Design of kdump in the kvmibm-installer

Tao Wu

September 1, 2015

1 Introduction

Kdump is a kernel crash dumping mechanism and is very reliable because the crash dump is captured from the context of a freshly booted kernel and not from the context of the crashed kernel. Kdump uses kexec to boot into a second kernel whenever system crashes. This second kernel, often called the crash kernel, boots with very little memory and captures the dump image.

The first kernel reserves a section of memory that the second kernel uses to boot. Kexec enables booting the capture kernel without going through an IPL sequence, so contents of the first kernel's memory are preserved, which is essential when kernel crash dump.

2 Introduction of design

2.1 Functions provided

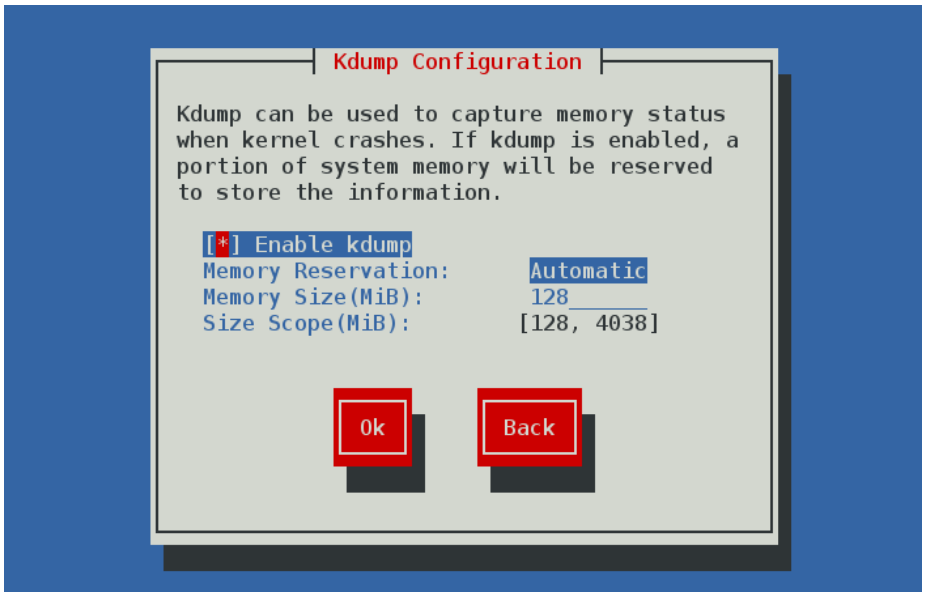
- Enable or disable kdump mechanism.
- Automatically or Manually reserve system memory for kdump.

2.2 Composition

The main form consists of

- A piece of brief introduction text.
- A check box to enable/disable kdump, enabled by default.
- A list box providing the choice of 'Automatic/Manual' mode of memory reservation, 'Automatic' by default.
- An entry for inputting memory size, '128' by default.

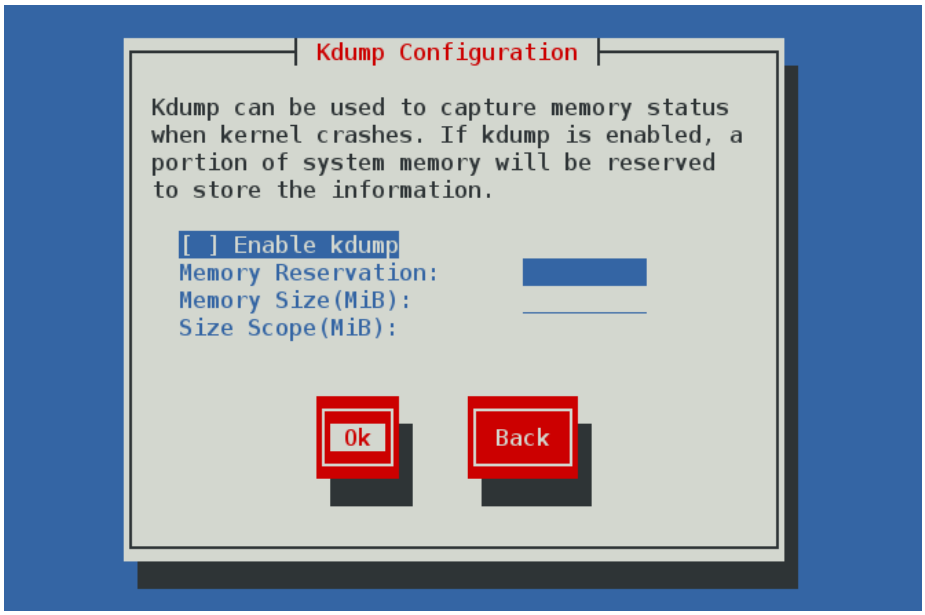
- And a text prompting the size scope which is in the form of [lower, upper], and lower is 128 by default, while upper is evaluated by running time.



2.3 Navigation

2.3.1 Enable/Disable kdump check box

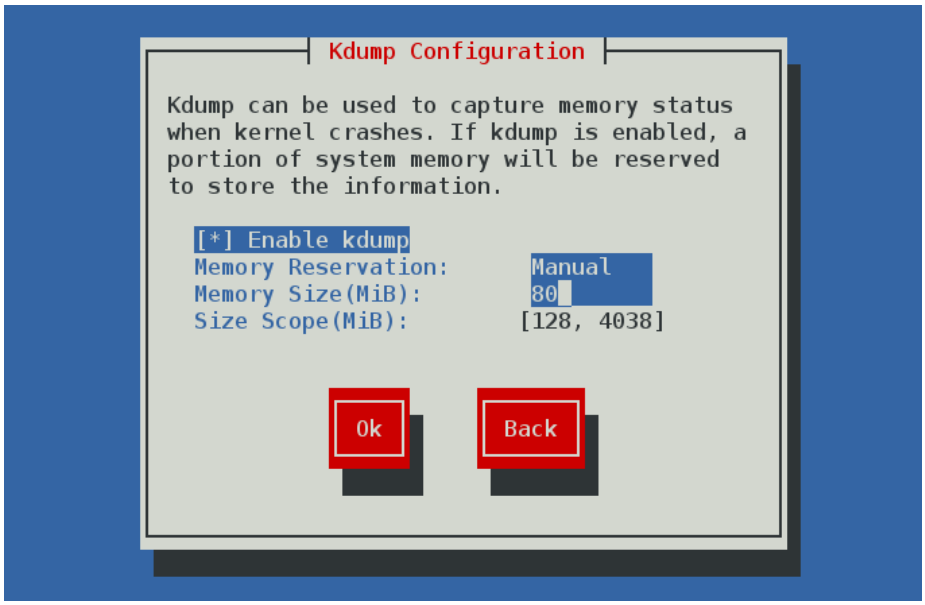
Unselecting the checkbox of “Enable kdump” will cause all values being turned to blank and transferring the focus to the ‘Ok’ button.



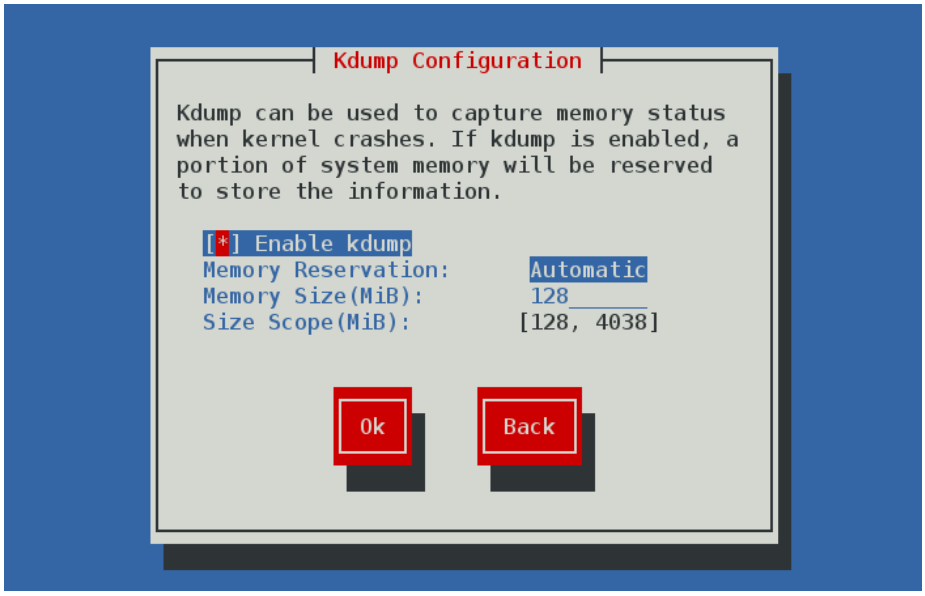
Alternatively, selecting the checkbox of "Enable kdump" will cause all values being restored to original values and setting the focus at the check box.

2.3.2 Switch among memory reservation list box

When 'Manual' item is selected, the check box of "Enable kdump" will be selected and the memory size entry being opened to accept input, and the default size scope emerges.

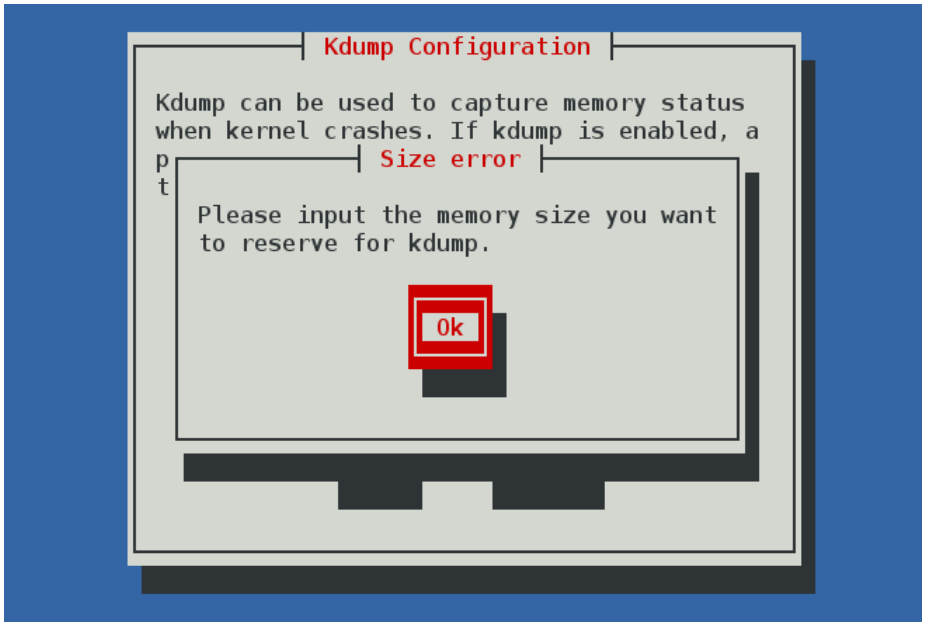


When 'Automatic item is selected, the check box of "Enable kdump" will be selected and the memory size entry changes to '128' and being closed to accept input, and the default size scope emerges.

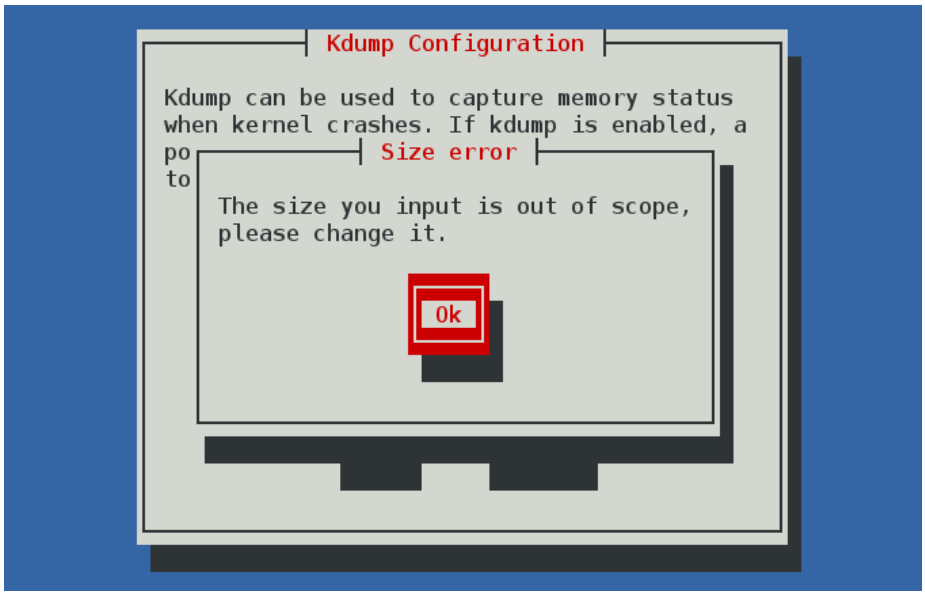


2.3.3 Memory size entry

When the memory reservation list box switches to 'Manual' mode, this entry must be inputted, or an error message box will pop up.



The value which user inputs needs to be inside the size scope, or an error message box will pop up.



2.4 Help screen

Once 'F3' on the keyboard is pressed, a help screen will pop up.

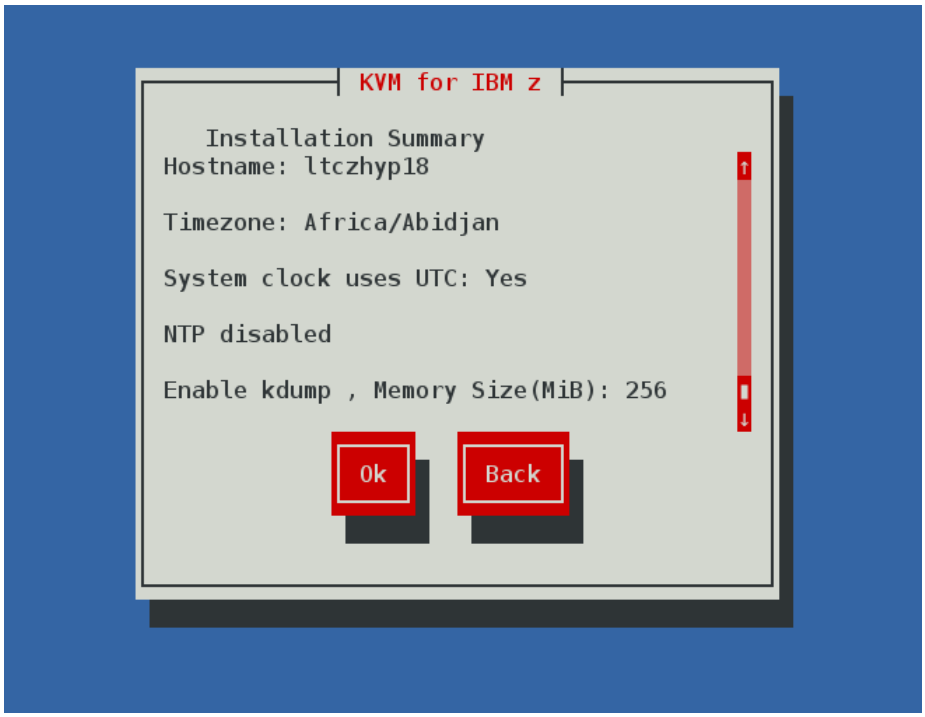
2.5 Deal with insufficient memory

At very rare circumstances, the system possesses less memory than minimal required size (384 MiB). But if that is the case, the configuration form will be tailored:

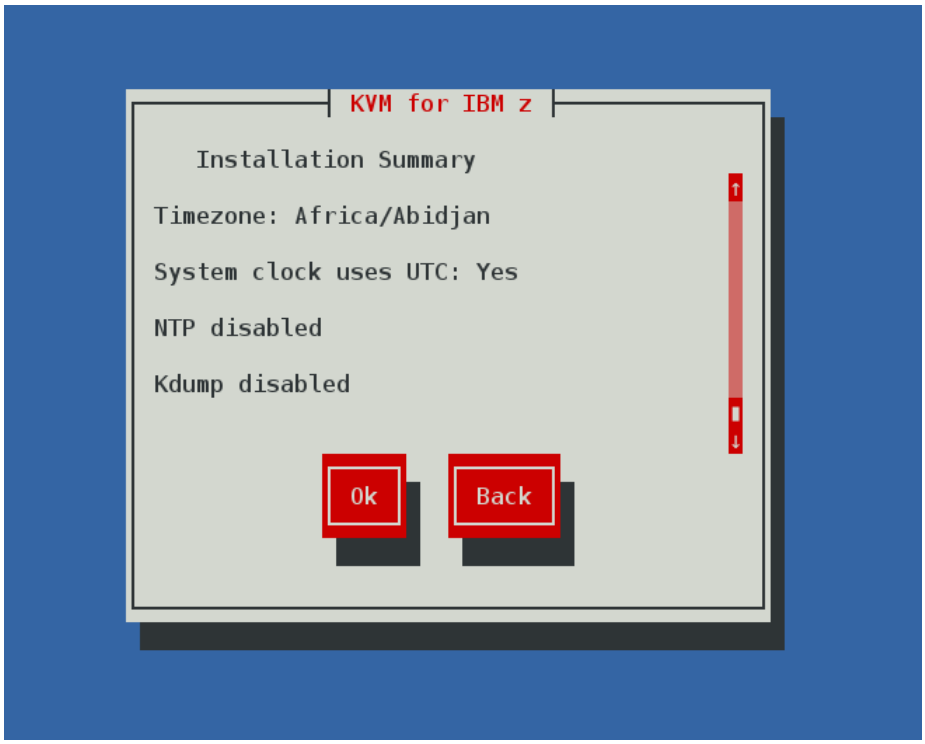
1. A piece of prompt text will be added below the introduction text.
2. The check box of 'Enable kdump' will be unselected, and will be unable to be selected.

2.6 Reflect in the summary form

1. If the kdump is enabled, the related part of summary form is as:



2. If the kdump is disabled, the related part of summary form is as:



3 Reserved memory size

It is necessary to give a proper memory size scope, and we set this range basing on:

A: The minimal memory size required for dumping memory when system crash. Referring to anaconda, it is 128 MiB.
B: The minimal memory size required by system for running kernel and services. Referring to anaconda, it is 256 MiB.
C: The total memory which the system possesses.

The proper size for kdump is expected to be no less than A and also no more than C minus B. So the scope of size of kdump should be [128, total memory - 256]. However if C is less than 384(A plus B, 128 + 256) MiB, kdump should not be enabled.

4 Capture the Dump

Normally kernel panic() will trigger booting into capture kernel but for testing purposes one can simulate the trigger in one of the following ways.

- Trigger through /proc interface

  echo c > /proc/sysrq-trigger

```
[root@ltczhyp18 ~]# echo c > /proc/sysrq-trigger Write failed: Broken pipe
[twu@localhost ~]$
```

- Trigger by inserting a module which calls panic()

The system will boot into the capture kernel. A kernel dump will be automatically saved in /var/crash/<dumpdir> and the system will boot back into the regular kernel. The name of the dump directory will depend on date and time of crash. For example, /var/crash/127.0.0.1-2015-07-29-09:31:38/vmcore. Right now we only support local disk (/var/crash) as dump target.

```
[twu@localhost ~]$ ssh root@9.152.166.41
root@9.152.166.41's password:
Last login: Wed Jul 29 09:25:54 2015 from 9.115.122.39
[root@ltczhyp18 ~]# ls /var/crash/
127.0.0.1-2015.07.29-09:31:38
[root@ltczhyp18 ~]# date
Wed Jul 29 09:32:28 GMT 2015
[root@ltczhyp18 ~]# ls /var/crash/127.0.0.1-2015.07.29-09\:31\:38/
vmcore  vmcore-dmesg.txt
[root@ltczhyp18 ~]#
```

5 Dump Analysis

Once the system has returned from recovering the crash, you may wish to analyse the kernel dump file using the crash tool.

- First, locate the recent vmcore dump file:

  find /var/crash -type f -mtime -1

- Once you have located a vmcore dump file, call crash:

crash /var/crash/127.0.0.1-2015-07-29-09:31:38/vmcore /usr/lib/debug/lib/modules/‘uname -r’/vmlinux

```
Copyright (C) 2002-2014  Red Hat, Inc.
Copyright (C) 2004, 2005, 2006, 2010  IBM Corporation
Copyright (C) 1999-2006  Hewlett-Packard Co
Copyright (C) 2005, 2006, 2011, 2012  Fujitsu Limited
Copyright (C) 2006, 2007  VA Linux Systems Japan K.K.
Copyright (C) 2005, 2011  NEC Corporation
Copyright (C) 1999, 2002, 2007  Silicon Graphics, Inc.
Copyright (C) 1999, 2000, 2001, 2002  Mission Critical Linux, Inc.
This program is free software, covered by the GNU General Public License,
and you are welcome to change it and/or distribute copies of it under
certain conditions.  Enter "help copying" to see the conditions.
This program has absolutely no warranty.  Enter "help warranty" for details.

GNU gdb (GDB) 7.6
Copyright (C) 2013 Free Software Foundation, Inc.
License GPLv3+: GNU GPL version 3 or later <http://gnu.org/licenses/gpl.html>
This is free software: you are free to change and redistribute it.
There is NO WARRANTY, to the extent permitted by law.  Type "show copying"
and "show warranty" for details.
This GDB was configured as "s390x-ibm-linux-gnu"...

      KERNEL: /usr/lib/debug/lib/modules/3.10.0-229.el7.ibm.11.s390x/vmlinux
    DUMPFILE: /var/crash/127.0.0.1-2015.07.29-09:31:38/vmcore  [PARTIAL DUMP]
        CPUS: 4
        DATE: Wed Jul 29 09:31:33 2015
      UPTIME: 00:07:45
LOAD AVERAGE: 0.01, 0.04, 0.05
       TASKS: 120
    NODENAME: ltczhyp18
     RELEASE: 3.10.0-229.el7.ibm.11.s390x
     VERSION: #1 SMP Mon May 18 11:29:14 CEST 2015
     MACHINE: s390x  (unknown Mhz)
      MEMORY: 3.6 GB
       PANIC: "Oops: 0004 [#1] SMP " (check log for details)
         PID: 11228
     COMMAND: "bash"
        TASK: dfa5bfa8  [THREAD_INFO: df448000]
         CPU: 3
       STATE: TASK_RUNNING (PANIC)

crash>
```