

Syd

Advanced Introduction to Secure Application Sandboxing

Ali Polatel



BalCCon2k25

whoami

Shine on you crazy diamond.

- Exherbo Linux dev, ex-Gentoo dev
- Main author of Sydbbox
- Chess trainer, Co-founder of `chesswob.org`
- Interests: Linux, BSD, Sandboxing, Security, Board games, Translation
- E-mail: `alip@chesswob.org`

Outline

You pick the place and I'll choose the time.

- History
- Threat model
- Features
- Hardenings
- Design
- Demo
- Q&A

History

And the echo of a distant time.

- Gentoo sandbox: C, LD_PRELOAD
- sydbox-1: C, ptrace(2), default on Exherbo since 2008
- sydbox-2: C, ptrace(2) & seccomp(2)
- sydbox-3: aka Syd, Rust, seccomp_unotify(2) & ptrace(2)

Threat model

Dogs of war don't negotiate.

- Trusted: Linux kernel + Syd
- Untrusted: All sandboxed apps (arbitrary code assumed)
- Goal: Restrict syscall surface; mediate FS/net; audit violations
- Out of scope: External network/physical attacks

Threat model (cont.)

Us and them, and after all we're only ordinary men.

- Unprivileged: No SETUID, EBPF, or LKM
- Multilayered: `seccomp(2)`, `ptrace(2)`, `landlock(7)`, `PR_SET_MDWE(2const)`
- General purpose: Secure by default with toggles for each hardening
- Sandbox lock and dynamic configuration:
 - `lock:on` Default, made immutable with `mseal(2)` after setup
 - `lock:off` All processes can configure, must lock after setup
 - `lock:exec` Initial thread can configure, used by `esandbox` on Exherbo
 - `lock:ipc` Configure sandbox through a UNIX socket

Features: Path sandboxing

All in all you're just another brick in the wall.

- Path sandboxing with refined categories: read, write, exec, ioctl, ...
- OpenBSD `pledge(2)` like category sets: `rpath`, `wpath`, `cpath`, `inet`, ...
- “Last matching rule wins” semantics simplifies stacking rules & profiles!

```
# Rules for /home
#
# Allow full access to $HOME.
# Forbid write ops on $HOME itself.
# rpath = stat,read,chmod,readdir
allow/rpath+${HOME}
allow/all+${HOME}/**

# Disallow all dotfiles except history.
deny/all+${HOME}/**/.*/**
allow/all+${HOME}/**/.*history*

# Mark shell-history files as append-only.
append+${HOME}/.*history
```

Features: Path hiding

There's someone in my head but it's not me.

- Path hiding: covers directory listings and path traversals

```
alip@hex / $ syd -m'allow/all+/***' -mdeny/stat+/boot -mlock:exec bash
alip@hex / $ ls /
bin dev etc home lib lib64 lost+found media mnt opt proc root run
alip@hex / $ cd /boot
bash: cd: /boot: No such file or directory
alip@hex / $ ls /etc/../../boot/../../etc/passwd
ls: cannot access '/etc/../../boot/../../etc/passwd': No such file or directory
alip@hex / $ test -c /dev/syd/allow/stat+/boot
alip@hex / $ ls /
bin boot dev etc home lib lib64 lost+found media mnt opt proc root
alip@hex / $ cd /boot
alip@hex /boot $ ls /etc/../../boot/../../etc/passwd
/etc/../../boot/../../etc/passwd
alip@hex /boot $
```


Features: Path masking

Wearing the inside out.

- Replace files and directories on the fly
- Like bind mounts but unprivileged!

```
alip@hex ~ $ cat /sys/devices/system/cpu/online
0-1
alip@hex ~ $ echo 0-16 > /tmp/online.mask
alip@hex ~ $ syd -m'allow/all+/**' -m'mask+/sys/**/cpu/online:/tmp/online.mask' bash
alip@hex ~ $ cat /sys/devices/system/cpu/online
0-16
alip@hex ~ $
exit
alip@hex ~ $ find /proc/sys 2>/dev/null | wc -l
1080
alip@hex ~ $ syd -m'allow/all+/**' -m'mask+/proc/sys/**:/dev/null:/var/empty' bash
alip@hex ~ $ find /proc/sys 2>/dev/null | wc -l
1
alip@hex ~ $ ls /proc/sys
alip@hex ~ $
```

Features: Network sandboxing

We're just two lost souls swimming in a fish bowl.

- Network sandboxing with rules based on address and port

```
# Excerpt from paludis profile
#
# Defaults for network sandboxing.
allow/net/bind+loopback!0
allow/net/bind+loopback!1024-65535
# Auto-allow successful binds for connect.
trace/allow_safe_bind:true
# Defaults for Landlock networking which is ABI>=4.
allow/lock/bind+0
allow/lock/bind+1024-65535
allow/lock/connect+1024-65535
```

Features: Learning mode

When I was a child I caught a fleeting glimpse.

- Flexible config interface: glob(3p), CIDR, ioctl(2) matching
- Learning mode with pandora(1) to automatically build profiles

```
# Syd profile generated by Pandora-0.15.0
```

```
# PROG: firefox
```

```
allow/stat,ioctl+/dev/dri/card1
```

```
ioctl/allow+DRM_IOCTL_GEM_CLOSE,DRM_IOCTL_GET_CAP
```

```
ioctl/allow+DRM_IOCTL_I915_GEM_CONTEXT_CREATE,DRM_IOCTL_I915_GEM_CONTEXT_CREATE_EXT
```

```
ioctl/allow+DRM_IOCTL_I915_GEM_CONTEXT_DESTROY,DRM_IOCTL_I915_GEM_CONTEXT_GETPARAM,DRM_IOCTL_I915_GEM_CONTEXT_CREATE_EXT
```

```
ioctl/allow+DRM_IOCTL_I915_GEM_CREATE,DRM_IOCTL_I915_GEM_CREATE_EXT
```

```
ioctl/allow+DRM_IOCTL_I915_GEM_GET_APERTURE,DRM_IOCTL_I915_GEM_GET_TILING
```

```
ioctl/allow+DRM_IOCTL_I915_GEM_MADVISE,DRM_IOCTL_I915_GEM_MMAP_OFFSET
```

```
ioctl/allow+DRM_IOCTL_I915_GEM_SET_DOMAIN,DRM_IOCTL_I915_GEM_SET_TILING
```

```
ioctl/allow+DRM_IOCTL_I915_GEM_BUSY,DRM_IOCTL_I915_GEM_EXECBUFFER2
```

```
ioctl/allow+DRM_IOCTL_I915_GETPARAM,DRM_IOCTL_I915_GET_RESET_STATS
```

```
ioctl/allow+DRM_IOCTL_I915_QUERY
```

```
ioctl/allow+DRM_IOCTL_PRIME_FD_TO_HANDLE,DRM_IOCTL_PRIME_HANDLE_TO_FD
```

```
ioctl/allow+DRM_IOCTL_SYNCOBJ_CREATE,DRM_IOCTL_SYNCOBJ_DESTROY,DRM_IOCTL_SYNCOBJ_WAIT
```

```
ioctl/allow+DRM_IOCTL_VERSION
```

```
ioctl/allow+TIOCGPGRP,TIOCGWINSZ
```

```
ioctl/allow+TCGETS
```

Features: landlock(7)

Set the controls for the heart of the sun.

- Lock sandboxing: Use `landlock(7)` to build multilayered sandboxes
- Use `syd-lock(1)` CLI to build lightweight sandboxes
- Support up to Landlock ABI 7 (current, Linux \geq 6.15)

```
alip@hex ~ $ syd-lock -h
```

```
Usage: syd-lock [-bchrwASUV] [-C level] [-E errata] [-F flag]... [-l cat[,cat...]{+|-}path|port[-port]]... {command [args...]}
```

```
Run a program under landlock(7).
```

```
Use -v to increase verbosity.
```

```
Use -A to exit with Landlock ABI version.
```

```
Use -V to check for Landlock support.
```

```
Use -l cat[,cat...]{+|-}path|port[-port] to specify categories with path or closed port range.
```

```
Use -C level to set ABI compatibility level, one of hard-requirement, soft-requirement, best-effort.
```

```
Use -E errata to check for fixes in current ABI. Argument may be a name or number.
```

```
Use -F flags to set landlock_restrict_self(2) flags.
```

```
Use -S to enable scoped signals.
```

```
Use -U to enable scoped UNIX abstract sockets.
```

```
Use 'list' with -l, -C, -E, -F to list categories, compat-levels, erratas and flags.
```

```
Use -r path as a shorthand for -l read,readdir,exec,iocctl+path.
```

```
Use -w path as a shorthand for -l all+path.
```

```
Use -b port as a shorthand for -l bind+port.
```

```
Use -c port as a shorthand for -l connect+port.
```

Features: Memory-Deny-Write-Execute (W^X)

I have become comfortably numb.

- Deny writable-executable memory with `PR_SET_MDWE(2const)`
- Harden W^X boundary with `seccomp(2)`
 - Deny `READ_IMPLIES_EXEC` personality(2)
 - Deny `PROT_READ|PROT_EXEC` mappings backed by writable files
 - Linux #219227, HardenedBSD #107
- `syd-mdwe(1)` CLI to build lightweight sandboxes coupled with `syd-lock(1)`

```
alip@hex ~ $ syd-mdwe -h
```

```
Usage: syd-mdwe [-hms] {command [args..]}
```

```
Run a command under Memory-Deny-Write-Execute protections.
```

```
Use -m to enable protections using prctl(2) PR_SET_MDWE (default).
```

```
Use -s to enable protections using seccomp(2) (use with -m to enable both).
```

```
alip@hex ~ $
```

Features: Crypt sandboxing

Careful with that axe, Eugene.

- AES-CTR + HMAC-SHA256 via AF_ALG sockets
 - Do not roll your own crypto, roll `socket(2)`s with `pipe(2)`s!
- Transparent & concurrent with Linux OFD locks (included in POSIX 2024)

```
alip@hex ~ $ export SYD_KEY=$(syd-key)
alip@hex ~ $ syd -m 'allow/all+/***' -m 'crypt/key:${SYD_KEY}' \
> -m 'crypt+/**/*.*crypt' bash
alip@hex ~ $ echo The meaning of life is 42. > test.crypt
alip@hex ~ $ file test.crypt
test.crypt: ASCII text
alip@hex ~ $ cat test.crypt
The meaning of life is 42.
alip@hex ~ $
exit
alip@hex ~ $ file test.crypt
test.crypt: SYD encrypted file, version 3, AES-256-CTR; HMAC-SHA256:0f60e75e49925
780ae6f00dc1290685005ad74cc0e4a128ca718190b8153fcf1, IV:77694655e1d5482fbd91ebd19
ffc8065, ciphertext 27 bytes
alip@hex ~ $ █
```

Features: Crypt sandboxing (cont.)

But the sun is eclipsed by the moon.

- Zero-copy with `splice(2)` and `pipe(2)`s:
 - Plaintext doesn't leak into Syd's address space!
- HMAC authentication detects corrupt data, returns `EBADMSG`!

```
alip@hex ~ $ tail -c +54 test.crypt | syd-aes -d -k${SYD_KEY} -i77694655e1d5482fb
d91ebd19ffc8065 -tSYD-ENC
The meaning of life is 42.
alip@hex ~ $ syd-bit -i $((53*8+185)) test.crypt
syd-bit: opening file test.crypt...
syd-bit: seeking to byte index 76...
syd-bit: read byte 0x04 at index 76.
syd-bit: flipped bit 1 resulting in 0x06.
syd-bit: moving back to file offset 76.
syd-bit: writing byte 0x06 at index 76...
syd-bit: flipped bit:609 of byte:76.
alip@hex ~ $ tail -c +54 test.crypt | syd-aes -d -k${SYD_KEY} -i77694655e1d5482fb
d91ebd19ffc8065 -tSYD-ENC
The meaning of life is 62.
alip@hex ~ $ syd -m 'allow/all+/**/' -m 'crypt/key:${SYD_KEY}' \
> -m 'crypt+/**/*.crypt' cat test.crypt
cat: test.crypt: Bad message
```

Features (cont.)

Haven't you heard it's a battle of words?

- Force sandboxing: Verify checksums at `exec(3)`, `mmap(2)` boundary
- Memory & PID sandboxing: Simple, unprivileged alternatives to cgroups
- Proxy sandboxing: Isolate network with a single exit through SOCKS5 proxy
- PTY sandboxing: Isolate TTY with zero-copy and unsafe ops `TI0CSTI` denied
- SafeSetID: Safely change UID/GID through predefined transition rules
- Ghost mode: Detach sandbox process from `syd`

Hardenings

Run like hell.

- Analysis-friendly: `ptrace(2)` made a no-op; `PR_SET_NAME` made a no-op with optional logging; `proc_pid_status(5)` normalized to prevent detection.
- Exec & memory: MDWE (W^X), PIE & non-exec stack enforced; secure-execution mode with `AT_SECURE`; SegvGuard/TPE for safe execution with crash throttling; ROP/SROP guards raising the bar for exploitation.
- `/proc` & devices: Hide other PIDs; block magic links of other processes; deny device-node creation/open.
- Signals & IPC: Guests cannot signal or interfere with the sandbox; remote-memory APIs denied; unsafe TTY ioctls blocked.

You rearrange me 'til I'm sane.



DEMO TIME!

Free games for May, see Emily play!

- Sacrifice a knight on f7 for DemoGods!
- Reproducible under `syd.git/doc/talks/2025-Syd-BalCCon/examples`
- Too easy? Try the CTF: <https://ctftime.org/event/2178>

Thanks for watching! Questions?

Wish you were here.

- Gitlab: <https://gitlab.exherbo.org/sydbox/sydbox.git>
- Manual: <https://man.exherbo.org>
- IRC: #sydbox at Libera
- Matrix: #sydbox:mailstation.de