Песочницы

АКОС, МФТИ

12 декабря, 2024

```
prctl(PR_SET_SECCOMP, /* ... */)
```

- Устанавливает настройки `seccomp` – модуля Linux, позволяющего ограничивать доступ к системным вызовам.

```
prctl(PR_SET_SECCOMP, SECCOMP_MODE_STRICT)
```

- Включает для текущего процесса режим seccomp, разрешающий только 4 системных вызова:

`read(...)` `write(...)` `exit(...)` `sigreturn(...)`

- Любой другой системный вызов убьёт процесс через **`SIGKILL`**.
- После включения **`SECCOMP_MODE_STRICT`** нельзя отключить.

А СЕЙЧАС МЫ БУДЕМ

ИЗОЛИРОВАТЬ КОД!

Практичность `SECCOMP_MODE_STRICT`

- Нельзя сделать `exec` – он заблокирован;
- Исполнять чужой код можно только вручную замаппив его в память;
- Нельзя расширить перечень разрешенных системных вызовов;
- **В общем, шляпа.** Поэтому 7 лет никто им не пользовался.

Через 7 лет люди озадачились перехватом трафика

```
sudo tcpdump
```

Трафика бывает очень много, его надо фильтровать

```
sudo tcpdump dst 142.250.74.46
```

(Эта команда покажет только пакеты, адресованные серверу Google)

Проблема

- Пакетов может быть по-настоящему очень много. Вспомним DPDK, который отправляет пакет за 80 тактов.
- Даже если фильтровать их быстро, передача пакета в userspace и обратно – уже минимум тысяча тактов.

Проблема

- Пакетов может быть по-настоящему очень много. Вспомним DPDK, который отправляет пакет за 80 тактов.
- Даже если фильтровать их быстро, передача пакета в userspace и обратно – уже минимум тысяча тактов.

Решение - BPF!

```
setsockopt(fd, SOL_SOCKET, SO_ATTACH_FILTER, filter, size)
```

- **Устаналивает BPF-фильтр на сокете**;
- BPF (Berkeley Packet Filter) – это байт-код, который исполняется в пространстве ядра;
- Программа на BPF может отфильтровать пакет до его доставки в userspace.

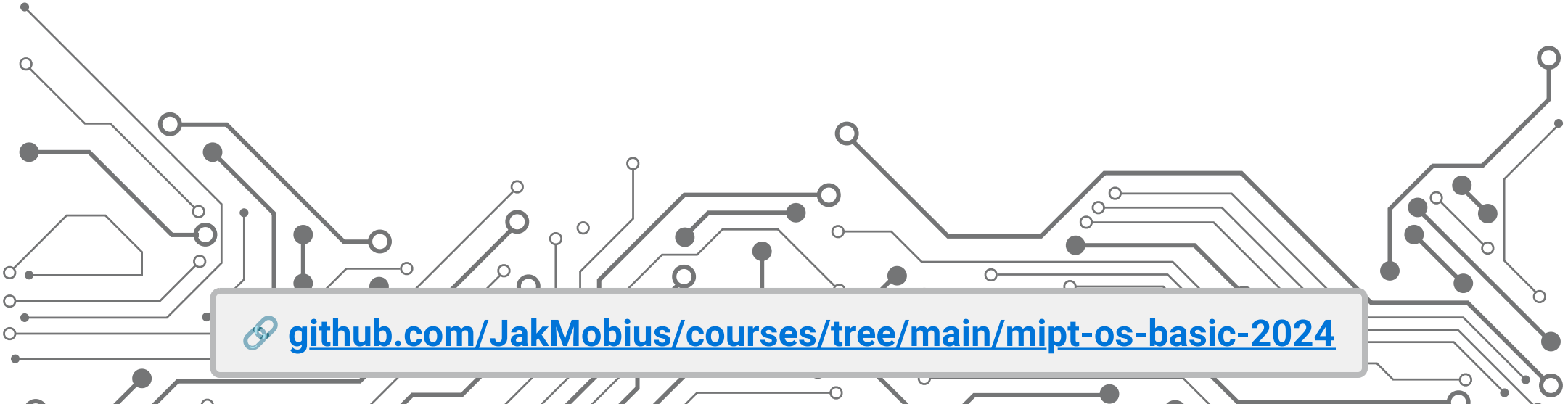
`setsockopt(fd, SOL_SOCKET, SO_ATTACH_FILTER, filter, size)`

- **Устаналивает BPF-фильтр на сокете**;
- BPF (Berkeley Packet Filter) – это байт-код, который исполняется в пространстве ядра;
- Программа на BPF может отфильтровать пакет до его доставки в userspace.

`prctl(PR_SET_SECCOMP, SECCOMP_MODE_STRICT)`

- **Устанавливает BPF-фильтр на системные вызовы.**
- Позволяет задавать сложные правила фильтрации;
- Может взаимодействовать с `ptrace()`
- Позволяет блокировать, трассировать или модифицировать системные вызовы.

Спасибо за внимание!



🔗 github.com/JakMobius/courses/tree/main/mipt-os-basic-2024