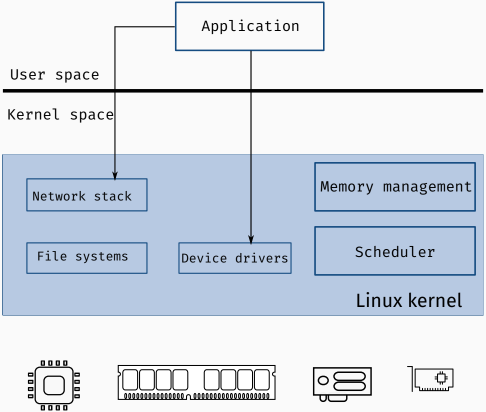
LXDs: Towards Isolation of Kernel Subsystems

Vikram Narayanan[1], Abhiram Balasubramanian[2], Charlie Jacobsen[2], Sarah Spall[2], Scott Bauer[2], Michael Quigley[2], Aftab Hussain[1], Abdullah Younis[1], Junjie Shen[1], Moinak Bhattacharyya[1], Anton Burtsev[1]

[1]University of California, Irvine

[2]University of Utah

Commodity kernels



Application

User space

Kernel space

Network stack

Memory management

File systems

Device drivers

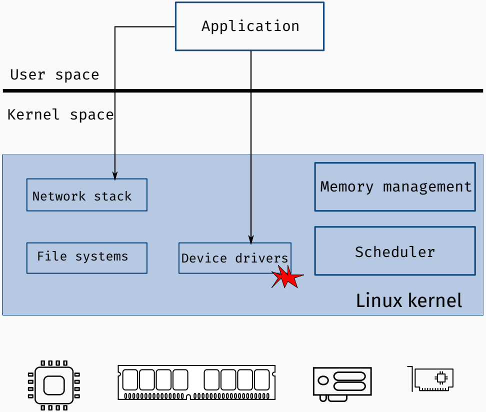
Scheduler

Linux kernel

Commodity kernels are monolithic

- Kernel extensions (fs, network stacks, drivers) run in the same address space

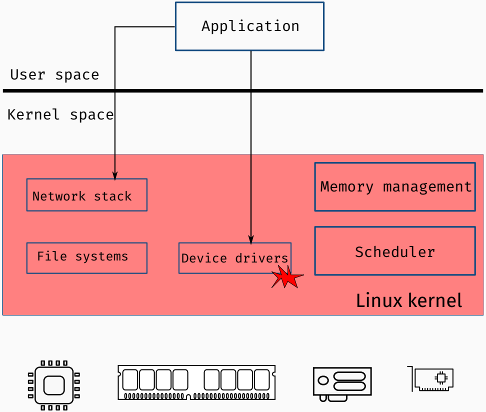
Commodity kernels



Commodity kernels are monolithic

- Kernel extensions (fs, network stacks, drivers) run in the same address space
- Vulnerability in a single component propagates to the entire kernel

Commodity kernels



Application

User space

Kernel space

Network stack

Memory management

File systems

Device drivers

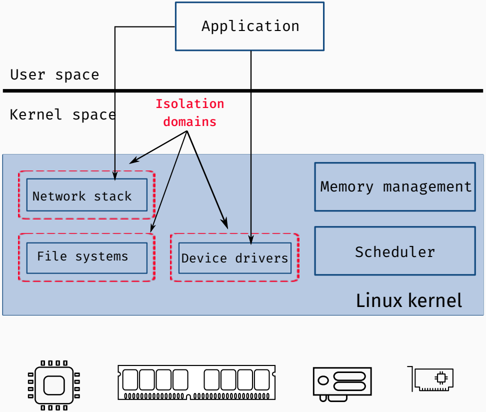
Scheduler

Linux kernel

Commodity kernels are monolithic

- Kernel extensions (fs, network stacks, drivers) run in the same address space
- Vulnerability in a single component propagates to the entire kernel

Isolation



Application

User space

Kernel space

Isolation domains

Network stack

File systems

Device drivers

Memory management

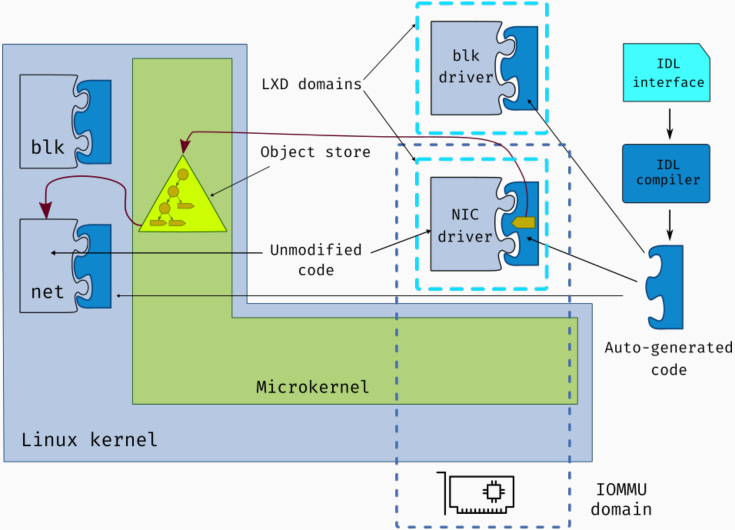
Scheduler

Linux kernel

Split monolithic kernel into isolated components

- to confine faults
- to improve reliability

Lightweight eXecution Domains: Architecture



LXDs:Design goals

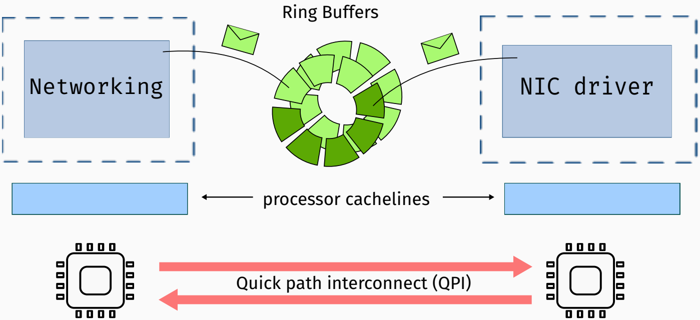
- backward compatibility with unmodified code

LXDs:Design goals

- backward compatibility with unmodified code
- transparent object synchronization across domain boundaries

LXDs Framework : fIPC

- fast inter process communication (fIPC)



LXDs Framework : IDL

```
int register_netdev(struct net_device *dev);

/* Projections */
projection <struct net_device> net_device {
  ...
  /* [modifier] <data_type> <struct_member_name> */;
  [in] unsigned int flags;
  [in] unsigned long long hw_features;
  [in] unsigned long long features;
  ...
  projection net_device_ops [alloc(caller)] *netdev_ops;
};
```

- Interface definition language

LXDs Framework : IDL

```
int register_netdev(struct net_device *dev);

/* Projections */
projection <struct net_device> net_device {
  ...
  /* [modifier] <data_type> <struct_member_name> */;
  [in] unsigned int flags;
  [in] unsigned long long hw_features;
  [in] unsigned long long features;
  ...
  projection net_device_ops [alloc(caller)] *netdev_ops;
};
```

- Interface definition language
- asynchronous runtime (async threads)

Evaluation

- Software-only device
  - network (dummy)
  - block device (null-blk)
- Hardware device
  - Intel 82599 10 Gbps ethernet controller (ixgbe)
  - iperf tx benchmarks: within 6-13% of the native driver

Visit us!

Usenix ATC'19
July 10, Track II - Security #1: Kernel
(4:10 - 5:30 PM)