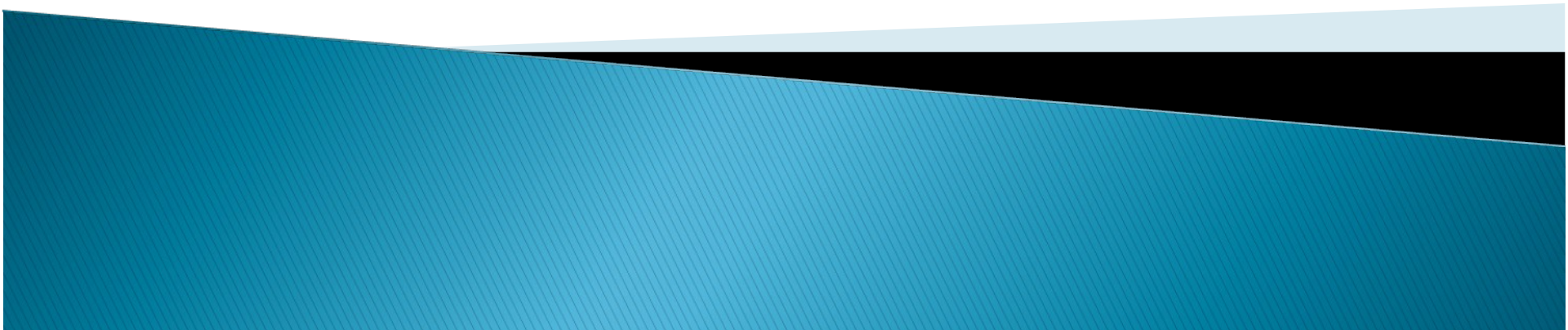
Wirtualna maszyna XEN

Szymon Acedański, Piotr Hofman, Łukasz Rekucki

SO @ MIMUW, 2006



XEN – cóż to takiego?

- Monitor maszyny wirtulnej
- Opracowany na Uniwersytecie Cambridge
- Projekt Open Source (licencja GPL)

- Alternatywne podejście do wirtualizacji: *parawirtualizacja*

O wirtualizacji ogólnie

- Historia
- Po co to komu
- Jak się ma do tego XEN

Historia wirtualizacji

- 1972 – IBM VM/CMS – Pierwsze wydanie “Hypervisor” bazującego na OS
- 1974 – idea wirtualizacji opisana poraz pierwszy
- 1988 – Mainframe PR/SM (Processor Resource/System Manager)
- 1999 – VMware Virtual Platform

Ekonomia (1)

- Oczekiwania klienta (np. firmy hostingowej):
  - Prawa roota
  - Bezpieczeństwo zasobów
  - Nieprzerwana praca systemu
  - Niska cena!
  - Potencjalnie duża moc maszyny,
    choć w zwykłej pacy nie wykorzystywana w pełni

Ekonomia (2)

- Co kosztuje?
  - Miejsce
  - Prąd (chłodzenie!)
  - Sprzęt
- No to mamy 100 klientów, kupiliśmy dla każdego komputer ...
  - Czyż to nie było bez sensu?

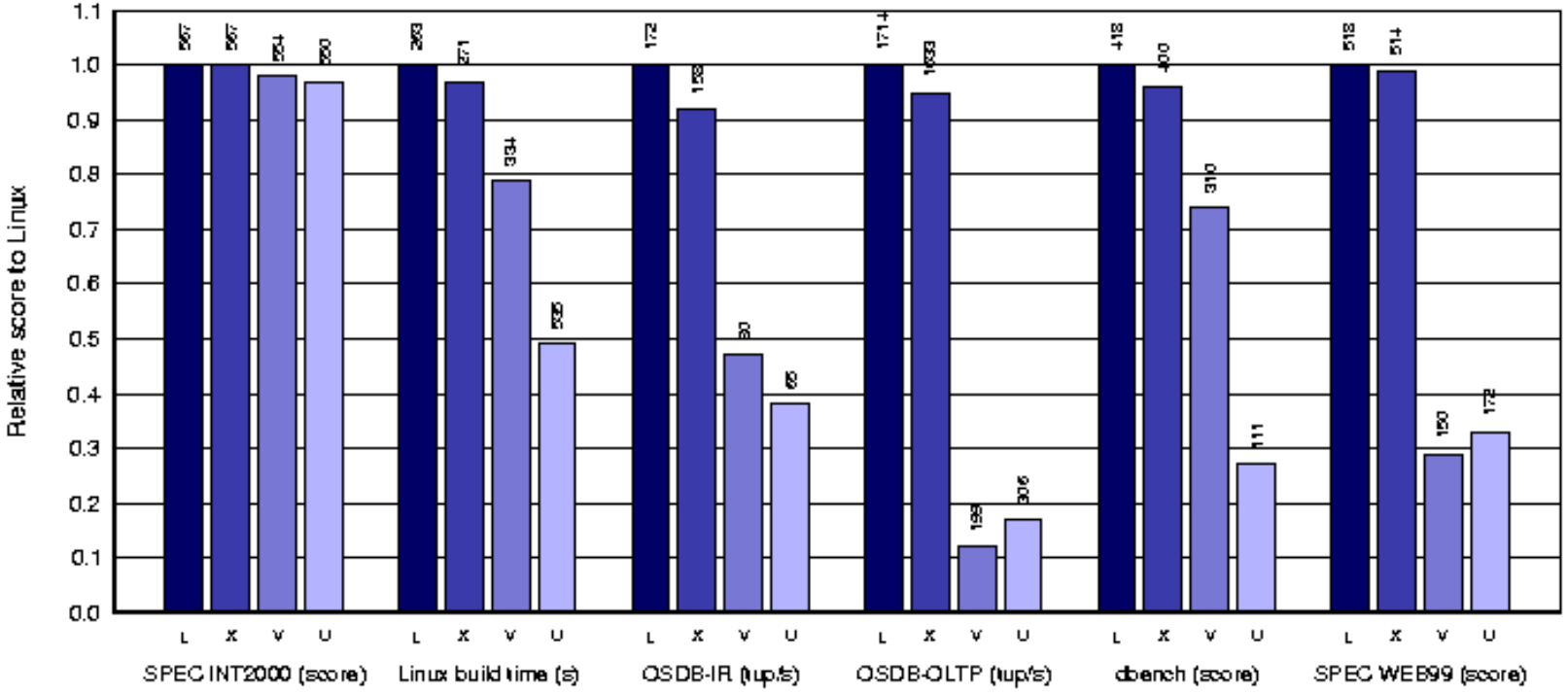
Przechodzimy na wirtualizację

- Mniej sprzętu więc łatwiej o niego dbać
- Jeśli sprzęt się zepsuje, to można cały system przenieść na inny komputer
- Łatwiejsze tworzenie kopii zapasowych
- Rośnie bezpieczeństwo, bo włamanie się na jedną maszynę nie implikuje włamania się na inne

Czemu XEN i parawirtualizacja?

- Lepsza wydajność
- Skuteczniejsza izolacja maszyn wirtualnych
- Redukcja czasu awarii, dzięki *Live Migration*

Porównanie wydajności



Relative performance on native Linux (L), Xen/Linux (X), VMware Workstation 3.2 (V), and User Mode Linux (U).

Trochę policzmy

- Szacuje się, że Google ma około 450 tysięcy serwerów
- Szacuje się, że jedno duże centrum danych, pożera tyle energii, jaka wystarczyłaby dla zaopatrzenia w prąd ok. 40-tysięcznego miasta
- Moja 4-os. rodzina płaci około 100 zł. miesięcznie
- Czyli firma Google płaciłaby w Polsce miesięcznie 450*10.000*100 zł = 450 mln zł (150 mln $)

Trochę pownioskujmy

- To nie jest świetne szacowanie, bo ceny prądu na świecie są różne, a i zapewne przy zakupie hurtowym cena ostro spada
- Ale w skali roku mówimy o sumie rzędu miliarda dolarów
- Redukcja kosztów o 5% to niemało, nie uważacie?
- A już na pewno wystarczająco dużo, żeby zamortyzować wydatek na oprogramowanie

Jest wsparcie dla XENa

Parawirtualizacja i technikalia

- Skąd ta para?
  Co jest tu innego niż w VMware?

„Zwykły proszek” vs. XEN

Pełna wirtualizacja

- Emulacja realnych urządzeń
  - z punktu widzenia systemu-gościa karta sieciowa wygląda jakby była zwykłą kartą PCI (model AMD PCNET w przypadku VMware)
- System nie wie, że działa na maszynie wirtualnej, ale działa ☺

Parawirtualizacja

- Własne, specjalne urządzenia
  - brak emulowanych urządzeń (PCI, zegar, ...)
  - wirtualny sprzęt, wymagający nowych sterowników
- System trzeba nauczyć (przerobić) działania jako wirtualny
- Ale aplikacji przerabiać nie trzeba

Domeny XEN’a

Domena XEN’a = Pojedyncza wirtualna maszyna

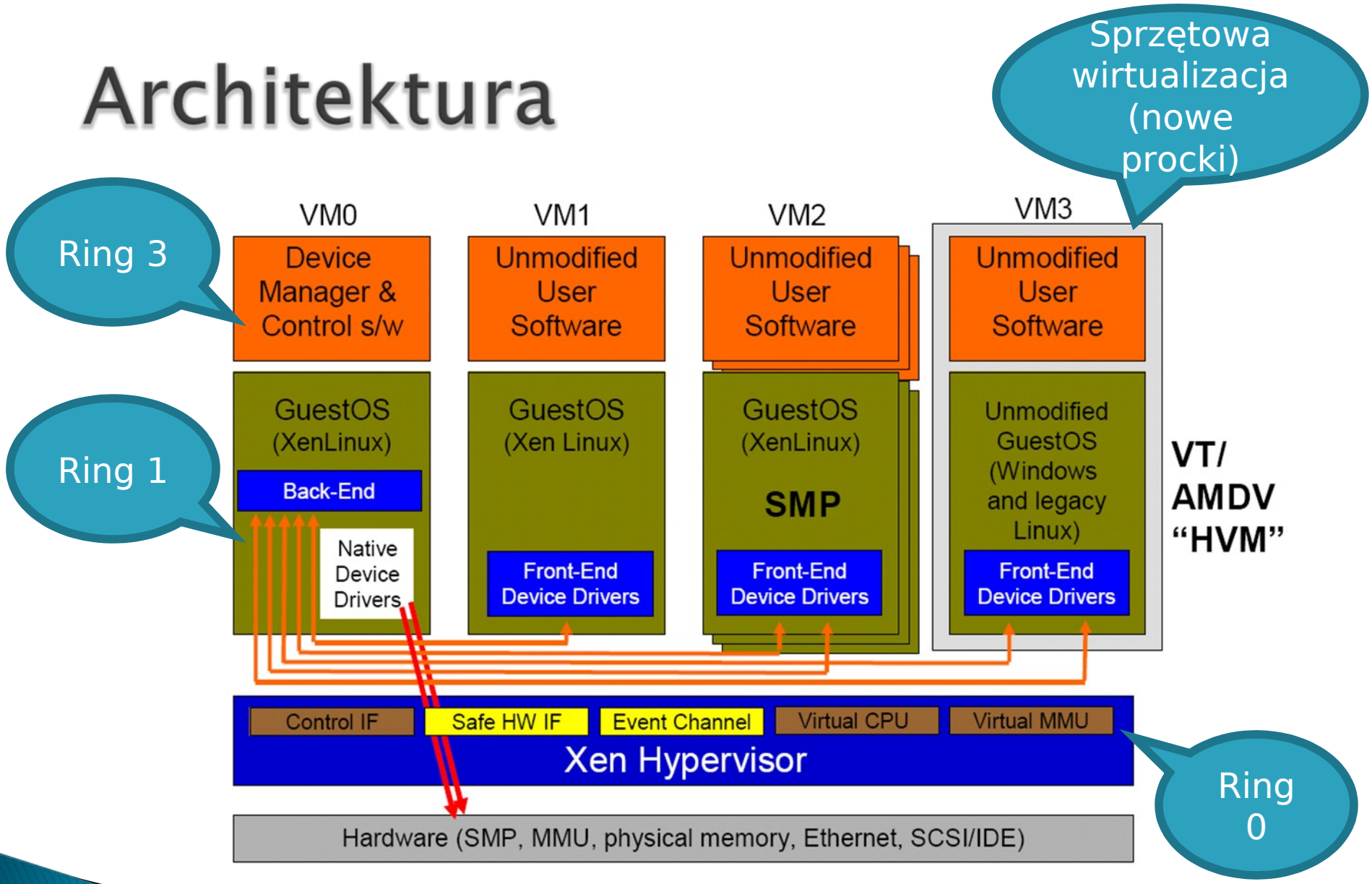
Dom0

- Domena administracyjna: tylko z niej można kontrolować (włączać/wyłączać) inne
- Jako jedyna ma bezpośredni dostęp do sprzętu, który dzielimy
- Zazwyczaj jako jedyna ma dostęp do rzeczywistego sprzętu (PCI, konsola)

Dom1, ... (DomU)

- Domena dla klienta
- Sterowniki urządzeń wirtualnych komunikują się z Dom0
- Całkowicie odizolowana od innych DomU

Architektura



Co muszą wiedzieć domeny?

- W jaki sposób dogadać się z hypervisorem?
  - To jest akurat proste – podstawowy mechanizm jest analogiczny do zwykłych syscall'i (`int 0x82`)
- Jak nie pogryźć się o pamięć?
  - Szczególnie, że nie jest tak, że każda domena dostaje swój spójny kawałek pamięci i koniec.
- Jak uzyskać dostęp do dysku i sieci?
- Jak mierzyć czas?

Stronicowanie

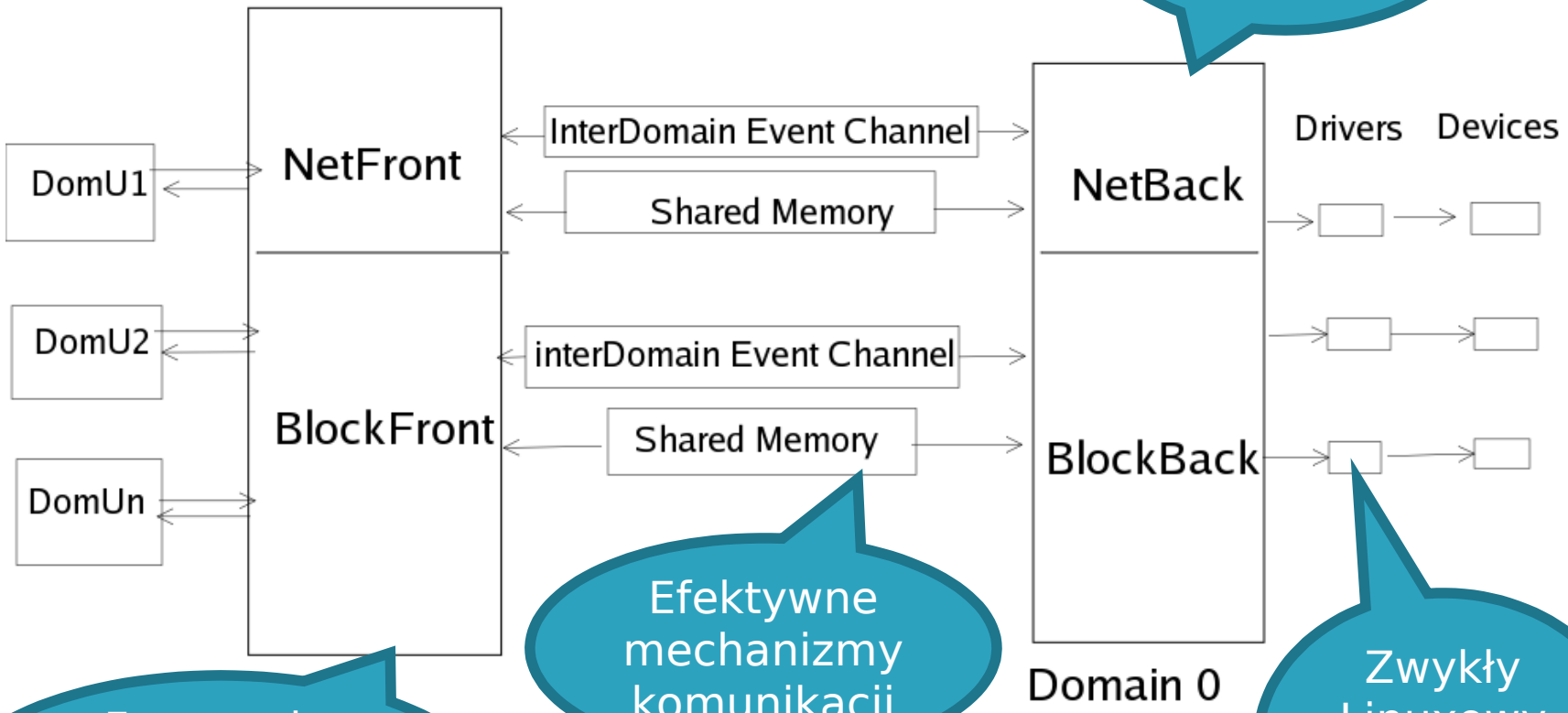
VMware

- Musi stworzyć iluzję ciągłej pamięci fizycznej
- Utrzymuje własne kopie wszystkich tablic stron nadzorowanej maszyny
- Przechwytuje ich zmiany i emuluje je na hoście

XEN

- Domeny są świadome nieciągłości pamięci
- Domena wie, jakie kawałki ma do dyspozycji i sama nimi zarządza
- Hypervisor może zmieniać przydział pamięci do domen (memory hotplug!)

Wirtualne sterowniki



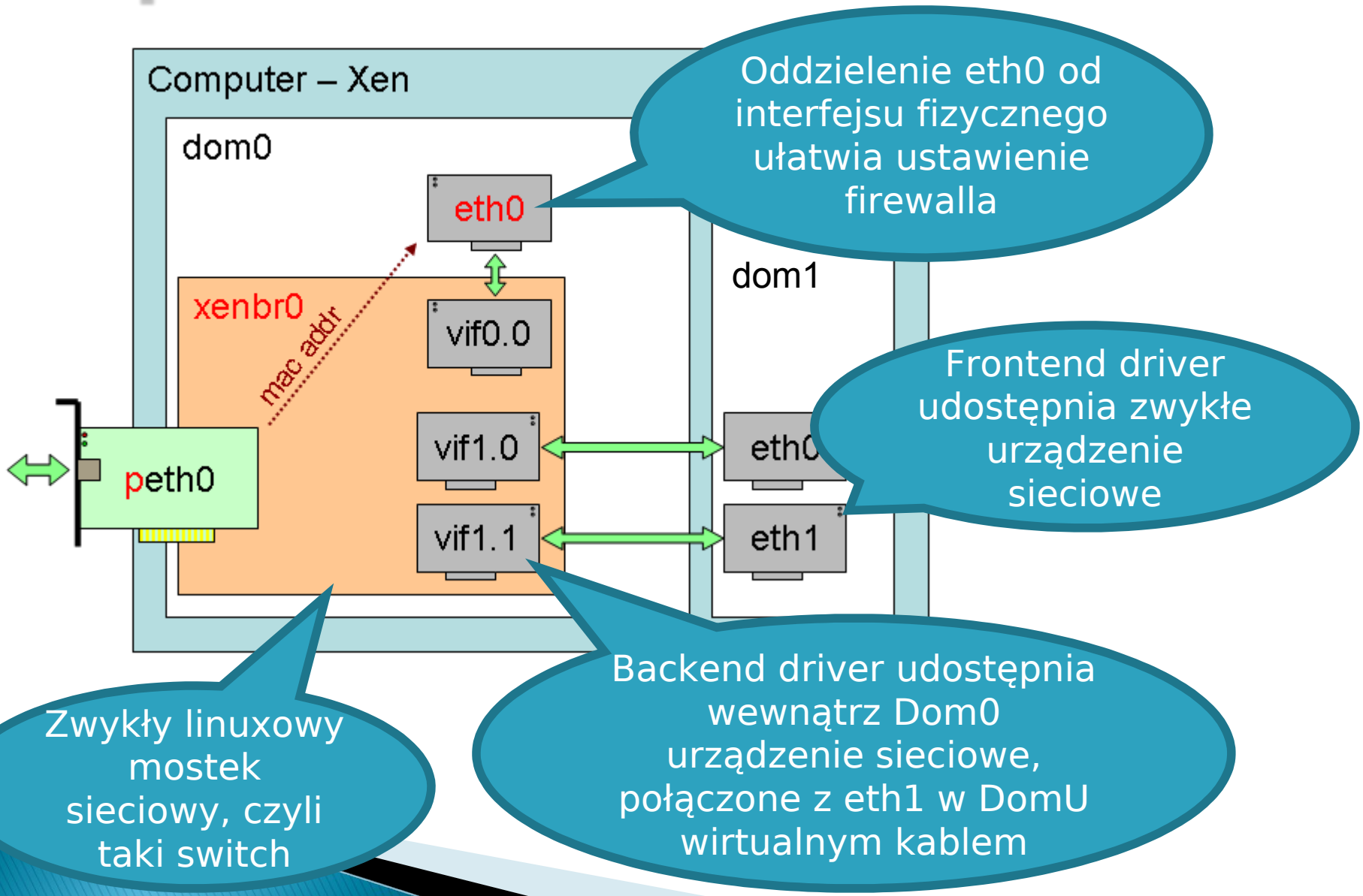
Backend – część działająca w Dom0

Frontend – część działająca w jednej z DomU

Efektywne mechanizmy komunikacji

Zwykły Linuxowy sterownik np. dysku

Współdzielenie sieci



Współdzielenie dysku

- Wirtualny sterownik pozwala Dom0 na udostępnienie dysku, partycji lub woluminu LVM domenom DomU
- Jednak jeden system plików nie może być (jeszcze) współdzielony przez wiele maszyn
- Opracowywany specjalny system plików XenFs ma rozwiązać ten problem

Czas – co za problem?

VMware

- Pozwólcie, że zademonstruję w praktyce...

XEN

- Czas rzeczywisty
  - Nanosekundy od startu systemu (użyteczny przy dostępnie do urządzeń)
- Czas wirtualny
  - Płynie tylko wtedy, gdy dana domena jest wykonywana (użyteczny dla schedulera procesów)
- Czas zegarowy
  - Taki, jakiego użytkownik oczekuje na zegarze

Izolacja – gdzie ona jest?

- W schedulerze procesora i dysku, który zapewnia równe traktowanie domen
- W mechanizmach bezpiecznego I/O, które określa uprawnienia domen przy dostępie do pamięci i portów I/O
- W wirtualnej jednostce pamięci, która sprawdza wszystkie zmiany tablic stron, jak również deskryptorów segmentów

Izolacja – co to oznacza?

- W izolacji chodzi nie tylko o to, żeby włamanie na jeden wirtualny komputer nie implikowało opanowania innych
- To również zapewnienie, że w przypadku obciążenia pewnych maszyn (brak pamięci, intensywne swapowanie, ping flood, fork bomb, ataki DoS), inne tego raczej nie odczuwają

Pytanie kontrolne...

- Które z wcześniej poznanych rozwiązań wirtualizacji wykorzystuje parawirtualizację?

Odpowiedź: UML

Wreszcie trochę praktyki

- Jak zainstalować?
  Jak skonfigurować?
  Jak fajnie działa Live Migration?

Instalacja

- Demostracyjny LiveCD
  Zawiera: OpenSUSE, Debian, Centos
- Dostępny jako pakiet w większości dystrybucji:
  OpenSuse, Debian Etch, Fedora Core 5+
- Dostępne źródła:
  http://www.xensource.com/download/
- XenEnterprise:
  30 dniowy trial
  Dla systemów Linux: $500+
  Dla systemów Windows – faza Beta

Konfiguracja

- Pliki konfiguracyjne = skrypty w Pythonie
- Xend - /etc/xen/xend-config.sxp
  - Konfiguracja serwera HTTP i relokacji (migracji)
  - Konfiguracja sieci: bridge, NAT, routed
- Konfiguracja domeny
  - Pliki jądra i (opcjonalnie) initrd
  - Dyski: physical, file, copy-on-write
  - Interfejsy sieciowe (statyczne lub dhcp)
  - Inne argumenty jądra („extra”)

Przykładowa konfiguracja

```
kernel = '/root/domU-2.6.17'
ramdisk = '/root/domU-2.6.17-local.img'

memory = 48
name = "XenCast"

dhcp='off'
root= '/dev/hda1'

vif   = [ '' ]
disk = [ 'phy:/dev/nbd0,hda1,w', 'phy:/dev/nbd1,hda2,w' ]

ip='192.168.0.101'
netmask = '255.255.255.0'
hostname = 'xen-cast'
gateway = '192.168.0.1'

# nfs_server = "192.168.0.2"
# nfs_root   = '/srv/etch'

extra= "4"
debian-server:~# _
```

Administracja – narzędzie xm

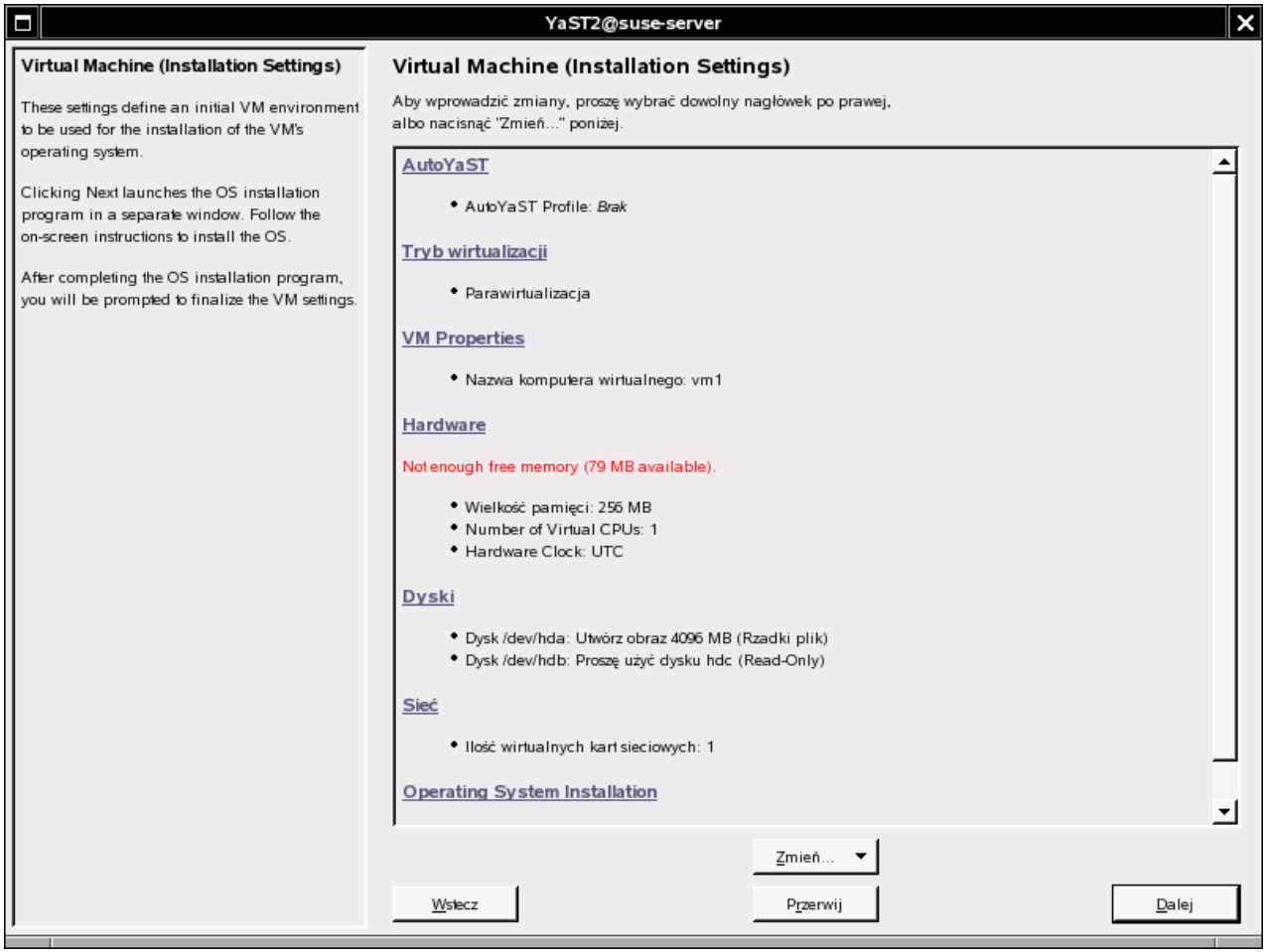
```
create <plik konfiguracyjny> [-c] [name=xxx]
console <ID lub nazwa domeny>

top
list [--long]

save <ID lub nazwa domeny> <nazwa pliku>
restore <nazwa pliku>
(un)pause <ID lub nazwa domeny>
shutdown | reboot | destroy <ID lub nazwa>
mem-set | vcpu-set <ID lub nazwa> <liczba>
migrate <domena> <host docelowy> [--live]

help [--long]
```

XEN pod SuSE


YaST2@suse-server
Virtual Machine (Installation Settings)
These settings define an initial VM environment to be used for the installation of the VM's operating system.
Clicking Next launches the OS installation program in a separate window. Follow the on-screen instructions to install the OS.
After completing the OS installation program, you will be prompted to finalize the VM settings.
Virtual Machine (Installation Settings)
Aby wprowadzić zmiany, proszę wybrać dowolny nagłówek po prawej, albo nacisnąć "Zmień..." poniżej.
AutoYaST
AutoYaST Profile: Brak
Tryb wirtualizacji
Parawirtualizacja
VM Properties
Nazwa komputera wirtualnego: vm1
Hardware
Not enough free memory (79 MB available).
Wielkość pamięci: 256 MB
Number of Virtual CPUs: 1
Hardware Clock: UTC
Dyski
Dysk /dev/hda: Utwórz obraz 4096 MB (Rzadki plik)
Dysk /dev/hdb: Proszę użyć dysku hdc (Read-Only)
Sieć
Ilość wirtualnych kart sieciowych: 1
Operating System Installation
Zmień...
Wstecz
Przerwij
Dalej

Typy migracji pod XEN 3.0

- Standardowa - „stop-and-copy”
- „Live Migration” - migracja żywego systemu
  1. Faza przed migracją
  2. Rezerwacja zasobów
  3. Iteracyjne kopiowanie
  4. „Stop-and-copy”
  5. Zobowiązanie
  6. Aktywacja

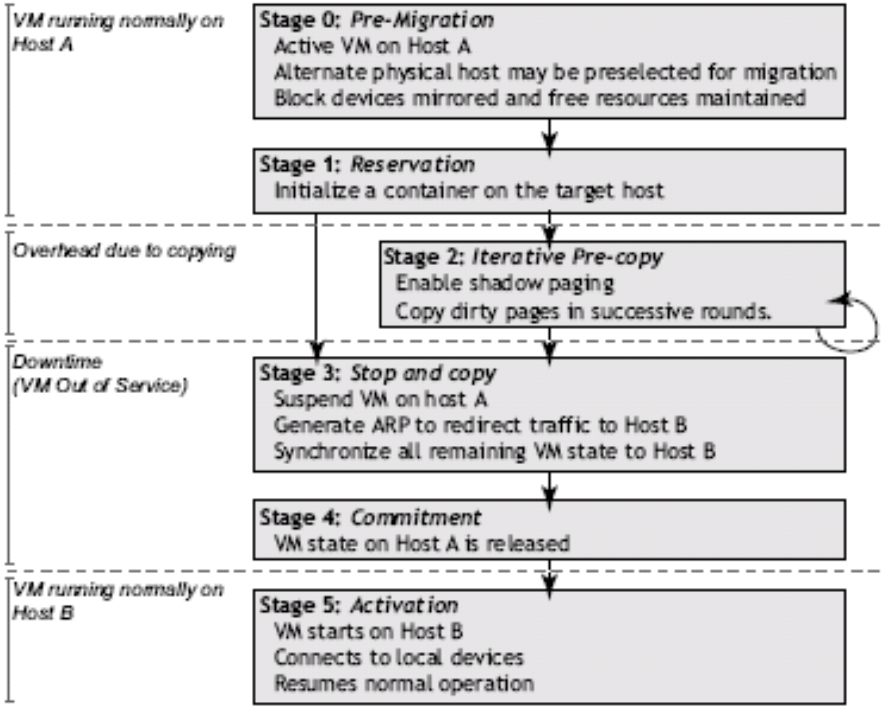
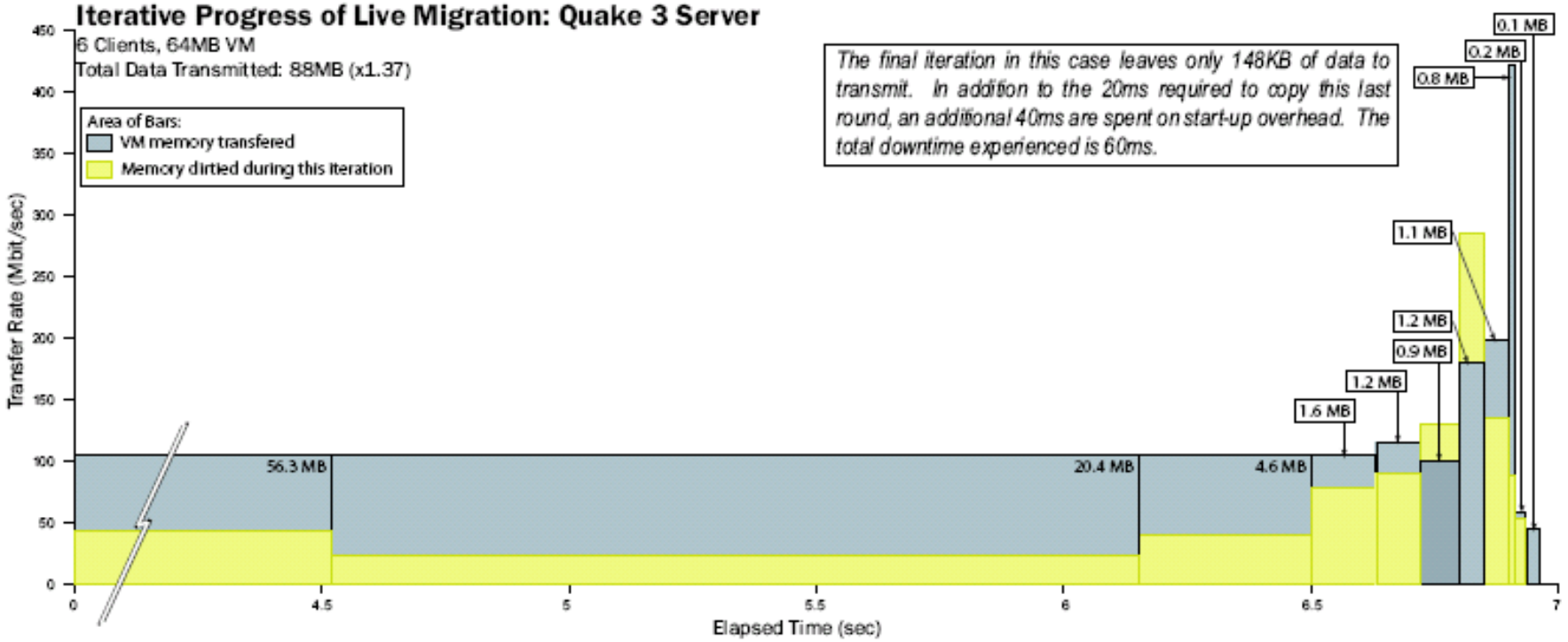


Figure 1: Migration timeline

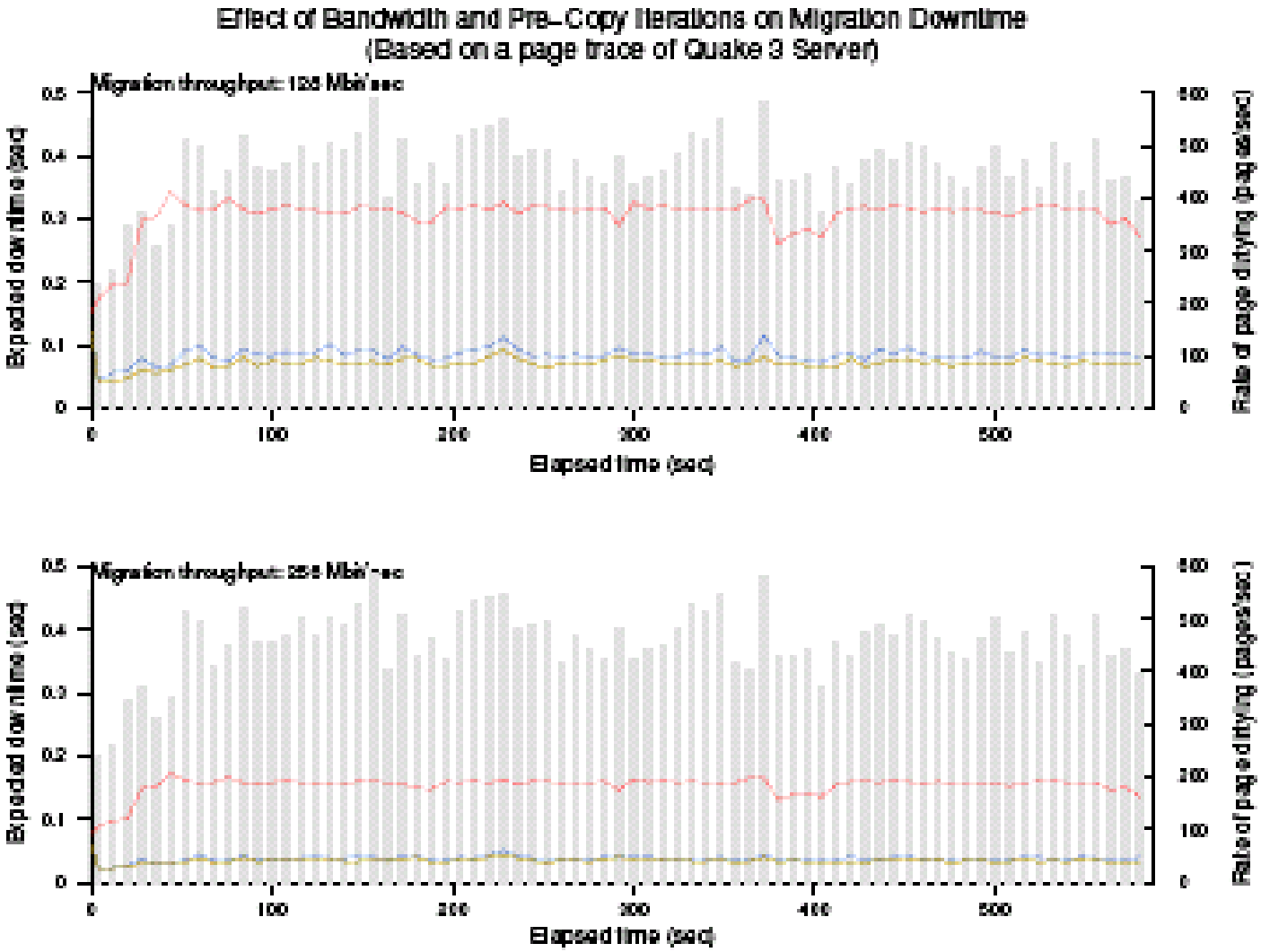
Cechy Live Migration

- Bardzo krótki czas niedostępności
- Zwiększony ruch w sieci (ale niewiele!)
- Zmienna prędkość kopiowania zależna od prędkości „brudzenia” stron
- Monitorowanie zmian przy pomocy „Shadow Page Tables”
- Wymaga współdzielonej przestrzeni dyskowej (np. NFS, NBD, iSCSI)

Przykład migracji (Quake 3.0)



Przykład migracji (Quake 3.0)



Przykład migracji (SPECweb99)

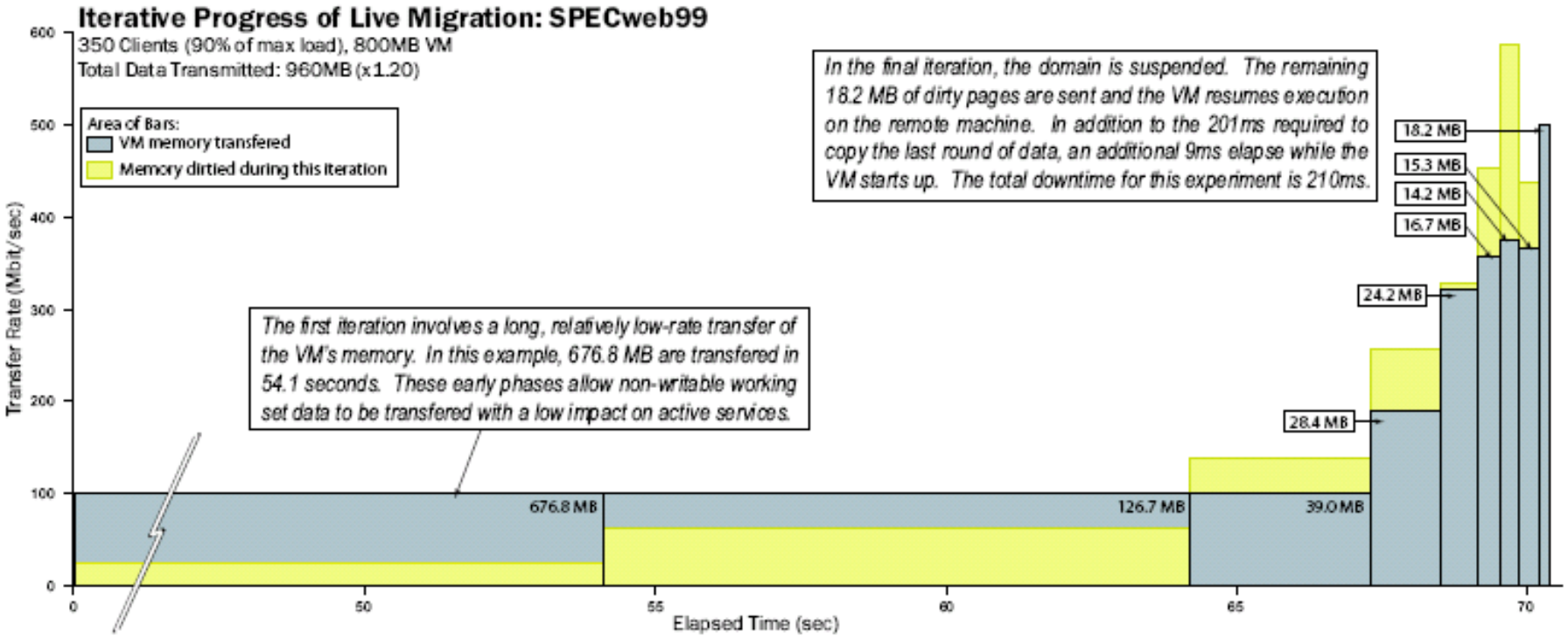
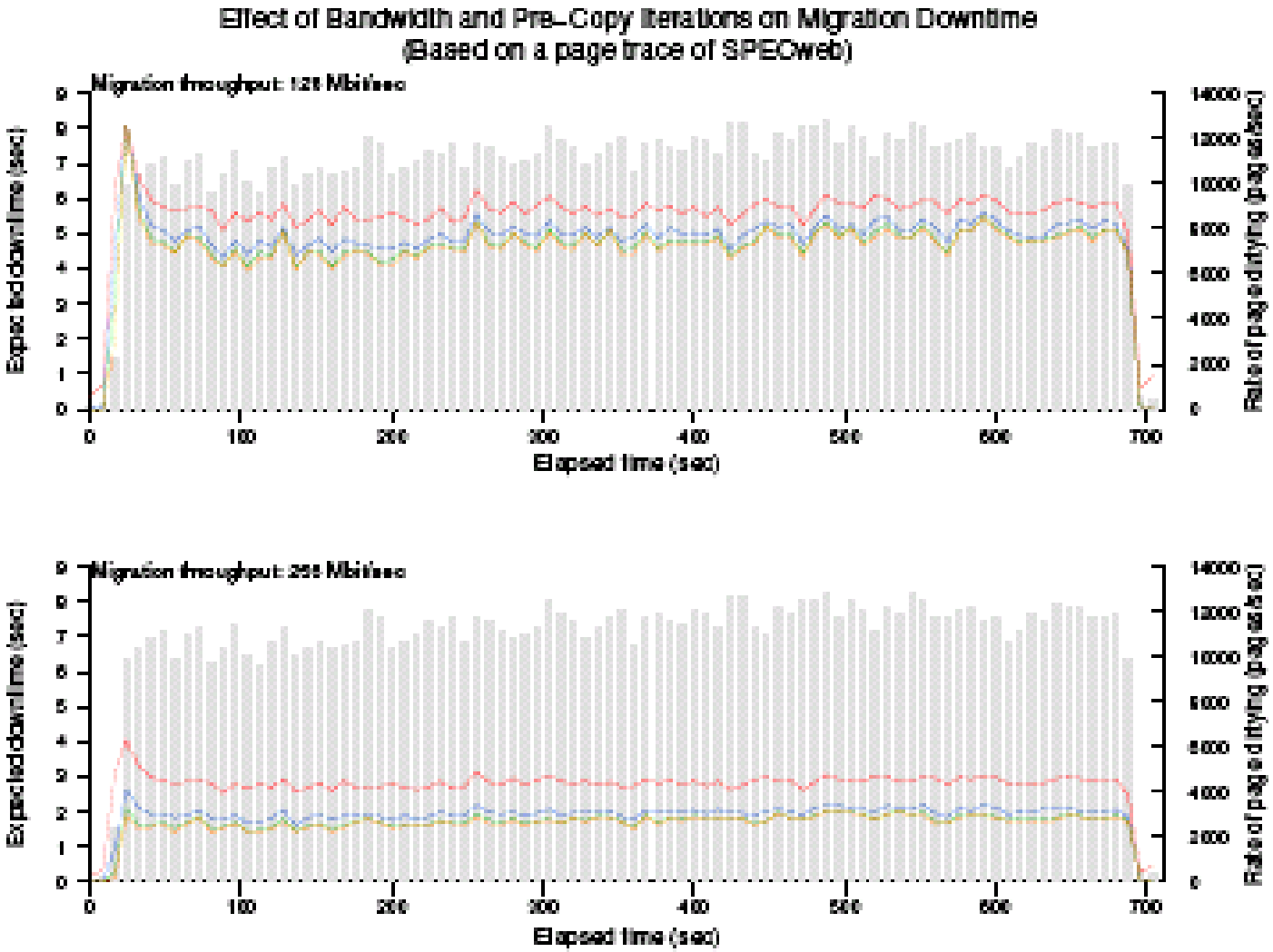


Figure 9: Results of migrating a running SPECweb VM.

Przykład migracji (SPECweb99)



Downtime z dla migracji z jedną iteracją

Downtime dla migracji z dwoma iteracjami

Downtime dla migracji z trzema iteracjami

Downtime dla migracji z czterema iteracjami

Nasz przykład migracji

- Sprzęt: VMware Server na Celeron M 1.4 Ghz, 760MB RAM.
- Virtual Host A: Debian Etch, 192 MB RAM, Xen 3.0.3, serwer NBD
- Virtual Host B: Debian Etch, 192 MB RAM, Xen 3.0.3
- Guest: Debian Etch, 48MB, partycja root i swap po NBD, server Icecast2 (64kbps, 44kHz, stereo)
- Prędkość sieci: ~1MB/s

Gdyby jeszcze był czas...

- Historia się tworzy
  Ludzie stojący za XENem
  Kto używa XENa?

O tworzącej się historii

- Projekt rozpoczęto i nadal jest rozwijany na uniwersytecie w Cambridge.
- W międzyczasie została powołana do życia firma

Do ciekawych idei należy Xeno servers. Projekt mający na celu stworzenie systemu ogólno dostępnych serwerów opartych na Xenie, na których możnaby wykonywać własny kod. Opłaty byłyby pobierane za czas kompilacji i wykonania. Zaangażowani w ten projekt są:

Evangelos Kotsovinos

Ian Pratt

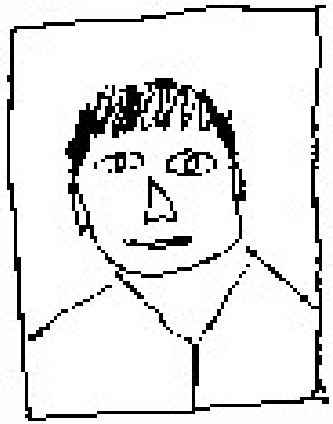
Steve Hand

Jon Crowcroft

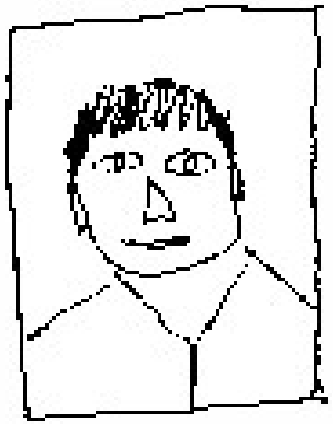
Keir Fraser

Theodore Hong

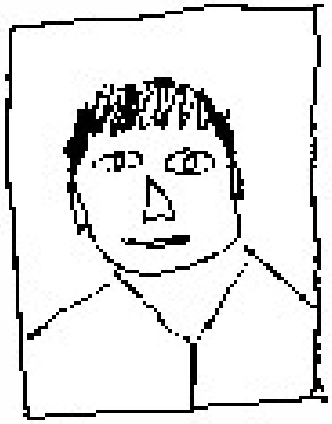
Shen Lin



Rob Bradford

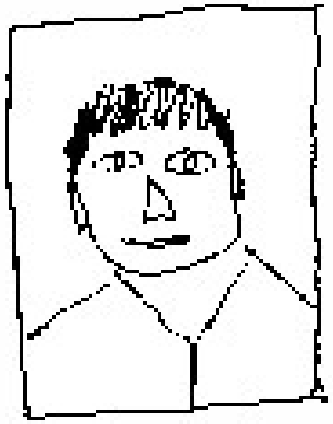


Lutz Schneider



Kiril Hristov

Thomas Fricke



Tom Wilkie

Kto używa XENa?

- Wygląda na to, że jeszcze nikt...
- Powodem jest zbyt, krótka obecność na rynku i nie potwierdzona stabilność. Ale pojawiają się doniesienia, że coraz to nowe firmy chcą przejść na technologię Xen source.

Dziękujemy.

Źródła danych i obrazków

- http://en.wikipedia.org/
- http://www.xensource.com/
- http://wiki.xensource.com/xenwiki/
- http://www.cl.cam.ac.uk/research/srg/netos/xen/performance.html
- http://continuum.asi.pwr.wroc.pl/linuxacademy/images/4/40/LinuxAcademy_XEN_23.pdf
- http://www.cl.cam.ac.uk/research/srg/netos/papers/2003-xensosp.pdf
- Ian Pratt xen 3.0 ststus report
- http://www.cl.cam.ac.uk/research/srg/netos/papers/2005-migration-nsdi-pre.pdf
- http://www.cl.cam.ac.uk/research/srg/netos/papers/ian-status.ppt