

Αποδοτικός διαμοιρασμός αρχείων μεταξύ host και unikernel

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Φώτης Ξενάκης

Επιβλέπων: Νεκτάριος Κοζύρης

30/10/2020

Cloud computing & virtualization

Εν αρχή ην το **cloud**:

- Utility computing
- Everything as a Service
- Μεγάλη κλίμακα

Βασίζεται στο **virtualization**:

- Ελαστικότητα στον πάροχο
- Απομόνωση στον χρήστη

Λειτουργικά συστήματα

Guest OS γενικού σκοπού:

- Επιβάλλει διαχωρισμό

Στην πράξη:

- Τρέχει **μία εφαρμογή**

Θα μπορούσε γίνεται πιο **αποδοτικά**. Bloat!

GUS	Application
GKS	Guest OS
HUS	Hypervisor
HKS	Host OS
	Hardware

Unikernels

“Λειτουργικό” + εφαρμογή:

- Library OS
- Single address space

Συνεπώς:

- **Αποδοτικότητα** (μέγεθος, μνήμη, χρόνος)
- Επιδόσεις(?)

OSv: Linux compatibility, πολλές λειτουργίες,
πολλά(!) συστήματα αρχείων

GKS	Unikernel
HUS	Hypervisor
HKS	Host OS
	Hardware

Συστήματα αρχείων & virtualization

Παραδοσιακά:

- Guest: **δίσκος** (συσκευή block)
- Host: **αρχείο** (opaque)
- **Indirection**

Πρόσβαση στο fs του host:

- High level
- **Ευκολότερο** για τον guest
- Πρόσβαση **και** από τον host

Virtio-fs

Μέχρι τώρα:

- NFS, SMB, VirtFS...
- **Δίκτυο**

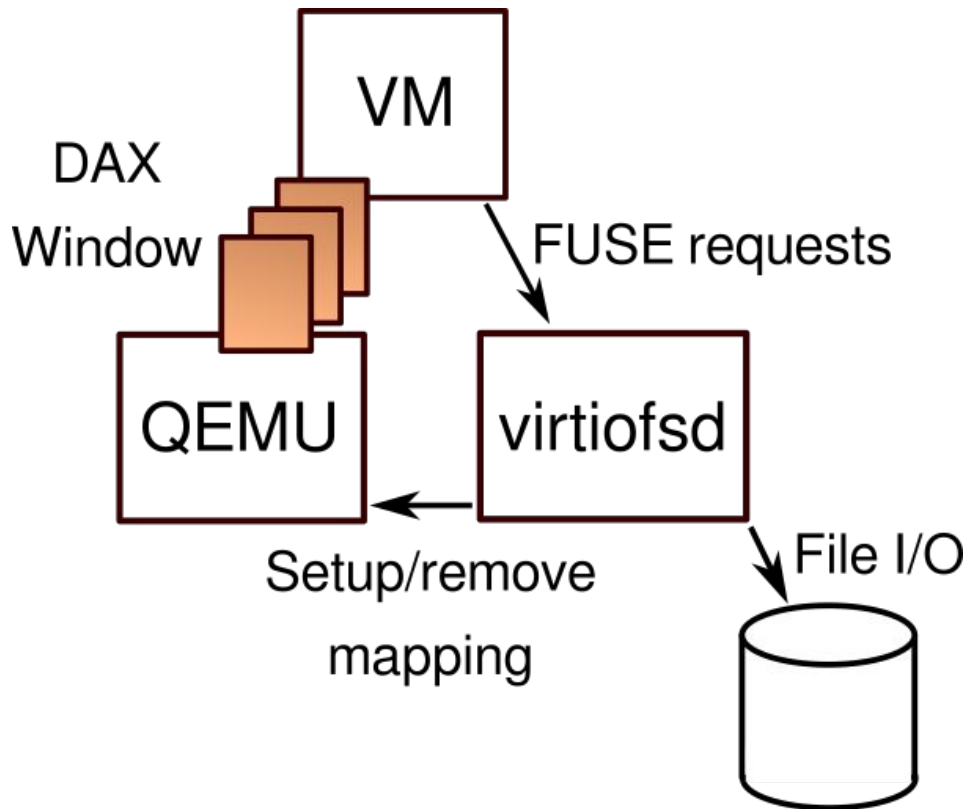
virtio-fs:

- Virtualization-native
- **virtio, FUSE**
- **virtiofsd**

Virtio-fs DAX window

Κοινή μνήμη guest \Leftrightarrow host:

- **mmap()**
- Όχι VMEXIT
- Καλύτερες επιδόσεις



Στόχοι εργασίας

Τεχνικοί:

- virtio-fs **DAX** window στο **OSv**
- **Read-only**
- Boot από virtio-fs

...και μη:

- **Συνεισφορά** σε έργο FOSS
- Πρακτική εργασία \Rightarrow πρακτική αξία

Virtio-fs στο OSv

Προϋπάρχουσα, στοιχειώδης υλοποίηση

Οδηγός:

- Low-level χειρισμός PCI συσκευής
- **Αγνωστικός** ως προς FUSE

Σύστημα αρχείων:

- High-level fs λειτουργίες
- Ξέρει **FUSE**

DAX window στο OSν

Οδηγός:

- PCI

Σύστημα αρχείων:

- map alignment
- FUSE_SETUPMAPPING
- FUSE_REMOVEMAPPING

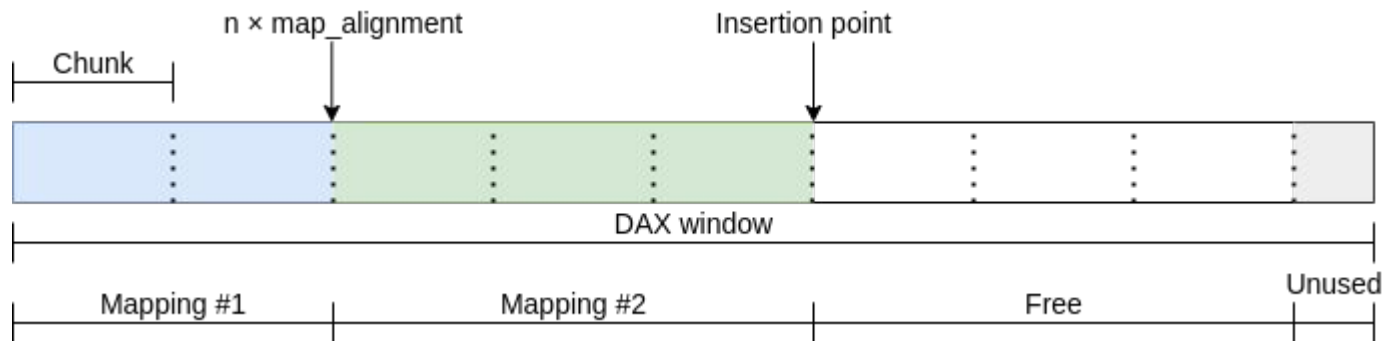
DAX window manager

Πρόβλημα:

- Πολλά mappings

Λύση:

- Chunks
- Στοίβα (LIFO)



Boot από virtio-fs

Αρχικά:

- **Boot** fs: πάντα ramfs
- **Root** fs: δυναμικά rofs \Rightarrow ZFS \Rightarrow ramfs

Τελικά:

- **Προαιρετικό** kernel command line option
- Root fs: rofs, ZFS, **virtio-fs**, ramfs
- Backwards compatible

Αλλαγές σε πυρήνα και εργαλεία

Αξιολόγηση

Διαφορετικά σενάρια:

- **Microbenchmark**
- Χρόνος **εκκίνησης**
- Πραγματική **εφαρμογή**

Microbenchmark

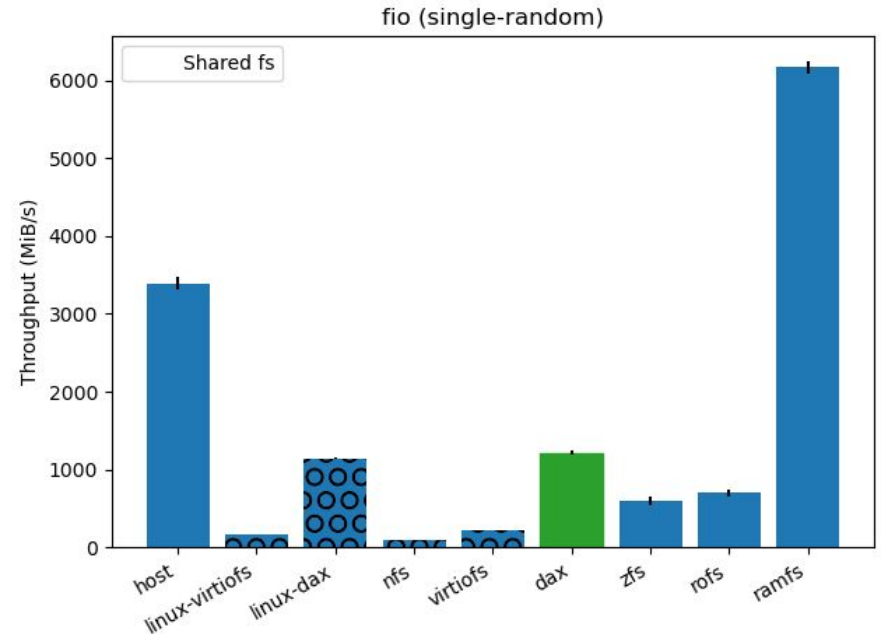
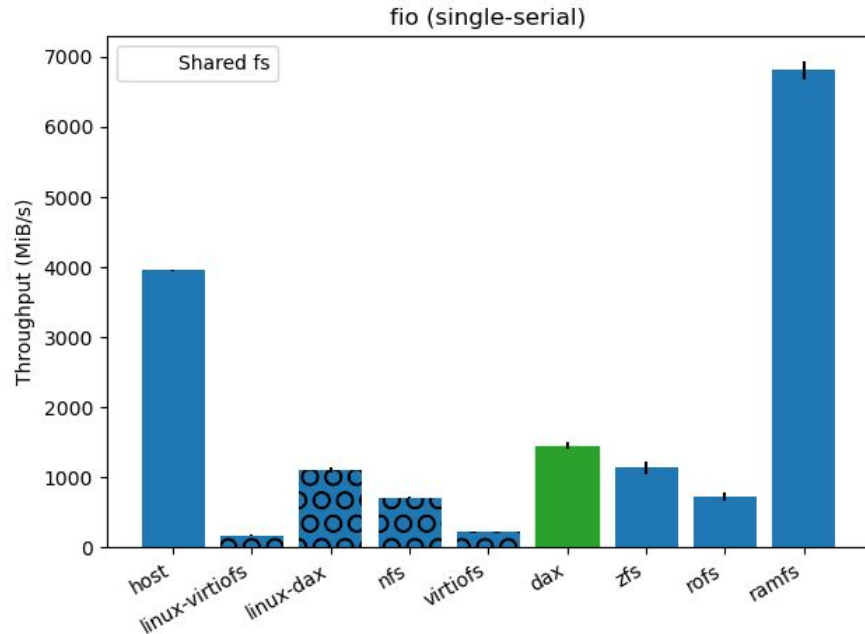
Flexible I/O tester (fio)

- 1 αρχείο / πολλά αρχεία
- Συνολικό μέγεθος 1 GiB
- Σειριακή / τυχαία ανάγνωση
- **Throughput**

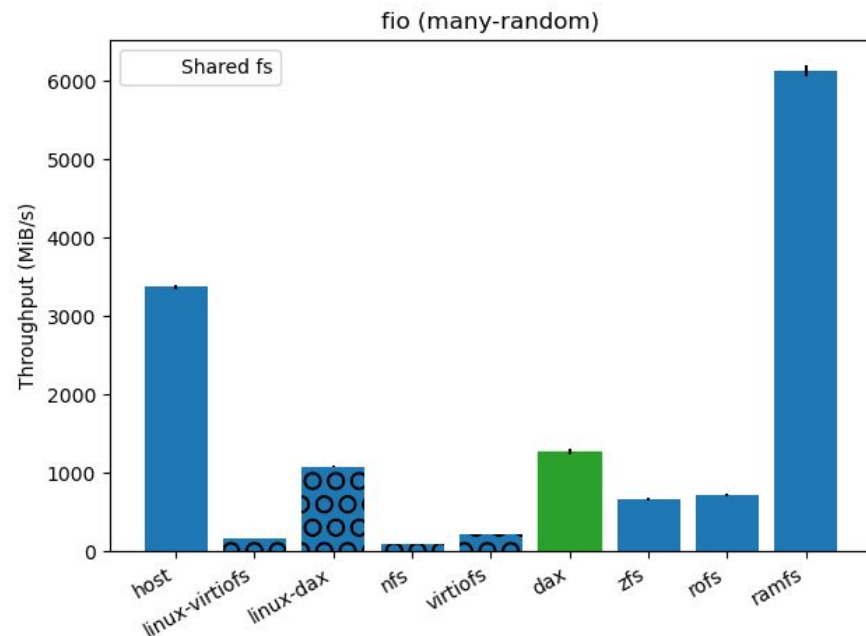
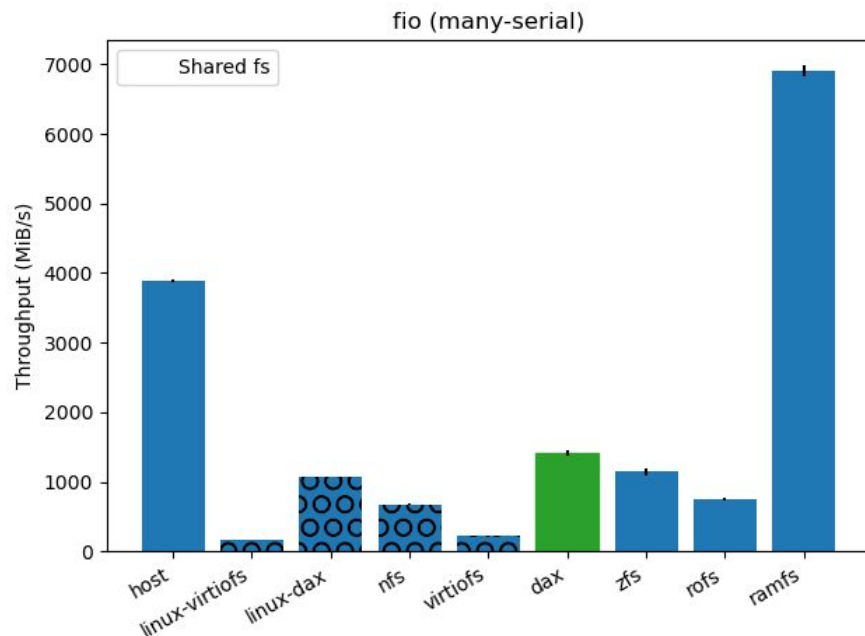
Σύγκριση:

- Host (baseline)
- Linux: [virtio-fs](#), [virtio-fs-DAX](#)
- OSv: [virtio-fs](#), [virtio-fs-DAX](#), ZFS, rofs, ramfs, [NFS](#)

Αποτελέσματα microbenchmark (ένα αρχείο)



Αποτελέσματα microbenchmark (πολλά αρχεία)



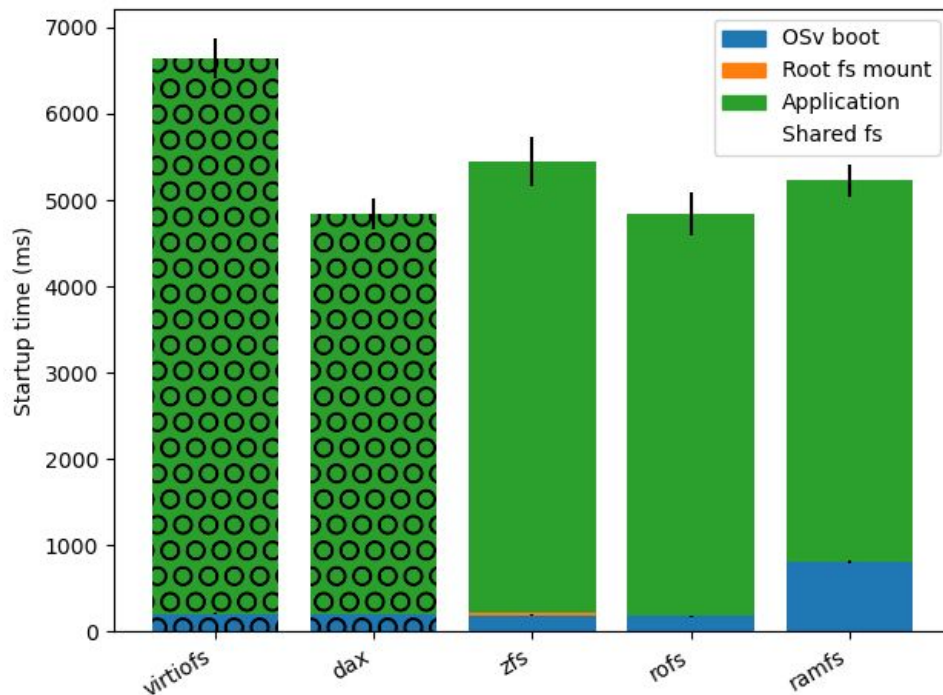
Startup time

Απλή **web** (spring) εφαρμογή

- Εκκίνηση OSν
- Root fs mount
- Εκκίνηση εφαρμογής

Σύγκριση:

- OSν: [virtio-fs](#), [virtio-fs-DAX](#), ZFS, rofs, ramfs



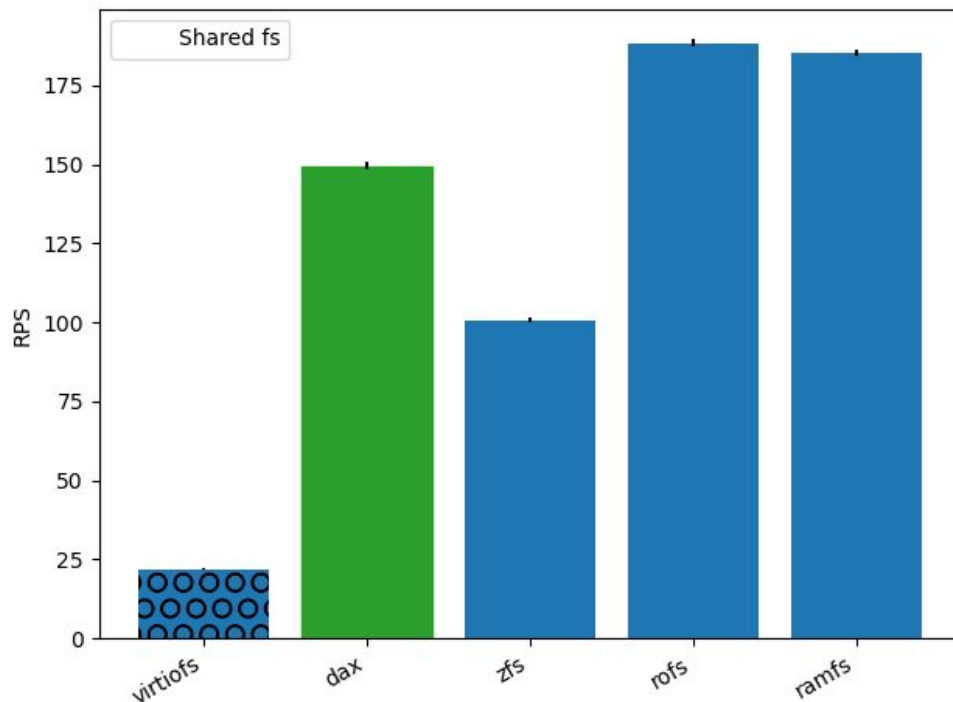
Application benchmark

Nginx static **web server**:

- Πολλά αρχεία ~MiB
- **Vegeta HTTP** load client
- **Requests / δευτερόλεπτο**

Σύγκριση:

- OSv: [virtio-fs](#), [virtio-fs-DAX](#), ZFS, rofs, ramfs



Σύνοψη

Τεχνικά συμπεράσματα:

- virtio-fs στο OSν **χρήσιμο** (αλλαγές χωρίς εκ νέου χτίσιμο)
- DAX window \Rightarrow άριστες **επιδόσεις**
- Ευχάριστη **ανάπτυξη** στο OSν (σαν ανάπτυξη εφαρμογής)

Μη τεχνικά συμπεράσματα:

- Ανοιχτές, υποστηρηκτικές **κοινότητες**
- Συνεισφορά

Συνεισφορά

OSv OSv Unikernel
@OSv_unikernel

OSv has had read-only DAX support since this commit in June - [github.com/cloudius-syste...](https://github.com/cloudius-systems/osv)

...


You are welcome to write a guest post on the QEMU blog (<https://qemu.org/blog/>) that gives an overview of OSv and the virtio-fs device interface. I imagine people would be interested in learning about both these topics.

The git repo for the QEMU website and blog is here:
<https://gitlab.com/qemu-project/qemu-web>

You can create a new blog post by dropping a Markdown file into `_posts/`. Images can be added to `screenshots/` directory.

The blog post can be sent to the QEMU mailing list using `git-send-email(1)` with Thomas Huth <thuth@redhat.com> CCed. Please also CC `virtio-fs@redhat.com` and we'll review it.

Stefan

 Thorsten 'the Linux kernel logger' Leemhuis(6/6) @kernellogger · Oct 20
#DAX support for #virtiofs was merged for #Linux #kernel 5.10. "[...] This can improve I/O performance for various workloads, as well as reducing the memory requirement by eliminating double caching [...]": git.kernel.org/torvalds/c/694...

Cover letter with details: lore.kernel.org/lkml/202008192...



index : [kernel/git/torvalds/linux.git](https://git.kernel.org/torvalds/linux.git)

Linux kernel source tree

about summary refs log tree **commit** diff stats log msg ▾

Merge tag 'fuse-update-5.10' of [git://git.kernel.org/pub/scm/linux/kernel/git/mszeredi/fuse](https://git.kernel.org/pub/scm/linux/kernel/git/mszeredi/fuse)

Ευχαριστώ!

Ερωτήσεις;