Optimize Kernel Configuration Application Note

80-PJ035-1 A

May 9, 2018

Qualcomm Technologies, Inc.
5775 Morehouse Drive
San Diego, CA 92121
U.S.A.

Revision history

Revision	Date	Description
A	May 2018	Initial release

Contents

1 Introduction

1.1 Purpose

This document describes how to optimize kernel configuration for perf-defconfig and performance builds.

1.2 Conventions

Function declarations, function names, type declarations, attributes, and code samples appear in a different font, for example, `cp armcc armcpp`.

Code variables appear in angle brackets, for example, `<number>`.

1.3 Technical assistance

For assistance or clarification on information in this document, submit a case to Qualcomm Technologies, Inc. (QTI) at https://createpoint.qti.qualcomm.com/.

If you do not have access to the CDMATech Support website, register for access or send email to support.cdmatech@qti.qualcomm.com.

2 Recommended kernel configuration file

It is recommended to use <chipset>-perf_defconfig as the user build reference configuration. It is tuned for better system performance and demonstrates improved kernel memory footprint.

2.1 LA-based environment

The corresponding perf-defconfig reference is declared in AndroidBoard.mk file as shown in the following snippet. The chipset-specific AndroidBoard.mk file must be taken as the reference. For example, the perf-defconfig reference for the MSM8953 chipset is as follows:

```
#----------------------------------------------------------------------
# Compile Linux Kernel
#-----------------------------------------------------------
------------
ifeq ($(KERNEL_DEFCONFIG),) ifeq ($(TARGET_BUILD_VARIANT),user)
KERNEL_DEFCONFIG:= msmcortex-perf_defconfig else
KERNEL_DEFCONFIG:= msmcortex_defconfig
endif
endif
```

- For any chipset, the corresponding AndroidBoard.mk reference file can be found in the following paths for the MSM8953 and MSM8909W chipsets:

  ```
  device/qcom/<(chipset_name)_32/64>/AndroidBoard.mk
  ```

- To prepare the image for userdebug build with perf-defconfig, make the following changes, for example, for the MSM8909W chipset:

  a. Go to device/qcom/msm8909w/AndroidBoard.mk

  b. Use of KERNEL_DEFCONFIG as perf-defconfig:

     ```
     ifeq ($(KERNEL_DEFCONFIG),) ifeq ($(TARGET_BUILD_VARIANT),user)
     KERNEL_DEFCONFIG := msmXXXX-perf_defconfig
     else
     KERNEL_DEFCONFIG := msmXXXX-perf_defconfig
     endif
     endif
     ```

2.2 LE-based environment

2.2.1 APQ8053

The corresponding perf-defconfig reference is declared in the apq8053.conf file. The corresponding apq8053.conf reference file can found in the following path: “/poky/meta-qti-bsp/conf/machine/apq8053.conf” #Perf Kernel config

```
KERNEL_PERF_IMAGETYPE = “Image-dtb”
KERNEL_PERF_DEFCONFIG = “apq8053_IoE-perf_defconfig”
```

2.2.2 APQ8009

The corresponding perf-defconfig reference is declared in the apq8009.conf file. The corresponding apq8009.conf reference file can be found in the following path: “/poky/meta-qti-bsp/conf/machine/apq8009.conf”

```
KERNEL_DEFCONFIG = “apq8909_defconfig”
KERNEL_PERF_DEFCONFIG = “apq8909-perf_defconfig”
```

2.3 Get kernel config from device

Use the following command:

```
adb pull /proc/config.gz
```

2.4 Extract kernel config

- To extract kernel config when pulling /proc/config.gz does not work, use the following steps:
  a. Get the vmlinux file and copy it to kernel source folder.
  b. Run the the following script from any kernel source folder on the vmlinux file: ./kernel/scripts/extract-ikconfig vmlinux > kernel_config.txt
- At times, extracting of kernel_config file from vmlinux may not work. In this case, use the following snippet to get the .config file from the source build. Check for the file as shown in the following:

  ```
  ls -a /out/target/product/msm<chipset>/obj/KERNEL_OBJ/.config
  ```

  The file is a hidden file ".config" <.dot config>

NOTE: To get the kernel config from the device and extract the kernel config from vmlinux, enable the following config in defconfig:

```
CONFIG_IKCONFIG_PROC=Y
CONFIG_IKCONFIG=Y
```

2.5 Disable configs

- To disable the configs, either change ‘y’ to ‘n’ or delete the config from perf-defconfig
- For example, for the MSM8953 chipset:
  a. Go to /kernel/arch/arm64/configs/msmcortex-perf_defconfig file
  b. Disable the CONFIG_CGROUP_DEBUG config:

```
+CONFIG_CGROUP_DEBUG=n
```

or

```
- CONFIG_CGROUP_DEBUG=y
```

2.6 Review kernel config

1. Pull the kernel config from the device using “adb pull /proc/config.gz”.
2. Review the kernel config whether changes are reflected or not.
   - If changes are made in defconfig to enable or disable the config, the following changes are seen in defconfig:

```
CONFIG_IKCONFIG=y              //If this CONFIG is enabled
# CONFIG_IKCONFIG is not set    //If this CONFIG is disabled
```

2.7 Configs disabled in perf-defconfig

Ensure the following configs are disabled in perf-defconfig:

- CONFIG_SERIAL_CORE=y
  - Initializes serial driver for early printk
- CONFIG_SERIAL_CORE_CONSOLE=y
  - Used for UART console write (write a console message to a serial port)
- CONFIG_SERIAL_MSM_HS=y
- CONFIG_SERIAL_MSM_HSL=y
- CONFIG_SERIAL_MSM_HSL_CONSOLE=y
  - By enabling these configs, it is possible to enable it’s onboard high-speed serial port for machine, based on the MSM™ chipset family of SoCs

NOTE: By disabling SERIAL configs, boot time and performance is improved.

- CONFIG_CGROUP_DEBUG=y
  - Enables a simple controller that exports debugging information about the control group (cgroup) framework
  - Disable this config for performance
- CONFIG_SLUB_DEBUG=y
- CONFIG_SLUB_DEBUG_ON=y

  - SLUB has extensive debug support features; disabling these can result in significant savings in code size
  - This also disables SLUB sysfs support (/sys/slab will not exist and there will be no support for cache validation, and so on)
- CONFIG_MSM_SMD_DEBUG=y
  - Disable this DEBUG config for performance
- CONFIG_DYNAMIC_DEBUG=y
  - Enables dynamic printk() support
  - Dynamic debugging is controlled via the 'dynamic_debug/control' file, which is contained in the 'debugfs' file system
  - The debugfs file system must first be mounted before making use of this feature
- CONFIG_DEBUG_PAGEALLOC=y
  - Unmap pages from the kernel linear mapping after free_pages()
  - Helps to find certain types of memory corruptions
  - Results in a large slowdown, ensure to disable this config.
- CONFIG_DEBUG_KMEMLEAK=y
  - If set to ‘y’, this Config enables the kernel memory leak detector
  - Enabling this feature introduces an overhead to memory allocations
  - See Documentation/dev-tools/kmemleak.rst for details
- CONFIG_DEBUG_KMEMLEAK_EARLY_LOG_SIZE=400
  - Used for maximum kmemleak early log entries
  - If CONFIG_DEBUG_KMEMLEAK is not set, this config is not required to be set
- CONFIG_DEBUG_KMEMLEAK_DEFAULT_OFF=y
  - Default kmemleak to OFF
- CONFIG_DEBUG_SPINLOCK=y
  - Used for debug spinlock usage
  - Disable this config for performance
- CONFIG_DEBUG_MUTEXES=y
  - Used for mutex debugging and deadlock detection
- CONFIG_DEBUG_ATOMIC_SLEEP=y
  - Sleep inside atomic section checking
  - Various routines which may sleep will become noisy, if called inside atomic sections when a spinlock is held, inside an RCU read side critical section, inside preempt disabled sections, inside an interrupt, and so on
- CONFIG_DEBUG_STACK_USAGE=y
  - Stacks utilization instrumentation

- CONFIG_DEBUG_LIST=y
  - Debugs linked list manipulation
- CONFIG_LOCKUP_DETECTOR=y
  - Detects hard and soft lockups
- CONFIG_SCHED_DEBUG=y
  - Collects scheduler debugging information
  - /proc/sched_debug file is provided for debug
- CONFIG_SYSRQ_SCHED_DEBUG=y
  - Results in verbose printk output when the "show-taskstates(T)" or "show-blocked-tasks(W)" sysrq triggers are invoked
  - Since this can increase the chances of a watchdog bark when the system is under heavy load, disable these debug prints
- CONFIG_DEBUG_MEMORY_INIT=y
  - Enables additional check during memory initialization
- CONFIG_DEBUG_PREEMPT=y
  - Enable kernel will use a debug variant of the commonly used smp_processor_id() function and will print warnings if kernel code uses it in a preemption-unsafe way
  - Kernel will detect preemption count underflows
- CONFIG_PAGE_POISONING=y
  - Poison page after freeing
  - Fill the pages with poison patterns after free_pages() and verify the patterns before alloc_pages()
  - Results in a large slowdown, but helps to find types of memory corruption
  - Cannot be enabled in combination with hibernation as it results in incorrect warnings of memory corruption after a resume since free pages are not saved to the suspend image
- CONFIG_RMNET_DATA_DEBUG_PKT=y
  - Enable debug for RmNet data to support MAP data feature for data connectivity
- CONFIG_BLK_DEV_LOOP=y
  - Loopback device support
- CONFIG_RCU_CPU_STALL_VERBOSE=y
  - Enables additional checks to print additional per-task information for RCU_CPU_STALL_DETECTOR
  - Causes RCU to invoke pirntk() per task, for any task that stalls the current RCU grace period

- CONFIG_TRACEPOINTS=y
  - Enables the kernel trace points for debugging
- CONFIG_DEBUG_BUGVERBOSE=y
  - Verbose BUG() reporting
  - If set to ‘y’, make BUG() panics output the file name and line number of the BUG call as well as the EIP and OOPS trace
  - Aids debugging, but costs about 70 to 100 K of memory
- CONFIG_CC_OPTIMIZE_FOR_SIZE=y
  - By disabling this config, boot time improvement of ~ 100 ms
  - As it causes GCC to generate jump-to-jump code, which causes cache line bouncing and hurting performance, disable this config
- CONFIG_DEBUG_SET_MODULE_RONX=y
  - Sets loadable kernel module data as no execute (NX) and text as read only (RO)
- CONFIG_CORESIGHT=y
- CONFIG_CORESIGHT_EVENT=y
- CONFIG_CORESIGHT_FUSE=y
- CONFIG_CORESIGHT_CTI=y
- CONFIG_CORESIGHT_TMC=y
- CONFIG_CORESIGHT_TPIU=y
- CONFIG_CORESIGHT_FUNNEL=y
- CONFIG_CORESIGHT_REPLICATOR=y
- CONFIG_CORESIGHT_STM=y
- CONFIG_CORESIGHT_HWEVENT=y
- CONFIG_CORESIGHT_ETMV4=y
- CONFIG_CORESIGHT_MODEM_ETM=y
- CONFIG_CORESIGHT_WCN_ETM=y
- CONFIG_CORESIGHT_RPM_ETM=y
  - CORESIGHT configs are used for CoreSight tracing support
  - Disable all CORESIGHT configs for performance
- CONFIG_RELAY=y-
  - User space relay support
  - Enables support for relay interface support in certain file systems (such as debugfs)
- CONFIG_FAULT_INJECTION=y
  - Provides fault-injection framework
  - For details, see Documentation/fault-injection/

- CONFIG_FAULT_INJECTION_STACKTRACE_FILTER=y
  - Provides stacktrace filter for fault-injection capabilities
- CONFIG_ALLOC_BUFFERS_IN_4K_CHUNKS=y
  - Allocates buffer in page order 0, if size >= 64 KB (need to disable in case of page order >=4)
- CONFIG_IOMMU_DEBUG=y
  - Enables IOMMU debugging
  - Forces the IOMMU to ON, even when you have less than 4 GB of memory and add debugging code
  - On overflow, always panic and allow to enable IOMMU leak tracing
  - Ensure to have a large IOMMU/AGP aperture

NOTE: There is a compilation error while disabling CONFIG_CC_OPTIMIZE_FOR_SIZE=y

- To disable CONFIG_CC_OPTIMIZE_FOR_SIZE=y, config requires changes in the main Makefile for compilation diff --git a/Makefile b/Makefile index 4045f28..326ca08 100644

```
--- a/Makefile
+++ b/Makefile
     @@ -619,13 +619,15 @@ KBUILD_CFLAGS   += $(call cc-disable-
warning,frame-address,)
        KBUILD_CFLAGS         += $(call cc-disable-
                              warning, format-
                              truncation)
                              += $(call cc-disable-
        KBUILD_CFLAGS         warning, format-
                              overflow)
                              += $(call cc-disable-
        KBUILD_CFLAGS         warning, int-in-bool-
                              context)
                              += $(call cc-disable-
       +KBUILD_CFLAGS         warning, array-bounds)
                              += $(call cc-option,-
        KBUILD_CFLAGS         fno-PIE)
                              += $(call cc-option,-
        KBUILD_AFLAGS         fno-PIE)

 ifdef CONFIG_CC_OPTIMIZE_FOR_SIZE
        KBUILD_CFLAGS         += -Os $(call cc-disable-
        else                  warning,maybe-
                              uninitialized,)

       -KBUILD_CFLAGS         += -O2
```

```
                             += -O2 $(call cc-disable-
                             warning,maybe-
      +KBUILD_CFLAGS         uninitialized,)
+
 Endif

diff --git a/arch/arm/configs/msmcortex-perf_defconfig
b/arch/arm/configs/msmcortexperf_defconfig index 1d28f37..264f22c 100644
--- a/arch/arm/configs/msmcortex-perf_defconfig
+++ b/arch/arm/configs/msmcortex-perf_defconfig
@@ -27,7 +27,6 @@ CONFIG_NAMESPACES=y
 CONFIG_BLK_DEV_INITRD=y
 CONFIG_RD_BZIP2=y
 CONFIG_RD_LZMA=y
-CONFIG_CC_OPTIMIZE_FOR_SIZE=y
 CONFIG_KALLSYMS_ALL=y
 CONFIG_EMBEDDED=y
 # CONFIG_SLUB_DEBUG is not set
```

A References

A.1 Acronyms and terms

Acronym or term	Definition
Cgroup	Control group
NX	No execute
RO	Read only