My VM is Lighter (and Safer) than your Container

SOSP 2017

Background

	Virtual Machine	Container
Pro:	1. Secure. 2. Resource isolation via CPU and memory	1. Light-weight. 2. Fast boot-time. 3. Can provision thousands of containers on a single physical server.
Cons:	1. Relatively heavy-weight. 2. Long boot-time. 3. Large memory footprint.	1. Insecure. 2. Isolation at process level. 3. Vulnerable to attacks.

Background

- VM is evolving.

- Unikernels, (Mirage, Osv, Rampkernel).
  - They are light-weight, they bootup faster.
  - They are hard to manage.

LightVM

- Decrease the size of the VM image.
    - Use unikernel.
    - Design a dediated tool (Tinyx) for creating minimalistic Linux VM images.

- Remove overhead in VM create/boot.
    - Remove XenStore.
    - Pre-initialize some parts of the VM.
    - Remove script execution during VM bootup.

Decrease VM Image Size

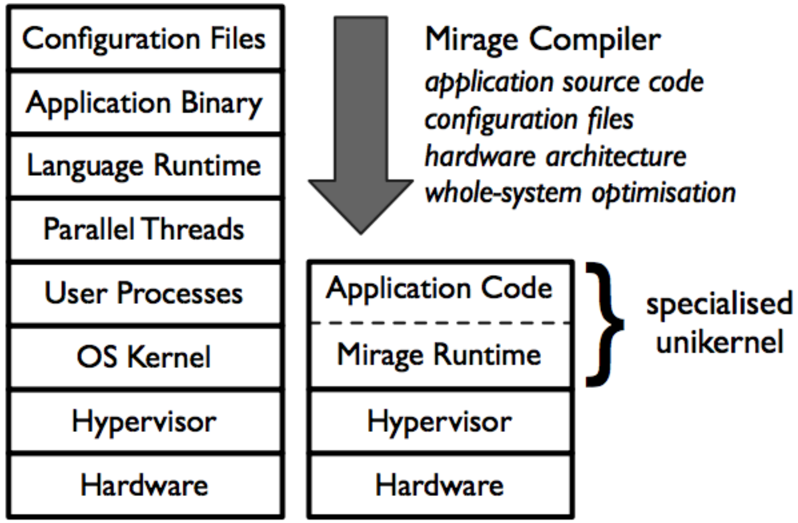


Figure 1: Contrasting software layers in existing VM appliances *vs*. unikernel's standalone kernel compilation approach.

Decrease VM Image Size

- Mirage is great!
- But you need to learn OCaml.
- And you need to learn LWT.
- And you need to learn to write functional programs.

Decrease VM Image Size

- The authors propose Tinyx.

- A tool for creating lightweight VM images.

- Currently only support Debian distributions (Ubuntu…).

Xen Background

- A hypervisor developed at Cambridge.

- Important notations:
  - Dom0: a monitoring VM for management task
  - XenStore: Store important information about the VM
  - Paravirtualization: VM runs native code on physical CPU.
  - Virtual device: virtualized devices used by VM kernels, virtual NIC, virtual disk.

Remove Overhead in Xen VM Create/Boot

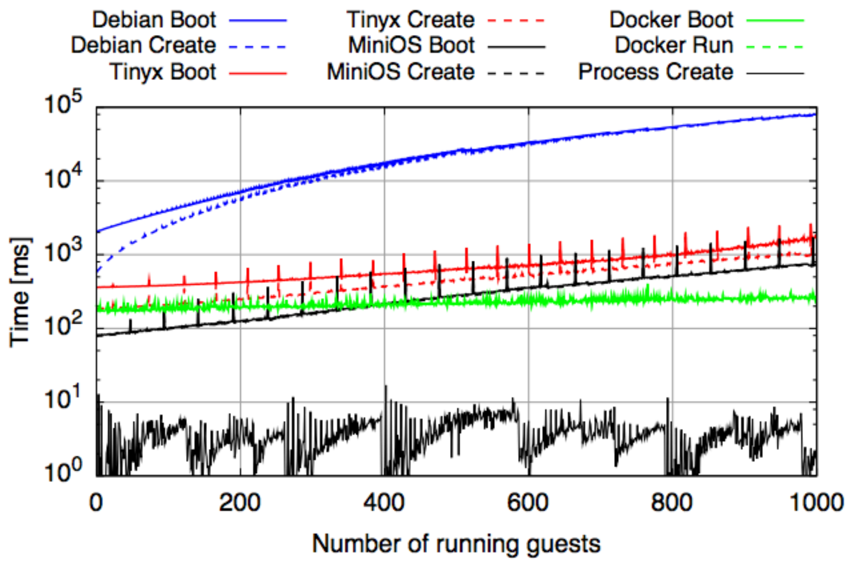


Figure 4: Comparison of domain instantiation and boot times for several guest types. With small guests, instantiation accounts for most of the delay when bringing up a new VM.

Remove Overhead in Xen VM Create/Boot

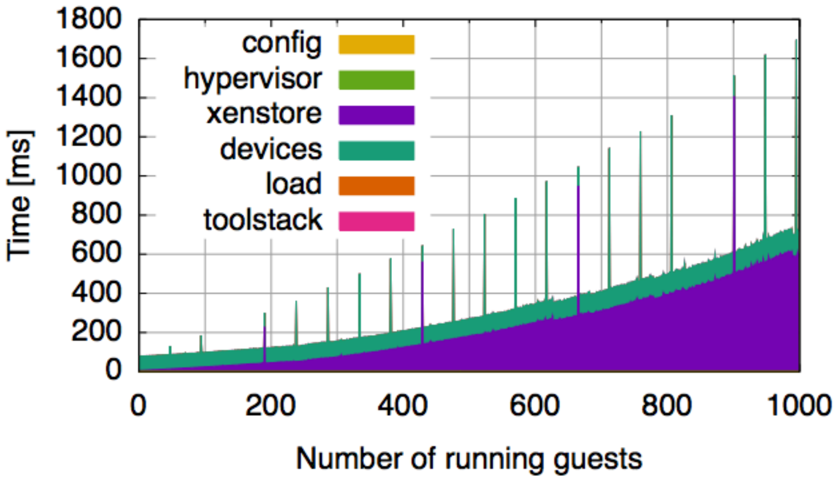


Figure 5: Breakdown of the VM creation overheads shows that the main contributors are interactions with the XenStore and the creation of virtual devices.

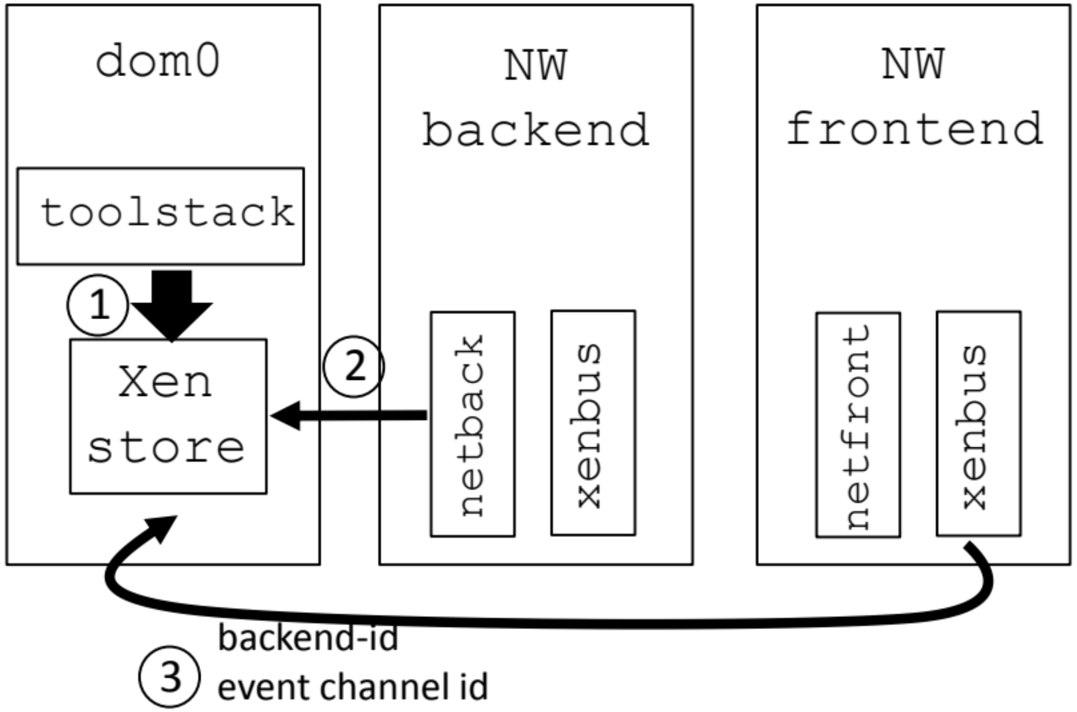
Overhead analysis

- XenStore Interaction during creation
  - XenStore uses complicated protocol, multiple read-write, multiple context-switches.
  - Linearly scan all the names to prevent duplicating names.
  - Concurrent updating records leads to failed transactions.
- Virtual device creation.

LightVM

- Remove XenStore interaction during VM creation and migration.
- Pre-calculate VM templates.
- Remove script execution when VM boots.

VM Creation with XenStore

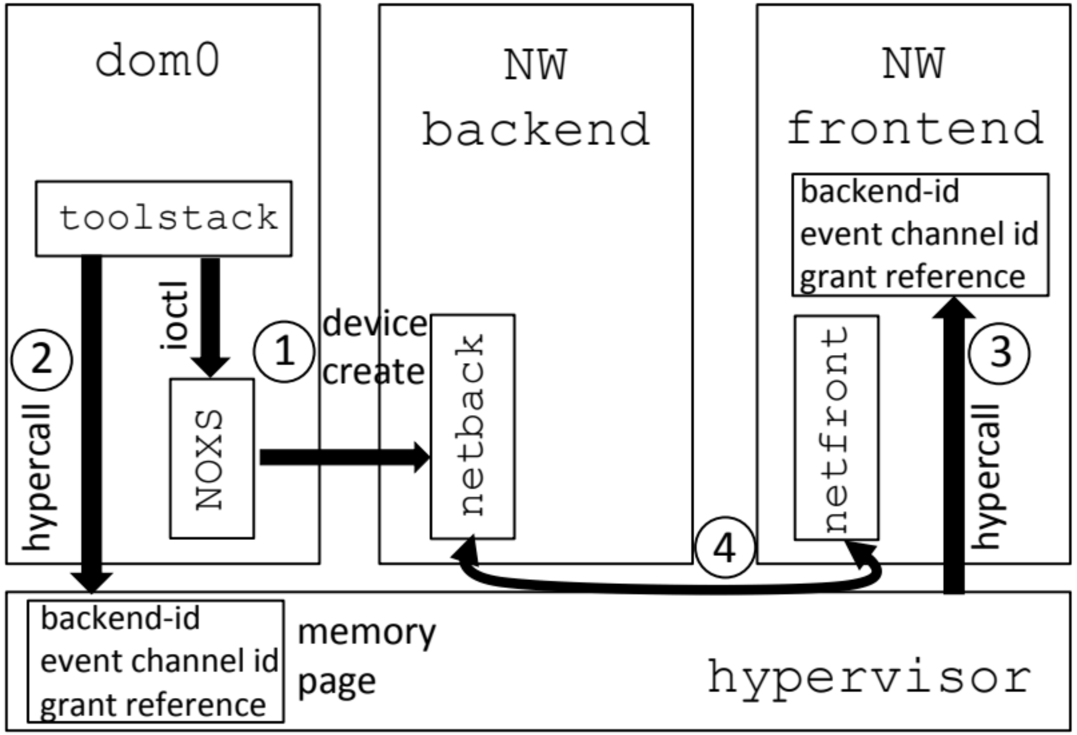


backend-id
event channel id
grant reference

VM Creation without XenStore

- Xen hypervisor already stores useful information for VM creation.
- Use shared-memory and hypercall to speed up VM creation.

VM Creation without XenStore



(b) noxs

Split ToolStack

- Pre-calculate VM templates.

- Initialize over VM templates.

Split ToolStack

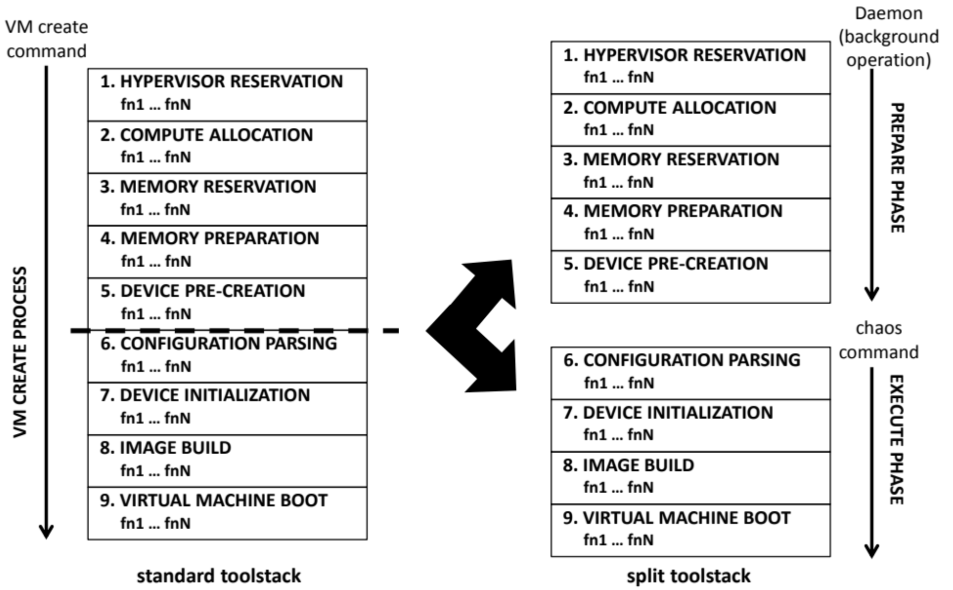


Figure 8: Toolstack split between functionality belonging to the prepare phase, carried out periodically by the chaos daemon, and an execute phase, directly called by chaos when a command is issued.

Remove Script Execution

- After a VM is created, it needs to boot.

- When booting, VM kernel needs to execute some scripts, which is slow.

- Merge script execution into the boot process.

Performance evaluation

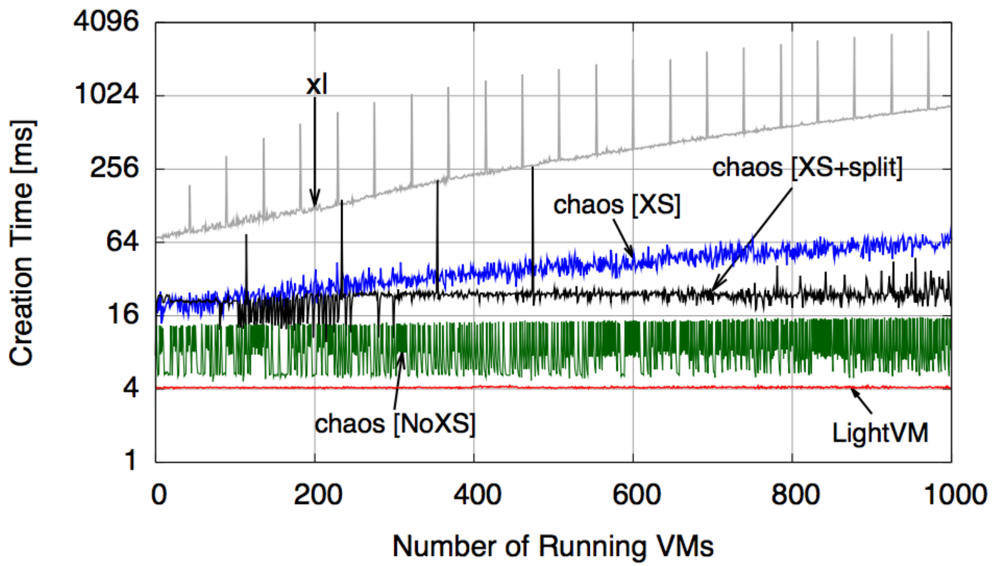


Figure 9: Creation times for up to 1,000 instances of the daytime unikernel for all combinations of LightVM's mechanisms. "xl" denotes standard Xen with no optimizations.

Performance evaluation

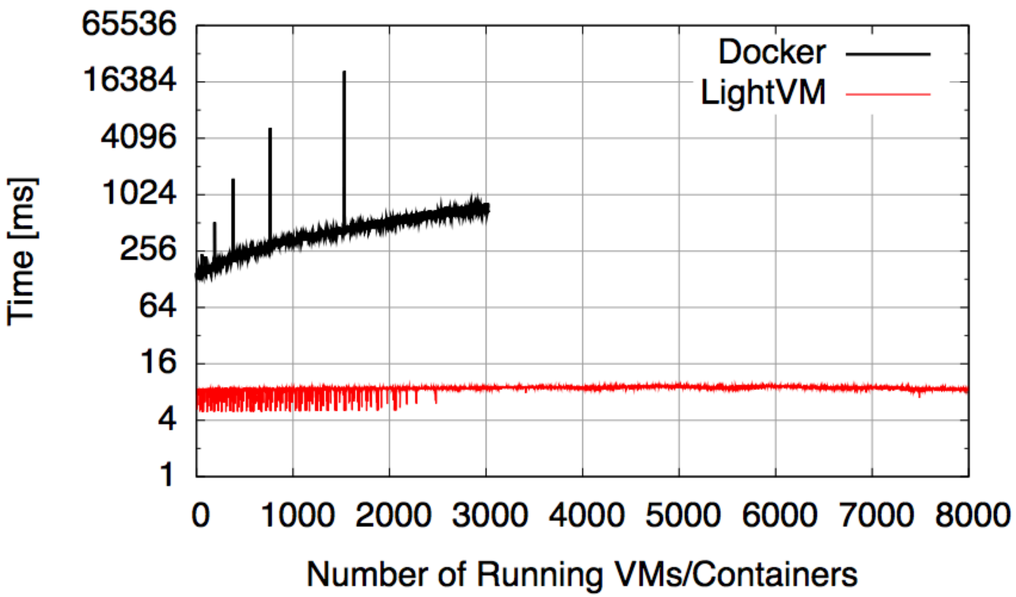


Figure 10: LightVM boot times on a 64-core machine versus Docker containers.

Performance Evaluation

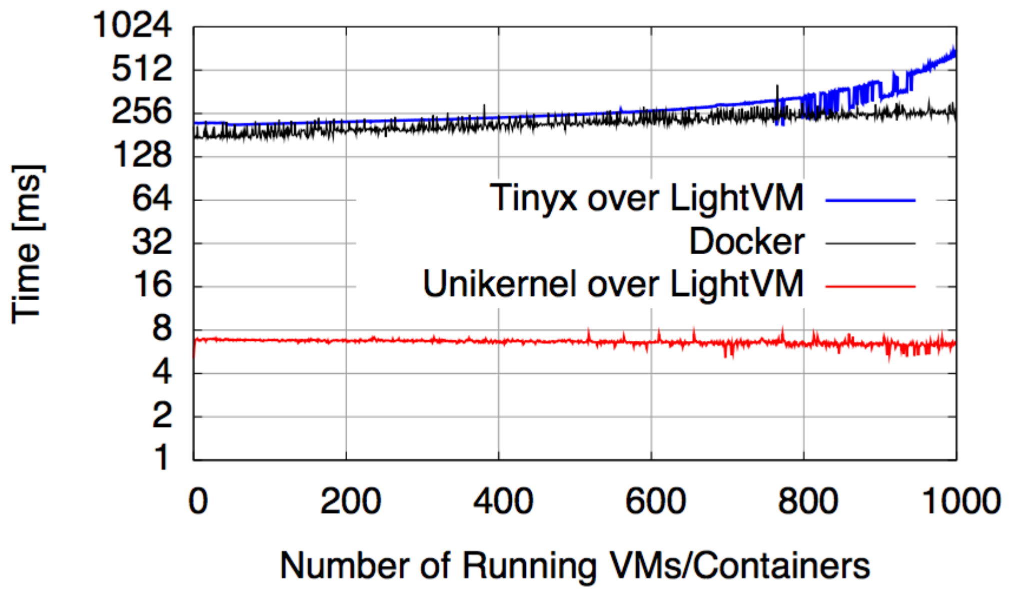


Figure 11: Boot times for unikernel and Tinyx guests versus Docker containers.

Performance Evaluation

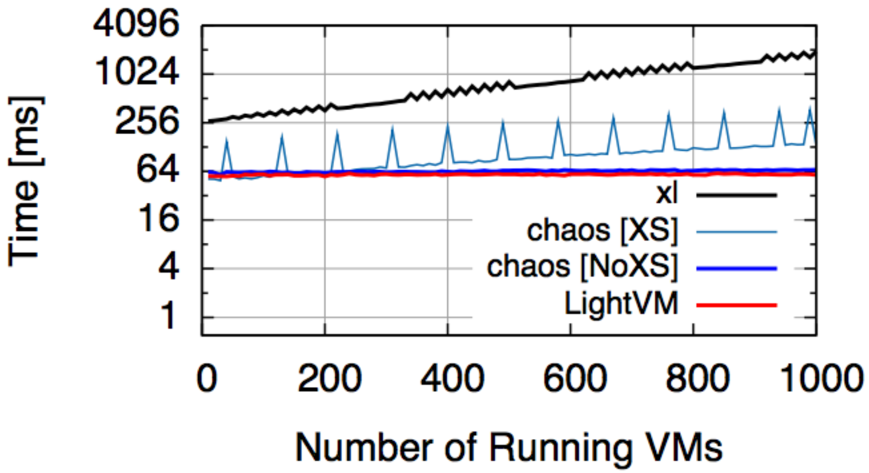


Figure 13: Migration times for the daytime unikernel.

Limitations

- Xen only, how about KVM/QEMU
- Is it really faster than container?
  - If VM runs unikernel, then yes.
  - If VM runs a minimalistic Linux distribution, then probably not.
- Intrusive modification to Xen toolstacks.
  - How about optimizing Xenstore?