درس آز سیستمعامل

استاد حمید بیگی



على رضا ديزجي، اميرحسن فتحي، مصطفى قديمي

گزارش آزمایش دوم

سؤال ۱. مشاهده فراخوانیهای سیستمی تعریف شده

در این قسمت با توجه به این که ساختار سخت افزار ما با سوال فرق دارد، فایل asm را به کمک دستور find پیدا کردیم:

- cd /usr/include/
- > find . -name asm
- ./x86_64-linux-gnu/asm
- > cd ./x86_64-linux-gnu/asm/
- > cat unistd.h



سؤال ۲. اجرای یک فراخوانی سیستمی

• دستورات زیر را درون terminal اجرا میکنیم.

- > mkdir 2 && cd 2
- > sudo nano testsyscall.cpp
- کد داده شده در سوال را در این قسمت کپی کرده و سپس با استفاده از دستور x + trl + x آن را ذخیره میکنیم.
 - به کمک دستورات زیر کد را کامپایل کرده و اجرا میکنیم.
- > sudo gcc -o testsyscall testsyscall.cpp
- > ./testsyscall

- نتیجهی اجرای آن ساخته شدن یک پوشه به نام testdir در مسیر ۱ فعلی است و در آخر پیام .0 The result is را چاپ میکند.
- همانطور که در توضیحات فراخوانی سیستمی آمده است، هر فراخوانی سیستم با یک شمارهی ثابت شناخته می شود. نقش NR_mkdir_ (که به طور سراسری ۲ تعریف شده است) در این جا این است تا عدد مربوط به این فراخوانی جایگرین آن شود.

```
mostafa@mostafa-UX303UB: /usr/include/x86_64-linux-gnu/asm

Eile Edit View Search Ierminal Help
#define __NR_semop 65
#define __NR_semctl 66
#define __NR_semctl 67
#define __NR_msgget 68
#define __NR_msgget 69
#define __NR_msgrcv 70
#define __NR_msgrcv 71
#define __NR_fcntl 72
#define __NR_fflock 73
#define __NR_fsync 74
#define __NR_fsync 75
#define __NR_ftuncate 76
#define __NR_ftruncate 77
#define __NR_getdents 78
#define __NR_getdents 78
#define __NR_getchd 79
#define __NR_fchdir 80
#define __NR_fchdir 81
#define __NR_fchdir 83
#define __NR_rename 82
#define __NR_rename 82
#define __NR_rename 85
#define __NR_creat 85
#define __NR_creat 85
#define __NR_Link 86
#define __NR_Link 86
#define __NR_Symlink 88
```

• تابع ()syscall یک تابع کوچک است که رابط زبان assembly آن دارای شماره و آرگومانهای مشخص است. تابع پر کاربردی است اگر از آن بدون wrapperها استفاده کنیم، رجسیترهای پردازنده قبل از فراخوانی سیستم ذخیره میکند و بعدا آنها را بازگردانی میکند. خروجی آن در صورتی که با موفقیت انجام گیرد، « • » و در غیر این صورت « ۱ – » خواهد بود.

سؤال ٣. اجراي سادهتر فراخوانيهاي سيستمي

```
#include <stdio.h>
#include <unistd.h>
#include <sys/stat.h>
int main() {
    long result;
    result = mkdir("testdir", 0777);
    printf("The result is %ld.\n", result);
    return 0;
}
```

سؤال ۴. آشنایی با چند فراخوان سیستمی پرکاربرد

• فراخوان سیستمی access

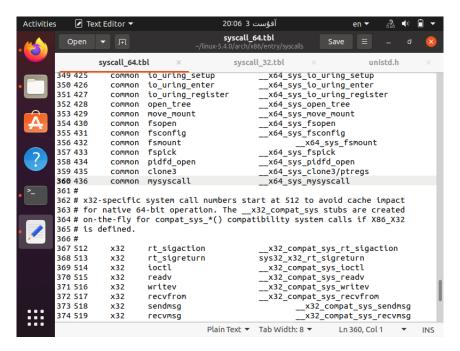
directory\ global\

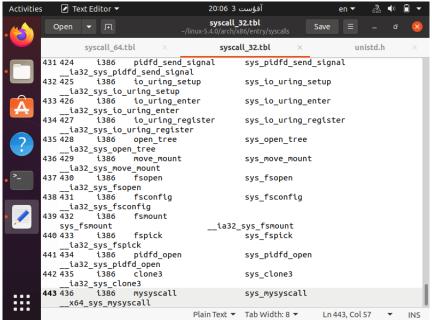
```
#include <errno.h>
#include <stdio.h>
#include <unistd.h>
int main(int argc, char *argv[])
{
    int result;
    char *path = argv[1];
    result = access(path, F_OK);
    if (result == 0)
    printf("%s exists\n", path);
    }
    else
    {
    printf("%s doesn't exist\n", path);
    result = access(path, R_OK);
    if (result == 0) {
    printf("read permission is granted\n");
    }
    else {
    printf("read permission isn't granted\n");
    result = access(path, W_OK);
    if (result == 0) {
    printf("write permission is granted\n");
    }
    else {
    printf("write permission isn't granted\n");
    result = access(path, X_OK);
    if (result == 0) {
    printf("execute permission is granted\n");
    }
    else {
    printf("execute permission isn't granted\n");
    }
    return 0;
}
                                                      • فراخوانهای سیستمی open, close, write
#include <sys/types.h>
#include <sys/stat.h>
#include <fcntl.h>
#include <unistd.h>
#include <string.h>
int main() {
    int open_result;
    int write_result;
    open_result = open("oslab2.txt", O_CREAT | O_WRONLY, 0777);
```

```
write_result = write(open_result, "Mostafa Ghadimi\n", strlen("Mostafa Ghadimi\n"));
    close(open_result);
    return 0;
}
                                                                     • فراخوان سیستمی sysinfo
#include <linux/kernel.h>
#include <stdio.h>
#include <sys/sysinfo.h>
int main()
{
    const double megabyte = 1024 * 1024;
    struct sysinfo si;
    sysinfo(&si);
    \label{lem:lem:mass}  \mbox{printf("total RAM: 1f.\%5 MB\n", si.totalram / megabyte);} 
    printf("free RAM: 1f.%5 MB\n", si.freeram / megabyte);
    return 0;
}
                                                                   • فراخوان سيستمي getrusage
#include <sys/time.h>
#include <sys/resource.h>
#include <stdio.h>
int main() {
    struct rusage ru;
    getrusage(RUSAGE_SELF, &ru);
    printf("maximum resident set size: %ld\n", ru.ru_maxrss);
    printf("integral shared memory size: %ld\n", ru.ru_ixrss);
    printf("integral unshared stack size: %ld\n", ru.ru_isrss);
}
```

سؤال ۵. اضافه کردن یک فراخوانی سیستمی به سیستمعامل

- با توجه به لینکی که در کلاس داده شد، میخواهیم یک فراخوانی سیستمی اضافه کنیم که تعداد page fault را نشان دهد.
- ۱. در ابتدا به ۲ فایل arch/x86/entry/syscalls/syscall_64.tbl و arch/x86/entry/syscall_64.tbl . ۱ فراخوانی سیستمی دلخواه با شماره ۴۳۶ را اضافه میکنیم.
 - rmicro به فایل include/uapi/asm-generic/unistd.h آن را اضافه می کنیم.
- ۳. حال می خواهیم کد خود را در kernel اضافه کنیم. در فایل include/linux/mm.h، در انتهای فایل یک متغیر سراسری به نام pfcount که نشاندهنده تعداد page fault متغیر سراسری به نام میکنیم.
- ۴. در فایل include/linux/sched.h، به task_struct (که یک abstraction برای پردازههای لینوکس می باشد) یک شمارنده page fault اضافه میکنیم.





- در فایل kernel/fork.c، کد kernel/fork.c، کد در فایل tsk-pf=0
- ۶. حال برای اینکه هر بار page fault رخ داد، متغیر خود را افزایش دهیم، در فایل page fault رخ داد، متغیر خود را خطوط ۱۵۱۵، ۱۵۱۵ و ۱۵۲۲ را اضافه میکنیم.
 - ۷. برای پیاده سازی تابع systemcall، ابتدا کد زیر را به فایل kernel/sys.c اضافه می کنیم.
 - ۸. تعریف تابع systemcall موردنظر را به فایل include/linux/syscalls.h اضافه می کنیم.
 - ۹. در نهایت برای compatibility کد زیر را به فایل kernel/sys_ni.c اضافه میکنیم.
- ۱۰. حال با استفاده از دستور nohup make -j12 & مرنل را کامپایل کرده و با استفاده از دستور nohup.out خروجی آنرا ملاحظه میکنیم.
- occ البینیم که با اجرای فایل test.c میخواهیم خروجی آن را ببینیم که با اجرای فایل خروجی test حاصل از ۱۱. با استفاده از test.c -o test آن را بدست می آوریم.

```
✓ Text Editor ▼
                                                               unistd.h
   Open ▼ 🗊
                                                                                                Save ≡
                                                                                                                unistd.h
               syscall 64.tbl
                                                               syscall 32.tbl
844 __SYSCALL(__NR_TSMOUNT, Sys_TSMOUNT)
845 #define __NR_fspick 433
846 __SYSCALL(__NR_fspick, sys_fspick)
847 #define __NR_pidfd_open 434
848 __SYSCALL(__NR_pidfd_open, sys_pidfd_open)
849 #lfdef ARCH_WANT_SYS_CLONE3
850 #define NR_clone3 435
851 SYSCALL(NR_clone3, sys_clone3)
 852 #endif
853 #define __NR_mysyscall 436
854 __SYSCALL __NR_mysyscall, sys_mysyscall
 856
857 #undef __NR_syscalls
858 #define __NR_syscalls 437
860 /*
860 * 32 bit systems traditionally used different
862 * syscalls for off_t and loff_t arguments, while
863 * 64 bit systems only need the off_t version.
864 * For new 32 bit platforms, there is no need to
865 * implement the old 32 bit off_t syscalls, so
866 * they take different names.
      * Here we map the numbers so that both versions
* use the same syscall table layout.
 867
869
                                               C/ObjC Header ▼ Tab Width: 8 ▼
                                                                                                    Ln 854, Col 40
     ✓ Text Editor ▼
                                                           .
آقۇست 3 11:20
                                                                 mm.h
                                                                                                         | | ≡ |
   syscall_64.tbl
                                                 syscall_32.tbl
                                                                                        unistd.h
                                                                                                                        mm.h
2902
                                  return false;
2903
2904
                     return PageGuard(page);
2905 }
2900 #etse
2907 static inline unsigned int debug_guardpage_minorder(void) { return 0; }
2908 static inline bool debug_guardpage_enabled(void) { return false; }
2909 static inline bool page_is_guard(struct page *page) { return false; }
2910 #endif /* CONFIG_DEBUG_PAGEALLOC */
2911
2912 #if MAX_NUMNODES > 1
2913 void __init setup_nr_node_ids(void);
2914 #else
2915 static inline void setup_nr_node_ids(void) {}
2916 #endif
2918 extern int memcmp_pages(struct page *page1, struct page *page2);
2920 static inline int pages_identical(struct page *page1, struct page *page2)
2921 {
2922
                     return !memcmp_pages(page1, page2);
2923 }
2924 extern unsigned long pfcount;
2925
2926 #endif /* __KERNEL__ */
2927 #endif /* _LINUX_MM_H */
```

۱۲. اگر خروجی را با dmesg چک کنیم مشاهده میکنیم که به ترتیب مجموع تعداد page fault ها و تعداد .۱۲ ها و تعداد page fault

C/ObjC Header ▼ Tab Width: 8 ▼ Ln 2925, Col 1 ▼ INS

```
✓ Text Editor ▼
                                                   .
آقۇست 3 20:15
                                                      sched.h
          Open ▼ 🗐
                                                                              Save ≡
             syscall_64.tbl
                                    syscall_32.tbl
                                                             unistd.h
                                                                                               sched.h
        1261 #ifdef CONFIG SECURITY
        1262
                       /* Used by LSM modules for access restriction: */
                       void
                                                              *security:
        1263
        1264 #endif
        1265
        1266 #ifdef CONFIG_GCC_PLUGIN_STACKLEAK
1267 unsigned long
                                                              lowest stack;
                       unsigned long
        1268
                                                             prev_lowest_stack;
        1269 #endif
                       unsigned long pf;
        1270
                      /*

* New fields for task_struct should be added above here, so that

* they are included in the randomized portion of task_struct.
        1271
        1272
        1273
        1274
                       randomized_struct_fields_end
        1276
        1277
                       /* CPU-specific state of this task: */
                       struct thread_struct
        1278
                                                              thread:
        1279
                       /*
    * WARNING: on x86, 'thread_struct' contains a variable-sized
    * structure. It *MUST* be at the end of 'task_struct'.
        1280
        1281
        1282
        1283
        1284
                        * Do not put anything below here!
:::
        1285
        1206 1 .
         Saving file "/home/alireza/linux... C/ObjC Header ▼ Tab Width: 8 ▼ Ln 1270, Col 26 ▼ INS
```

```
✓ Text Editor ▼
                                                    fork.c
          Open ▼ 🗐
                                                                           Save ≡
          syscall_64.tbl
                             syscall_32.tbl
                                                  unistd.h
                                                                 mm.h
                                                                              sched.h
                                                                                             fork.c
         854 {
                      unsigned long *stackend;
        855
                      stackend = end_of_stack(tsk);
*stackend = STACK_END_MAGIC; /* for overflow detection */
        857
        858
        859 }
        861 static struct task_struct *dup_task_struct(struct task_struct *orig, int
            node)
        862 {
                      struct task_struct *tsk;
        863
                      unsigned long *stack;
struct vm_struct *stack_vm_area __maybe_unused;
         864
        865
                      int err;
        867
        868
                      if (node == NUMA_NO_NODE)
        869
                      node = tsk_fork_get_node(orig);
tsk = alloc_task_struct_node(node);
        870
        871
                      if (!tsk)
                      return NULL;

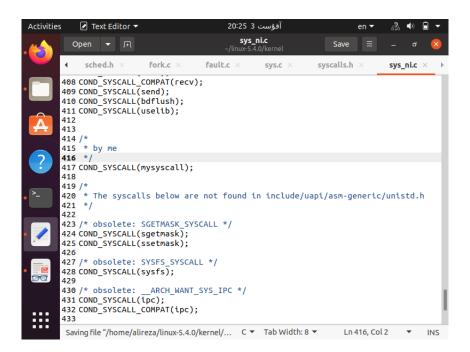
tsk->pf = 0;

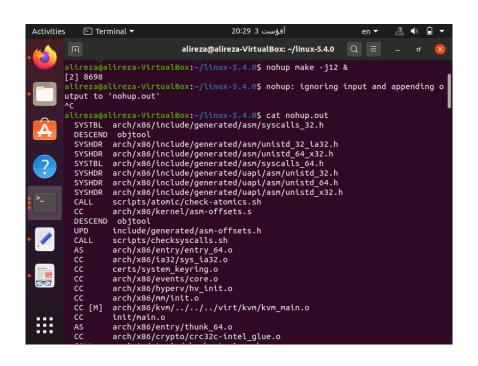
stack = alloc_thread_stack_node(tsk, node);
        872
        873
        874
        875
                      if (!stack)
                               goto free_tsk;
        876
:::
        878
                      if (memcg_charge_kernel_stack(tsk))
        Saving file "/home/alireza/linux-5.4.0/kernel/... C ▼ Tab Width: 8 ▼ Ln 873, Col 21 ▼ INS
```

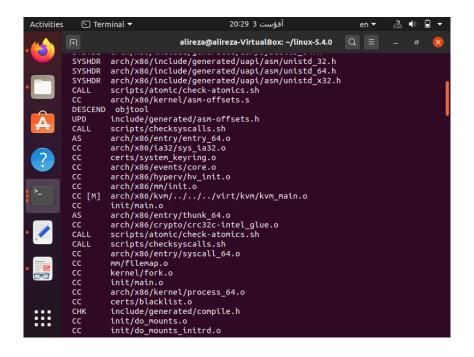
```
Activities
           ✓ Text Editor ▼
                                               .
آقۇست 3 21:20
                                                  fault.c
                                                                        Save ≡
          Open ▼ 🗐
                          *AZ_OS2.txt
                                                                           fault.c
                              perf_sw_event(PERF_COUNT_SW_PAGE_FAULTS_MIN, 1, regs,
        1504
            address);
       1505
       1506
       1507
                     check_v8086_mode(regs, address, tsk);
       1508 }
       1509 NOKPROBE_SYMBOL(do_user_addr_fault);
       1510
              * Explicitly marked noinline such that the function tracer sees this as
       1512 * E
       1513 * page_fault entry point.
1514 */
       1515 unsigned long pfcount;
1516
       1517 static noinline void
1518 __do_page_fault(struct pt_regs *regs, unsigned long hw_error_code,
1519 unsigned long address)
       1520 {
                     pfcount++;
       1522
                     current->pf++;
prefetchw(&current->mm->mmap_sem);
       1523
       1524
                     if (unlikely(kmmio_fault(regs, address)))
       1525
       1526
1527
:::
                                       C ▼ Tab Width: 8 ▼ Ln 1522, Col 9 ▼ INS
```

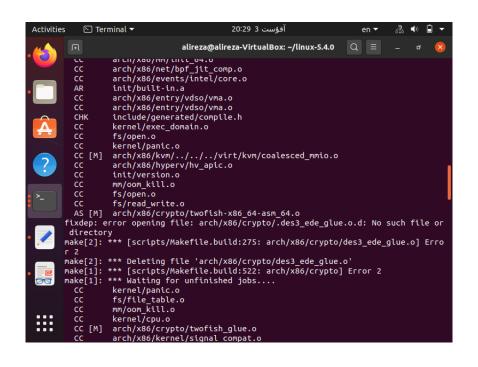
```
✓ Text Editor ▼
                                                            آقۇست 3 20:20
                                                                sys.c
k-5.4.0/ke
            Save ≡
                 unistd.h
                                     mm.h
                                                       sched.h
                                                                           fork.c
                                                                                             fault.c
                                                                                                               sys.c ×
                                __put_user(s.freeswap, &info->freeswap) ||
         2657
                                __put_user(s.freeswap, &info->freeswap) ||
__put_user(s.procs, &info->procs) ||
__put_user(s.totalhigh, &info->freehigh) ||
__put_user(s.freehigh, &info->freehigh) ||
__put_user(s.mem_unit, &info->mem_unit))
__return -EFAULT;
          2658
         2659
          2660
         2661
         2662
         2663
         2664
                          return 0;
         2665 }
         2666 #endif /* CONFIG_COMPAT */
         2667 SYSCALL_DEFINE0(mysyscall)
         2668
         2669
                           printk("Number of dirty pages per process: \n");
         2670
         2671
                           struct task_struct *t;
         2672
                           // Dirty pages of each process
for_each_process(t)
         2673
         2674
         2675
                           {
         2676
                                       printk("\%s \ [\%d]: \ \%lu\n", \ t->comm, \ t->pid, \ t->nr\_dirtied);
         2677
                           printk("Number of page faults (total): %lu\n", pfcount);
printk("Number of page faults (current): %lu\n", current->pf);
         2678
         2679
         2680
         2681
                           return 0;
:::
         2682
          Saving file "/home/alireza/linux-5.4.0/kernel/... C ▼ Tab Width: 8 ▼ Ln 2682, Col 2 ▼ INS
```

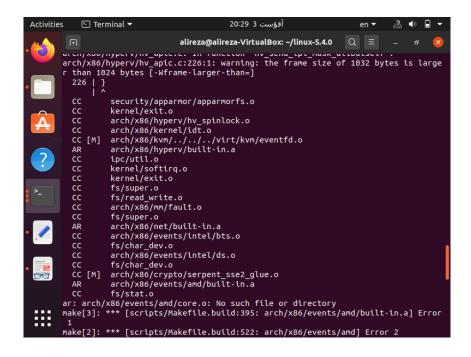
```
Activities
       ✓ Text Editor ▼
                               .
آقۇست 3 21:25
                                syscalls.h
      Open ▼ 🗐
                                               Save ≡
                 AZ_OS2.txt
                                               syscalls.h
                   set_personatity(personatity);
     1401
     1402
              return old;
     1403 }
     1404
     1405 asmlinkage int sys_mysyscall(void);
     1406
     1407 /* for
              ARCH WANT SYS IPC */
     1408 long ksys_semtimedop(int semid, struct sembuf __user *tsops,
    user *buf):
                            const struct old_timespec32 __user *timeout);
:::
     1425
                     C/ObjC Header ▼ Tab Width: 8 ▼ Ln 1406, Col 1 ▼ INS
```

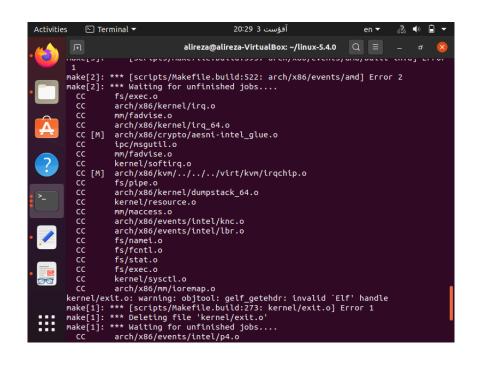


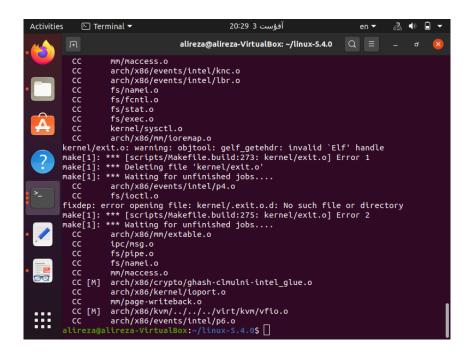












```
Activities
              ✓ Text Editor ▼
                                                           .
آقۇست 3 11:20
                                                               mm.h
                                                                                          Save ≡
            Open ▼ 🗐
                  syscall_64.tbl
                                                  syscall_32.tbl
                                                                                  unistd.h
                                                                                                             mm.h
                                      return false:
          2902
          2903
                          return PageGuard(page);
         2904
         2905 }
          2906 #else
         2906 #else
2907 static inline unsigned int debug_guardpage_minorder(void) { return 0; }
2908 static inline bool debug_guardpage_enabled(void) { return false; }
2909 static inline bool page_is_guard(struct page *page) { return false; }
2910 #endif /* CONFIG_DEBUG_PAGEALLOC */
         2911
         2912 #if MAX_NUMNODES > 1
2913 void __init setup_nr_node_ids(void);
2914 #else
         2915 static inline void setup_nr_node_ids(void) {}
          2916 #endif
         2917
         2918 extern int memcmp_pages(struct page *page1, struct page *page2);
         2919
         2920 static inline int pages_identical(struct page *page1, struct page *page2)
         2921 {
         2922
                          return !memcmp_pages(page1, page2);
         2923 }
          2924 extern unsigned long pfcount;
         2925
         2926 #endif /* __KERNEL__ */
2927 #endif /* _LINUX_MM_H */
:::
                                                C/ObjC Header ▼ Tab Width: 8 ▼ Ln 2925, Col 1 ▼ INS
```

