GPUMODESWITCH
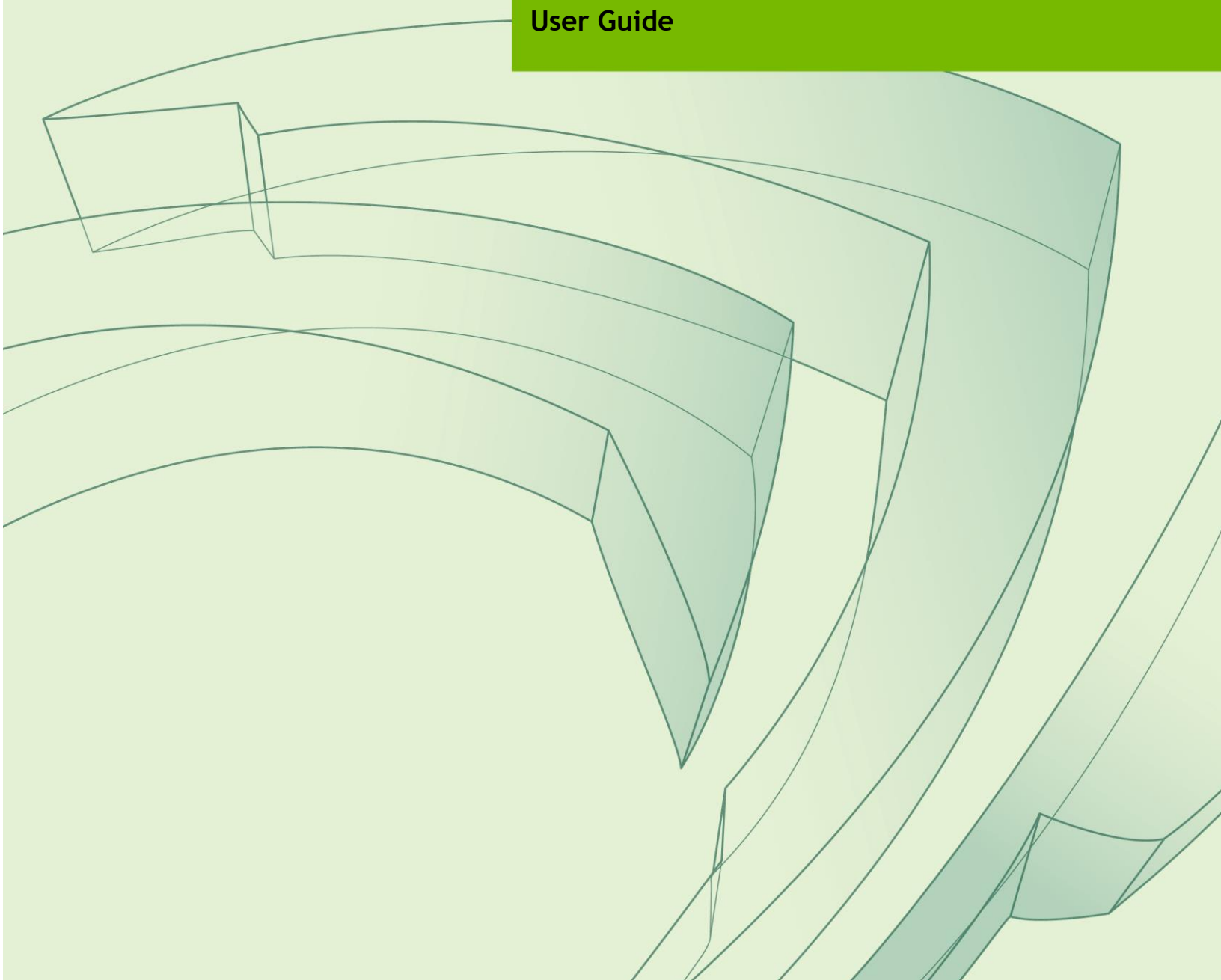
DU-07764-001 | September 2015

User Guide

DOCUMENT CHANGE HISTORY

DU-07764-001

Version	Date	Authors	Description of Change
1.0	8/30/2015	AC	Initial release

LIST OF FIGURES

LIST OF TABLES

Chapter 1.   INTRODUCTION

gpumodeswitch is a command-line tool that is used to switch supported NVIDIA GPUs between *compute* and *graphics* mode. This chapter describes these modes and when to use them. Chapter 2 describes how to use gpumodeswitch.

1.1   COMPUTE AND GRAPHICS MODE

NVIDIA Tesla GPUs are shipped in a configuration optimized for high-performance compute (HPC) applications:

Setting	Value	Notes
Classcode	3D Controller	This classcode indicates to operating systems (OS) that the GPU is not intended for use as a primary display device.
Memory BAR	8 gigabytes	Tesla GPUs expose a large memory *base address register* (BAR) for direct access to the framebuffer from the CPU, and other PCI Express devices.
I/O base BAR	Disabled	The GPU need not consume any legacy I/O resources when used as a non-display device.
ECC protection	Enabled	Error Correcting Code (ECC) is enabled on the GPU framebuffer to protect against single- and multi-bit memory errors.

Table 1 Compute mode settings

Collectively this GPU configuration is referred to as *compute mode*. While compute mode is optimal for HPC usage, it can cause compatibility problems with OS and hypervisors when the GPU is used primarily as a graphics device:

▶ Some OS require that the GPU advertise a VGA display controller classcode in order for the GPU to be used as a primary graphics device.

▶ Some hypervisors cannot support passthrough of GPUs with large memory BARs to guest virtual machines.

To address these problems, certain NVIDIA Tesla GPUs support setting the GPU into *graphics mode*:

Setting	Value	Notes
Classcode	VGA Controller	This classcode indicates to OS that the GPU can function as a primary display device.
Memory BAR	256 megabytes	The GPUs a smaller memory BAR for direct access to the framebuffer.
I/O base BAR	Enabled	The GPU exposes an I/O BAR to claim the resources require to operate as a VGA controller.
ECC protection	Disabled	ECC protection is disabled by default, though it can still be enabled by use of the *nvidia-smi* management tool

Table 2 Graphics mode settings

The mode of the GPU is established directly at power-on, from settings stored in the GPU's non-volatile memory. `gpumodeswitch` is used to program the mode of the GPU by updating the GPU's non-volatile memory settings.

1.2   WHEN TO USE GRAPHICS MODE

We recommend that graphics mode be used whenever supported Tesla products are used in the following scenarios:

▶ GPU passthrough with hypervisors that do not support large BARs. At the time of publication, this includes Citrix XenServer 6.2, 6.5, VMware ESXi 5.1, 5.5, 6.0, Red Hat Enterprise Linux 7.0, 7.1.
▶ GPU passthrough to Windows VMs on Xen and KVM hypervisors.
▶ GRID Virtual GPU deployments.
▶ VMware vSGA deployments.

> 💬 **Note:** consult individual NVIDIA product release notes for the latest information on compatibility with compute and graphics modes.

1.3   SUPPORTED PRODUCTS

At the time of writing, `gpumodeswitch` is supported on the following products:

▶ Tesla M60

▶ Tesla M6

Chapter 2. USING GPUMODESWITCH

gpumodeswitch is a command line utility that runs on Windows or Linux. This chapter describes how to to use gpumodeswitch, and the optional Linux boot packages included with gpumodeswitch.

2.1 SUPPORTED ENVIRONMENTS

gpumodeswitch can be run in the following environments

▶ Windows 64-bit command prompt
▶ Linux 32/64-bit shell (including Citrix XenServer dom0)

> 💬 **Note:** If the server platform hosting the Tesla GPUs does not natively run any of the supported environments, we recommend temporarily booting Linux on the server; the gpumodeswitch release package includes bootable Linux images for this purpose. These images are described in section 2.4.

2.2 WHAT'S IN THE PACKAGE

File	Description
gpumodeswitch.exe	Windows executable
nvflsh64.sys	Windows 64-bit driver module
gpumodeswitch	Linux executable, also usable on XenServer dom0
gpumodeswitch.iso	Bootable Linux ISO image
gpumodeswitch.zip	Bootable Linux image for use with USB storage (e.g. a flash key)

2.3   RUNNING GPUMODESWITCH

`gpumodeswitch` supports two operations: listing the current mode of GPUs in the platform, and changing the mode of individual or all GPUs.

When running on Linux, `gpumodeswitch` requires root permissions.

When running on Windows, `gpumodeswitch` requires administrator permissions (see Figure 2 on page 11), and the `nvflsh64.sys` driver should be in the same directory as the `gpumodeswitch` executable.

2.3.1 Listing current GPU modes

To list the current mode of all GPUs in the system, use `--listgpumodes`:

```
# gpumodeswitch --listgpumodes

NVIDIA GPU Mode Switch Utility Version 1.02
Copyright (C) 2015, NVIDIA Corporation. All Rights Reserved.

PLX (8747h)          (10B5,8747,10B5,8747) H:--:NRM S:00,B:81,PCI,D:00,F:00
Adapter: PLX (8747h)          (10B5,8747,10B5,8747) H:--:NRM S:00,B:81,PCI,D:00,F:00

Identifying EEPROM...
EEPROM ID (47,8747) : uChip 25AA320A 1.8-5.5V 4Kx8S, page
GPU Mode: N/A

Tesla M60          (10DE,13F2,10DE,113A) H:82:SP8 S:00,B:83,PCI,D:00,F:00
Adapter: Tesla M60          (10DE,13F2,10DE,113A) H:82:SP8 S:00,B:83,PCI,D:00,F:00

Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page
GPU Mode: Graphics

Tesla M60          (10DE,13F2,10DE,113A) H:82:SP16 S:00,B:84,PCI,D:00,F:00
Adapter: Tesla M60          (10DE,13F2,10DE,113A) H:82:SP16 S:00,B:84,PCI,D:00,F:00

Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page
GPU Mode: Compute
#
```

2.3.1.1   Log file output

The `--listgpumode` command also writes GPU mode information to a log file named `listgpumodes.txt`. By default `gpumodeswitch` creates the log file in the current

working directory (if the current working directory is writable). To change the path on which the log file is written, use the `--outfilepath` option:

```
# gpumodeswitch --listgpumodes --outfilepath /tmp/
```

> 💬 **Note:** when using the bundled Linux bootable images (see section 2.4), the system boots into a read-only file system. Log files can be generated by using the `outfilepath` option to write the log files into `/tmp`.

The log file summarizes GPU modes:

```
# more /tmp/listgpumodes.txt
GPU ID:
PLX (8747h)            (10B5,8747,10B5,8747) H:--:NRM S:00,B:81,PCI,D:00,F:00
GPU Mode: N/A
GPU ID:
Tesla M60              (10DE,13F2,10DE,113A) H:82:SP8 S:00,B:83,PCI,D:00,F:00
GPU Mode: Graphics
GPU ID:
Tesla M60              (10DE,13F2,10DE,113A) H:82:SP16 S:00,B:84,PCI,D:00,F:00
GPU Mode: Graphics
#
```

2.3.2 Switching GPU modes

To switch a GPU's mode, use the `--gpumode` command: `--gpumode graphics` switches to graphics mode, `--gpumode compute` switches to compute mode.

2.3.2.1   Switching all GPUs

By default, the command works on all supported GPUs in the platform; confirm the mode switch by entering 'y':

```
# gpumodeswitch --gpumode graphics

NVIDIA GPU Mode Switch Utility Version 1.02
Copyright (C) 2015, NVIDIA Corporation. All Rights Reserved.

Update GPU Mode of all adapters to "graphics"?
Press 'y' to confirm or 'n' to choose adapters or any other key to abort:
y
Updating GPU Mode of all eligible adapters to "graphics"

PLX (8747h)            (10B5,8747,10B5,8747) H:--:NRM S:00,B:81,PCI,D:00,F:00
Adapter: PLX (8747h)           (10B5,8747,10B5,8747) H:--:NRM S:00,B:81,PCI,D:00,F:00

Identifying EEPROM...
EEPROM ID (47,8747) : uChip 25AA320A 1.8-5.5V 4Kx8S, page
Cannot set GPU mode for this adapter
```

```
Tesla M60              (10DE,13F2,10DE,113A) H:82:SP8 S:00,B:83,PCI,D:00,F:00
Adapter: Tesla M60              (10DE,13F2,10DE,113A) H:82:SP8 S:00,B:83,PCI,D:00,F:00

Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page

Programming UPR setting for requested mode..
License image updated successfully.

Programming ECC setting for requested mode..
The display may go *BLANK* on and off for up to 10 seconds or more during the update
process depending on your display adapter and output device.

Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page
NOTE: Preserving straps from original image.
Clearing original firmware image...
Storing updated firmware image...
................
Verifying update...
Update successful.

Firmware image has been updated from version 84.04.7C.00.00 to 84.04.7C.00.00.

A reboot is required for the update to take effect.

InfoROM image updated successfully.

Tesla M60              (10DE,13F2,10DE,113A) H:82:SP16 S:00,B:84,PCI,D:00,F:00
Adapter: Tesla M60              (10DE,13F2,10DE,113A) H:82:SP16 S:00,B:84,PCI,D:00,F:00

Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page


Programming UPR setting for requested mode..
License image updated successfully.

Programming ECC setting for requested mode..
The display may go *BLANK* on and off for up to 10 seconds or more during the update
process depending on your display adapter and output device.

Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page
NOTE: Preserving straps from original image.
Clearing original firmware image...
Storing updated firmware image...
................
Verifying update...
Update successful.

Firmware image has been updated from version 84.04.7C.00.00 to 84.04.7C.00.00.

A reboot is required for the update to take effect.
```

```
InfoROM image updated successfully.
#
```

> 💬 **Note:** after a GPU mode switch, the server platform should be rebooted to ensure that the modified resources of the GPU are correctly accounted for by any OS or hypervisor running on the platform.

2.3.2.2   Switching individual GPUs

To switch the mode of an individual GPU, press 'n' when prompted, then enter the index of the GPU you wish to switch:

```
# gpumodeswitch --gpumode graphics

NVIDIA GPU Mode Switch Utility Version 1.02
Copyright (C) 2015, NVIDIA Corporation. All Rights Reserved.

Update GPU Mode of all adapters to "graphics"?
Press 'y' to confirm or 'n' to choose adapters or any other key to abort:
n
Select display adapter:
<0> PLX (8747h)          (10B5,8747,10B5,8747) H:--:NRM S:00,B:81,PCI,D:00,F:00
<1> Tesla M60            (10DE,13F2,10DE,113A) H:82:SP8 S:00,B:83,PCI,D:00,F:00
<2> Tesla M60            (10DE,13F2,10DE,113A) H:82:SP16 S:00,B:84,PCI,D:00,F:00
<3> PLX (8747h)          (10B5,8747,10B5,8747) H:--:NRM S:00,B:85,PCI,D:00,F:00
<4> PLX (8747h)          (10B5,8747,10B5,8747) H:86:SP8 S:00,B:87,PCI,D:00,F:00
<5> GRID K520            (10DE,118A,10DE,100D) H:88:SP8 S:00,B:89,PCI,D:00,F:00
<6> GRID K520            (10DE,118A,10DE,100D) H:88:SP16 S:00,B:8A,PCI,D:00,F:00
Select a number (ESC to quit): 1


Tesla M60          (10DE,13F2,10DE,113A) H:82:SP8 S:00,B:83,PCI,D:00,F:00
Adapter: Tesla M60           (10DE,13F2,10DE,113A) H:82:SP8 S:00,B:83,PCI,D:00,F:00


Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page


Programming UPR setting for requested mode..
License image updated successfully.


Programming ECC setting for requested mode..
The display may go *BLANK* on and off for up to 10 seconds or more during the update
process depending on your display adapter and output device.


Identifying EEPROM...
EEPROM ID (EF,3013) : WBond W25X40A 2.7-3.6V 4096Kx1S, page
NOTE: Preserving straps from original image.
Clearing original firmware image...
Storing updated firmware image...
```

```
.................
Verifying update...
Update successful.

Firmware image has been updated from version 84.04.7C.00.00 to 84.04.7C.00.00.

A reboot is required for the update to take effect.

InfoROM image updated successfully.
#
```

2.3.2.3   Log file output

The `--gpumode` command writes GPU mode update information to a log file named `setgpumode.txt`. By default `gpumodeswitch` creates the log file in the current working directory (if the current working directory is writable). To change the path on which the log file is written, use the `--outfilepath` option:

```
# gpumodeswitch --gpumode graphics --outfilepath /tmp
```

> **Note:** when using the bundled Linux bootable images (see section 2.4), the system boots into a read-only file system. Log files can be generated by using the `outfilepath` option to write the log files into `/tmp`.

The log file summarizes the command execution:

```
# more /tmp/setgpumode.txt
GPU ID:
PLX (8747h)          (10B5,8747,10B5,8747) H:--:NRM S:00,B:81,PCI,D:00,F:00
Cannot set GPU mode for this adapter
GPU ID:
Tesla M60            (10DE,13F2,10DE,113A) H:82:SP8 S:00,B:83,PCI,D:00,F:00
Successfully updated GPU mode to graphics.
GPU ID:
Tesla M60            (10DE,13F2,10DE,113A) H:82:SP16 S:00,B:84,PCI,D:00,F:00
Successfully updated GPU mode to graphics.
#
```

2.4   LINUX BOOTABLE IMAGES

If your server platform is not running one of the `gpumodeswitch` supported environments, use one the Linux bootable images included in the release to boot your server to a minimal Linux environment that includes `gpumodeswitch`.

2.4.1 Booting directly from ISO

The `gpumodeswitch.iso` file in the release package is intended for direct boot on a server platform, using the server's remote management capability. Connect the ISO file as an emulated storage device on the server, reboot the server, and use the BIOS boot menu to select the emulated device for boot.
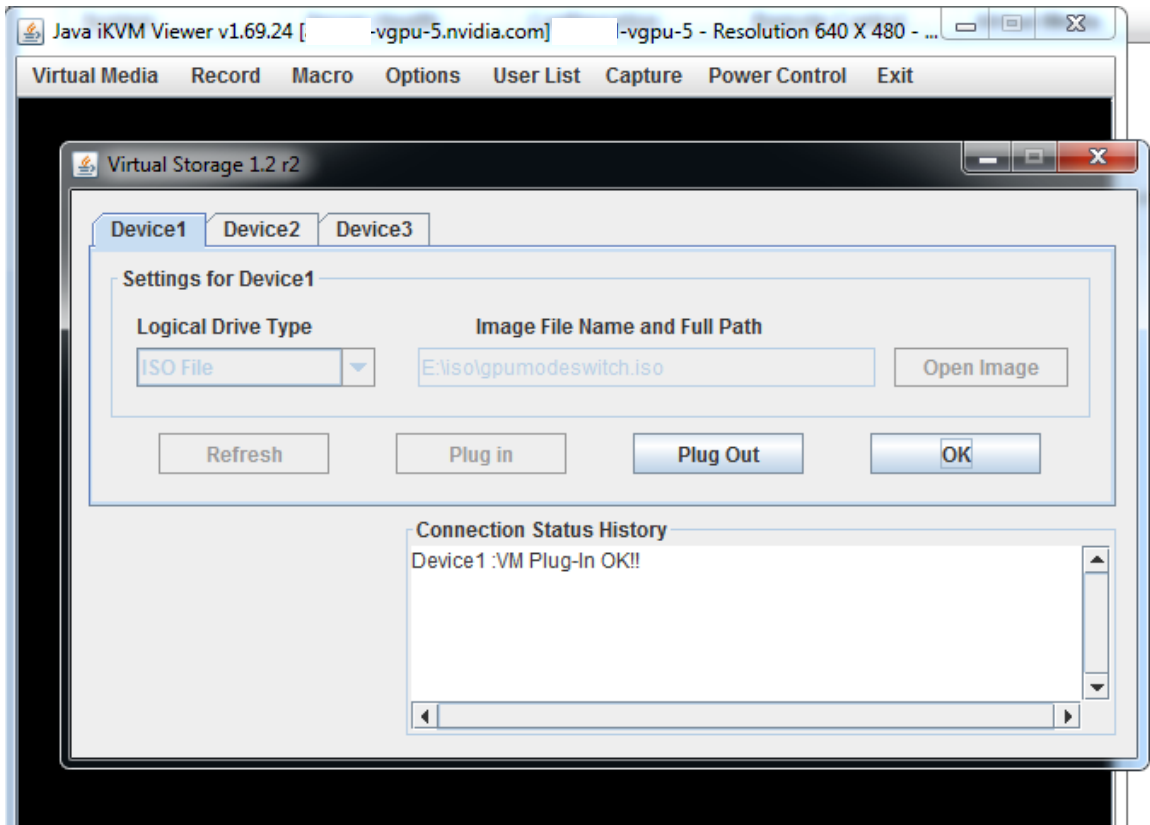

Figure 1 Connecting gpumodeswitch.iso via server remote management

The ISO image boots to a Linux shell prompt from which `gpumodeswitch` can be run directly. Refer to section 2.3 for instructions on running `gpumodeswitch` from the shell.

2.4.2 Booting from a USB flash key

The `gpumodeswitch.zip` file in the release package is intended to be unpacked onto a USB flash key, which can then be used to boot the server.

On a Windows system, follow these steps to create a bootable USB key:

▶ Connect a USB flash key of at least 64 megabytes in size.
▶ In Windows Explorer, right-click on the USB drive and select `Format` to format it. **Note the driver letter that is assigned to the USB drive** (for example, `E:`).

▶ Unzip `gpumodeswitch.zip` to the freshly formatted USB drive. Do this by right-clicking on the zip file, select "`Extract All…`", browse to the USB drive's letter, and then select `OK` to unzip the contents.

▶ Open a Windows command prompt with Administrator privileges. Do this by searching for '`cmd`', then right-click on the `cmd` program icon and selecting '`Run as administrator`':
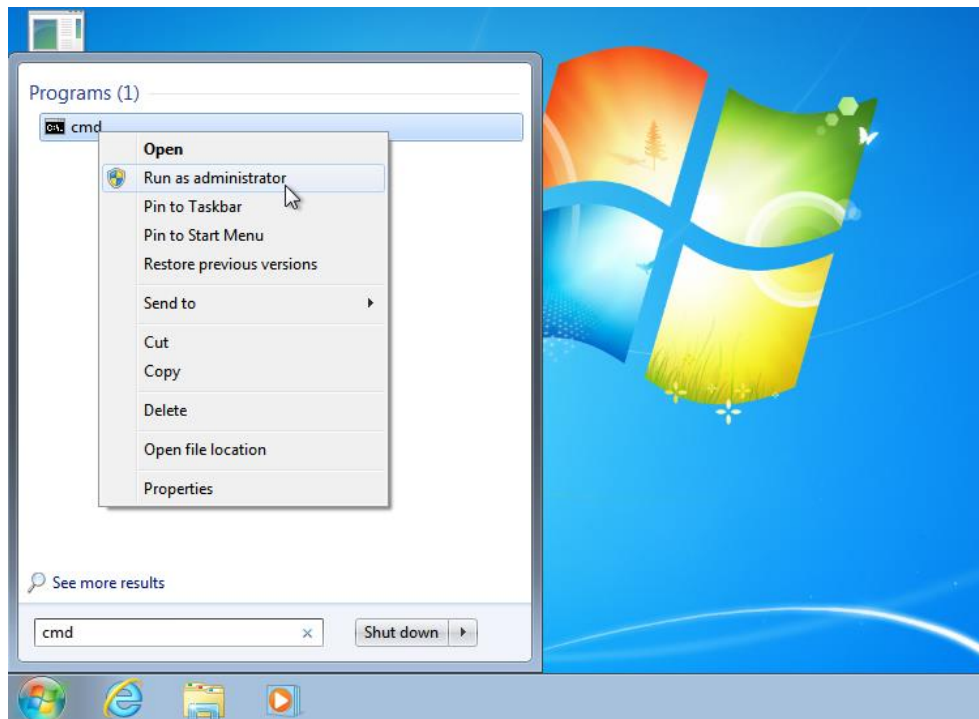

Figure 2 Launching an administrator command prompt

▶ In the command prompt, change to the USB drive by typing its drive letter and then hitting return; for example, '`e:`')

▶ Run the following command, substituting your USB drive's letter for '`e:`' –

```
syslinux –m –a e:
```

▶ Close the command prompt's window, eject / unplug the USB drive.

Boot the server from the USB drive. The image boots to a Linux shell prompt from which `gpumodeswitch` can be run directly. Refer to section 2.3 for instructions on running `gpumodeswitch` from the shell.

2.5 TROUBLESHOOTING

2.5.1 Unloading an existing NVIDIA driver

If `gpumodeswitch` is run in non-Windows environment where an existing NVIDIA driver is already loaded on the GPU, an error is reported:

```
[root@xenserver ~]# ./gpumodeswitch --listgpumodes

NVIDIA GPU Mode Switch Utility Version 1.02
Copyright (C) 2015, NVIDIA Corporation. All Rights Reserved.
...

ERROR: In order to avoid the irreparable damage to your graphics
adapter it is necessary to unload the NVIDIA kernel driver first:

            rmmod nvidia

[root@xenserver ~]#
```

2.5.1.1 Linux

On Linux, halt any services using the GPU, then use `rmmod` to unload the NVIDIA driver:

```
[root@linux ~]# rmmod nvidia
```

2.5.1.2 XenServer

On XenServer, you must halt any VMs that are using the GPU, and XenServer's `gpumon` service before you can run `gpumodeswitch`.

To halt `gpumon`, use 'service xcp-rrdd-gpumon stop':

```
[root@xenserver ~]# service xcp-rrdd-gpumon stop
Stopping XCP RRDD plugin xcp-rrdd-gpumon:                [  OK  ]
```

Then use 'rmmod' to unload the NVIDIA kernel driver:

```
[root@xenserver ~]# rmmod nvidia
```