Syscall Tracing für libiotrace

Gerrit Klein - HPC Projekt (Prof. Keller)

Gliederung

Gliederung

1. Vorstellung libiotrace

Gliederung

1. Vorstellung libiotrace
2. Filename Resolution

Gliederung

1. Vorstellung libiotrace
2. Filename Resolution
3. Motivation - Syscall Tracing

Gliederung

1. Vorstellung libiotrace
2. Filename Resolution
3. Motivation - Syscall Tracing
4. `ministrace`

Gliederung

1. Vorstellung libiotrace
2. Filename Resolution
3. Motivation - Syscall Tracing
4. `ministrace`
5. Todo: Integration

Vorstellung libiotrace

Motivation

Motivation

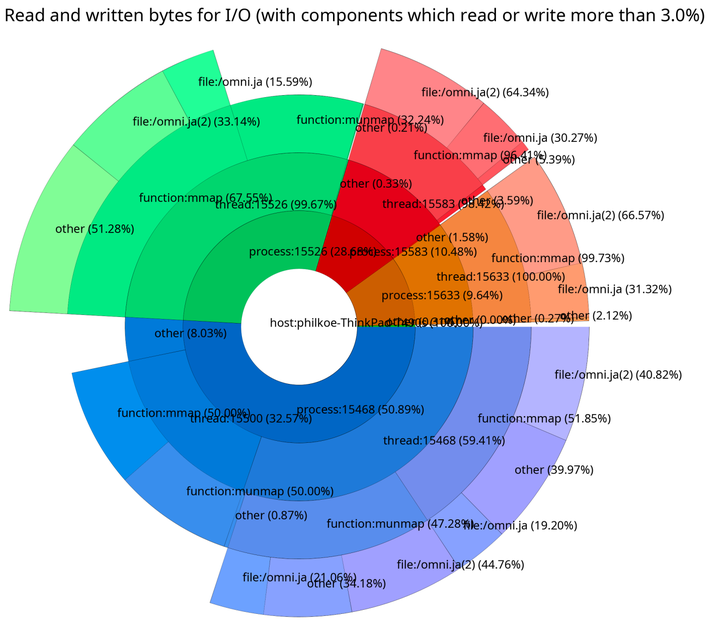
- Problem: File-I/O = Bottleneck

Motivation

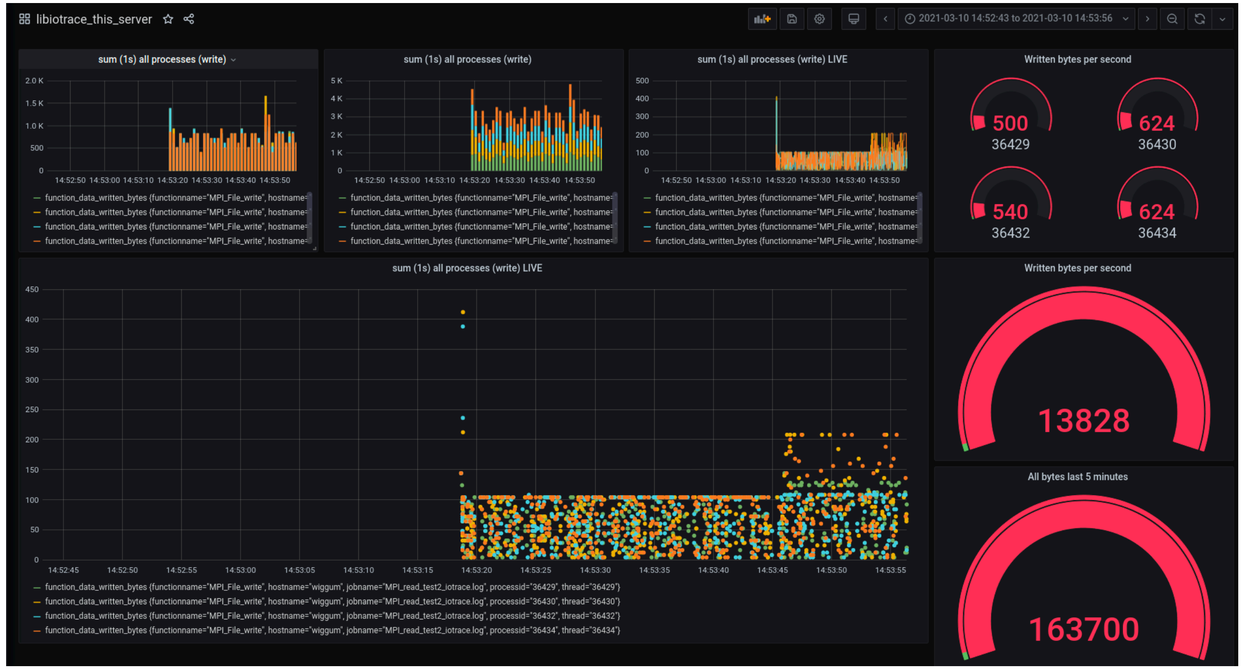
- Problem: File-I/O = Bottleneck
  - **Tracing** (POSIX-/ MPI-I/O): libiotrace

Motivation

- Problem: File-I/O = Bottleneck
  - **Tracing** (POSIX-/ MPI-I/O): libiotrace
  - **Auswertung**: Java Tool (Post mortem) / Grafana (Live Tracing)



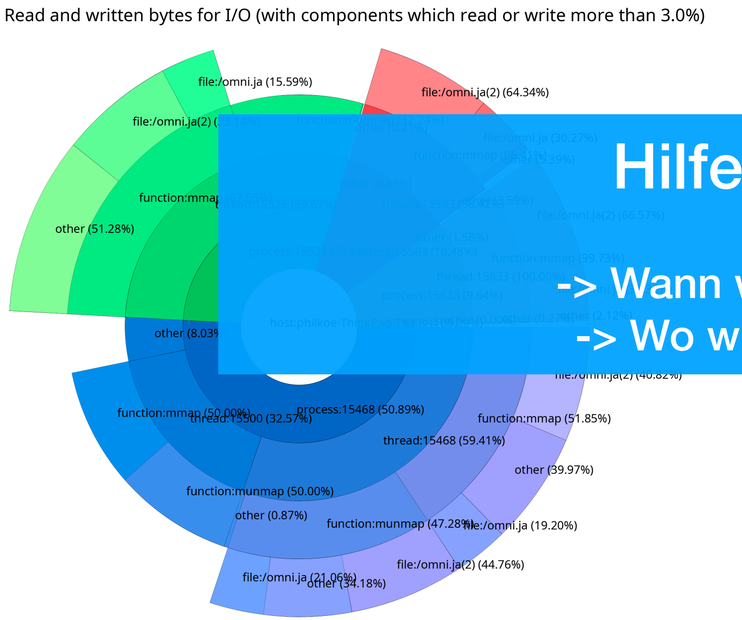
Quelle: GitHub



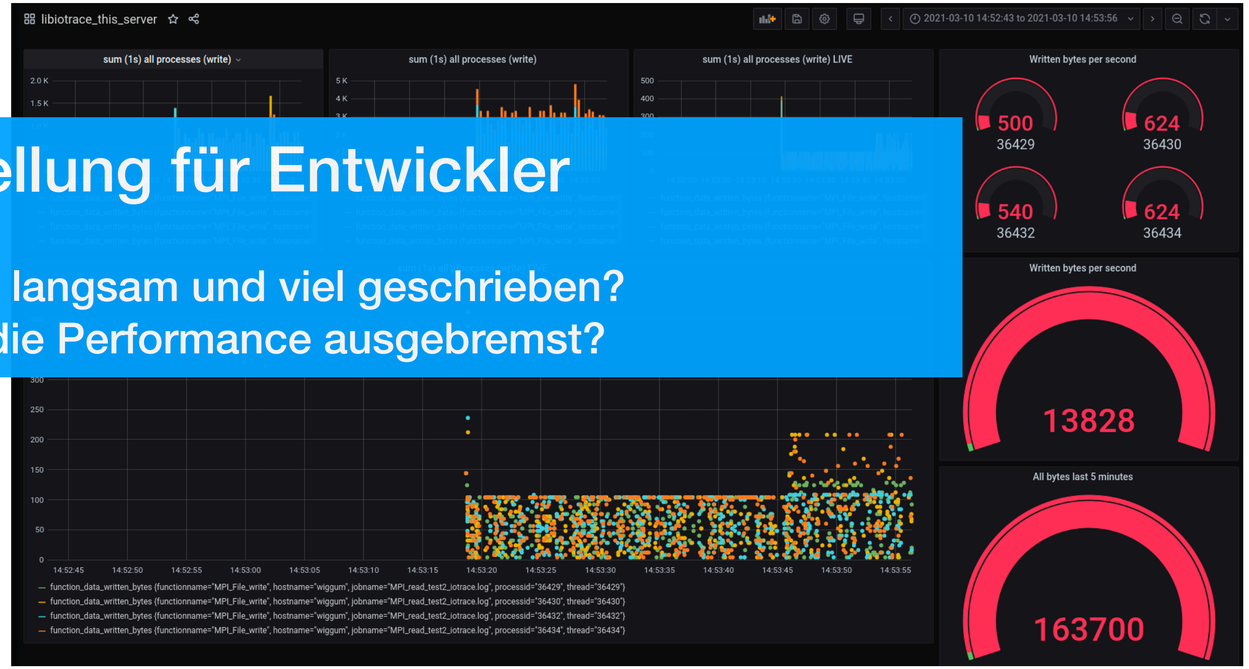
Quelle: GitHub

Motivation

- Problem: File-I/O = Bottleneck
  - **Tracing** (POSIX-/ MPI-I/O): libiotrace
  - **Auswertung**: Java Tool (Post mortem) / Grafana (Live Tracing)



Quelle: GitHub



Quelle: GitHub

Hilfestellung für Entwickler

-> Wann wird langsam und viel geschrieben?
-> Wo wird die Performance ausgebremst?

Funktionsweise

```
➜  ~ LD_PRELOAD=path/to/libiotrace_shared.so ./path/to/exec
```

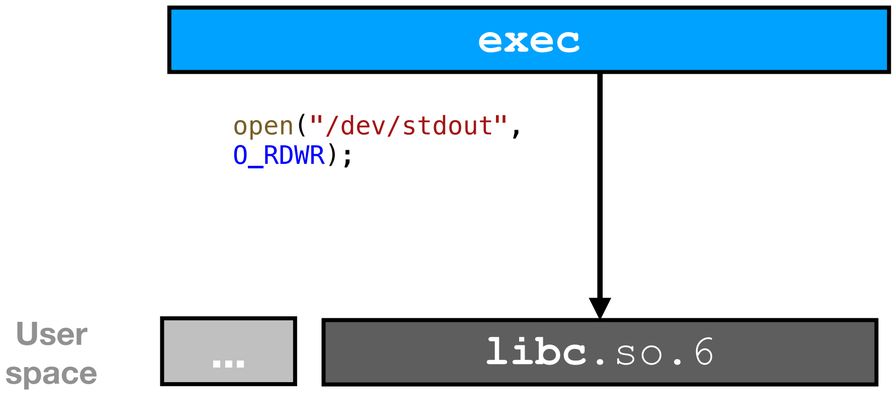
Funktionsweise

```
➜  ~ LD_PRELOAD=path/to/libiotrace_shared.so ./path/to/exec
```

User
space

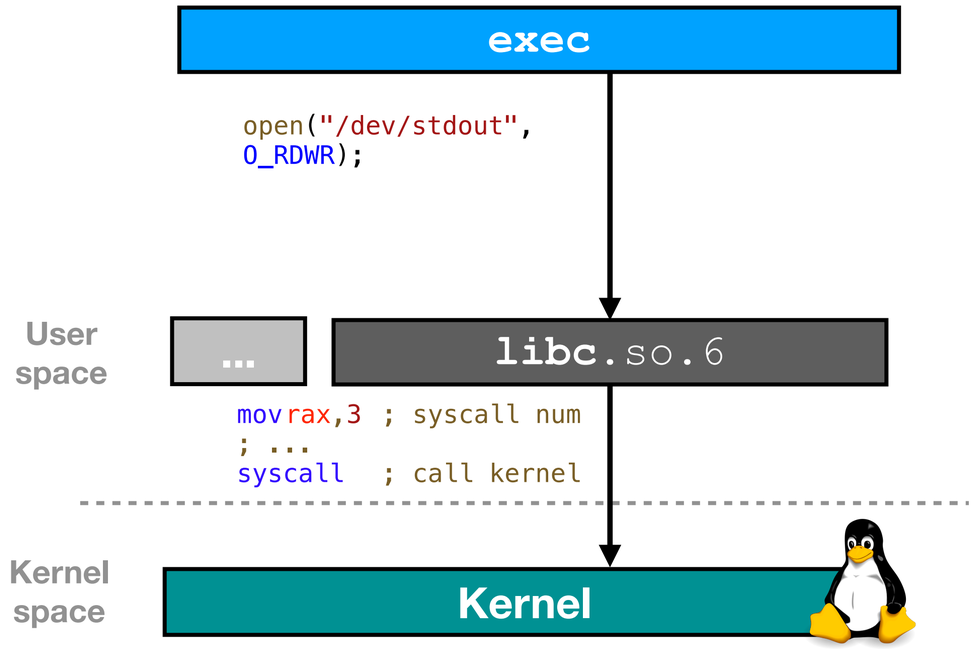
Funktionsweise

```
➜  ~ LD_PRELOAD=path/to/libiotrace_shared.so ./path/to/exec
```



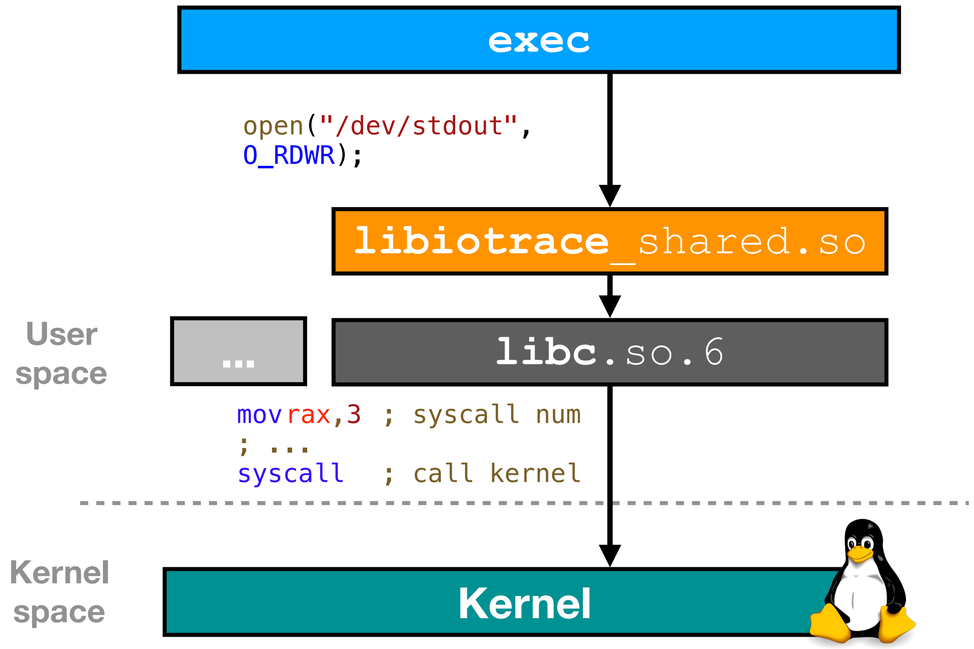
Funktionsweise

```
➜  ~ LD_PRELOAD=path/to/libiotrace_shared.so ./path/to/exec
```



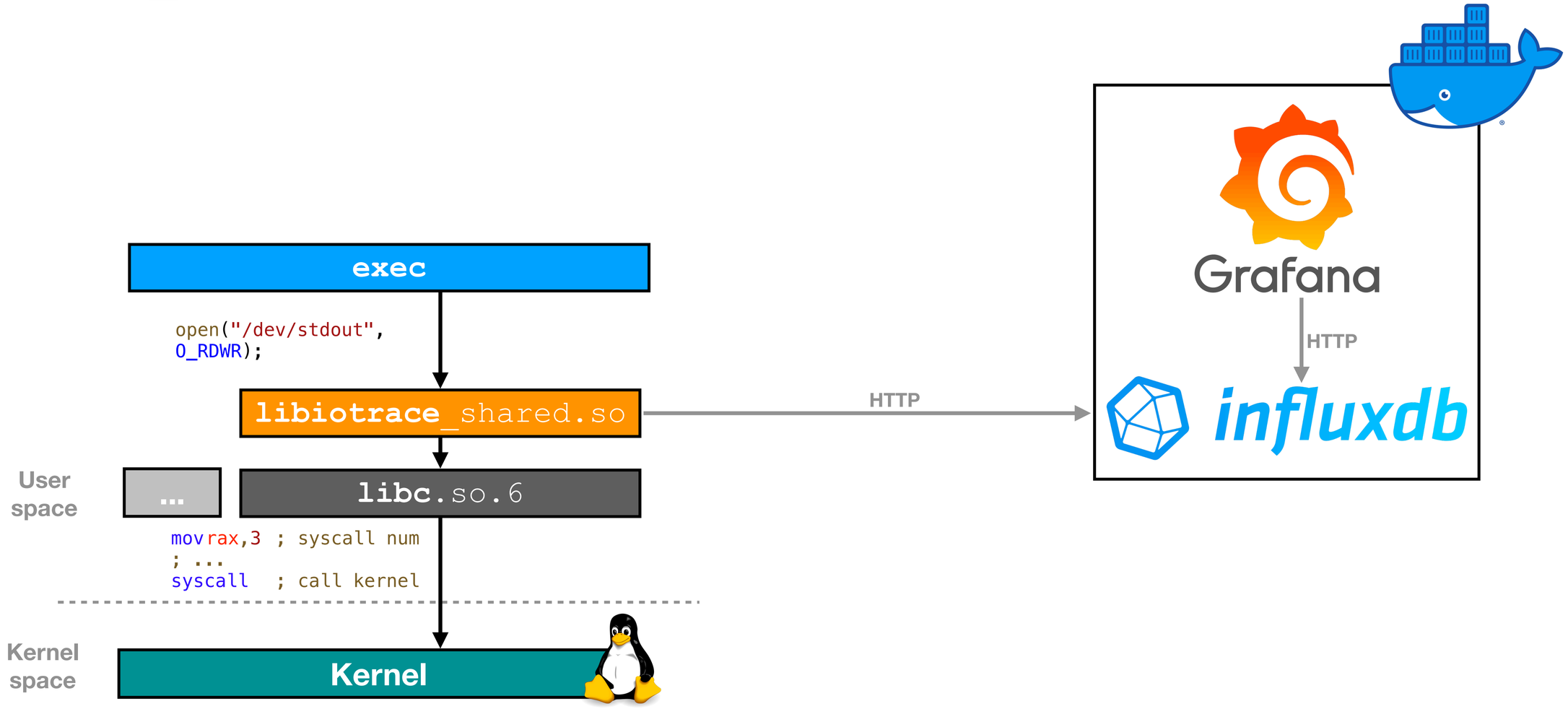
Funktionsweise

```
➜  ~ LD_PRELOAD=path/to/libiotrace_shared.so ./path/to/exec
```



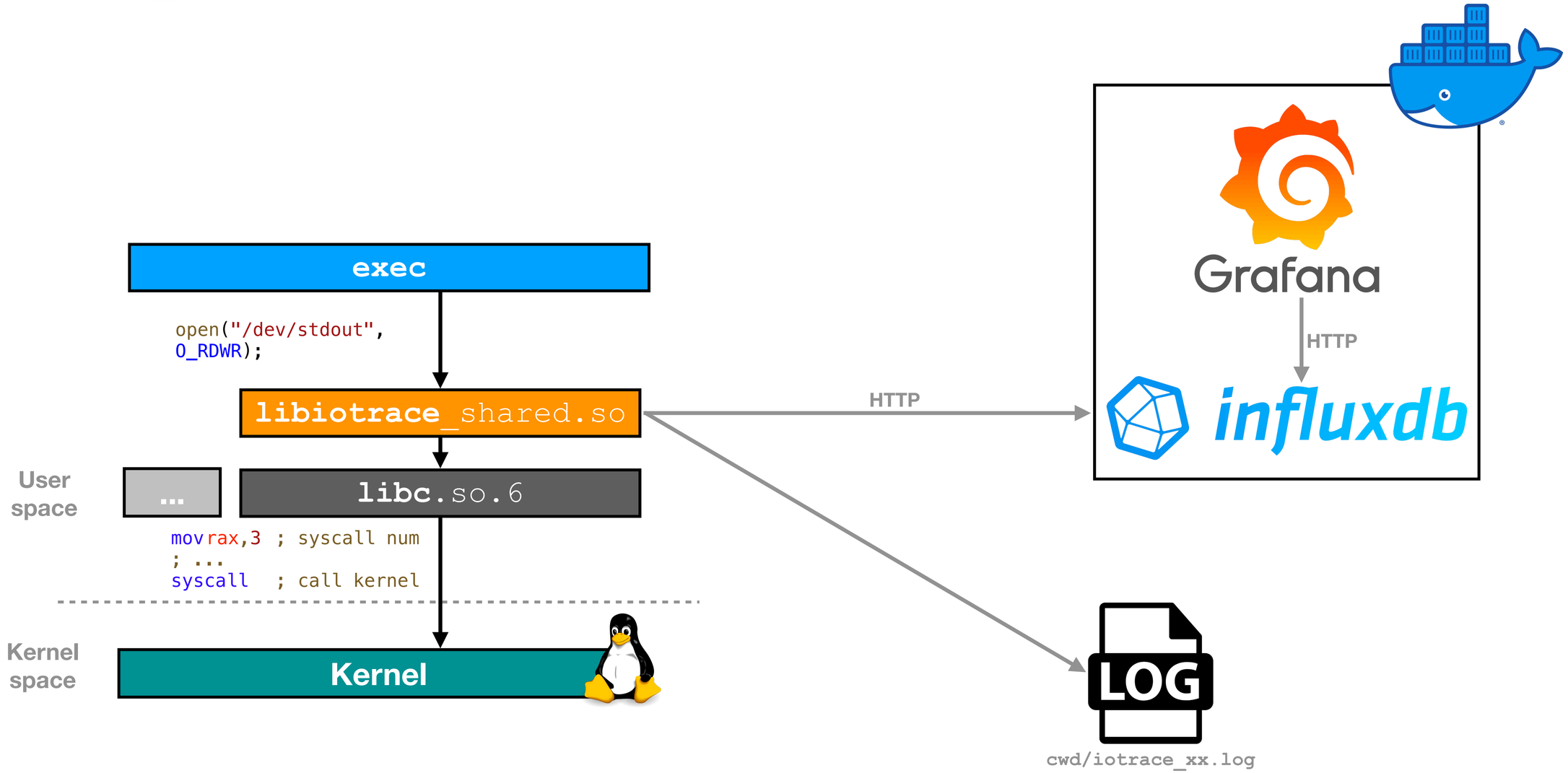
Funktionsweise

```
➜  ~ LD_PRELOAD=path/to/libiotrace_shared.so ./path/to/exec
```



Funktionsweise

```
➜  ~ LD_PRELOAD=path/to/libiotrace_shared.so ./path/to/exec
```



Filename Resolution

Motivation

Motivation

- Problem für Tracing:

Motivation

- Problem für Tracing:
  - Initiales öffnen: **Filename**

```
int fd = open("path/to/file",
O_RDONLY);
```

Motivation

- Problem für Tracing:
  - Initiales öffnen: **Filename**
  - Folgende Operationen: **Handles**

```
int fd = open("path/to/file",
O_RDONLY);

(void)read(fd, buf, sizeof buf
-1);
```

Motivation

- Problem für Tracing:
  - Initiales öffnen: **Filename**
  - Folgende Operationen: **Handles**
- Auswertung - Situation:

```
int fd = open("path/to/file",
O_RDONLY);

(void)read(fd, buf, sizeof buf
-1);
```

Motivation

- Problem für Tracing:
  - Initiales öffnen: **Filename**
  - Folgende Operationen: **Handles**

```
int fd = open("path/to/file",
O_RDONLY);

(void)read(fd, buf, sizeof buf
-1);
```

- Auswertung - Situation:
  - Post mortem: Java-Programm

Motivation

- Problem für Tracing:
  - Initiales öffnen: **Filename**
  - Folgende Operationen: **Handles**

```
int fd = open("path/to/file",
O_RDONLY);

(void)read(fd, buf, sizeof buf
-1);
```

- Auswertung - Situation:
  - Post mortem: Java-Programm

Motivation

- Problem für Tracing:
  - Initiales öffnen: **Filename**
  - Folgende Operationen: **Handles**

```
int fd = open("path/to/file",
O_RDONLY);

(void)read(fd, buf, sizeof buf
-1);
```

- Auswertung - Situation:
  - Post mortem: Java-Programm
  - Live Tracing

Motivation

- Problem für Tracing:
  - Initiales öffnen: **Filename**
  - Folgende Operationen: **Handles**

```
int fd = open("path/to/file",
O_RDONLY);

(void)read(fd, buf, sizeof buf
-1);
```

- Auswertung - Situation:
  - Post mortem: Java-Programm ✓
  - Live Tracing ✗

Funktionsweise

Handle	Filename
...	...

"Filename Map"
Handle -> Filename

Funktionsweise


```
{... "function_name":"open","file_type":
{"descriptor":3},"function_data":{"mode":"read_only","creation":
[],"status":[],"file_mode":[],"file_name":"/etc/fstab","id":
{"device_id":64768,"inode_nr":799128}}}
```


Handle	Filename
...	...

"Filename Map"

Handle -> Filename

Funktionsweise


```
{... "function_name":"open","file_type":
{"descriptor":3},"function_data":{"mode":"read_only","creation":
[],"status":[],"file_mode":[],"file_name":"/etc/fstab","id":
{"device_id":64768,"inode_nr":799128}}}
```


Handle	Filename
3	"/etc/fstab"
...	...

"Filename Map"

Handle -> Filename

Funktionsweise


```
{... "function_name":"open","file_type":
{"descriptor":3},"function_data":{"mode":"read_only","creation":
[],"status":[],"file_mode":[],"file_name":"/etc/fstab","id":
{"device_id":64768,"inode_nr":799128}}}

{... ,"function_name":"mmap","file_type":
{"descriptor":3},"function_data":
{"address":"0xffff9b6f3000","length":100,"protection_flags":
["read"],... }                                          }
```


Handle	Filename
3	"/etc/fstab"
...	...

"Filename Map"

Handle -> Filename

Funktionsweise


```
{... "function_name":"open","file_type":
{"descriptor":3},"function_data":{"mode":"read_only","creation":
[],"status":[],"file_mode":[],"file_name":"/etc/fstab","id":
{"device_id":64768,"inode_nr":799128}}}

{... ,"function_name":"mmap","file_type":
{"descriptor":3},"function_data":
{"address":"0xffff9b6f3000","length":100,"protection_flags":
["read"],... }, "traced_filename":"/etc/fstab"}
```


Handle	Filename
3	"/etc/fstab"
...	...

"Filename Map"

Handle -> Filename

Funktionsweise


```
{... "function_name":"open","file_type":
{"descriptor":3},"function_data":{"mode":"read_only","creation":
[],"status":[],"file_mode":[],"file_name":"/etc/fstab","id":
{"device_id":64768,"inode_nr":799128}}}
```

```
{... ,"function_name":"mmap","file_type":
{"descriptor":3},"function_data":
{"address":"0xffff9b6f3000","length":100,"protection_flags":
["read"],... }, "traced_filename":"/etc/fstab"}
```


Handle	Filename
3	"/etc/fstab"
"0xffff9b6f3000"	"/etc/fstab"
...	...

"Filename Map"

Handle -> Filename

Funktionsweise


```
{... "function_name":"open","file_type":
{"descriptor":3},"function_data":{"mode":"read_only","creation":
[],"status":[],"file_mode":[],"file_name":"/etc/fstab","id":
{"device_id":64768,"inode_nr":799128}}}

{... ,"function_name":"mmap","file_type":
{"descriptor":3},"function_data":
{"address":"0xffff9b6f3000","length":100,"protection_flags":
["read"],... }, "traced_filename":"/etc/fstab"}

{...,"function_name":"munmap","wrapper":{...,"file_type":
{"address":"0xffff9b6f3000"}
                                        }
```


Handle	Filename
3	"/etc/fstab"
"0xffff9b6f3000"	"/etc/fstab"
...	...

"Filename Map"

Handle -> Filename

Funktionsweise


```
{... "function_name":"open","file_type":
{"descriptor":3},"function_data":{"mode":"read_only","creation":
[],"status":[],"file_mode":[],"file_name":"/etc/fstab","id":
{"device_id":64768,"inode_nr":799128}}}

{... ,"function_name":"mmap","file_type":
{"descriptor":3},"function_data":
{"address":"0xffff9b6f3000","length":100,"protection_flags":
["read"],... }, "traced_filename":"/etc/fstab"}

{...,"function_name":"munmap","wrapper":{...,"file_type":
{"address":"0xffff9b6f3000"},
 "traced_filename":"/etc/fstab"}
```


Handle	Filename
3	"/etc/fstab"
"0xffff9b6f3000"	"/etc/fstab"
...	...

"Filename Map"

Handle -> Filename

Funktionsweise


```
{... "function_name":"open","file_type":
{"descriptor":3},"function_data":{"mode":"read_only","creation":
[],"status":[],"file_mode":[],"file_name":"/etc/fstab","id":
{"device_id":64768,"inode_nr":799128}}}

{... ,"function_name":"mmap","file_type":
{"descriptor":3},"function_data":
{"address":"0xffff9b6f3000","length":100,"protection_flags":
["read"],... }, "traced_filename":"/etc/fstab"}

{...,"function_name":"munmap","wrapper":{...,"file_type":
{"address":"0xffff9b6f3000"},
 "traced_filename":"/etc/fstab"}
```


Handle	Filename
3	"/etc/fstab"
~~"0xffff9b6f3000"~~	~~"/etc/fstab"~~
...	...

"Filename Map"

Handle -> Filename

Funktionsweise


```
{... "function_name":"open","file_type":
{"descriptor":3},"function_data":{"mode":"read_only","creation":
[],"status":[],"file_mode":[],"file_name":"/etc/fstab","id":
{"device_id":64768,"inode_nr":799128}}}

{... ,"function_name":"mmap","file_type":
{"descriptor":3},"function_data":
{"address":"0xffff9b6f3000","length":100,"protection_flags":
["read"],... }, "traced_filename":"/etc/fstab"}

{...,"function_name":"munmap","wrapper":{...,"file_type":
{"address":"0xffff9b6f3000"},
 "traced_filename":"/etc/fstab"}

{...,"function_name":"close","return_state":"ok","file_type":
{"descriptor":3}                                          }
```


Handle	Filename
3	"/etc/fstab"
~~"0xffff9b6f3000"~~	~~"/etc/fstab"~~
...	...

"Filename Map"

Handle -> Filename

Funktionsweise


```
{... "function_name":"open","file_type":
{"descriptor":3},"function_data":{"mode":"read_only","creation":
[],"status":[],"file_mode":[],"file_name":"/etc/fstab","id":
{"device_id":64768,"inode_nr":799128}}}

{... ,"function_name":"mmap","file_type":
{"descriptor":3},"function_data":
{"address":"0xffff9b6f3000","length":100,"protection_flags":
["read"],... }, "traced_filename":"/etc/fstab"}

{...,"function_name":"munmap","wrapper":{...,"file_type":
{"address":"0xffff9b6f3000"},
 "traced_filename":"/etc/fstab"}

{...,"function_name":"close","return_state":"ok","file_type":
{"descriptor":3}, "traced_filename":"/etc/fstab"}
```


Handle	Filename
3	"/etc/fstab"
~~"0xffff9b6f3000"~~	~~"/etc/fstab"~~
...	...

"Filename Map"

Handle -> Filename

Funktionsweise


```
{... "function_name":"open","file_type":
{"descriptor":3},"function_data":{"mode":"read_only","creation":
[],"status":[],"file_mode":[],"file_name":"/etc/fstab","id":
{"device_id":64768,"inode_nr":799128}}}

{... ,"function_name":"mmap","file_type":
{"descriptor":3},"function_data":
{"address":"0xffff9b6f3000","length":100,"protection_flags":
["read"],... }, "traced_filename":"/etc/fstab"}

{...,"function_name":"munmap","wrapper":{...,"file_type":
{"address":"0xffff9b6f3000"},
 "traced_filename":"/etc/fstab"}

{...,"function_name":"close","return_state":"ok","file_type":
{"descriptor":3}, "traced_filename":"/etc/fstab"}
```


Handle	Filename
~~3~~	~~"/etc/fstab"~~
~~"0xffff9b6f3000"~~	~~"/etc/fstab"~~
...	...

"Filename Map"

Handle -> Filename

Integration: Modul `fnres`

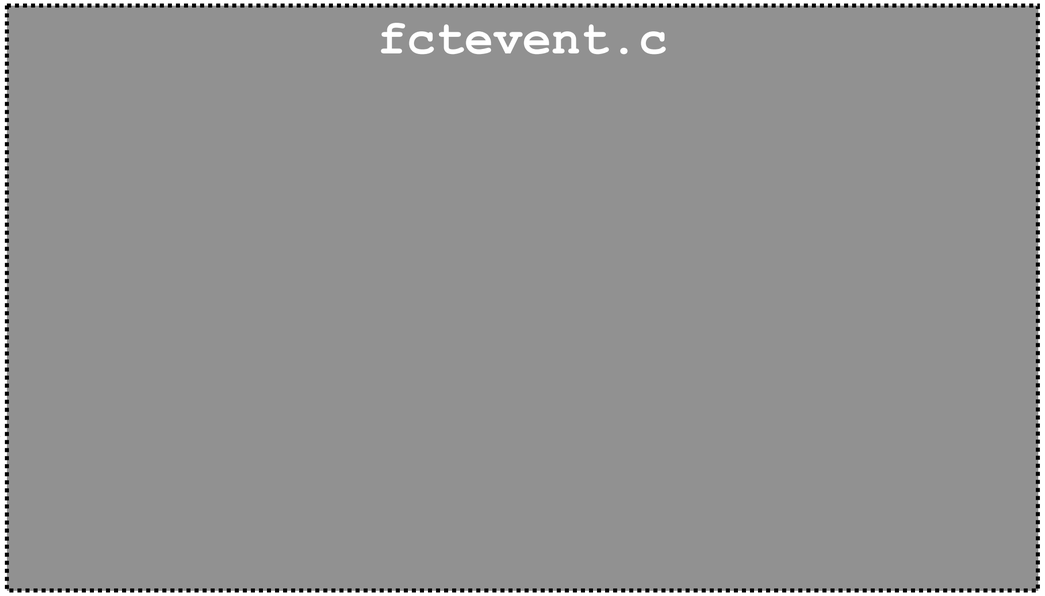
BEFORE:

```
struct basic b;

traced_filename: NULL
function_name: "open";
function_data: {...}
...
```

Integration: Modul `fnres`

BEFORE:

```
struct basic b;

traced_filename: NULL
function_name: "open";
function_data: {...}
...

fnres_trace_fc
tevent(&b);
```



Integration: Modul `fnres`

BEFORE:

```
struct basic b;

traced_filename: NULL
function_name: "open";
function_data: {...}
...

fnres_trace_fc
tevent(&b);
```

fctevent.c

1. **switch-case**(hashed_fct_name)

Integration: Modul `fnres`

BEFORE:

```
struct basic b;

traced_filename: NULL
function_name: "open";
function_data: {...}
...

fnres_trace_fc
tevent(&b);
```

fctevent.c

1. **switch-case** (hashed_fct_name)
2. Map Key ableiten von `basic` struct

Integration: Modul `fnres`

BEFORE:

```
struct basic b;

traced_filename: NULL
function_name: "open";
function_data: {...}
...

fnres_trace_fc
tevent(&b);
```

fctevent.c

1. **switch-case** (hashed_fct_name)
2. Map Key ableiten von `basic` struct

```
typedef struct {
    union id handle;
    enum id_type handle_type;
} fnmap_key;
```

Integration: Modul `fnres`

BEFORE:

```
struct basic b;

traced_filename: NULL
function_name: "open";
function_data: {...}
...

fnres_trace_fc
tevent(&b);
```

fctevent.c

1. **switch-case** (hashed_fct_name)
2. Map Key ableiten von `basic` struct
3. Key hashen + Pointer zu filename C String speichern

```
typedef struct {
    union id handle;
    enum id_type handle_type;
} fnmap_key;
```

Integration: Modul `fnres`

BEFORE:

```
struct basic b;

traced_filename: NULL
function_name: "open";
function_data: {...}
...

fnres_trace_fc
tevent(&b);
```

fctevent.c

1. **switch-case** (hashed_fct_name)
2. Map Key ableiten von `basic` struct
3. Key hashen + Pointer zu filename C String speichern
4. `traced_filename` setzen

```
typedef struct {
    union id handle;
    enum id_type handle_type;
} fnmap_key;
```

Integration: Modul `fnres`

BEFORE:

```
struct basic b;

traced_filename: NULL
function_name: "open";
function_data: {...}
...

fnres_trace_fc
tevent(&b);
```

fctevent.c

1. **`switch-case`** (`hashed_fct_name`)
2. Map Key ableiten von `basic` struct
3. Key hashen + Pointer zu filename C String speichern
4. `traced_filename` setzen

AFTER:

```
struct basic b;

traced_filename: 0x0300..
function_name: "open";
function_data: {...}
...
```

```
typedef struct {
    union id handle;
    enum id_type handle_type;
} fnmap_key;
```

Integration: Modul `fnres`

BEFORE:

```
struct basic b;

traced_filename: NULL
function_name: "open";
function_data: {...}
...

fnres_trace_fc
tevent(&b);
```

fctevent.c

1. **`switch-case`** (`hashed_fct_name`)
2. Map Key ableiten von `basic` struct
3. Key hashen + Pointer zu filename C String speichern
4. `traced_filename` setzen

AFTER:

```
struct basic b;

traced_filename: 0x0300..
function_name: "open";
function_data: {...}
...
```

```
typedef struct {
    union id handle;
    enum id_type handle_type;
} fnmap_key;
```

- `atomic_hash`: Lock-free, C
  - Map: Hashed `fnmap_key` -> `char* filename`

Motivation - Syscall Tracing

Limitierungen `fnres`

Limitierungen `fnres`


```
filename":"_ NOT FOUND _","hostname":"fpj-vm","process_id":16215,"thread_id":16215,"function_name":"mmap",'
filename":"_ NOT FOUND _","hostname":"fpj-vm","process_id":16215,"thread_id":16215,"function_name":"close",
```

Limitierungen `fnres`


```
filename":"_ NOT FOUND _","hostname":"fpj-vm","process_id":16215,"thread_id":16215,"function_name":"mmap",'
filename":"_ NOT FOUND _","hostname":"fpj-vm","process_id":16215,"thread_id":16215,"function_name":"close",
```

Limitierungen `fnres`


```
filename":"_ NOT FOUND _","hostname":"fpj-vm","process_id":16215,"thread_id":16215,"function_name":"mmap",'
filename":"_ NOT FOUND _","hostname":"fpj-vm","process_id":16215,"thread_id":16215,"function_name":"close",
```


- Fortified functions `open` -> `__open_2`

Limitierungen `fnres`


```
filename":"_ NOT FOUND _","hostname":"fpj-vm","process_id":16215,"thread_id":16215,"function_name":"mmap",'
filename":"_ NOT FOUND _","hostname":"fpj-vm","process_id":16215,"thread_id":16215,"function_name":"close",
```


- Fortified functions `open` -> `__open_2`
- Kein libc call ?!

Limitierungen `fnres`

```
filename":"_ NOT FOUND _","hostname":"fpj-vm","process_id":16215,"thread_id":16215,"function_name":"mmap",'
filename":"_ NOT FOUND _","hostname":"fpj-vm","process_id":16215,"thread_id":16215,"function_name":"close",
```

- Fortified functions `open` -> `__open_2`
- Kein libc call ?!

```
[pid 35846] openat(AT_FDCWD, "/dev/shm/UC8GnG", O_RDWR|O_CREAT|O_EXCL, 0644) = 17
    /lib/x86_64-linux-gnu/libpthread-2.31.so(__open64+0xd4) [0x14ad4]
    /lib/x86_64-linux-gnu/libpthread-2.31.so(sem_open+0x2b5) [0x12df5]
    ...
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(MPI_File_open+0x20a) [0xa8899]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(main+0xc8) [0x1391]
    /lib/x86_64-linux-gnu/libc-2.31.so(__libc_start_main+0xf3) [0x270b3]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(_start+0x2e) [0x120e]
...
[pid 35846] mmap(NULL, 32, PROT_READ|PROT_WRITE, MAP_SHARED, 17, 0) = 0x7f57ac8ee000
    /lib/x86_64-linux-gnu/libc-2.31.so(mmap64+0x26) [0x11ba46]
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(mmap+0x197) [0x943d7]
    /lib/x86_64-linux-gnu/libpthread-2.31.so(sem_open+0x41e) [0x12f5e]
    ...
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(MPI_File_open+0x20a) [0xa8899]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(main+0xc8) [0x1391]
    /lib/x86_64-linux-gnu/libc-2.31.so(__libc_start_main+0xf3) [0x270b3]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(_start+0x2e) [0x120e]
```

Limitierungen `fnres`

```
filename":"_ NOT FOUND _","hostname":"fpj-vm","process_id":16215,"thread_id":16215,"function_name":"mmap",'
filename":"_ NOT FOUND _","hostname":"fpj-vm","process_id":16215,"thread_id":16215,"function_name":"close",
```

- Fortified functions `open` -> `__open_2`
- Kein libc call ?!

```
[pid 35846] openat(AT_FDCWD, "/dev/shm/UC8GnG", O_RDWR|O_CREAT|O_EXCL, 0644) = 17
    /lib/x86_64-linux-gnu/libpthread-2.31.so(__open64+0xd4) [0x14ad4]
    /lib/x86_64-linux-gnu/libpthread-2.31.so(sem_open+0x2b5) [0x12df5]
    ...
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(MPI_File_open+0x20a) [0xa8899]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(main+0xc8) [0x1391]
    /lib/x86_64-linux-gnu/libc-2.31.so(__libc_start_main+0xf3) [0x270b3]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(_start+0x2e) [0x120e]
...
[pid 35846] mmap(NULL, 32, PROT_READ|PROT_WRITE, MAP_SHARED, 17, 0) = 0x7f57ac8ee000
    /lib/x86_64-linux-gnu/libc-2.31.so(mmap64+0x26) [0x11ba46]
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(mmap+0x197) [0x943d7]
    /lib/x86_64-linux-gnu/libpthread-2.31.so(sem_open+0x41e) [0x12f5e]
    ...
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(MPI_File_open+0x20a) [0xa8899]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(main+0xc8) [0x1391]
    /lib/x86_64-linux-gnu/libc-2.31.so(__libc_start_main+0xf3) [0x270b3]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(_start+0x2e) [0x120e]
```

<- Kein "libc call" = Kein libiotrace wrapper

Limitierungen `fnres`

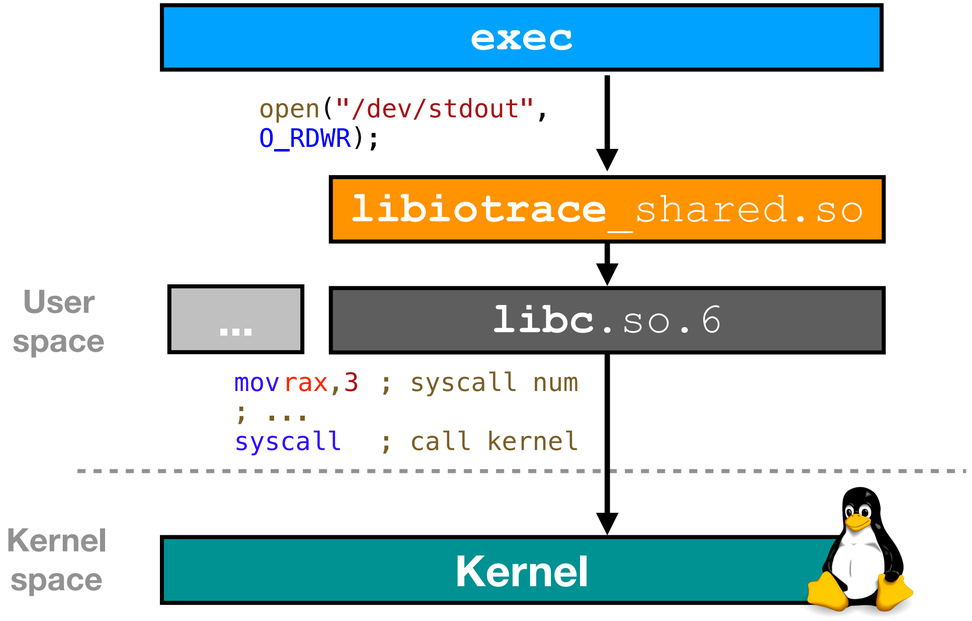
```
filename":"_ NOT FOUND _","hostname":"fpj-vm","process_id":16215,"thread_id":16215,"function_name":"mmap",'
filename":"_ NOT FOUND _","hostname":"fpj-vm","process_id":16215,"thread_id":16215,"function_name":"close",
```

- Fortified functions `open` -> `__open_2`
- Kein libc call ?!

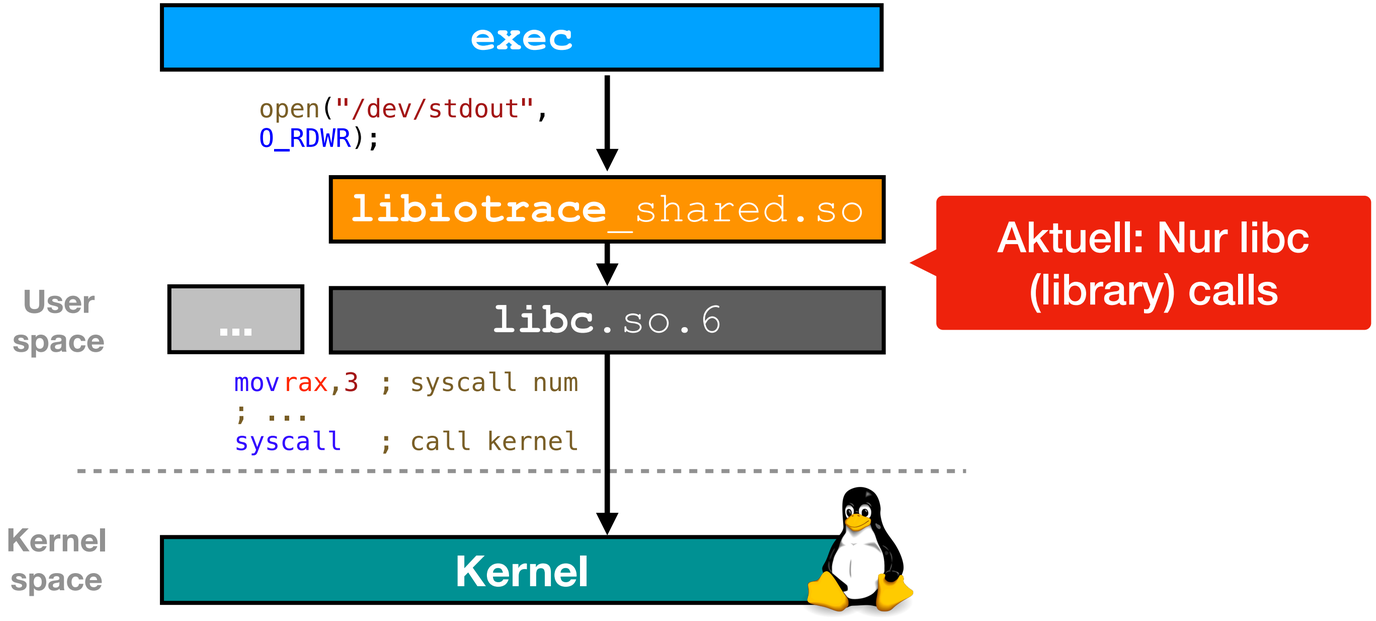
```
[pid 35846] openat(AT_FDCWD, "/dev/shm/UC8GnG", O_RDWR|O_CREAT|O_EXCL, 0644) = 17
    /lib/x86_64-linux-gnu/libpthread-2.31.so(__open64+0xd4) [0x14ad4]
    /lib/x86_64-linux-gnu/libpthread-2.31.so(sem_open+0x2b5) [0x12df5]
    ...
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(MPI_File_open+0x20a) [0xa8899]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(main+0xc8) [0x1391]
    /lib/x86_64-linux-gnu/libc-2.31.so(__libc_start_main+0xf3) [0x270b3]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(_start+0x2e) [0x120e]
...
[pid 35846] mmap(NULL, 32, PROT_READ|PROT_WRITE, MAP_SHARED, 17, 0) = 0x7f57ac8ee000
    /lib/x86_64-linux-gnu/libc-2.31.so(mmap64+0x26) [0x11ba46]
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(mmap+0x197) [0x943d7]
    /lib/x86_64-linux-gnu/libpthread-2.31.so(sem_open+0x41e) [0x12f5e]
    ...
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(MPI_File_open+0x20a) [0xa8899]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(main+0xc8) [0x1391]
    /lib/x86_64-linux-gnu/libc-2.31.so(__libc_start_main+0xf3) [0x270b3]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(_start+0x2e) [0x120e]
```

<- Kein "libc call" = Kein libiotrace wrapper

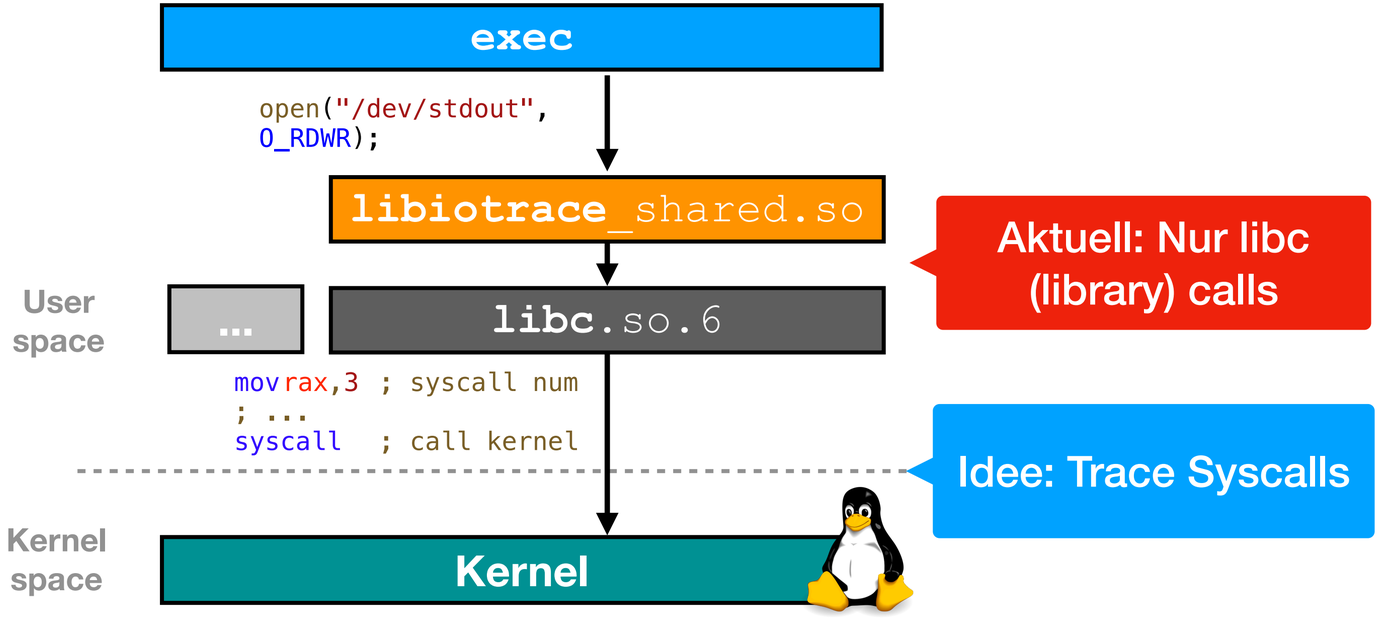
"Remedy"


exec
open("/dev/stdout",
O_RDWR);
libiotrace_shared.so
User space
...
libc.so.6
mov rax,3 ; syscall num
; ...
syscall ; call kernel
Kernel space
Kernel

"Remedy"

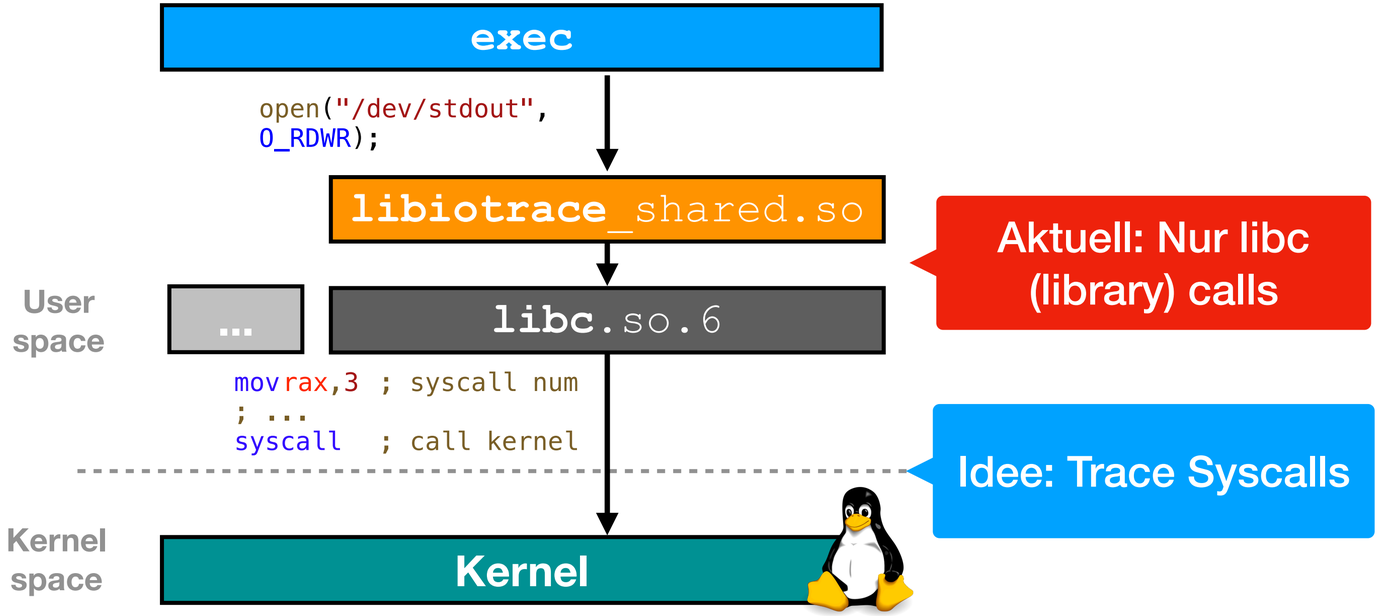

exec
open("/dev/stdout", O_RDWR);
libiotrace_shared.so
User space
...
libc.so.6
Aktuell: Nur libc (library) calls
mov rax,3 ; syscall num
; ...
syscall ; call kernel
Kernel space
Kernel

"Remedy"

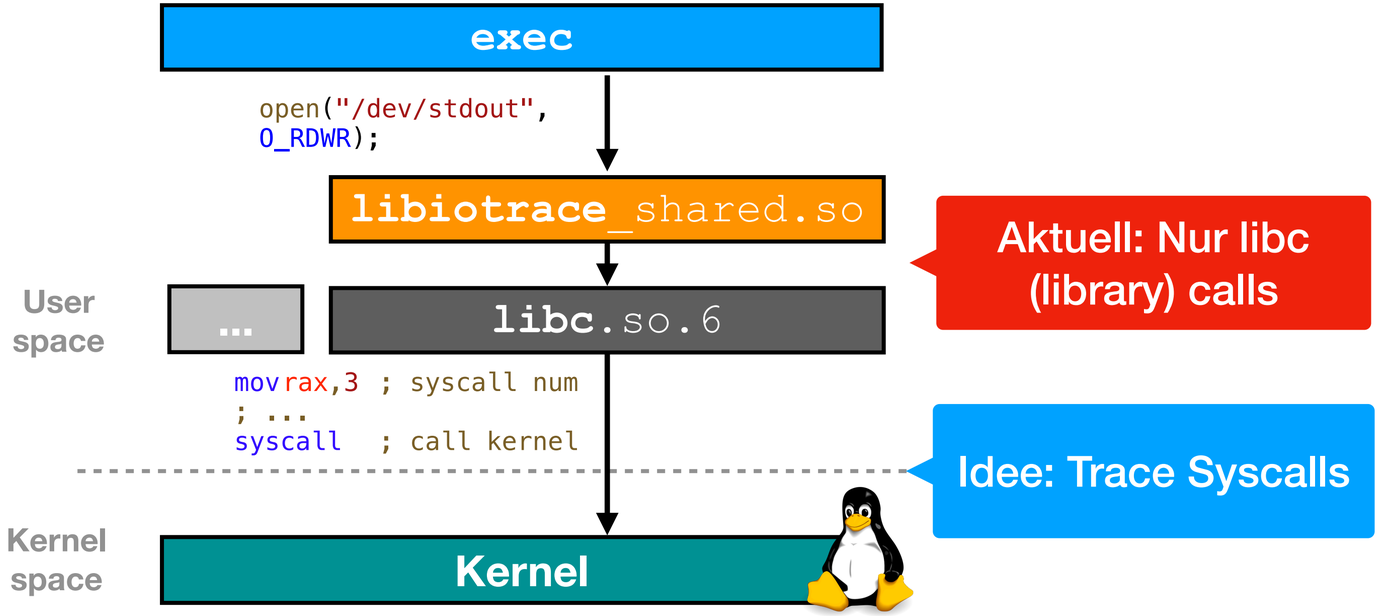

exec
open("/dev/stdout", O_RDWR);
libiotrace_shared.so
User space
...
libc.so.6
mov rax,3 ; syscall num
; ...
syscall ; call kernel
Kernel space
Kernel
Aktuell: Nur libc (library) calls
Idee: Trace Syscalls

"Remedy"

- **Warum?**

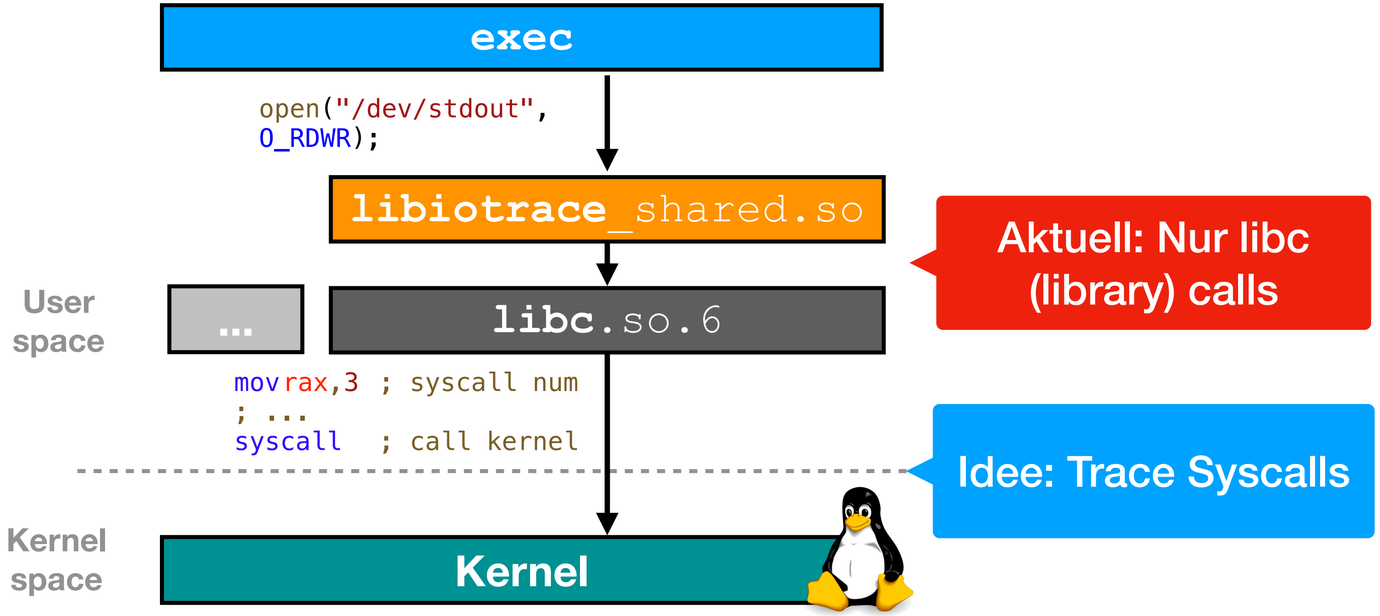


"Remedy"



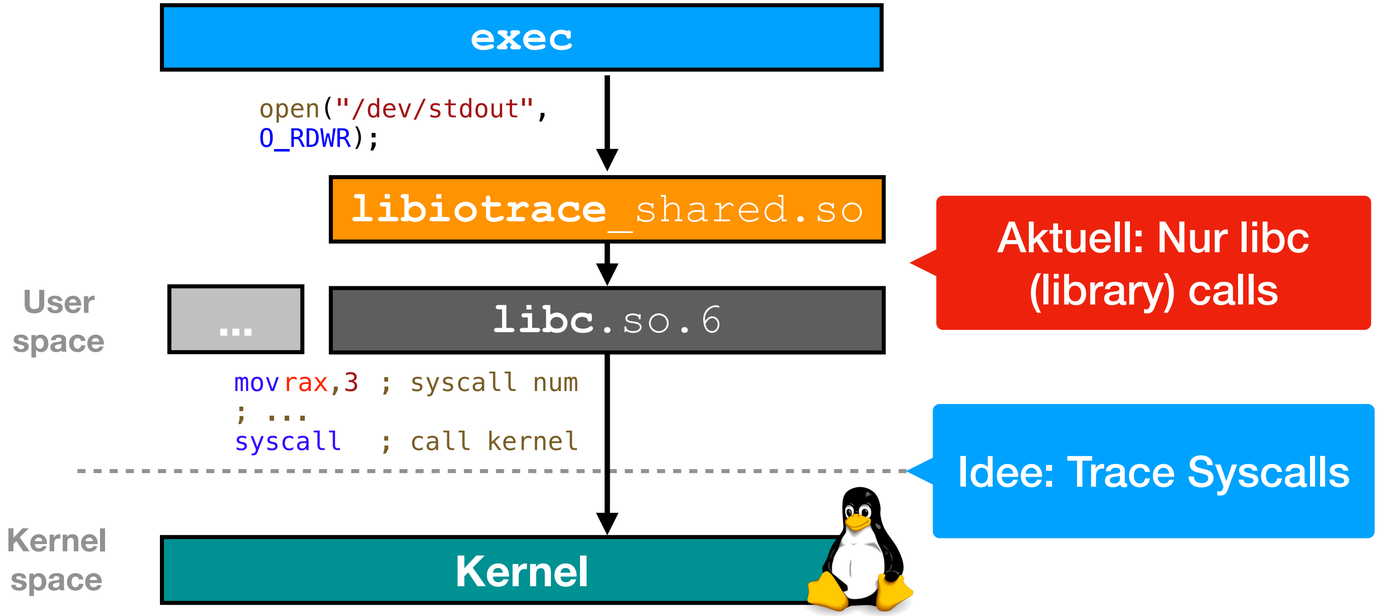
- **Warum?**
  - Jeder `open` -> Syscall

"Remedy"



- **Warum?**
  - Jeder `open` -> Syscall
- **Wie?**

"Remedy"



- **Warum?**
  - Jeder `open` -> Syscall
- **Wie?**
  - `ptrace`(2)

Ansatz für Umsetzung

Ansatz für Umsetzung

- **Execution Stack unwinding**

Ansatz für Umsetzung

- **Execution Stack unwinding**
  - Von Remote Target !!

Ansatz für Umsetzung

- **Execution Stack unwinding**
  - Von Remote Target !!

```
[pid 35846] mmap(NULL, 32, PROT_READ|PROT_WRITE, MAP_SHARED, 17, 0) = 0x7f57ac8ee000
    /lib/x86_64-linux-gnu/libc-2.31.so(mmap64+0x26) [0x11ba46]
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(mmap+0x197) [0x943d7]
    /lib/x86_64-linux-gnu/libpthread-2.31.so(sem_open+0x41e) [0x12f5e]
    ...
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(MPI_File_open+0x20a) [0xa8899]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(main+0xc8) [0x1391]
    /lib/x86_64-linux-gnu/libc-2.31.so(__libc_start_main+0xf3) [0x270b3]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(_start+0x2e) [0x120e]
```

Ansatz für Umsetzung

- **Execution Stack unwinding**
  - Von Remote Target !!

```
[pid 35846] mmap(NULL, 32, PROT_READ|PROT_WRITE, MAP_SHARED, 17, 0) = 0x7f57ac8ee000
    /lib/x86_64-linux-gnu/libc-2.31.so(mmap64+0x26) [0x11ba46]
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(mmap+0x197) [0x943d7]
    /lib/x86_64-linux-gnu/libpthread-2.31.so(sem_open+0x41e) [0x12f5e]
    ...
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(MPI_File_open+0x20a) [0xa8899]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(main+0xc8) [0x1391]
    /lib/x86_64-linux-gnu/libc-2.31.so(__libc_start_main+0xf3) [0x270b3]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(_start+0x2e) [0x120e]
```

Ansatz für Umsetzung

- **Execution Stack unwinding**
  - Von Remote Target !!

```
[pid 35846] mmap(NULL, 32, PROT_READ|PROT_WRITE, MAP_SHARED, 17, 0) = 0x7f57ac8ee000
    /lib/x86_64-linux-gnu/libc-2.31.so(mmap64+0x26) [0x11ba46]
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(mmap+0x197) [0x943d7]
    /lib/x86_64-linux-gnu/libpthread-2.31.so(sem_open+0x41e) [0x12f5e]
    ...
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(MPI_File_open+0x20a) [0xa8899]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(main+0xc8) [0x1391]
    /lib/x86_64-linux-gnu/libc-2.31.so(__libc_start_main+0xf3) [0x270b3]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(_start+0x2e) [0x120e]
...
```

Ansatz für Umsetzung

- **Execution Stack unwinding**
  - Von Remote Target !!

```
[pid 35846] mmap(NULL, 32, PROT_READ|PROT_WRITE, MAP_SHARED, 17, 0) = 0x7f57ac8ee000
    /lib/x86_64-linux-gnu/libc-2.31.so(mmap64+0x26) [0x11ba46]
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(mmap+0x197) [0x943d7]
    /lib/x86_64-linux-gnu/libpthread-2.31.so(sem_open+0x41e) [0x12f5e]
    ...
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(MPI_File_open+0x20a) [0xa8899]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(main+0xc8) [0x1391]
    /lib/x86_64-linux-gnu/libc-2.31.so(__libc_start_main+0xf3) [0x270b3]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(_start+0x2e) [0x120e]
```

```
[pid 35846] openat(AT_FDCWD, "/dev/shm/UC8GnG", O_RDWR|O_CREAT|O_EXCL, 0644) = 17
    /lib/x86_64-linux-gnu/libpthread-2.31.so(__open64+0xd4) [0x14ad4]
    /lib/x86_64-linux-gnu/libpthread-2.31.so(sem_open+0x2b5) [0x12df5]
    ...
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(MPI_File_open+0x20a) [0xa8899]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(main+0xc8) [0x1391]
    /lib/x86_64-linux-gnu/libc-2.31.so(__libc_start_main+0xf3) [0x270b3]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(_start+0x2e) [0x120e]
```

Ansatz für Umsetzung

- **Execution Stack unwinding**
  - Von Remote Target !!

```
[pid 35846] mmap(NULL, 32, PROT_READ|PROT_WRITE, MAP_SHARED, 17, 0) = 0x7f57ac8ee000
    /lib/x86_64-linux-gnu/libc-2.31.so(mmap64+0x26) [0x11ba46]
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(mmap+0x197) [0x943d7]
    /lib/x86_64-linux-gnu/libpthread-2.31.so(sem_open+0x41e) [0x12f5e]
    ...
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(MPI_File_open+0x20a) [0xa8899]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(main+0xc8) [0x1391]
    /lib/x86_64-linux-gnu/libc-2.31.so(__libc_start_main+0xf3) [0x270b3]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(_start+0x2e) [0x120e]
```

```
[pid 35846] openat(AT_FDCWD, "/dev/shm/UC8GnG", O_RDWR|O_CREAT|O_EXCL, 0644) = 17
    /lib/x86_64-linux-gnu/libpthread-2.31.so(__open64+0xd4) [0x14ad4]
    /lib/x86_64-linux-gnu/libpthread-2.31.so(sem_open+0x2b5) [0x12df5]
    ...
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(MPI_File_open+0x20a) [0xa8899]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(main+0xc8) [0x1391]
    /lib/x86_64-linux-gnu/libc-2.31.so(__libc_start_main+0xf3) [0x270b3]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(_start+0x2e) [0x120e]
```

Ansatz für Umsetzung

- **Execution Stack unwinding**
  - Von Remote Target !!

```
[pid 35846] mmap(NULL, 32, PROT_READ|PROT_WRITE, MAP_SHARED, 17, 0) = 0x7f57ac8ee000
    /lib/x86_64-linux-gnu/libc-2.31.so(mmap64+0x26) [0x11ba46]
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(mmap+0x197) [0x943d7]
    /lib/x86_64-linux-gnu/libpthread-2.31.so(sem_open+0x41e) [0x12f5e]
    ...
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(MPI_File_open+0x20a) [0xa8899]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(main+0xc8) [0x1391]
    /lib/x86_64-linux-gnu/libc-2.31.so(__libc_start_main+0xf3) [0x270b3]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(_start+0x2e) [0x120e]
```

```
[pid 35846] openat(AT_FDCWD, "/dev/shm/UC8GnG", O_RDWR|O_CREAT|O_EXCL, 0644) = 17
    /lib/x86_64-linux-gnu/libpthread-2.31.so(__open64+0xd4) [0x14ad4]
    /lib/x86_64-linux-gnu/libpthread-2.31.so(sem_open+0x2b5) [0x12df5]
    ...
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(MPI_File_open+0x20a) [0xa8899]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(main+0xc8) [0x1391]
    /lib/x86_64-linux-gnu/libc-2.31.so(__libc_start_main+0xf3) [0x270b3]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(_start+0x2e) [0x120e]
```

Ansatz für Umsetzung

- **Execution Stack unwinding**
  - Von Remote Target !!

```
[pid 35846] mmap(NULL, 32, PROT_READ|PROT_WRITE, MAP_SHARED, 17, 0) = 0x7f57ac8ee000
    /lib/x86_64-linux-gnu/libc-2.31.so(mmap64+0x26) [0x11ba46]
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(mmap+0x197) [0x943d7]
    /lib/x86_64-linux-gnu/libpthread-2.31.so(sem_open+0x41e) [0x12f5e]
    ...
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(MPI_File_open+0x20a) [0xa8899]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(main+0xc8) [0x1391]
    /lib/x86_64-linux-gnu/libc-2.31.so(__libc_start_main+0xf3) [0x270b3]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(_start+0x2e) [0x120e]
```

```
[pid 35846] openat(AT_FDCWD, "/dev/shm/UC8GnG", O_RDWR|O_CREAT|O_EXCL, 0644) = 17
    /lib/x86_64-linux-gnu/libpthread-2.31.so(__open64+0xd4) [0x14ad4]
    /lib/x86_64-linux-gnu/libpthread-2.31.so(sem_open+0x2b5) [0x12df5]
    ...
    /home/philkoe/git/fsprj2/libiotrace/build/src/libiotrace_shared.so(MPI_File_open+0x20a) [0xa8899]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(main+0xc8) [0x1391]
    /lib/x86_64-linux-gnu/libc-2.31.so(__libc_start_main+0xf3) [0x270b3]
    /home/philkoe/git/fsprj2/libiotrace/build/test/MPI_read(_start+0x2e) [0x120e]
```

Warnung,
Eintrag ergänzen,
....

`ptrace`(2)

`ptrace`(2)

- Syscall (`sys_ptrace`)

`ptrace`(2)

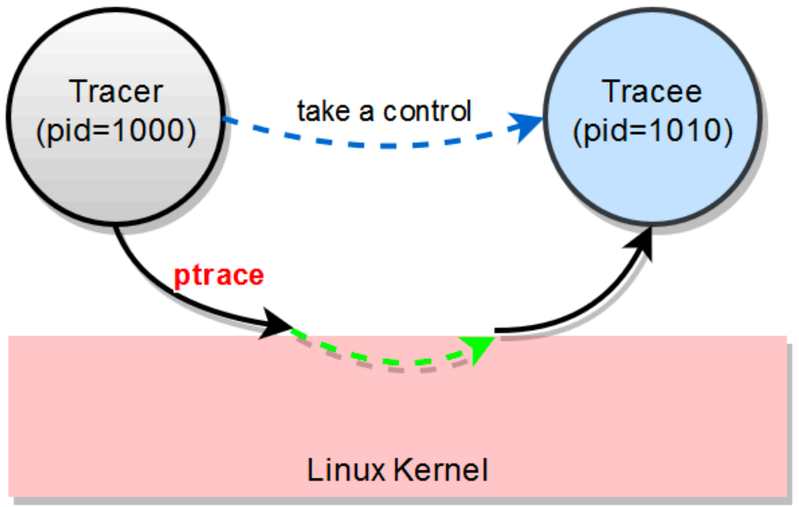
- Syscall (`sys_ptrace`)
  - Unix-like OS (Linux, macOS, ...)

`ptrace`(2)

- Syscall (`sys_ptrace`)
  - Unix-like OS (Linux, macOS, ...)
- "**p**rocess **trace**"

`ptrace`(2)

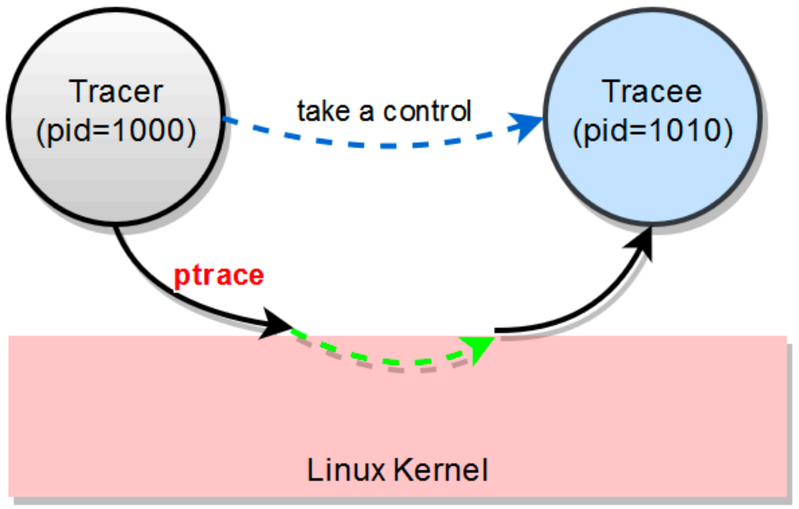
- Syscall (`sys_ptrace`)
  - Unix-like OS (Linux, macOS, ...)
- "**p**rocess **trace**"
  - Rollen: **Tracer** & **Tracee**



Quelle: fadeevab.com

`ptrace`(2)

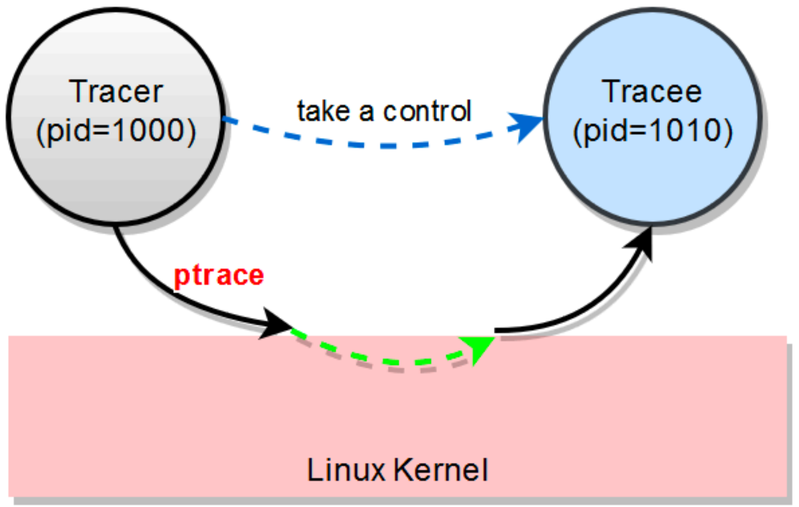
- Syscall (`sys_ptrace`)
  - Unix-like OS (Linux, macOS, ...)
- "**p**rocess **trace**"
  - Rollen: **Tracer** & **Tracee**
  - State **inspection** & modification



Quelle: fadeevab.com

`ptrace`(2)

- Syscall (`sys_ptrace`)
  - Unix-like OS (Linux, macOS, ...)
- "**p**rocess **trace**"
  - Rollen: **Tracer** & **Tracee**
  - State **inspection** & modification
- Use-cases: Breakpoint Debugging -> `gdb`(1) & **Syscall Tracing** -> `strace`(1)



Quelle: fadeevab.com

Syscall Tracing

Syscall Tracing

- Syscalls verwenden CPU Register:

%rax	System call	%rdi	%rsi	%rdx	%r10	%r8	%r9
0	sys_read	unsigned int fd	char *buf	size_t count			
1	sys_write	unsigned int fd	const char *buf	size_t count			
2	sys_open	const char *filename	int flags	int mode			
3	sys_close	unsigned int fd					
4	sys_stat	const char *filename	struct stat *statbuf				
5	sys_fstat	unsigned int fd	struct stat *statbuf				
6	sys_lstat	fconst char *filename	struct stat *statbuf				

Quelle: blog.rchapman.org

Syscall Tracing

- Syscalls verwenden CPU Register:

%rax	System call	%rdi	%rsi	%rdx	%r10	%r8	%r9
0	sys_read	unsigned int fd	char *buf	size_t count			
1	sys_write	unsigned int fd	const char *buf	size_t count			
2	sys_open	const char *filename	int flags	int mode			
3	sys_close	unsigned int fd					
4	sys_stat	const char *filename	struct stat *statbuf				
5	sys_fstat	unsigned int fd	struct stat *statbuf				
6	sys_lstat	fconst char *filename	struct stat *statbuf				

Quelle: blog.rchapman.org

- Für tracee auslesen: `PTRACE_GETREGSET`

ministrace

`strace`**(1)**

`strace`**(1)**

- `man`: "trace system calls and signals"

`strace`(1)

- `man`: "trace system calls and signals"
  - Verwendet `ptrace`(2)

`strace`(1)

- `man`: "trace system calls and signals"
  - Verwendet `ptrace`(2)

```
→  ~ strace sleep 5
execve("/usr/bin/sleep", ["sleep", "5"], 0xfffff1e5f048 /* 25 vars */) = 0
brk(NULL)                                   = 0xaaaad0698000
faccessat(AT_FDCWD, "/etc/ld.so.preload", R_OK) = -1 ENOENT (No such file or directory)
openat(AT_FDCWD, "/etc/ld.so.cache", O_RDONLY|O_CLOEXEC) = 3
fstat(3, {st_mode=S_IFREG|0644, st_size=29405, ...}) = 0
mmap(NULL, 29405, PROT_READ, MAP_PRIVATE, 3, 0) = 0xffff88deb000
close(3)                                    = 0
...
clock_nanosleep(CLOCK_REALTIME, 0, {tv_sec=5, tv_nsec=0},
```

`strace`(1)

- `man`: "trace system calls and signals"
  - Verwendet `ptrace`(2)

Syscall-Name (NR)

```
→  ~ strace sleep 5
execve("/usr/bin/sleep", ["sleep", "5"], 0xfffff1e5f048 /* 25 vars */) = 0
brk(NULL)                                   = 0xaaaad0698000
faccessat(AT_FDCWD, "/etc/ld.so.preload", R_OK) = -1 ENOENT (No such file or directory)
openat(AT_FDCWD, "/etc/ld.so.cache", O_RDONLY|O_CLOEXEC) = 3
fstat(3, {st_mode=S_IFREG|0644, st_size=29405, ...}) = 0
mmap(NULL, 29405, PROT_READ, MAP_PRIVATE, 3, 0) = 0xffff88deb000
close(3)                                    = 0
...
clock_nanosleep(CLOCK_REALTIME, 0, {tv_sec=5, tv_nsec=0},
```

`strace`(1)

- `man`: "trace system calls and signals"
  - Verwendet `ptrace`(2)

Syscall-Name (NR)

Syscall-ARGS

```
→  ~ strace sleep 5
execve("/usr/bin/sleep", ["sleep", "5"], 0xfffff1e5f048 /* 25 vars */) = 0
brk(NULL)                                   = 0xaaaad0698000
faccessat(AT_FDCWD, "/etc/ld.so.preload", R_OK) = -1 ENOENT (No such file or directory)
openat(AT_FDCWD, "/etc/ld.so.cache", O_RDONLY|O_CLOEXEC) = 3
fstat(3, {st_mode=S_IFREG|0644, st_size=29405, ...}) = 0
mmap(NULL, 29405, PROT_READ, MAP_PRIVATE, 3, 0) = 0xffff88deb000
close(3)                                    = 0
...
clock_nanosleep(CLOCK_REALTIME, 0, {tv_sec=5, tv_nsec=0},
```

`strace`(1)

- `man`: "trace system calls and signals"
  - Verwendet `ptrace`(2)

Syscall-Name (NR) Syscall-ARGS Syscall-RTN-VAL

```
→  ~ strace sleep 5
execve("/usr/bin/sleep", ["sleep", "5"], 0xfffff1e5f048 /* 25 vars */) = 0
brk(NULL)                                   = 0xaaaad0698000
faccessat(AT_FDCWD, "/etc/ld.so.preload", R_OK) = -1 ENOENT (No such file or directory)
openat(AT_FDCWD, "/etc/ld.so.cache", O_RDONLY|O_CLOEXEC) = 3
fstat(3, {st_mode=S_IFREG|0644, st_size=29405, ...}) = 0
mmap(NULL, 29405, PROT_READ, MAP_PRIVATE, 3, 0) = 0xffff88deb000
close(3)                                    = 0
...
clock_nanosleep(CLOCK_REALTIME, 0, {tv_sec=5, tv_nsec=0},
```

`strace`(1)

- `man`: "trace system calls and signals"
  - Verwendet `ptrace`(2)

Syscall-Name (NR) Syscall-ARGS Syscall-RTN-VAL

```
→  ~ strace sleep 5
execve("/usr/bin/sleep", ["sleep", "5"], 0xfffff1e5f048 /* 25 vars */) = 0
brk(NULL)                                   = 0xaaaad0698000
faccessat(AT_FDCWD, "/etc/ld.so.preload", R_OK) = -1 ENOENT (No such file or directory)
openat(AT_FDCWD, "/etc/ld.so.cache", O_RDONLY|O_CLOEXEC) = 3
fstat(3, {st_mode=S_IFREG|0644, st_size=29405, ...}) = 0
mmap(NULL, 29405, PROT_READ, MAP_PRIVATE, 3, 0) = 0xffff88deb000
close(3)                                    = 0
...
clock_nanosleep(CLOCK_REALTIME, 0, {tv_sec=5, tv_nsec=0},
```

- `ministrace` = Minimaler `strace` Nachbau

Syscall Parsing

Syscall Parsing

- Von Kernel Source:

Syscall Parsing

- Von Kernel Source:
  - **Nr**, **Name**: tbl file

```
# <number> <abi> <name> ...
0   common restart_syscall      sys_...
1   common exit          sys_exit
2   common fork          sys_fork
3   common read          sys_read
...
```

Syscall Parsing

- Von Kernel Source:
  - **Nr**, **Name**: tbl file

```
# <number> <abi> <name> ...
0   common restart_syscall      sys_...
1   common exit          sys_exit
2   common fork          sys_fork
3   common read          sys_read
...
```

Syscall Parsing

- Von Kernel Source:
  - **Nr**, **Name**: tbl file

```
# <number> <abi> <name> ...
0   common restart_syscall      sys_...
1   common exit          sys_exit
2   common fork          sys_fork
3   common read          sys_read
...
```

Syscall Parsing

- Von Kernel Source:
  - **Nr**, **Name**: tbl file
  - **Args**: Macros

```
# <number> <abi> <name> ...
0   common restart_syscall      sys_...
1   common exit          sys_exit
2   common fork          sys_fork
3   common read          sys_read
...
```

```
SYSCALL_DEFINE1(exit, int, error_code)
```

Syscall Parsing

- Von Kernel Source:
  - **Nr**, **Name**: tbl file
  - **Args**: Macros

```
# <number> <abi> <name> ...
0   common restart_syscall      sys_...
1   common exit          sys_exit
2   common fork          sys_fork
3   common read          sys_read
...
```

```
SYSCALL_DEFINE1(exit, int, error_code)
```

Syscall Parsing

- Von Kernel Source:
  - **Nr**, **Name**: tbl file
  - **Args**: Macros
- Parsing script:

```
# <number> <abi> <name> ...
0   common restart_syscall      sys_...
1   common exit          sys_exit
2   common fork          sys_fork
3   common read          sys_read
...
```

```
SYSCALL_DEFINE1(exit, int, error_code)
```

Syscall Parsing

- Von Kernel Source:
  - **Nr**, **Name**: tbl file
  - **Args**: Macros
- Parsing script:
  - Via RegEx

```
# <number> <abi> <name> ...
0   common restart_syscall      sys_...
1   common exit          sys_exit
2   common fork          sys_fork
3   common read          sys_read
...
```

```
SYSCALL_DEFINE1(exit, int, error_code)
```

Syscall Parsing

- Von Kernel Source:
  - **Nr**, **Name**: tbl file
  - **Args**: Macros
- Parsing script:
  - Via RegEx
  - Output: Syscall table

```
# <number> <abi> <name> ...
0   common restart_syscall      sys_...
1   common exit          sys_exit
2   common fork          sys_fork
3   common read          sys_read
...
```

```
SYSCALL_DEFINE1(exit, int, error_code)
```

Syscall Parsing

- Von Kernel Source:
  - **Nr**, **Name**: tbl file
  - **Args**: Macros
- Parsing script:
  - Via RegEx
  - Output: Syscall table
    - Platform specific

```
# <number> <abi> <name> ...
0   common restart_syscall      sys_...
1   common exit          sys_exit
2   common fork          sys_fork
3   common read          sys_read
...
```

```
SYSCALL_DEFINE1(exit, int, error_code)
```

Syscall Parsing

- Von Kernel Source:
  - **Nr**, **Name**: tbl file
  - **Args**: Macros
- Parsing script:
  - Via RegEx
  - Output: Syscall table
    - Platform specific

```
# <number> <abi> <name> ...
0   common restart_syscall      sys_...
1   common exit          sys_exit
2   common fork          sys_fork
3   common read          sys_read
...
```

```
SYSCALL_DEFINE1(exit, int, error_code)
```

```
// kernel version 5.4.0-97-generic on aarch64

const syscall_entry syscalls[] = {
  [__SNR_restart_syscall] = {
    .name  = "restart_syscall",
    .nargs = 0,
    .args  = {-1, -1, -1, -1, -1, -1}},
  [__SNR_exit] = {
    .name  = "exit",
    .nargs = 1,
    .args  = {ARG_INT, -1, -1, -1, -1, -1}},
```

Syscall Parsing

- Von Kernel Source:
  - **Nr**, **Name**: tbl file
  - **Args**: Macros
- Parsing script:
  - Via RegEx
  - Output: Syscall table
    - Platform specific

```
# <number> <abi> <name> ...
0   common restart_syscall      sys_...
1   common exit                 sys_exit
2   common fork                 sys_fork
3   common read                 sys_read
...

SYSCALL_DEFINE1(exit, int, error_code)
```

```
// kernel version 5.4.0-97-generic on aarch64

const syscall_entry syscalls[] = {
  [__SNR_restart_syscall] = {
    .name  = "restart_syscall",
    .nargs = 0,
    .args  = {-1, -1, -1, -1, -1, -1}},
  [__SNR_exit] = {
    .name  = "exit",
    .nargs = 1,
    .args  = {ARG_INT, -1, -1, -1, -1, -1}},
```

Syscall Parsing

- Von Kernel Source:
  - **Nr**, **Name**: tbl file
  - **Args**: Macros
- Parsing script:
  - Via RegEx
  - Output: Syscall table
    - Platform specific

```
# <number> <abi> <name> ...
0   common restart_syscall      sys_...
1   common exit                 sys_exit
2   common fork                 sys_fork
3   common read                 sys_read
...

SYSCALL_DEFINE1(exit, int, error_code)
```

```
// kernel version 5.4.0-97-generic on aarch64

const syscall_entry syscalls[] = {
  [__SNR_restart_syscall] = {
    .name  = "restart_syscall",
    .nargs = 0,
    .args  = {-1, -1, -1, -1, -1, -1}},
  [__SNR_exit] = {
    .name  = "exit",
    .nargs = 1,
    .args  = {ARG_INT, -1, -1, -1, -1, -1}},
```

Syscall Parsing

- Von Kernel Source:
  - **Nr**, **Name**: tbl file
  - **Args**: Macros
- Parsing script:
  - Via RegEx
  - Output: Syscall table
    - Platform specific

```
# <number> <abi> <name> ...
0   common restart_syscall      sys_...
1   common exit                 sys_exit
2   common fork                 sys_fork
3   common read                 sys_read
...

SYSCALL_DEFINE1(exit, int, error_code)
```

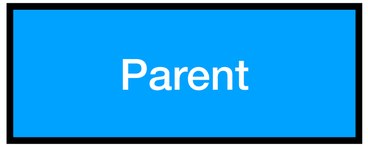
```
// kernel version 5.4.0-97-generic on aarch64

const syscall_entry syscalls[] = {
  [__SNR_restart_syscall] = {
    .name  = "restart_syscall",
    .nargs = 0,
    .args  = {-1, -1, -1, -1, -1, -1}},
  [__SNR_exit] = {
    .name  = "exit",
    .nargs = 1,
    .args  = {ARG_INT, -1, -1, -1, -1, -1}},
```

Tracing - Rollen

- "`-D` Run tracer [...] as a **grandchild**, not as parent"

Tracing - Rollen

- "`-D` Run tracer [...] as a **grandchild**, not as parent"

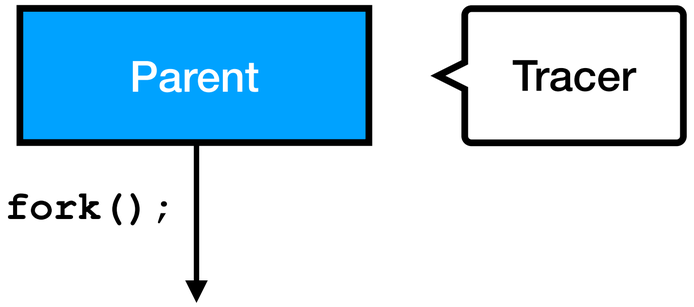


Tracing - Rollen

- "`-D` Run tracer [...] as a **grandchild**, not as parent"

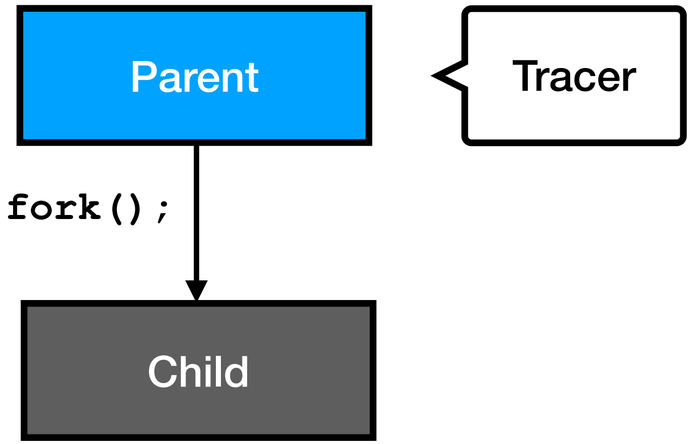
Tracing - Rollen

- "`-D` Run tracer [...] as a **grandchild**, not as parent"



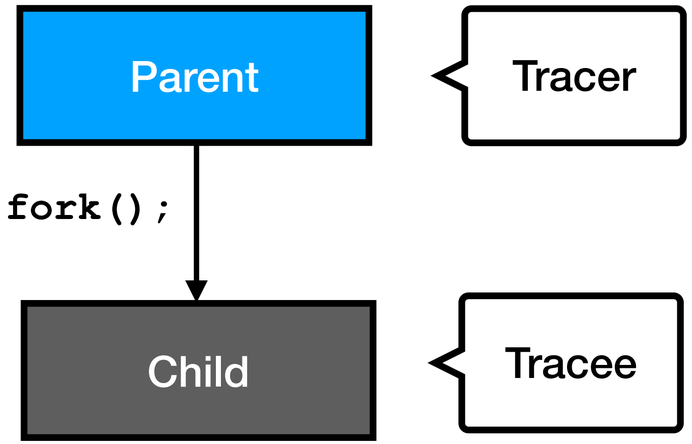
Tracing - Rollen

- "`-D` Run tracer [...] as a **grandchild**, not as parent"



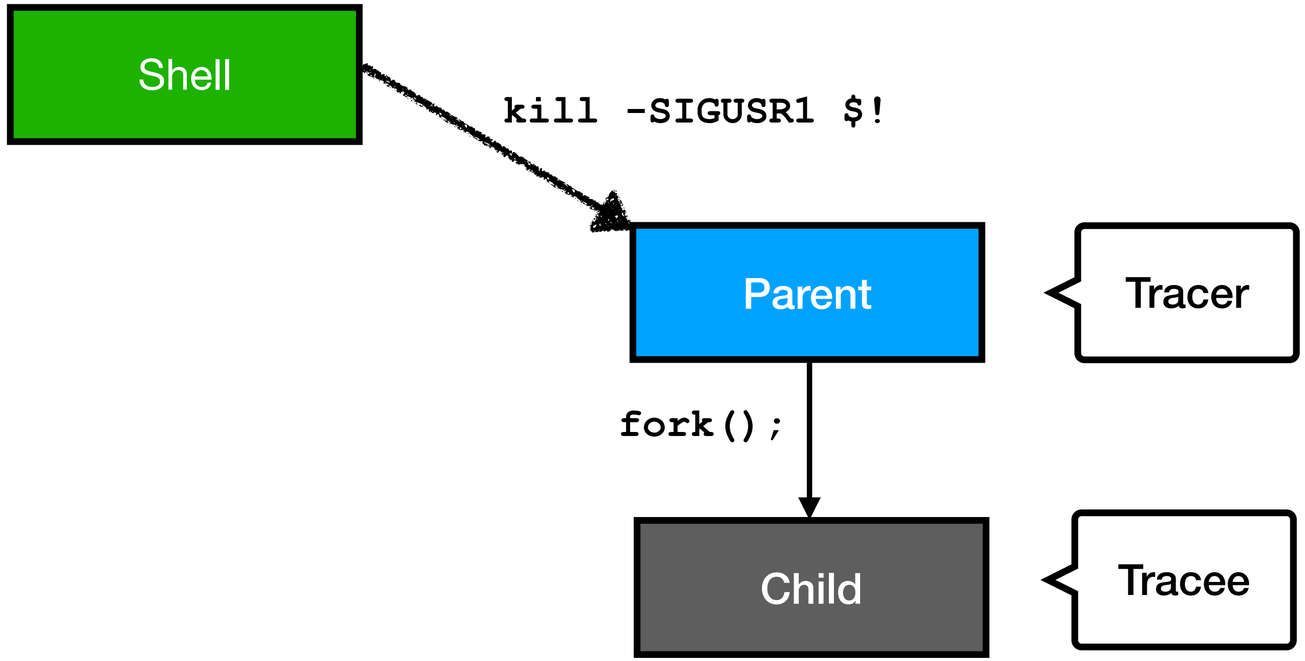
Tracing - Rollen

- "`-D` Run tracer [...] as a **grandchild**, not as parent"



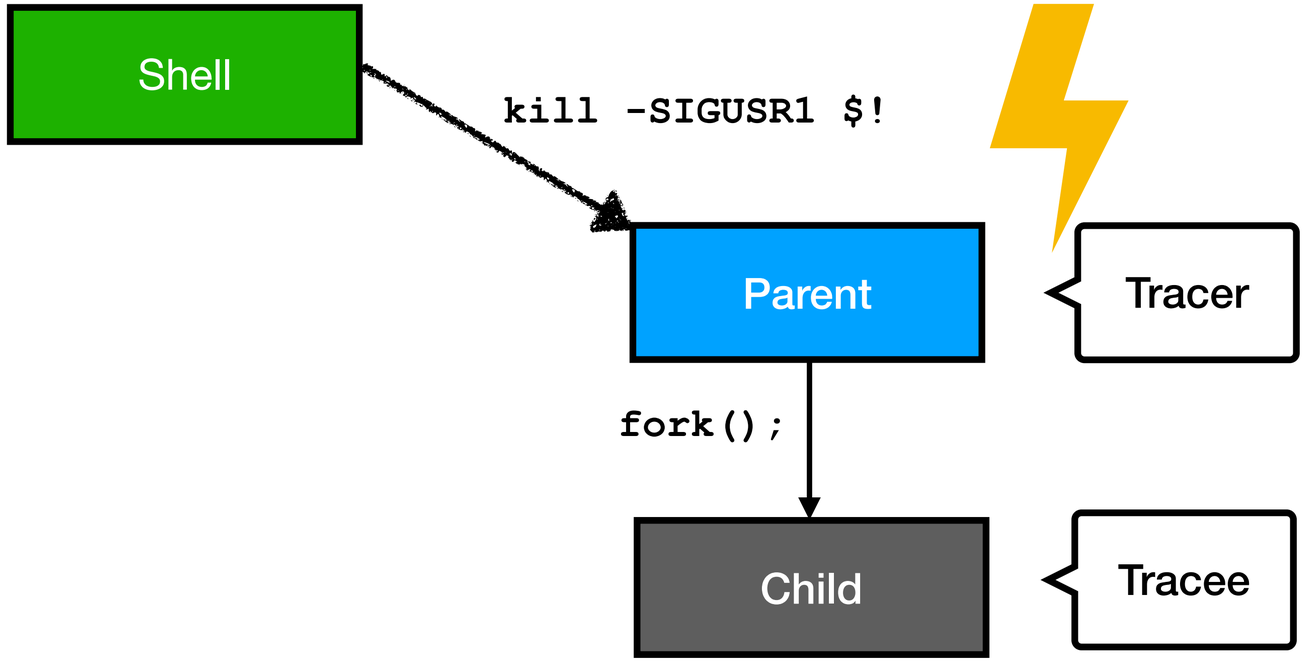
Tracing - Rollen

- "`-D` Run tracer [...] as a **grandchild**, not as parent"



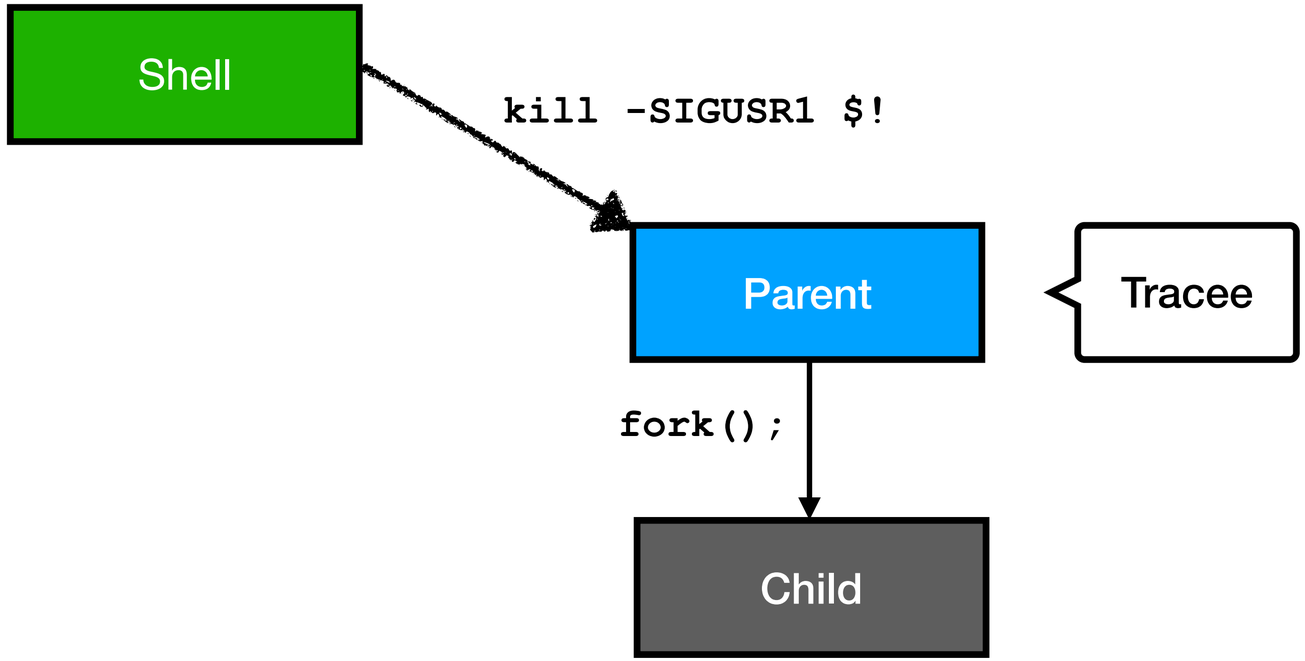
Tracing - Rollen

- "`-D` Run tracer [...] as a **grandchild**, not as parent"



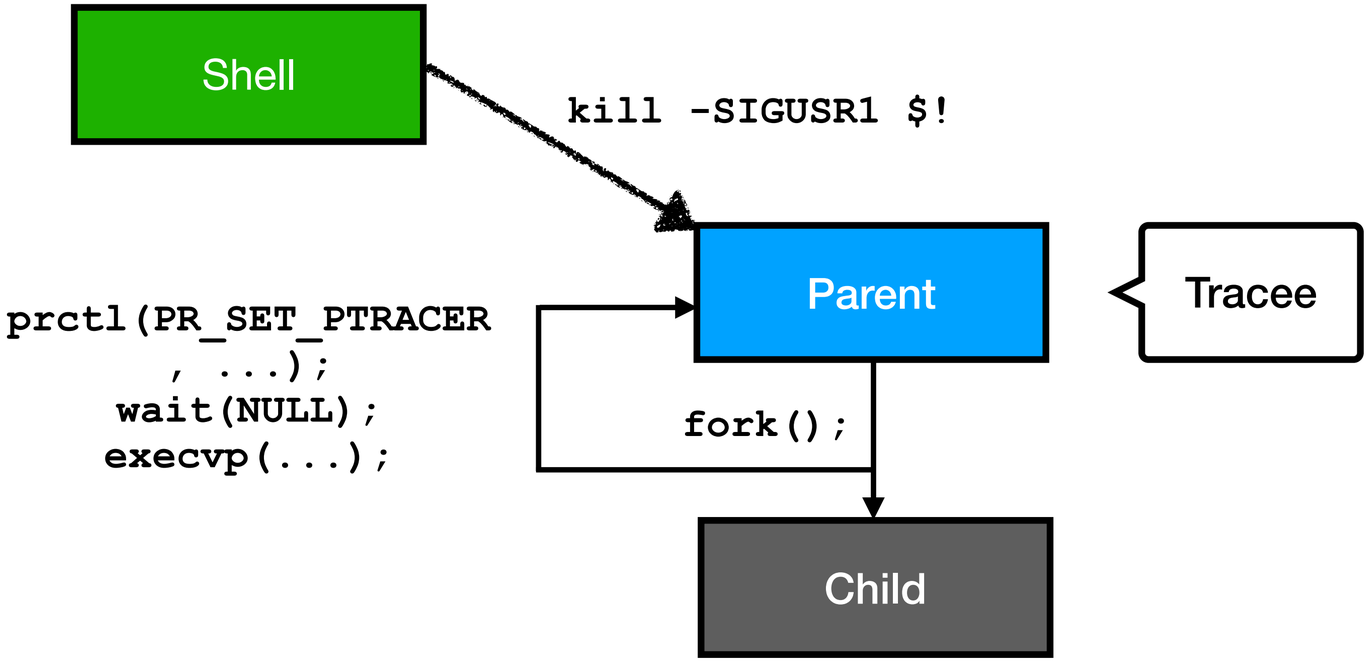
Tracing - Rollen

- "`-D` Run tracer [...] as a **grandchild**, not as parent"



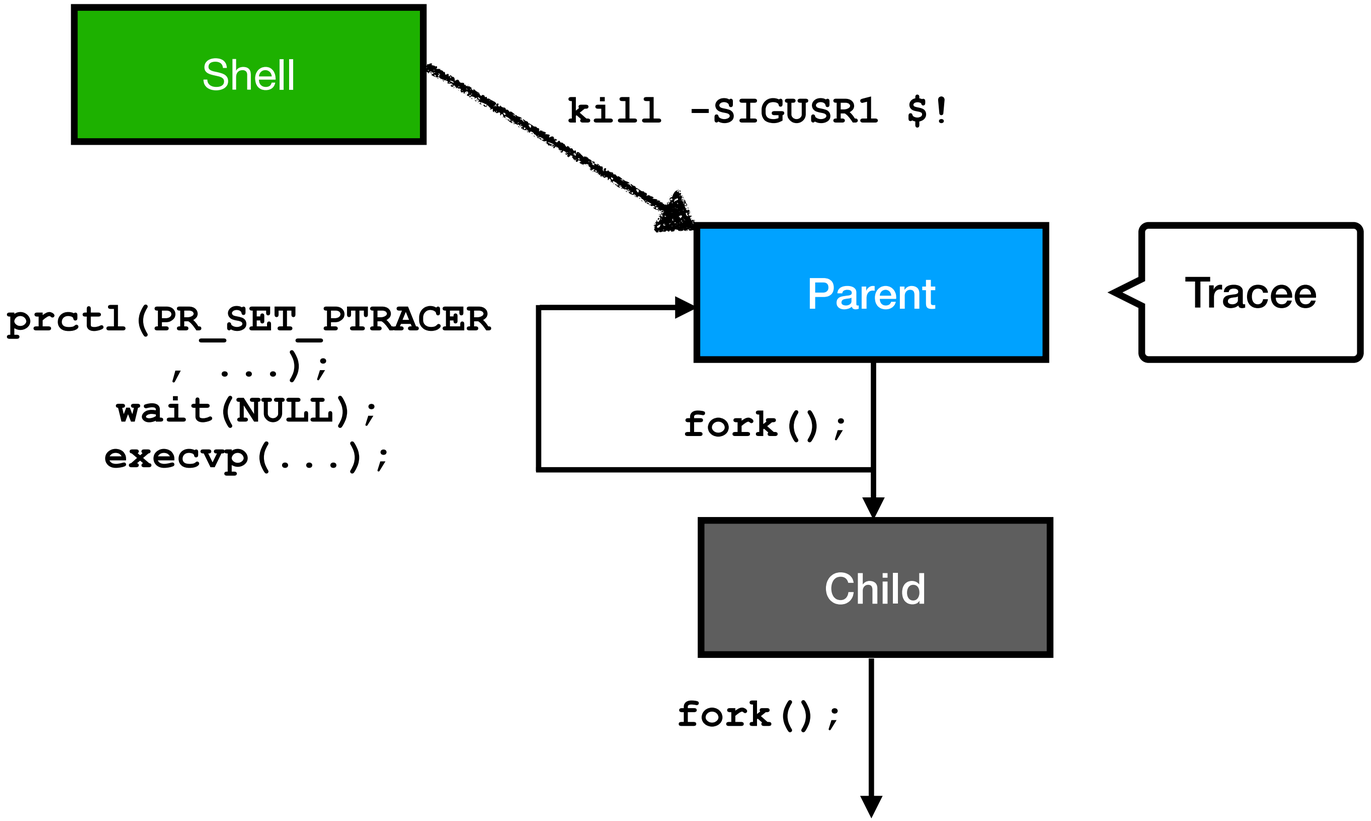
Tracing - Rollen

- "`-D` Run tracer [...] as a **grandchild**, not as parent"



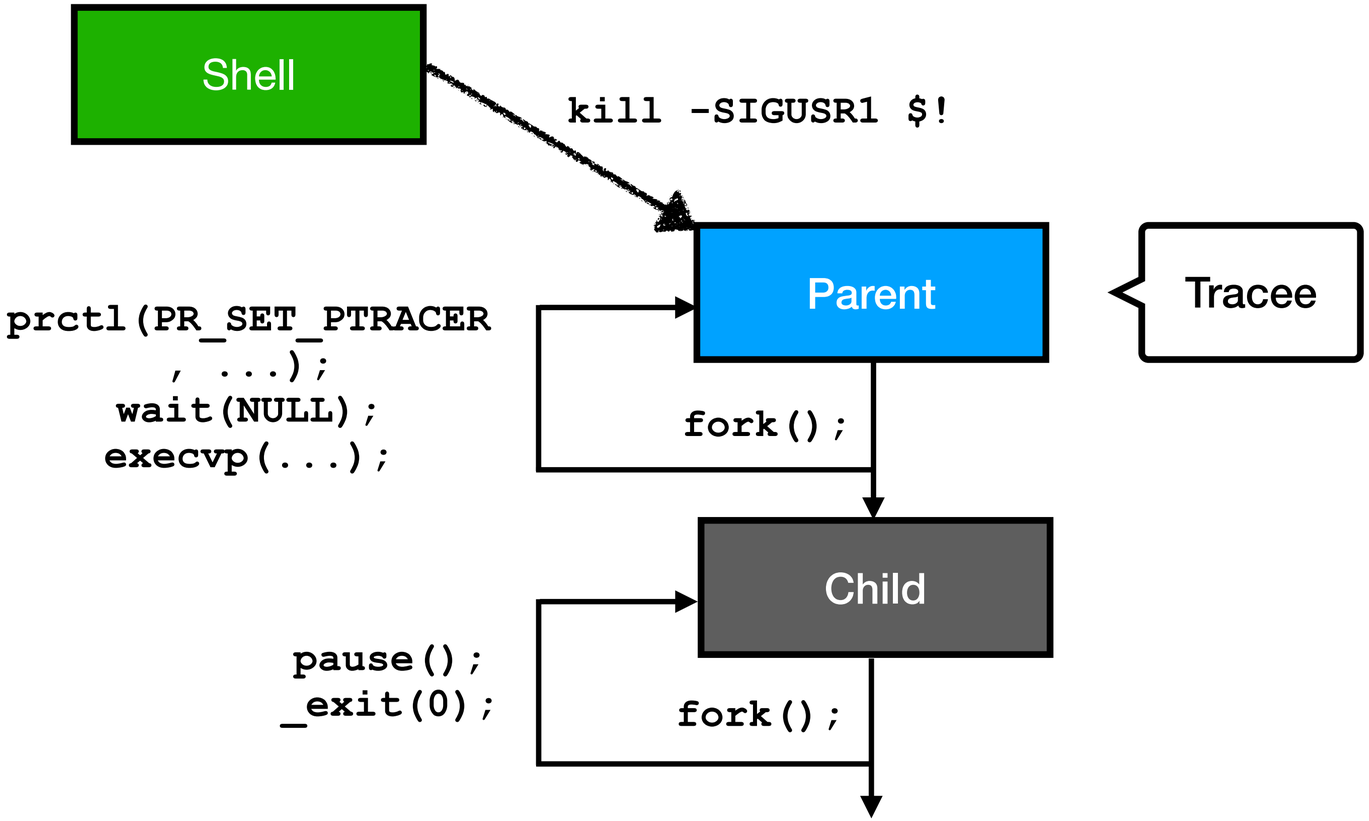
Tracing - Rollen

- "`-D` Run tracer [...] as a **grandchild**, not as parent"



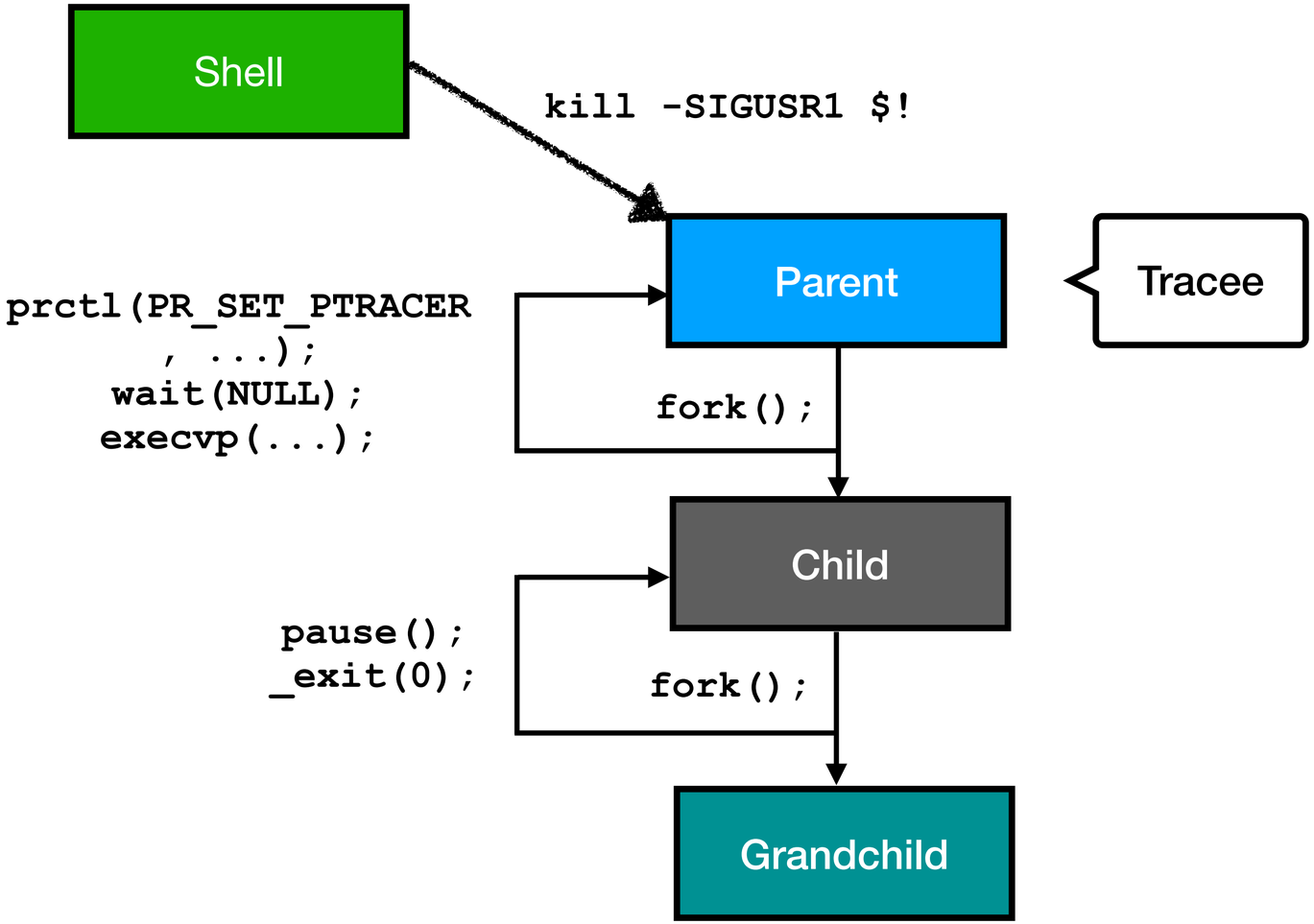
Tracing - Rollen

- "`-D` Run tracer [...] as a **grandchild**, not as parent"



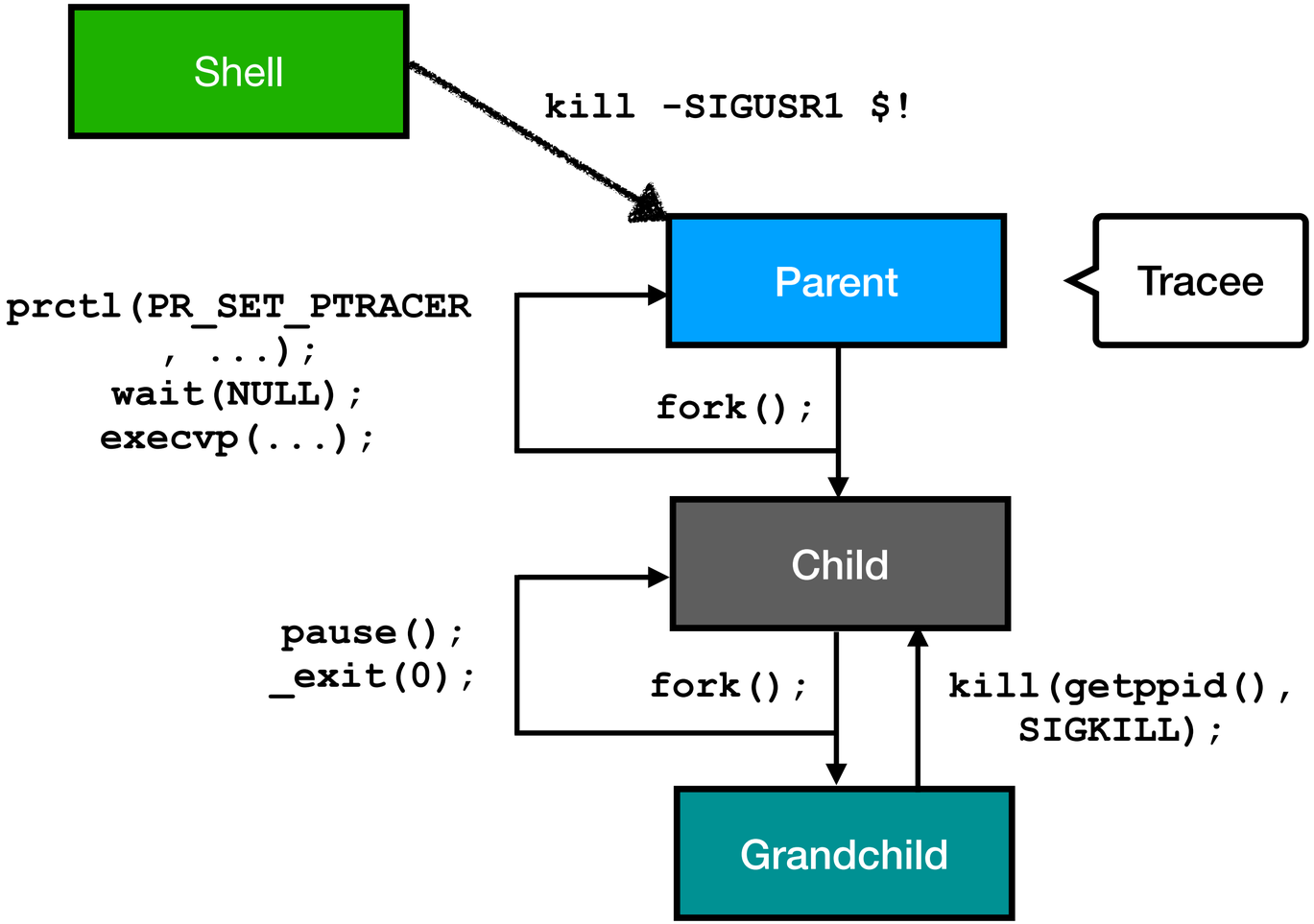
Tracing - Rollen

- "`-D` Run tracer [...] as a **grandchild**, not as parent"



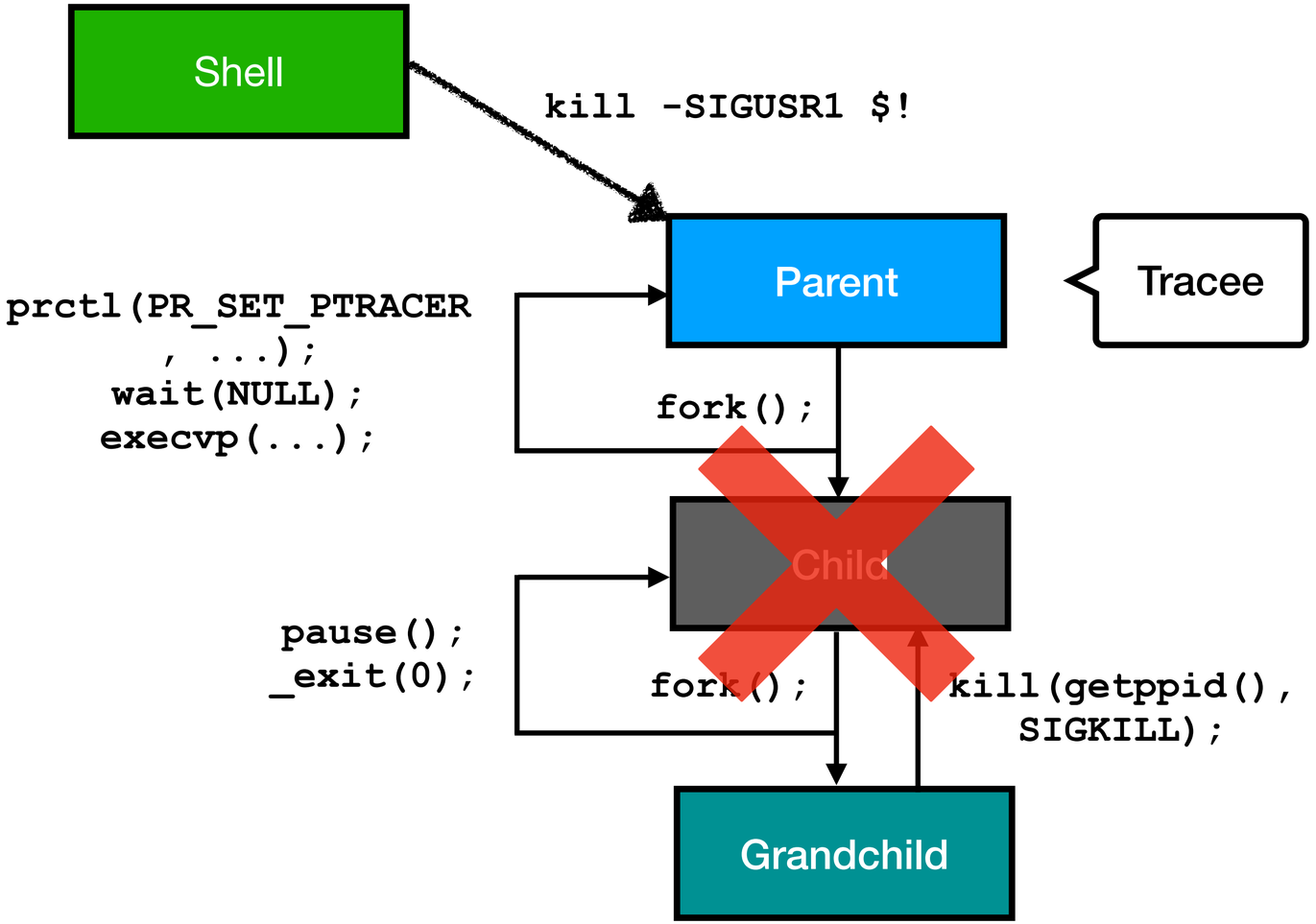
Tracing - Rollen

- "`-D` Run tracer [...] as a **grandchild**, not as parent"



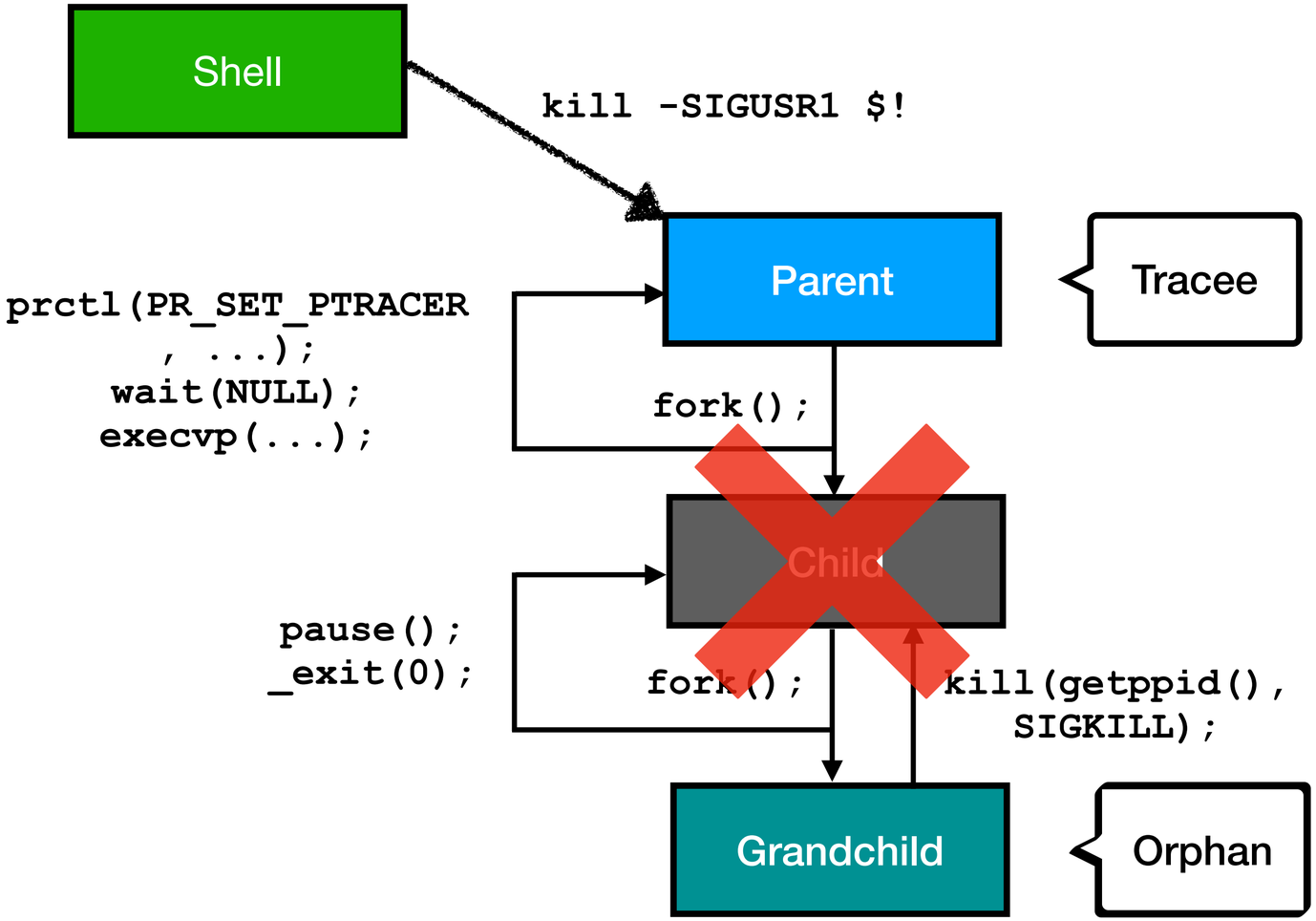
Tracing - Rollen

- "`-D` Run tracer [...] as a **grandchild**, not as parent"



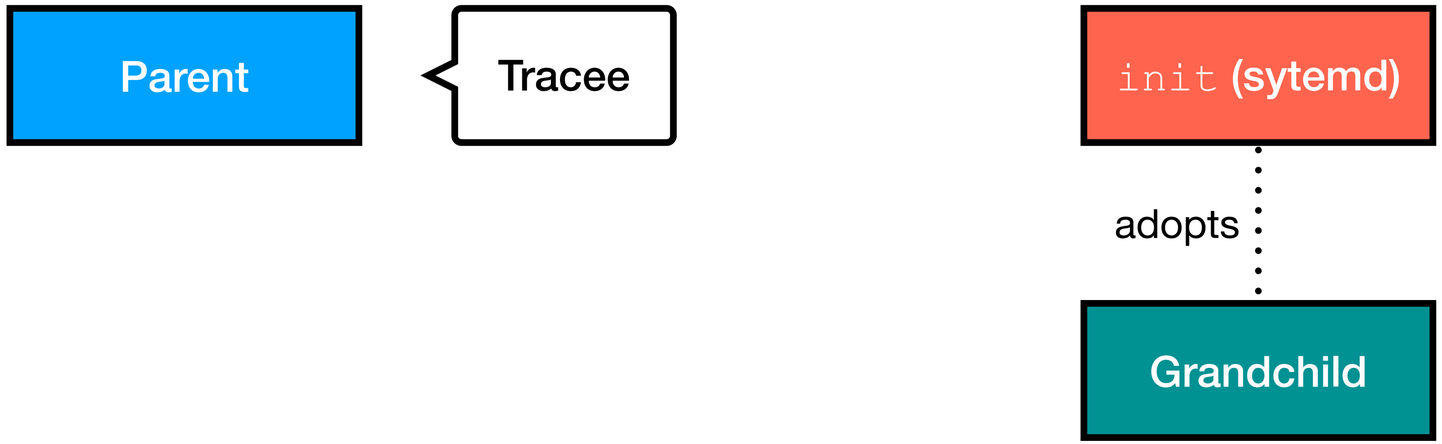
Tracing - Rollen

- "`-D` Run tracer [...] as a **grandchild**, not as parent"



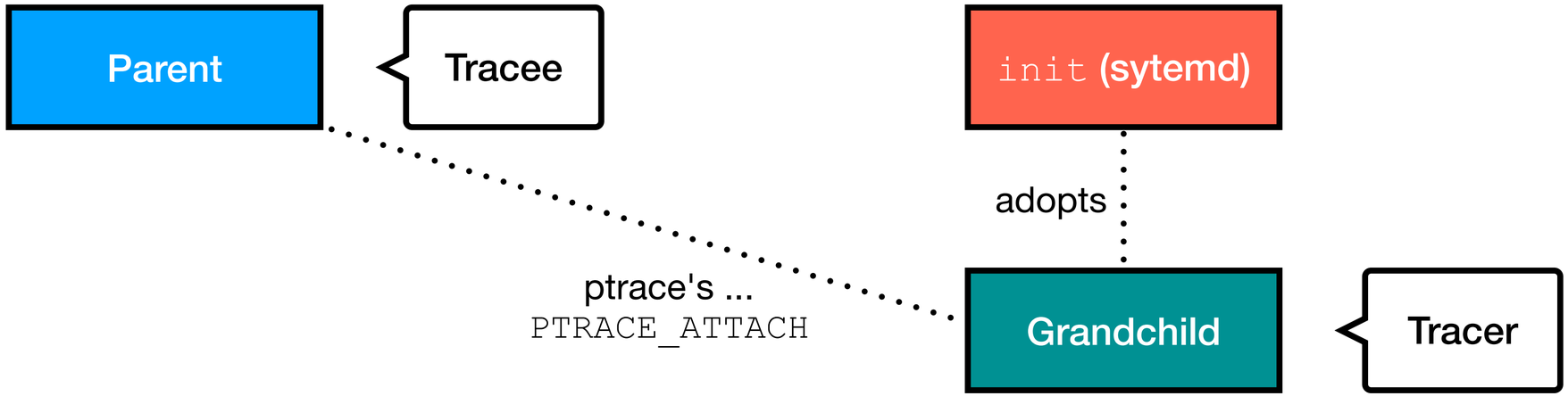
Tracing - Rollen

- "`-D` Run tracer [...] as a **grandchild**, not as parent"



Tracing - Rollen

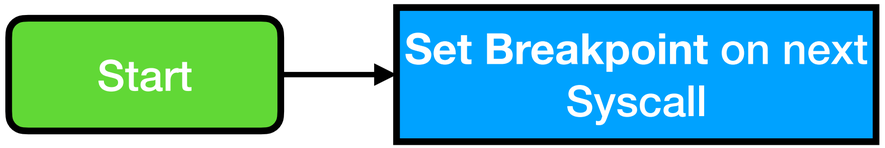
- "`-D` Run tracer [...] as a **grandchild**, not as parent"



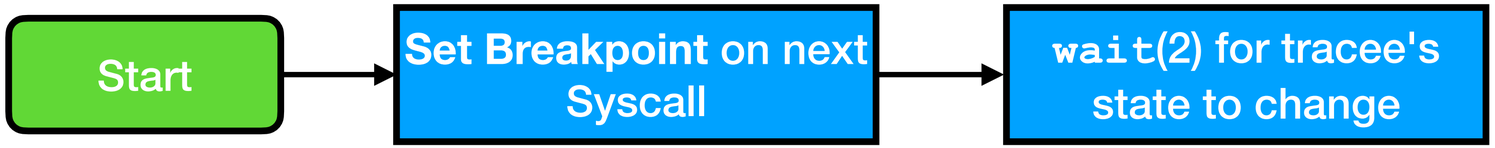
Tracing - Flow

Tracing - Flow

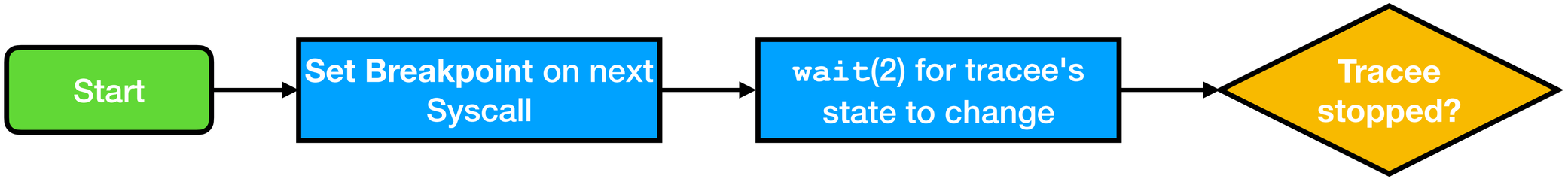
Tracing - Flow



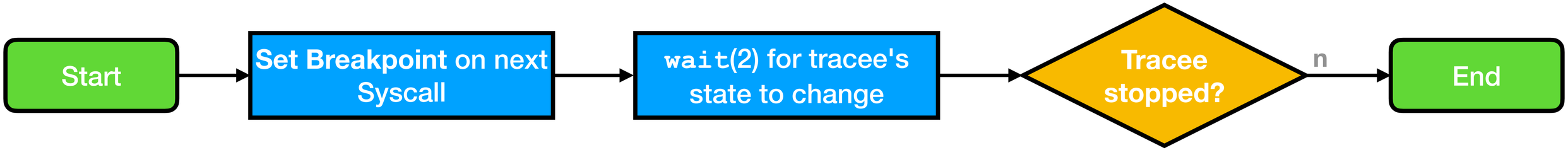
Tracing - Flow



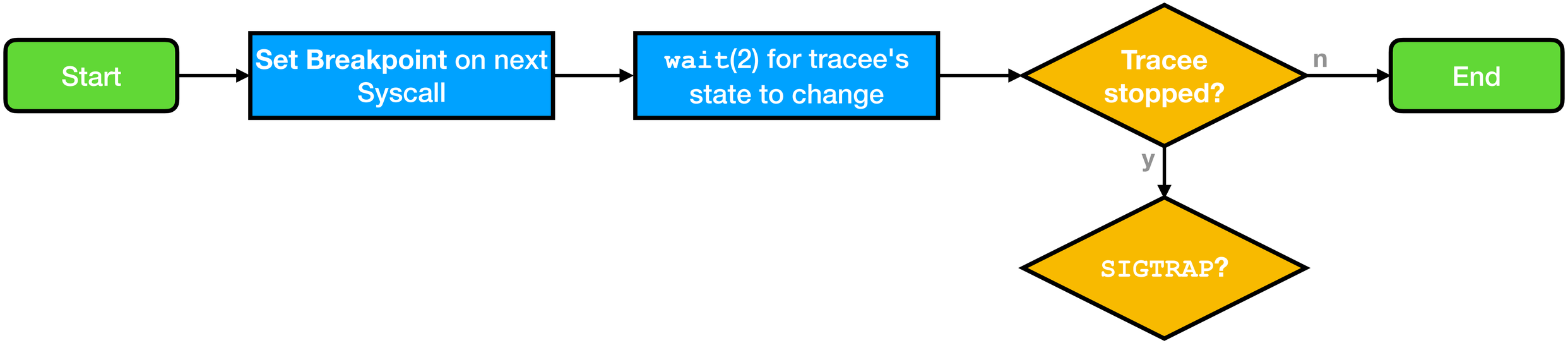
Tracing - Flow



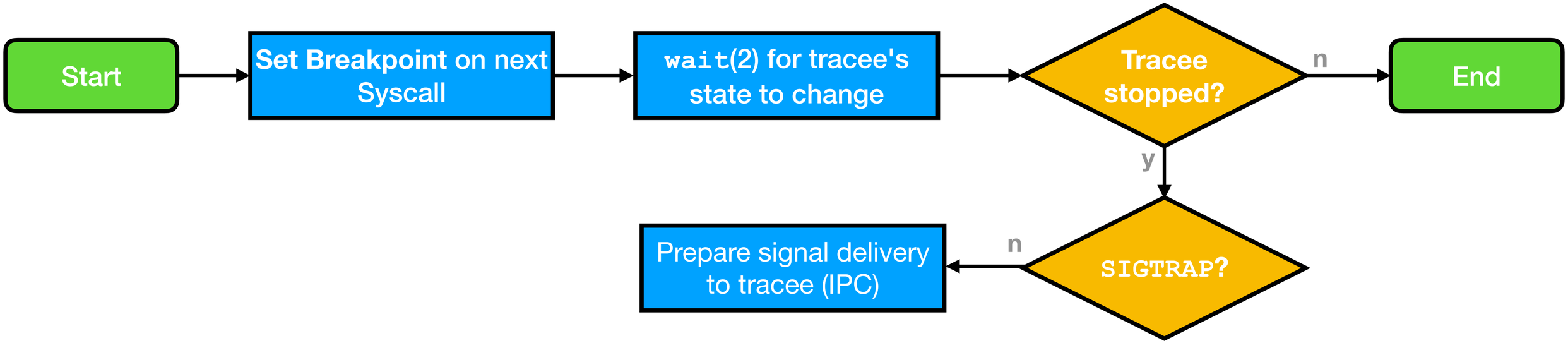
Tracing - Flow



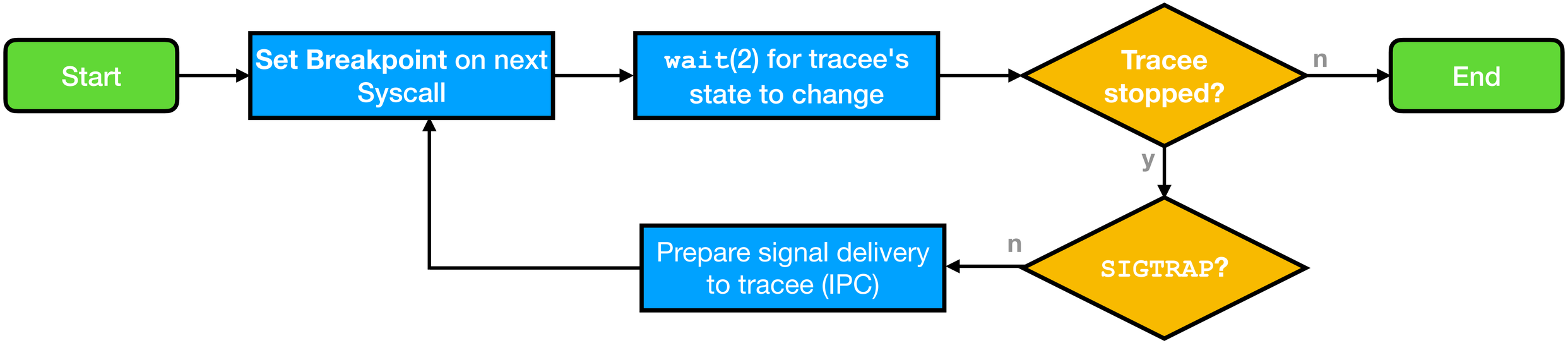
Tracing - Flow



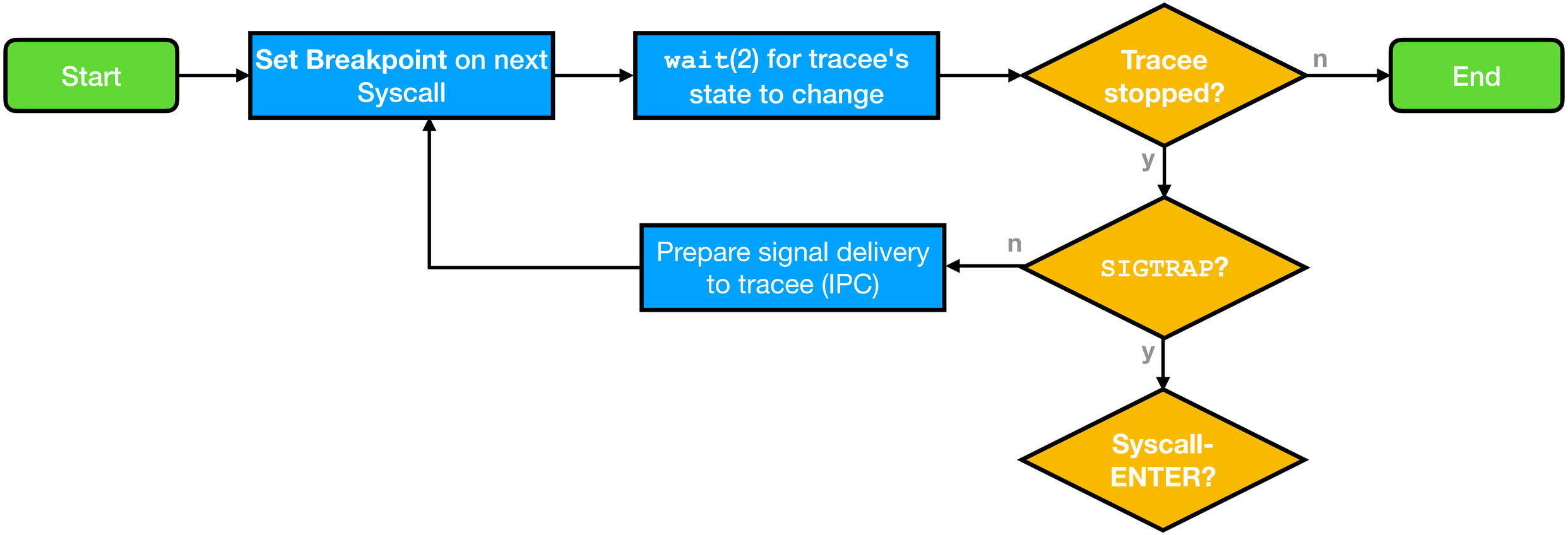
Tracing - Flow



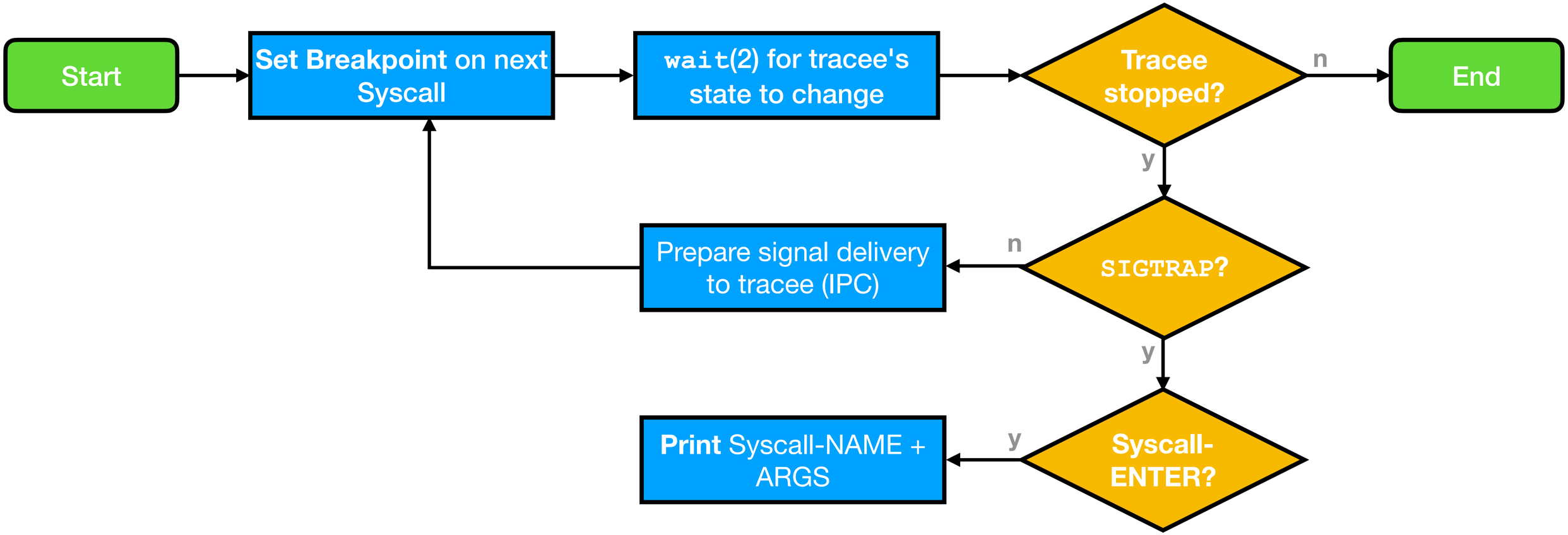
Tracing - Flow



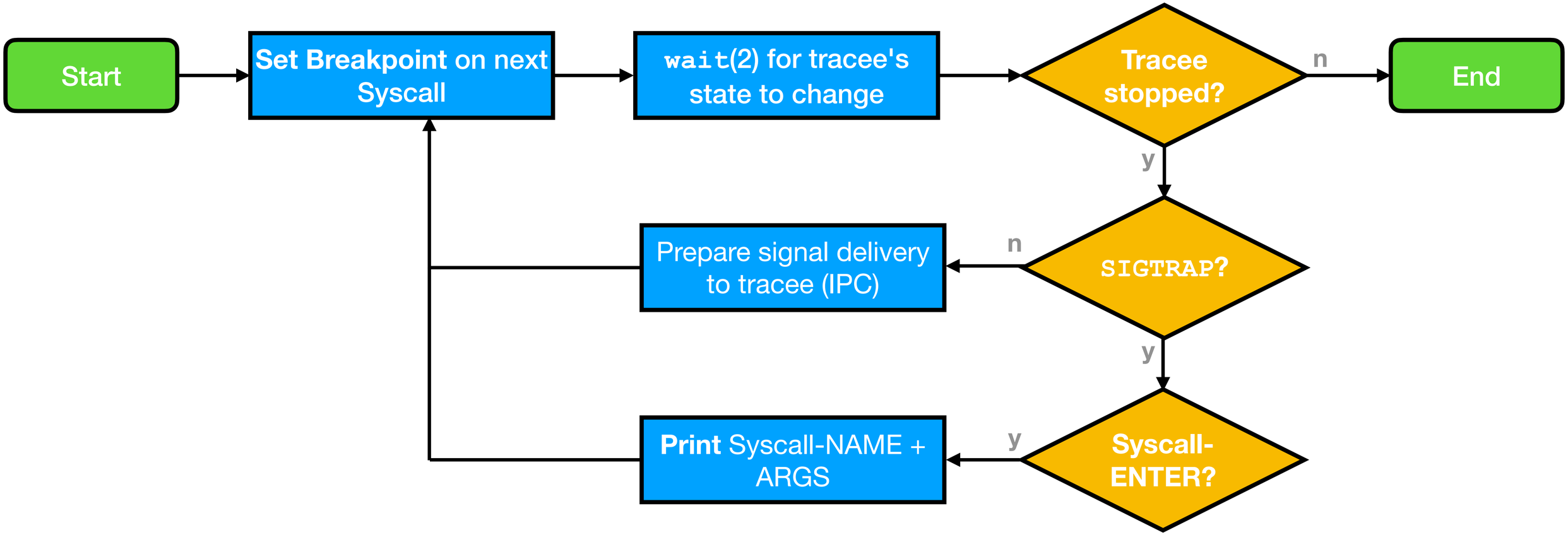
Tracing - Flow



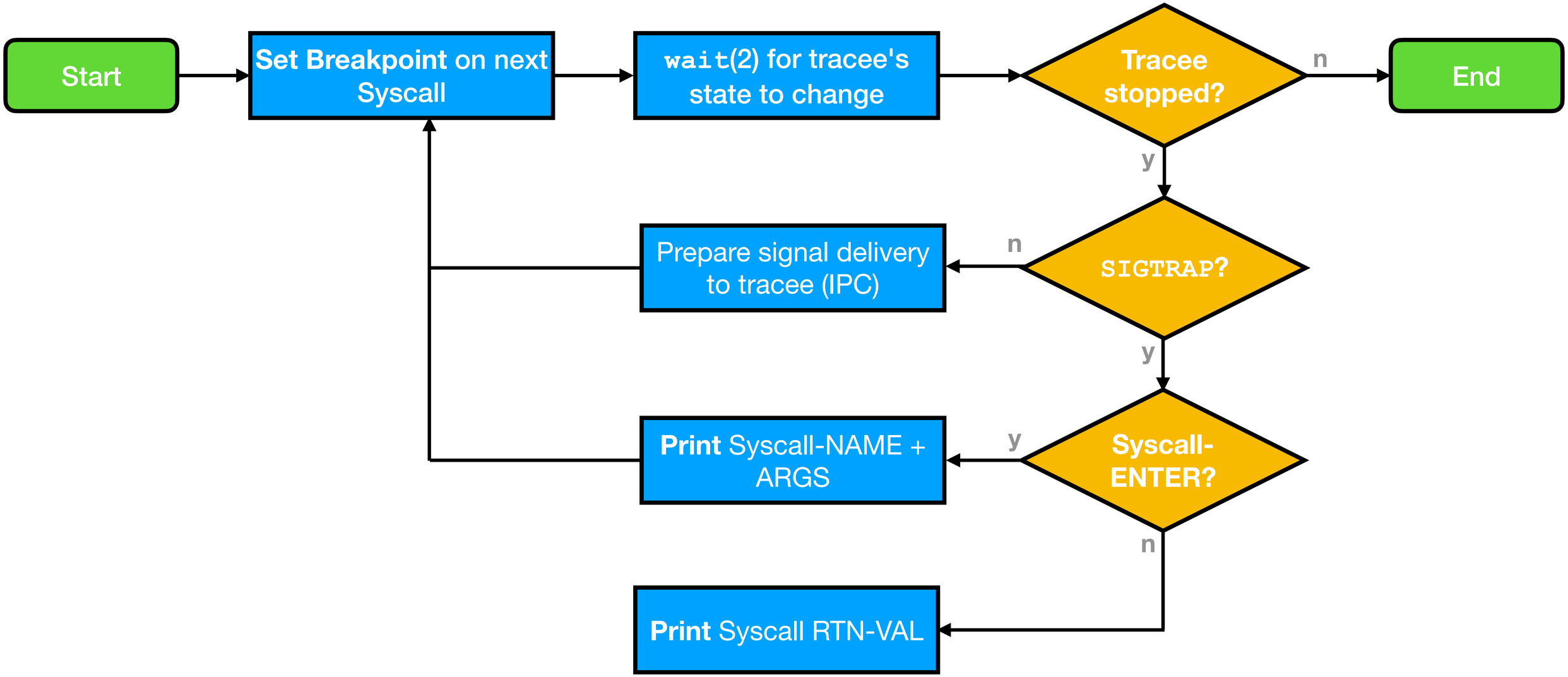
Tracing - Flow



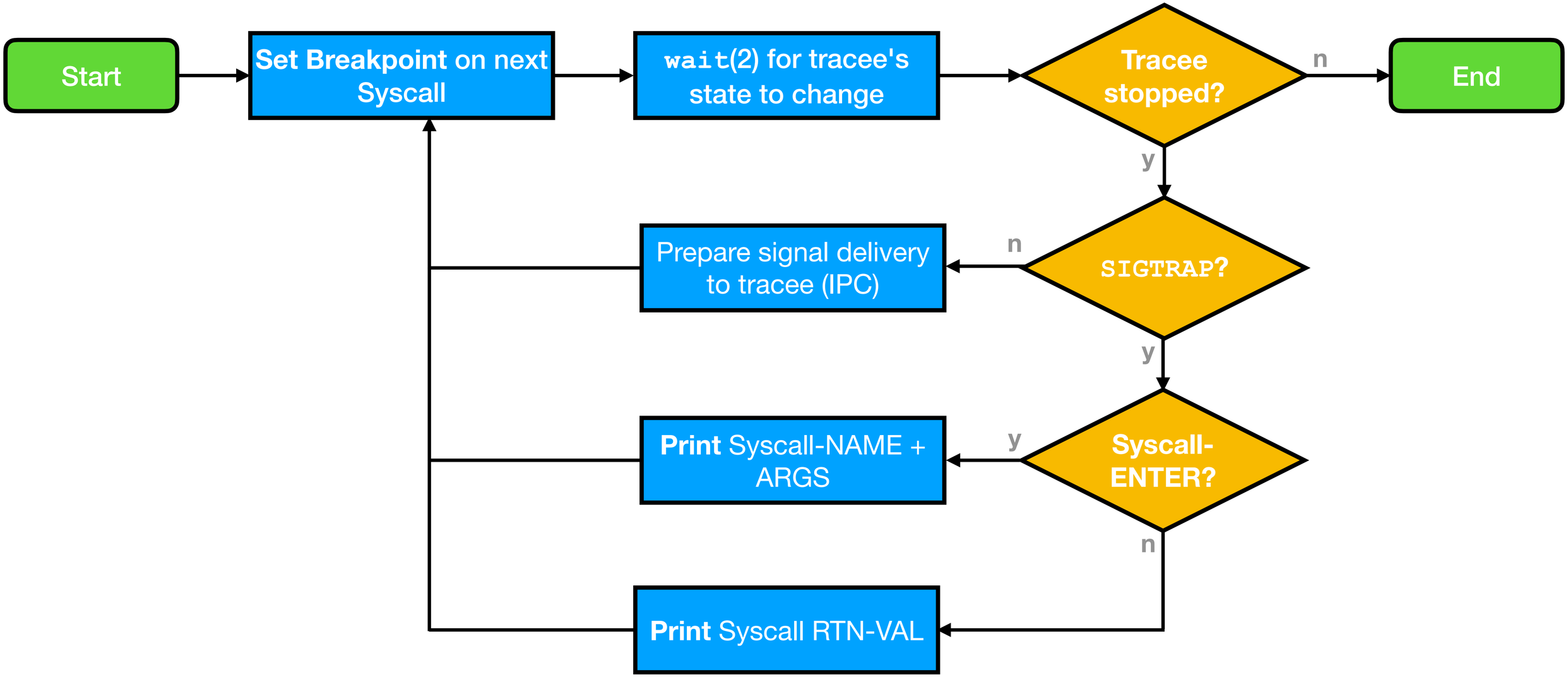
Tracing - Flow



Tracing - Flow



Tracing - Flow



Probleme

Probleme

- **aarch64 Support** -> Falsche Syscall-NR

Probleme

- **aarch64 Support** -> Falsche Syscall-NR
- **Tracing subset** Syscalls

Probleme

- **aarch64 Support** -> Falsche Syscall-NR
- **Tracing subset** Syscalls
  - Workaround: `-e dup2,execve` -> Lookup table

Probleme

- **aarch64 Support** -> Falsche Syscall-NR
- **Tracing subset** Syscalls
  - Workaround: `-e dup2,execve` -> Lookup table
- **Performance**

Probleme

- **aarch64 Support** -> Falsche Syscall-NR
- **Tracing subset** Syscalls
  - Workaround: `-e dup2,execve` -> Lookup table
- **Performance**
- **OR'ed Flags**

```
open("test.txt", O_WRONLY | O_APPEND);
```

Probleme

- **aarch64 Support** -> Falsche Syscall-NR
- **Tracing subset** Syscalls
  - Workaround: `-e dup2,execve` -> Lookup table
- **Performance**
- **OR'ed Flags**

```
open("test.txt", O_WRONLY | O_APPEND);
```

1025

Todo: Integration

Ansatz

Ansatz

- **Shared VM** (Thread-like)

Ansatz

- **Shared VM** (Thread-like)
  - Problem: Buffer Mutex => Deadlock

Ansatz

- **Shared VM** (Thread-like)
  - Problem: Buffer Mutex => Deadlock
  - Idee: Atomare Queue

Quellen

1. **How debuggers work: Part 2 - Breakpoints - Eli Bendersky's website**. (2011). Retrieved 6 February 2022, from https://eli.thegreenplace.net/2011/01/27/how-debuggers-work-part-2-breakpoints

2. **ptrace(2) - Linux manual page**. (2022). Retrieved 6 February 2022, from https://man7.org/linux/man-pages/man2/ptrace.2.html

3. **Playing with ptrace, Part I | Linux Journal**. (2002). Retrieved 6 February 2022, from https://www.linuxjournal.com/article/6100?page=0,1

4. **GitHub - nelhage/ministrace: A minimal toy implementation of strace(1)**. (2022). Retrieved 6 February 2022, from https://github.com/nelhage/ministrace

5. Valsorda, F. (2022). **Searchable Linux Syscall Table for x86 and x86_64** | PyTux. Retrieved 6 February 2022, from https://filippo.io/linux-syscall-table/