开启SELINUX导致热补丁加载失败问题定位

问题现象

执行 kpatch 命令加载热补丁文件 klp_dc.ko 时报错，如下所示：

```
[root@hotpatch ~]# kpatch load klp_dc.ko
loading patch module: klp_dc.ko
insmod: ERROR: could not insert module klp_dc.ko: Permission denied
kpatch: failed to load module klp_dc.ko
```

定位过程

kpatch 命令是 shell 脚本，打开 shell 调试信息，可以看到报错的位置如下：

```
out="$(LC_ALL=C insmod "$module" 2>&1)"
```

手动加载ko可以加载成功，没有报错。

```
LC_ALL=C insmod klp_dc.ko
```

通过 strace 跟踪 kpatch 命令，可以看到是系统调用 finit_module 执行失败，接下来继续定位系统调用 finit_module 失败的原因。

> 19128 finit_module(3, "", 0)   = -1 EACCES (Permission denied)
> 19128 write(2, "insmod: ERROR: could not insert "..., 68 <unfinished ...>

下面是 finit_module 的主函数，可以看到 3938 行有个 debug 日志。打开debug开关后，再复现，可以快速定界失败是发生在 3934 行 may_init_module 还是之后的逻辑中。

```
3927 SYSCALL_DEFINE3(finit_module, int, fd, const char __user *, uargs, int, flags)
3928 {
3929     struct load_info info = { };
3930     loff_t size;
3931     void *hdr;
3932     int err;
3933
3934     err = may_init_module();
3935     if (err)
3936         return err;
3937
3938     pr_debug("finit_module: fd=%d, uargs=%p, flags=%i\n", fd, uargs, flags);
3939
3940     if (flags & ~(MODULE_INIT_IGNORE_MODVERSIONS
3941                 |MODULE_INIT_IGNORE_VERMAGIC))
3942         return -EINVAL;
3943
3944     err = kernel_read_file_from_fd(fd, &hdr, &size, INT_MAX,
```

```
3945                            READING_MODULE);
3946     if (err)
3947         return err;
3948     info.hdr = hdr;
3949     info.len = size;
3950
3951     return load_module(&info, uargs, flags);
3952 }
```

打开 debug 开关

```
echo -n "file kernel/module.c +pmfl" >
/sys/kernel/debug/dynamic_debug/control
```

复现一次，查看 dmesg 日志，有打印 3938 行日志，说明报错发生在 3944 或者 3951 行。

> [4414.384518] module:__do_sys_finit_module:3938: finit_module: fd=3,
> uargs=000000006f085ba8, flags=0

接下来通过 ftrace 的 function_graph 跟踪 finit_module 函数确认报错位置（ftrace命令如下）。

```
echo function_graph > /sys/kernel/debug/tracing/current_tracer
echo __x64_sys_finit_module > /sys/kernel/debug/tracing/set_graph_function
echo 8 > /sys/kernel/debug/tracing/tracing_cpumask
echo 1 > /sys/kernel/debug/tracing/tracing_on
taskset -c 3 kpatch load klp_dc.ko
echo 0 > /sys/kernel/debug/tracing/tracing_on
cat /sys/kernel/debug/tracing/trace > /root/ftrace.log
echo > /sys/kernel/debug/tracing/set_graph_function
echo nop > /sys/kernel/debug/tracing/current_tracer
```

通过ftrace的日志看到，执行完 3944 行的 kernel_read_file_from_fd 之后就结束了，进一步查看，最后在执行安全相关（avc_has_perm）的流程 。

```
 3)               |    kernel_read_file_from_fd() {
 3)               |      __fdget() {
 3)   0.220 us    |        __fget_light();
 3)   0.595 us    |      }
 3)               |      kernel_read_file() {
 3)               |        security_kernel_read_file() {
 3)               |          selinux_kernel_read_file() {
 3)               |            selinux_kernel_module_from_file() {
 3)               |              __inode_security_revalidate() {
 3)               |                _cond_resched() {
 3)   0.198 us    |                  rcu_all_qs();
 3)   1.491 us    |                }
 3)   2.063 us    |              }
 3)               |              avc_has_perm() {
 3)   0.356 us    |                avc_denied();
 3)               |                slow_avc_audit() {
 3)               |                  common_lsm_audit() {
 3)               |                    audit_log_start() {
 3)   0.298 us    |                      audit_filter();
 3)               |                      kmem_cache_alloc() {
 3)   0.275 us    |                        should_failslab();
 3)   1.045 us    |                      }
```

第一反应，确认 SELINUX 状态，执行 getenforce， 状态是 Enforcing 表示开启。

```
[root@hotpatch ~]# getenforce
Enforcing
```

关闭 SELINUX 后再加载热补丁，结果成功。说明问题跟安全审计功能有关系。

```
[root@hotpatch ~]# setenforce 0
[root@hotpatch ~]# getenforce
Permissive
[root@hotpatch ~]# kpatch load klp_dc.ko
loading patch module: klp_dc.ko
```

先把 SELINUX 再次打开。直觉引导去查看 audit 日志（或者执行 sealert -a /var/log/audit/audit.log 看详细内容，需要安装 setroubleshoot-server 命令）。

发现确实有 insmod、kpatch 相关的日志。audit 日志可读性比较差，基于经验认为此问题跟 klp_dc.ko 文件的安全属性有关系。

> type=AVC msg=audit(1647955225.041:276): avc: denied { module_load } for pid=6221 comm="insmod" path="/root/klp_dc.ko" dev="dm-0" ino=526484 scontext=unconfined_u:unconfined_r:kpatch_t:s0-s0:c0.c1023 tcontext=system_u:object_r:admin_home_t:s0 tclass=system permissive=0
> type=SYSCALL msg=audit(1647955225.041:276): arch=c000003e syscall=313 success=no exit=-13 a0=3 a1=55e6ecebd410 a2=0 a3=3 items=0 ppid=6220 pid=6221 auid=0 uid=0 gid=0 euid=0 suid=0 fsuid=0 egid=0 sgid=0 fsgid=0 tty=pts1 ses=7 comm="insmod" exe="/usr/bin/kmod" subj=unconfined_u:unconfined_r:kpatch_t:s0-s0:c0.c1023 key=(null)ARCH=x86_64 SYSCALL=finit_module AUID="root" UID="root" GID="root" EUID="root" SUID="root" FSUID="root" EGID="root" SGID="root" FSGID="root"
> type=PROCTITLE msg=audit(1647955225.041:276): proctitle=696E736D6F64006B6C705F64632E6B6F

对比 klp_dc.ko 和内核自带ko 文件的安全属性，可以看到 USER 和 TYPE 有差异

```
[root@hotpatch ~]# ls -Z klp_dc.ko -l
-rw-------. 1 root root unconfined_u:object_r:default_t:s0 17952 Mar 22 17:03 klp_dc.ko
[root@hotpatch trace]# cd /lib/modules/4.19.90-2112.8.0.0131.u31.fos30.x86_64/kernel/kernel/trace
[root@hotpatch trace]# ls -Z * -l
-rw-r--r--. 1 root root system_u:object_r:modules_object_t:s0 28797 Mar 15 14:00 ring_buffer_benchmark.ko
[root@hotpatch trace]#
```

执行 chcon 命令修改 USER 和 TYPE 的属性分别为 system_u 和 modules_object_t 后，kpatch 执行成功。

```
[root@hotpatch ~]# chcon -u system_u -t modules_object_t klp_dc.ko
[root@hotpatch ~]# ls -Z klp_dc.ko  -l
-rw-------. 1 root root system_u:object_r:modules_object_t:s0 17952 Mar 22 17:03 klp_dc.ko
[root@hotpatch ~]#
[root@hotpatch ~]# kpatch load klp_dc.ko
loading patch module: klp_dc.ko
```

系统启用 SELINUX 且设置为 Enforcing 状态时，需要注意第三方文件的安全属性！