# skelix.net skelixos tutorial06\_en.html

http://skelix.net/skelixos/tutorial06 en.html

- Home
- Skelix OS
- · Xinyi Boxing
- About Me
- •
- 中文

| <u>Prev</u> | Tutorial 06: Multi-Tasking | Next |
|-------------|----------------------------|------|

#### **Our Goal**

In this tutorial, there will be several tasks running in parallel in Skelix, and each task use its own LDT, we are going to implement preemptively task switching.

Download source code

# **TSS**

At first let's clarify one opinion, running multiple tasks concurrently on one single processor(one core) is not going to happen, in reality, the processor switches tasks very fast, say 100 times per second, so it looks like several processes are running at the same time.

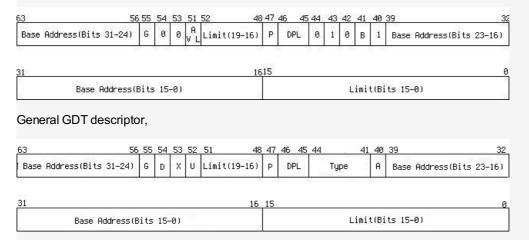
The 386 processor has hardware support for multitasking, but certainly it also can be accomplished manually, <a href="here is a program I wrote before">here is execution of several code snippets</a>. We all know there is only one set of registers in every single CPU, and they are used by all processes, so when a running processes is stopped and before it is switched to another process, all the informations about this process which are registers and some other information relate to them should be stored, because another process is going to use them. This information called "context" in general, and the 386 processor use the task state segments (TSS) to store this information, which is at least 104 bytes long, and a TSS descriptor refers to it. TSS descriptors only appear in the GDT, that means a user task can not switch to any specific process by itself via LDT. When we switch processes in the hardware way, the processor stores all informations in the TSS automatically.

The TSS defines the state of the execution environment for a task. Now let's take a look at the TSS structure,

| Not used   | Back link  |  |
|------------|------------|--|
| PLØ        | ESP        |  |
| Not used   | PLØ SS     |  |
| PL1        | ESP        |  |
| Not used   | PL1 SS     |  |
| PL2        | ESP        |  |
| Not used   | PL2 SS     |  |
| CI         | R3         |  |
| E          | IP         |  |
| EFL        | AGS        |  |
| E          | AX         |  |
| E          | CX         |  |
| E          | DX         |  |
| El         | BX         |  |
| E          | SP         |  |
| El         | BP         |  |
| E:         | SI         |  |
| El         | DI         |  |
| Not used   | ES         |  |
| Not used   | CS         |  |
| Not used   | SS         |  |
| Not used   | DS         |  |
| Not used   | FS         |  |
| Not used   | GS         |  |
| Not used   | LDT        |  |
| I/O bitmap | Not used T |  |

It is a big table, but it is not that scary if you look into it. For those "not used" fields, they are reserved by Intel. The 386 processor supports nested task, that means when task 1 switch to task 2, then the "Back link" field of task 2 is set to the selector of task 1 and the NT(Nested Task) bit in its EFLAGS is set as well, so when task 2 returns, the processor knows which task is going to take control. We know that the 386 processor supports 4 different privileges, and TSS allows each privilege uses different stacks when task switching, so there are 4 different stacks for one process can be used. We are going to talk about CR3 register in following tutorials. We are going to talk about LDT in a short while. About I/O bitmap an T bit, we don't care about them at this moment.

Like other descriptors in GDT, the TSS also need a descriptor in GDT refer to it, here is its format,



We can compare it to a general segment descriptor, we can see the D(data or code segment) and X(not used) bits are set to zero, the AVL bit is available to users. The type is 010B, the B(bit 41) is the busy bit of this TSS descriptor, because one process can only own one TSS, so once it is executing, the B bit of its descriptor is set to indicate its TSS can not be used any more just in case of the reentrant. A(access) is set to 1, but the TSS can neither be read nor written by processes even A bit is set to writable. Other field of the descriptor have the same meaning compares to normal data and code segments, just one thing we have to know, the limit field of the descriptor must can present a memory space that at least 104 bytes long which is the minimum length of TSS. There is another thing I have to mention, in reality most OS use software context switching because hardware context switching just save the registers we mentioned above, but not include FPU, MMX, SSE etc. Since we are not going to use them in Skelix, so it is safe to use TSS to do environment saving.

In this tutorial, DPL fields of TSS descriptors are going to be set to zero, so that only the kernel has the right to perform task switching. Just like GDTR and IDTR, TSS descriptors also has its own specific register for task switching, called TR, it can be loaded to this register by instruction LTR. And the selector refers to TSS descriptor in GDT can not be load into any other segment registers, or an exception will occur.

There are several ways to perform task switching, and we are going to use a far jump to a TSS descriptor in GDT to

achieve it. All those boring stuffs I mentioned above serve for this purpose. If you want to find out alternate ways to do it, you'd better read some code from wherever possible, because there are really only few books on this topic.

During task switching, the processor first stores the context of current task in the current TSS, then loads the task register of the new task, then loads the context of the new task from the new TSS with the new task's descriptor in GDT, finally executes the new task.

Let's take a look at the code, this is the TSS structure we are going to use,

### 06/include/task.h

```
struct TSS_STRUCT {
    unsigned int back_link;
    unsigned int esp0, ss0;
    unsigned int esp1, ss1;
    unsigned int esp2, ss2;
    unsigned int cr3;
    unsigned int eip;
    unsigned int eflags;
    unsigned int esp2, ebp;
    unsigned int esp, ebp;
    unsigned int esi, edi;
    unsigned int es, cs, ss, ds, fs, gs;
    unsigned int ldt;
    unsigned int trace_bitmap;
};
```

It is exactly the same as the TSS structure I mentioned above, but because this structure is going to be used by the processor, so it has to be 104 bytes long and no paddings in it, if you use IA-64 or other arches or other compilers, check documents youself.

For a working OS, those information are not enough, so anothing structure TASK STRUCT is used to wrap TSS structure.

```
#define TS RUNNING
                        0
#define TS RUNABLE
#define TS STOPPED
struct TASK_STRUCT {
        struct TSS STRUCT tss;
        unsigned long long tss entry;
        unsigned long long ldt[2];
        unsigned long long ldt_entry;
        int state;
        int priority;
        struct TASK STRUCT *next;
#define DEFAULT LDT CODE
                                 0x00cffa000000ffffULL
#define DEFAULT LDT DATA
                                 0x00cff2000000ffffULL
```

```
#define INITIAL PRIO
```

```
200
```

This is all we need at this moment, those  $\mathtt{TS}_*$  stuff defines all states that a process can be at. The  $\mathtt{tss}_\mathtt{entry}$  in  $\mathtt{TASK}_\mathtt{STRUCT}$  defines the descriptor of the tss field in this structure, I am going to explain it in a short while. Two \*ldt\* fields will be explained later on. state field store the current state of this process, which is one of those  $\mathtt{TS}_*$ . priority indicates the sequence of the execution of processes in system, and the new task will be given a initial priority INITIAL PRIO. All tasks in Skelix are managed as a link, the next field defines the next task in the link.

Now let's look at an example, this is the <code>TASK\_STRUCT</code> structure for task 0, that is the first task in system, when kernel finishes all initialization, it becomes task 0.

```
static unsigned long TASK0 STACK[256] = {0xf};
```

This is the stack used as the CPL0 stack for task 0. Whitout that 0xF, I encountered a problem during my compiling, if it is just initialized like static unsigned long TASKO\_STACK[256]; then this memory area just vanishes (Shen Feng explains, without 0xF, TASKO\_STACK[256] would be an zero-initialized static data, and it would be in .bss section, and the object file only record its length and start address etc.. That's why a nonzero value has to be used to keep it in .data section.)

Make esp0 points to the "bottom" of the stack. DATA\_SEL and CODE\_SEL appears in next few lines are defined at06/include/kernel.h, they are selectors of data and code segments in GDT.

```
/* esp1 ss1 esp2 ss2 */
0, 0, 0, 0,
/* cr3 */
Ο,
/* eip eflags */
0, 0,
/* eax ecx edx ebx */
0, 0, 0, 0,
/* esp ebp */
0, 0,
/* esi edi */
0, 0,
/* es
                CS
                               ds */
USER DATA SEL, USER CODE SEL, USER DATA SEL,
/* ss
                fs
                               gs */
USER DATA SEL, USER DATA SEL, USER DATA SEL,
/* ldt */
0x20,
/* trace bitmap */
0x00000000},
/* tss_entry */
```

```
0,
/* idt[2] */
{DEFAULT_LDT_CODE, DEFAULT_LDT_DATA},
/* idt_entry */
0,
/* state */
TS_RUNNING,
/* priority */
INITIAL_PRIO,
/* next */
0,
```

Now we have the TSS, we are going to create a TSS descriptor, remember there are two reserved place in GDT.

#### 06/bootsect.s

The forth entry(0x3) are reserved for TSS of current running task, so a macro <code>CURR\_TASK\_TSS = 3</code> has been defined as an index refers to this position in GDT. We are going to let the current task uses this place, once it gives up the control, it saves its descriptor in its <code>tss\_entry</code> of <code>TASK\_STRUCT</code>. When a new task takes over the control, it loads its own TSS descriptor from its<code>TASK\_STRUCT</code> to this place. In this way, we can allow unlimited tasks works in system. Because there is a length limit about GDT, it can only have 8096 descriptors, actually, Linux has the limit of tasks for quite long time, I don't know why it has this limit, because eliminating this limit seems to be quite simple.

```
unsigned long long
set_tss(unsigned long long tss) {
    unsigned long long tss_entry = 0x0080890000000067ULL;
    tss_entry |= ((tss)<<16) & 0xffffff0000ULL;
    tss_entry |= ((tss)<<32) & 0xff000000000000ULL;
    return gdt[CURR_TASK_TSS] = tss_entry;
}</pre>
```

set\_tss generates the TSS descriptor and put it into GDT, we can see it set the DPL of descriptor to 0, so only kernel can use this descriptor.

```
unsigned long long
get_tss(void) {
    return gdt[CURR_TASK_TSS];
}
```

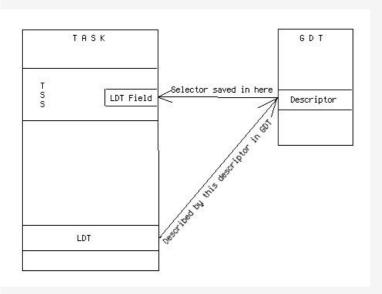
# LDT

It has been a while since we use GDT, LDT should be fairly easy at this point. LDT is just like GDT, but it is local, every task can have its own LDT, we are going to let every task in Skelix has two descriptors in LDT, the first one is for the code segment, and the second one is for data and stack segments. Now let's go over the selector format,



we are going to let all tasks work at privilege level 3, that means RPL=11b, and Tl=1 indicates this selector refers to a LDT. Unlike GDT, the first descriptor of LDT can be used by users, so the selector refers to code segment will be 0x7, and the selector refers to data and stack segments will be 0xF.

The LDT field in TSS is the selector which refers to an descriptor in GDT, this descriptor describes the LDT. I know it sounds confusing, so let's make it clear, first we let every task owns a LDT, and this LDT resides in somewhere in memory(we do not concern about virtual memory at this stage), and we need a descriptor to describe this memory area which contains this LDT, and this descriptor will be put in GDT, and the selector refers to this descriptor will be put into the LDT field in TSS of this task.



The third descriptor of GDT is for current running task's TSS, and the forth descriptor is quite clear, it is for current task's LDT. So theirs selector will be 0x18 and 0x20 respectively. Like functions relate to TSS, we also have,

# 06/task.c

```
unsigned long long long long ldt) {
    unsigned long long ldt_entry = 0x008082000000000fULL;

    ldt_entry |= ((ldt) <<16) & 0xffffff0000ULL;

    ldt_entry |= ((ldt) <<32) & 0xff00000000000ULL;

    return gdt[CURR_TASK_LDT] = ldt_entry;
}</pre>
```

It is similar to GDT entries except for the DPL is 3 instead 0.

```
unsigned long
get_tss(void) {
    return gdt[CURR_TASK_TSS];
}
```

At this moment we will let all tasks have the same LDT content, it means they share the same memory space, that is not a good idea, but I will give them separate memory space by using virtual memory, it will be introduced in later tutorial.

## 06/include/task.h

```
#define DEFAULT_LDT_CODE 0x00cffa000000ffffULL

#define DEFAULT_LDT_DATA 0x00cff2000000ffffULL
```

These are the LDT descriptors in tasks' LDT, they are almost the same as in GDT, except for the DPL field is set to 3.

I mentioned above, all tasks are linked as a single direction link structure, there are two important pointers, one is the <code>nextfield inTASKO</code>, it is the head of the whole link, and a <code>current</code> pointer is used for pointing to the current running task.

```
Creating and Scheduling
```

Before any tasks start, we define,

```
06/task.c
```

```
struct TASK STRUCT *current = &TASK0;
```

Then we take a look at how a new task is generated,

```
06/init.c
```

new\_task takes four arguments, the first one is the TASK\_STRUCT of new task, the second one is the start entry of the program, the eip field in TSS will be assigned to this value, so the processor can know where to start this program, the esp0 and esp3arguments are for the ESP pointer in privilege level 0 and 3, because their selectors have fixed value, 0x10 an 0xf, ss0 and ss are not needed.

```
memcpy(task, &TASKO, sizeof(struct TASK_STRUCT));

task->tss.esp0 = stack0;

task->tss.eip = eip;

task->tss.eflags = 0x3202;

task->tss.esp = stack3;

task->priority = INITIAL_PRIO;

task->state = TS_STOPPED;
```

TASKO are used as a template, we just modify some fields to suits our needs. We change new task state to  $\mathtt{TS\_STOPPED}$  before the task link is modified correctly. Otherwise, the link can be broken if another  $\mathtt{new\_task}$  is invoked at the same time.

```
task->next = current->next;
current->next = task;
task->state = TS_RUNABLE;
}
```

Puts new task in task link.

```
extern void task1_run(void);
extern void task2_run(void);
```

They are the entries of new tasks, will be introduced in a short while.

```
static long task1_stack0[1024] = {0xf, };
static long task1_stack3[1024] = {0xf, };
static long task2_stack0[1024] = {0xf, };
static long task2_stack3[1024] = {0xf, };
```

Because we didn't have the memory management done in this tutorial, so we just set up several arrays as task stacks.

```
void
init(void) {
```

Like you might guess, there are two new registers for TSS and LDT like GDT and IDT, called TR and LDTR, they are loaded by ltrand lldt. Before we trying to do any task manipulation, we have to load valid values to TR and LDTR, or you will get an general protection exception.

Creates two new tasks, please note that, before creating any new tasks, the interrupt had been enabled.

Now the kernel has became task TASK0, by a long return it loaded correct EIP and CS and SS and ESP from TASK0 structure, the stack looks like,

Now all tasks are running at privilege level 3, so all other parts that have privilege level higher that that can not be accessed, it gives you some idea of protection. But in Skelix, I am not going to separate the kernel space from the whole memory, the memory protection will be achieved by the memory mapping of virtual memory. In later tutorial, the user tasks can invoke system functions via system calls.

Now we have three tasks, then which part will do the switching? Well, as you may guess, we have the timer interrupt in a certain interval, so it is the good and clearly choice. We have to change the code in timer interrupt like this,

# 06/timer.c

Traversing the task link, to change priorities of all tasks.

if (v->state == TS RUNNING) {

if ((v->priority+=30) <= 0)

v->priority = 0xffffffff;

```
void do_timer(void) {
    struct TASK_STRUCT *v = &TASK0;
    int x, y;
    ++timer_ticks;
    get_cursor(&x, &y);
    set_cursor(71, 24);
    kprintf(KPL_DUMP, "%x", timer_ticks);
    set_cursor(x, y);
    outb(0x20, 0x20);
    cli();
    for (; v; v=v->next) {
```

```
} else
v->priority -= 10;
}
```

Well, the priority of all tasks will be changed during timer interrupt, the running task get higher numerical priority, that is the lower priority. And the waiting tasks get higher priority.

```
if (! (timer_ticks%1))
    scheduler();
```

We re-schedule all tasks in two timer ticking.

```
sti();
}
```

Let's check out the scheduler,

```
void scheduler(void) {
    struct TASK_STRUCT *v = &TASK0, *tmp = 0;
    int cp = current->priority;

for (; v; v = v->next) {
        if ((v->state==TS_RUNABLE) && (cp>v->priority)) {
            tmp = v;
            cp = v->priority;
        }
}
```

Traversing the task link, to find the task which has the highest priority to run, that is the smallest number.

```
if (tmp && (tmp!=current)) {
    current->tss_entry = get_tss();
    current->ldt_entry = get_ldt();
```

We save the TSS and LDT descriptors from GDT to TASK\_STRUCT, because some state bits in those descriptors might have been modified by the processor.

```
tmp->tss_entry = set_tss((unsigned long long)((unsigned int)&tmp-
>tss));

tmp->ldt_entry = set_ldt((unsigned long long)((unsigned int)&tmp-
>ldt));

current->state = TS_RUNABLE;

tmp->state = TS_RUNNING;

current = tmp;
```

Sets TSS and LDT descriptors of the task which is going to run to the GDT, and changes theirs state fields.

Skelix uses a far jump instruction to perform the task switching, the segment part is the TSS selector, and the offset part is simply ignored by the processor. The TSS\_SEL\_STR is defined as "\$0x18", the quotation marks are included, so C can take all those string fragments into one string.

At last, let's check out the code of task1\_run and task2\_run, their entries are written in assembly instead C, because I do not want to handle the stack change in C function calls. These two tasks actually call other C functions to do the real job.

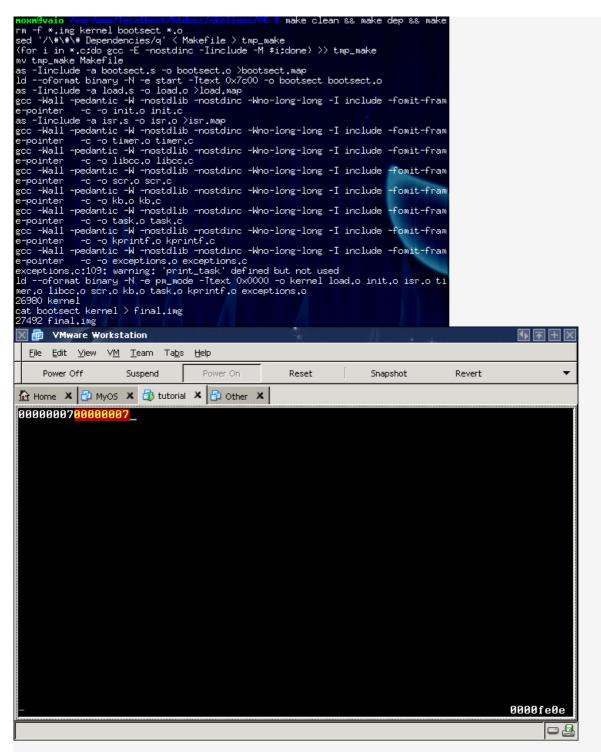
At first, let's check if tasks work in privilege 3 properly, so we call kprintf to print the current CS that task is using, then let it go to busy idle or it will keep printing it on the screen, and screen scrolls up too fast to see any result.

# 06/init.c

Don't forget to add new modules to KERNEL OBJS in Makefile.

# 06/Makefile

KERNEL\_OBJS= load.o init.o isr.o timer.o libcc.o scr.o kb.o task.o kprintf.o
exceptions.o



As we can see, first we are using LDT, and the CS has been set to 0x7 correctly, another thing is as we make the kprintf buffer global, so actually all tasks are sharing same buffer, so output of two tasks might be mixed together when another task starts printting before former task finishes.

Now let them do something fancier, we let two tasks keep displaying different letters,

# 06/init.c

```
void
do_task1(void) {
    print_c('A', BLUE, WHITE);
}
void
do_task2(void) {
    print_c('B', GRAY, BROWN);
```

```
f *.img kernel bootsect *.o
'/\#\#\ Dependencies/q' < Makefile > tmp_make
i in *.c;do gcc -E -nostdinc -Iinclude -M $i;done) >> tmp_make
sed /\*\*\* Dependencies/q < Makerile > tmp_make
(for i in *.c;do gcc -E -nostdinc -Iinclude -M $i;done) >> tmp_make
mv tmp_make Makefile
as -Iinclude -a bootsect,s -o bootsect.o >bootsect.map
ld --oformat binary -N -e start -Itext 0x7c00 -o bootsect bootsect.o
as -Iinclude -a load.s -o load.o >load.map
gcc -Wall -pedantic -W -nostdlib -nostdinc -Wno-long-long -I include -fomit-fram
e-pointer  -c -o init.o init.c
as -Iinclude -a isr.s -o isr.o >isr.map
gcc -Wall -pedantic -W -nostdlib -nostdinc -Wno-long-long -I include -fomit-fram
e-pointer -c -o timer,o timer,o
gcc -Wall -pedantic -W -nostdlib -nostdinc -Wno-long-long -I include -fomit-fram
e-pointer -c -o libcc.o libcc.c
    -Wall -pedantic -W -nostdlib -nostdinc -Wno-long-long -I include -fomit-fram
gcc -Wall -pedantic -W -nostdlib -nostdinc -Wno-long-long -I include -Fomit-fram
e-pointer -c -o scr.o scr.c
gcc -Wall -pedantic -W -nostdlib -nostdinc -Wno-long-long -I include -fomit-fram
e-pointer -c -o kb.o kb.c
gcc -Wall -pedantic -W -nostdlib -nostdinc -Wno-long-long -I include -fomit-fram
e-pointer -c -o task.c
gcc -Wall -pedantic -W -nostdlib -nostdinc -Wno-long-long -I include -fomit-fram
e-pointer -c -o kprintf.c kprintf.c
gcc -Wall -pedantic -W -nostdlib -nostdinc -Wno-long-long -I include -fomit-fram
e-cointer -c -o exceptions.o exceptions.c
e-pointer -c -o exceptions.o exceptions.o -Wno-long-long -I include -fomit-fram
e-pointer -c -o exceptions.o exceptions.o to exceptions.o:109; warning; 'print_task' defined but not used
ld --oformat binary -N -e pm_mode -Ttext 0x0000 -o kernel load.o init.o isr.o ti
mer.o liboc.o scr.o kb.o task.o kprintf.o exceptions.o
26948 kernel
cat bootsect keys :
cat bootsect kernel > final.img
27460 final.img
 🗙 📵 VMware Workstation
  <u>File Edit View VM Team Tabs</u>
    Power Off
                  Suspend
                                              Reset
                                                            Snapshot
                                                                          Revert
                               Power On
 🚹 Home 🗶 🗐 MyOS 🗶 👘 tutorial 🗶 🚰 Other 🗶
                                                  ВАААААААА
                                                                                    000097f1
                                                                                    000097f1
  000097f1
  00009c26
           BRERRRERRRERRERREP<mark>AAAA</mark>ARI
 00009cdd
                                                                                    00009cdd
 BAAAAAAAAAAAAAAAAB
                                                                                    00009e45
 00009f1a
 ааааааааааааааааа
                                                                                    0000a0e2
 0000a1e4
  0000a40f
 0000a533
   Prev
              Tutorial 06: Multi-Tasking
                                                   Next
```