-mr1)
          - AOSP O Kernel (android-4.0)

- OEM/ODM Kernel
  - Clone
  - Release/Update/Upgrade
  - Feature
  - Merge
  - Cherrypick/backport

https://source.android.com/devices/architecture/kernel/android-common
Android Common Kernel Testing

- Moved from a mere reference set of patchset dump to community driven build/boot tests and functional testing.
- KernelCI
  - KernelCI build loop and smoke test for android common kernels
    https://kernelci.org/job/android/
- Linux Kernel Functional Testing
  - Running VTS/CTS on AOSP builds
    - https://qa-reports.linaro.org/android-lkft/
  - Upstream Kselftests/LTP subset of tests on Open Embedded builds
    - https://qa-reports.linaro.org/lkft
  - ELC2018: Keeping up with LTS: Linux Kernel Functional Testing (LKFT) on Devices - Thomas Gall, Linaro
Android Kernel Patchset Evolution
What's in the Android Patches?

- Ashmem
- Binder
- Pmem
- Logger
- Early suspend
- Wakelocks
- Alarm Timer
- LowMemoryKiller
- Paranoid network
- Yaffs2 fs
- Ram_console
- Apanic
- Adb gadget driver
- Gpio patches
- Lots of other small fixes and hacks for arm, mmc, Bluetooth™, etc.

v4.4..android-4.4 git diff stats
- 523 files changed, 46634 insertions(+), 1634 deletions(-)
  - Including UPSTREAM / BACKPORT / FROM LIST fixes
- ~14% Networking
- ~09% Energy Aware Scheduling
- ~09% USB Gadgets
- ~09% Atomic Display Framework
- ~08% Verity Boot
- ~08% Sdcard FS
- ~04% FIQ debugger
- ~04% Input
- ~03% Cpufreq
- Rest: Documentation, include, kernel, arch, mm..
linux-4.14.20 vs android-4.14
v4.14.20..android-4.14 git diff stats

- 432 files changed, 39445 insertions(+), 2730 deletions(-)
  - Including UPSTREAM / BACKPORT / FROM LIST fixes
- ~12% Sdcard FS
- ~11% Netfilter
- ~11% Energy Aware Scheduling
- ~6% USB gadgets
- ~5% Input
- ~5% Fiq debugger
- Others: arch/, Documentation/, include/ et al.
v4.4..android-4.4 git diff stats

- 523 files changed, 46634 insertions(+), 1634 deletions(-)
  ○ Including UPSTREAM / BACKPORT / FROM LIST fixes
- ~14% Networking (UID Based Routing, PPP, SIOCKILLADDR)
- ~09% Energy Aware Scheduling
- ~09% USB Gadgets (MTP/PTP, Misc RNDIS hacks/fixes)
- 00% Atomic Display Framework
- ~08% Verity Boot
- ~08% Sdcard FS
- ~04% FIQ debugger
- ~04% Input
- 03% Cpusfreq (Interactive Governor)
- Rest: Documentation, include, kernel, arch, mm..

v4.14.20..android-4.14 git diff stats

- 432 files changed, 39445 insertions(+), 2730 deletions(-)
  ○ Including UPSTREAM / BACKPORT / FROM LIST fixes
- ~12% Sdcard FS
- ~11% Netfilter
- ~11% Energy Aware Scheduling
- ~06% USB gadgets
- ~05% Input
- ~05% Fiq debugger
- Others: arch/ , Documentation/ , include/ et al.
v4.14.20..android-4.14 git diff stats

- ~31% drivers/
  - ~06% usb gadgets
  - ~05% fiq_debugger
  - ~03% goldfish
  - ~05% input
    (keyreset, keychord, keycombo, gpio)
  - ~03% dm (dm-android-verity)
  - ~02% misc
    (uid sys stats, memory state time)

- ~23% fs/
  - ~12% sdcardfs
  - ~05% f2fs upstream backports
  - ~03% squashfs

- ~12% net/
  - ~11% netfilter (quota2, qtaguid)
  - Paranoid networking, NFC fixes

- ~13% kernel/
  - ~11% Energy Aware Scheduling
  - Wakeup reason logging/reporting

- ~05% arch/
  - Appended -dtb kernel Image build targets
  - Debug hooks, Ranchu defconfigs etc

- Others:
  - Documentation, include, mm, mmc, tracing hooks et al.
File System

- **SdcardFS**
  - What is sdcardsfs?
    - FUSE alternative for emulated storage in AOSP
      - Got rid of a lot of I/O performance overhead in Fuse
    - Not a traditional file system like ext4, fat32
      - File system wrapper derived from wrapfs
      - Implement in-kernel FAT32 emulation layer to manipulate permissions and case sensitivity
    - Shipped in Samsung devices for long
  - Upstream to staging tree is in progress
    - [LPC2017: SDCardFS Upstreaming - Daniel Rosenberg, Google](#)
    - Fuse support being completely dropped from AOSP broke testing/development efforts on mainline kernel
  - Huawei’s sdcardsfs implementation (hwsdcardfs)
    - Huawei’s in house sdcardsfs alternative, up for review on [AOSP Gerrit](#)
    - Already shipping in Mate 10/Pro and other devices.
File System

● F2FS (Flash-Friendly File System)
  ○ Developed for NAND flash memory-based storage devices from the ground up
  ○ Follow upstream development model and patches backported on Android kernel
  ○ No out of tree patches

● SquashFS
  ○ Read-only compressed filesystem for Linux devices
  ○ Out of tree Android kernel changes include:
    ■ Custom readpages() implementation
      ● Pack as much pages as possible in the same page actor so that only 1 read request is issued.
    ■ Optimize reading uncompressed data
    ■ lz4 being recommended compression algorithm
Networking

- Paranoid networking
  - Restrict network access to certain group of users
  - Largely perceived as Android kernel hacks with hardcoded AIDs mapped to userspace groupids
  - Hardcoded userIDs rejected upstream
    - Asked to move to n/w namespaces for n/w filtering based on control groups instead.
    - Require a fair bit of userland changes and unlikely to happen to save mere tens of lines of out of tree code.
  - On recent AOSP releases, access to services/daemons can be enabled via upstream CAP_NET_RAW and CAP_NET_ADMIN capabilities as well.
    - So we might be able to drop few paranoid networking checks in future.
Networking

● Netfilter: qtaguid, quota2, idletimer
  ○ Data usage tracking & limiting
    ■ qtaguid and quota2 modules to do per uid usage tracking and accounting
    ■ LPC2017: Replacing xt_qtaguid with an upstream eBPF implementation
  ○ IDLETIMER notifications
    ■ Help ConnectivityService deal with quiet interfaces
    ■ Track and send netlink messages when interface becomes active again after an idle period
      ● Functionality to be moved to userspace netd (NFLogListener)
Energy Aware Scheduler

- To make Linux scheduler fully aware of the CPU capabilities and optimize energy consumption
- Generic baseline design & Arch/SoC independent solution
- Under active development, testing and upstreaming phase
  - Patches/discussions can be tracked on lkml and ARM’s [linux-power.git](https://github.com/ARM-labs/linux-power.git) repo
  - LPC2017: Energy Aware scheduler development
USB

● **USB Gadgets**
  ○ USB device state changes
    ■ UEVENT notifications to userspace (UsbDeviceManager)
    ■ Custom device class interface (/sys/class/android_usb/android0)
      ● Used mostly for legacy/non-configfs gadgets to track device state changes, functions enabled etc
    ■ Rejected upstream
      ● Should read usb state changes from upstream interface /sys/class/udc/*/state instead
  ○ Android Accessory driver
    ■ USB accessory mode allows users to connect USB host hardware specifically designed for AOSP powered devices e.g. Kiosks
    ■ Audio Source driver
      ● USB Audio support in accessory mode.
      ● AOSP device serves as a data source to the host.
      ● Accessory mode audio has not been widely adopted, and is not currently recommended for new designs.
USB

- Dual Role USB Phy sysfs interface
  - Generic /sys/class/dual_role_usb/ interface to track and change the state of dual role usb ports

- OTG Wakelock
  - Grab a wakelock when vbus is present
    - Let user keep devices awake during charging
    - Or temporarily wake-up the device on charger connect/disconnect events
Android Verified Boot

- Block-level integrity protection and forward error correction
  - For AOSP system and vendor/oem read-only partitions
- dm-android-verity
  - dm-verity extension to implement verified boot for AOSP
  - Helps ensure AOSP users booting a verified and trusted software
- dm-verity-fec
  - Add sysfs attribute to account for correctable error stats
Input

● Keychord, Keyreset/Keycombo drivers
  ○ Driver to handle different key press combinations
  ○ Chunks of it already upstreamed as part of SYSRQ driver

● Generic GPIO input support
  ○ Supports keyboard matrices
  ○ Direct inputs/outputs
  ○ Axes connected to gpios
FIQ debugger

- Low level kernel debugger for ARM
- Intended to use robust Fast IRQ (FIQ) interface for debugging
  - FIQ is similar to an NMI on x86
  - Debugger fall backs to using IRQ otherwise
- Parts of it already integrated with upstream KDB
- In Progress:
  - KDB extensions
  - fiq_debugger like feature for ARMv8
  - Extend NMI watchdog
  - IPI FIQ for ARM
Others

- Private Anonymous memory
  - Anon memory tagged/named by userspace to track and debug physical memory usage

- Scheduling Cgroups for cpusets, cpuctrl and schedtune
  - Restrict cpu-usage per task by putting apps into different buckets "TOP_APP / FOREGROUND / SYSTEM / BACKGROUND".
  - Let processes (system_server, binder viz.) move other tasks if they have CAP_SYS_NICE in the affected task's user namespace.

- Goldfish
  - Add goldfish sync driver
  - ACPI based enumeration of framebuffer and audio
  - Ranchu defconfigs
Others

- **ARCH**
  - Appended -dtb (image.gz-dtb, image-dtb) support
  - CONFIG_CMDLINE_EXTEND support
  - Dump memory around registers when displaying regs

- **Tracing**
  - MMC, GPU, Min/Max cpufreq, sched uninterruptible sleep tracing

- **MMC**
  - Sysfs interface for IO latency histogram
  - Additional retries on SD detection
  - Embedded sdio support and other sdio fixes

- **Memory State Time driver**
  - New memory_state_time driver tracks time spent in different DDR frequency and bandwidth states
Others

- **Security / Perf**
  - Add an option to restrict all access to performance events by users without CAP_SYS_ADMIN, to reduce the attack surface of the kernel
  - Shot down upstream [https://patchwork.kernel.org/patch/9249919/](https://patchwork.kernel.org/patch/9249919/)

- **Net**
  - Sysfs based knobs for controlling TCP window size
  - Kconfig to keep RFKILL controlled devices awake during suspend
  - Sysctl knob to control the initial congestion window

- **init**
  - `skip_initramfs` option to allow choosing whether to boot using the initramfs or not at runtime.
Others

- **FS**
  - Tracepoints in ext4/f2fs/mpage to track readpages/buffered write()s.
  - Task I/O accounting: counter to track fsync

- **NFC**
  - Couple of buffer overflow and memory corruption fixes.

- **Power**
  - Log wakeup reason and source
  - Report suspend time

- **UID SYS stats**
  - Per UID based system statistics exported to /proc
    - For example: stats to be used by BatteryStats service
experimental/android-mainline-tracking
android-mainline-tracking Rationale

- Android kernel patchset / features rebased to latest Linux release or -rc
  - Major non-LTS release versions are tagged.
  - For example: experimental-android-4.15
- Find/Report/Fix AOSP regressions or ABI breakages in upstream kernel.
- Testbed for, yet to be submitted, upstream Android kernel patches.
- To be used as a reference or experimental preview tree by member partners for upcoming SoCs.
v4.15..android-mainline-tracking git diff stats

- 369 files changed, 36272 insertions(+), 1741 deletions(-)
- ~13% Energy Aware Scheduling
- ~13% Sdcard FS
- ~13% Netfilter
- ~08% USB
- ~06% ARCH
- ~06% Input
- ~05% FIQ debugger
- ~04% Goldfish
- ~03% Device Mapper
Android Patchset Delta

[Bar chart showing line of patch for different Android versions.]
Call To Action

- While a good chunk of Android kernel changes have made it upstream, there is a good chunk of static delta that doesn't have an owner actively trying to upstream it.
- On an average around 38K lines of code keep getting rebased on next Android common kernels.
Thank You

#HKG18

HKG18 keynotes and videos on: connect.linaro.org
For further information: www.linaro.org