HighLoad++

Управление памятью в гипервизоре

Все о виртуализации памяти в Parallels

Анна Воробьева

Parallels™

В облака за эффективностью

Утилизация памяти

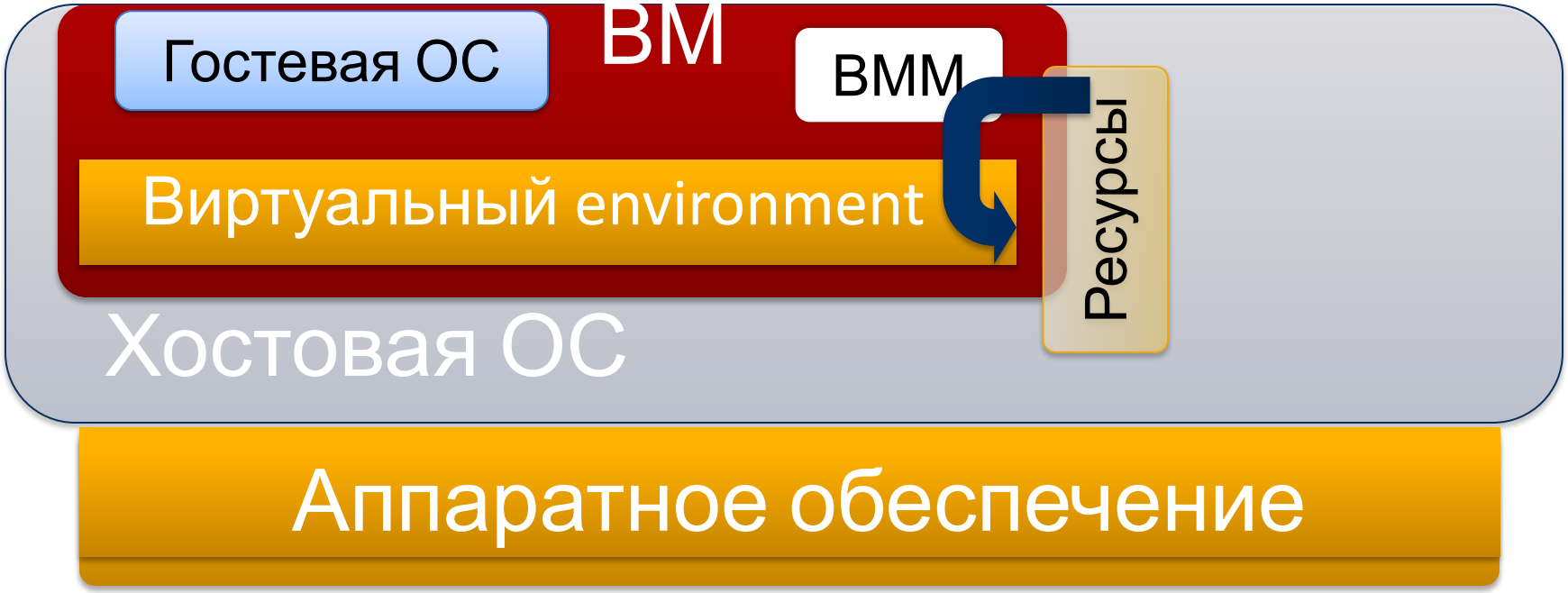
Мифы и страхи overcommit-а

Знания для безопасной виртуализации памяти

Содержание

- ✔ Постановка задачи
- ✔ Решения
  - ✔ Квоты, выбор backing store, алгоритма вытеснения
  - ✔ Balloon
  - ✔ Page sharing, compression
- ✔ Сравнение по продуктам

Немного терминологии



Задача распределения памяти

Разграничим термины

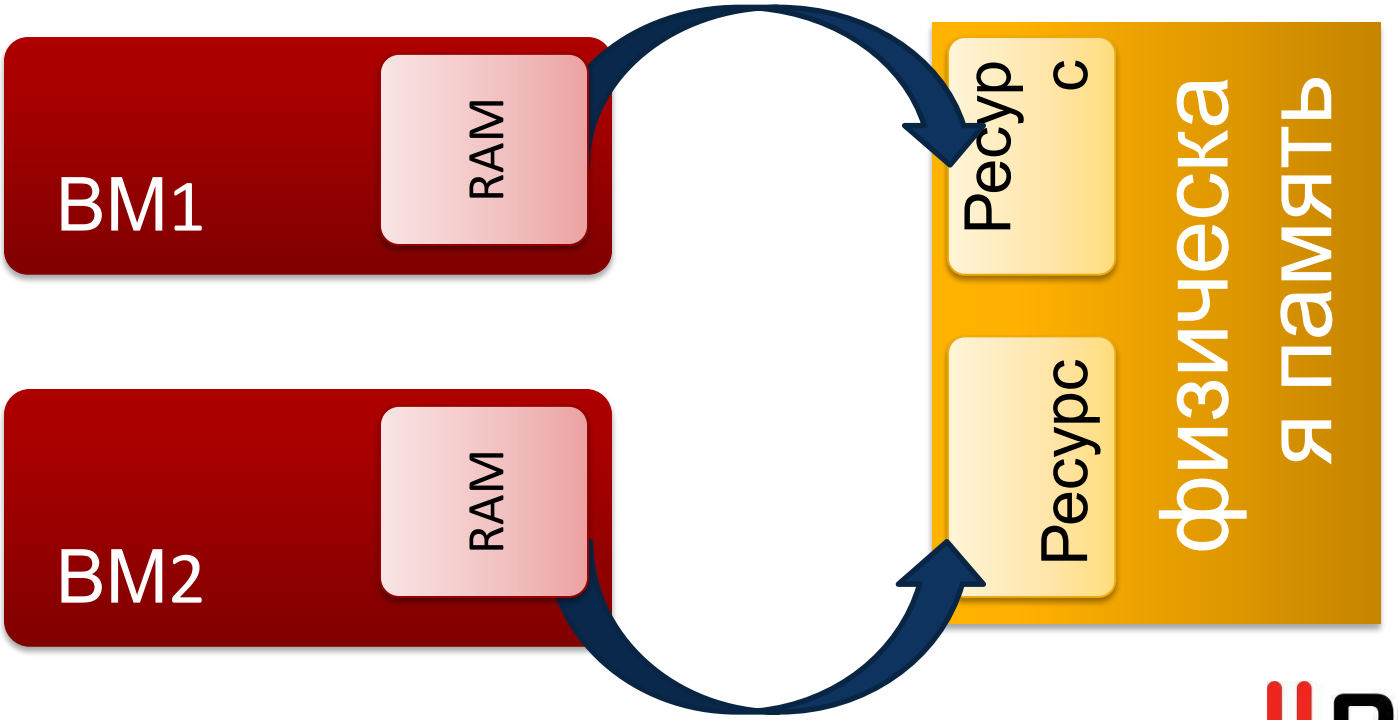
Overcommitment

- ✔ Σ(сконфигурированной памяти) + накладные расходы ≥ разрешеный лимит ноды
- ✔ ВМ подлежат всем действиям, описанным в докладе

Overload

- ✔ Σ(используемой памяти) + накладные расходы ≥ разрешенный лимит ноды
- ✔ ВМ подлежат миграции

Распределение памяти: шаг 1



Алгоритмы вытеснения

- ✔ LRU (last recently used)
- ✔ FIFO (first in first out)
- ✔ Aging (+NFU)
- ✔ NRU (not recently used – A-/D- bits)
  - а ведь еще можно дать всем второй шанс
- ✔ Clock
- ✔ Random

Алгоритмы вытеснения не работают

- ✔ Гостевая ОС вытесняет страницы по своим алгоритмам (semantic gap)
- ✔ Отсутствие локальности обращений
- ✔ ОС не может поместить в процесс своего агента, а мы можем

Office-битва (Windows 2008 x64)

Вытеснение (swapping)

- ✔ Avg Cycle Time = 345000
- ✔ Overcommit = 42%

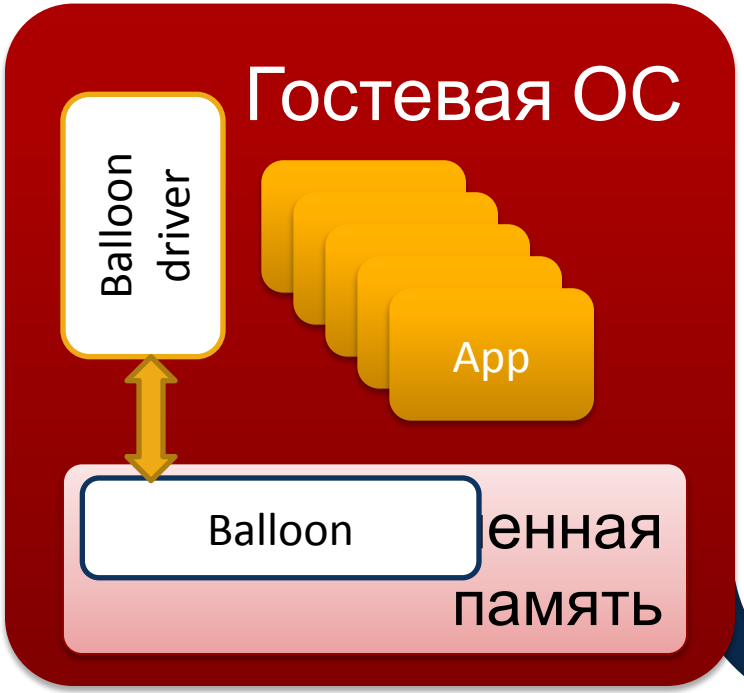
Ballooning

- ✔ Avg Cycle Time = 222000
- ✔ Overcommit = 93%

В 1.5 раза эффективнее

Parallels™

Ballooning



Ресурс памяти

Balloon – это гостевой драйвер

Страницы, отданные ВММу balloon-ом, не потребуются гостю и не содержат информации

Ballooning

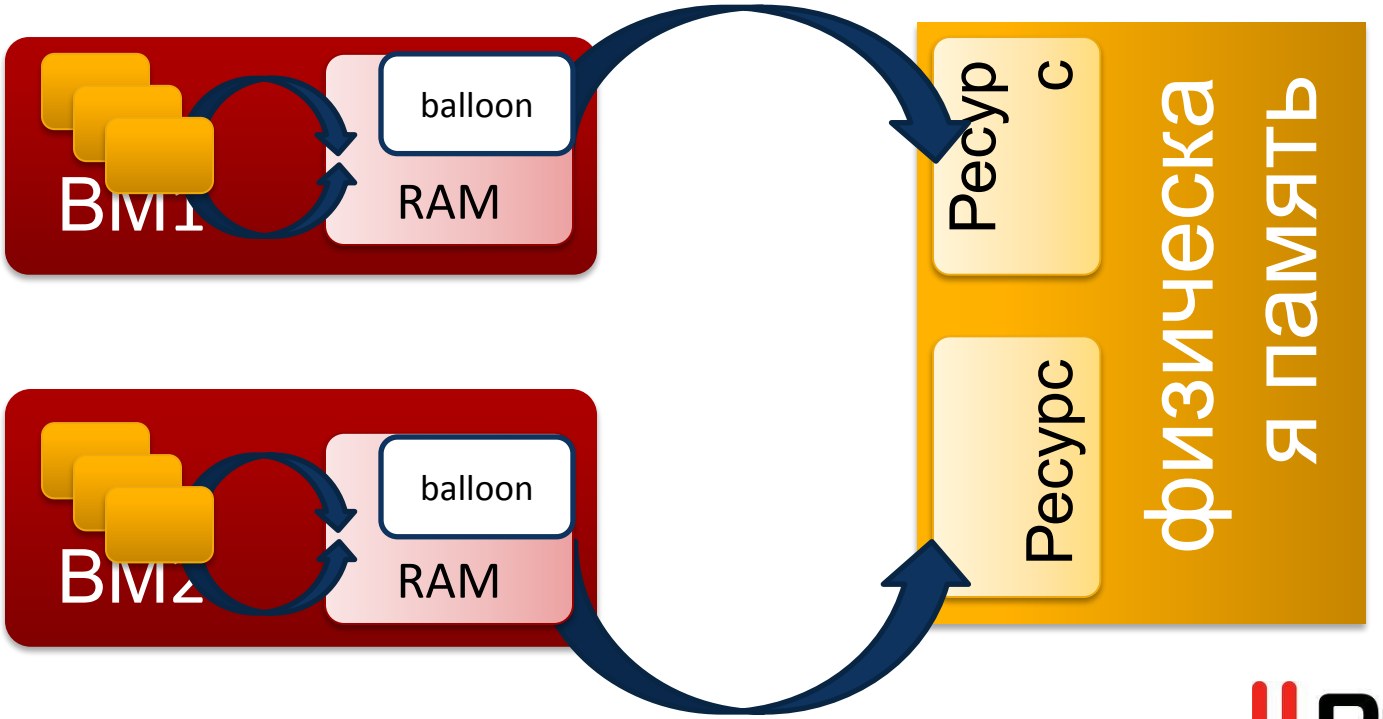
Плюс

- ✔ Сокращение подкачки между ВММ и гостем

Минусы

- ✔ Гостевой своппинг вплоть до гостевых крешей
- ✔ Неуниверсальность
- ✔ Отсутствие гарантий

Распределение памяти: шаг 2



Но откуда известен объем ресурса?

Конфигурационные данные

- ✔ Гарантия
- ✔ Лимит
- ✔ Приоритет/доля
- ✔ Разрешенный лимит ноды

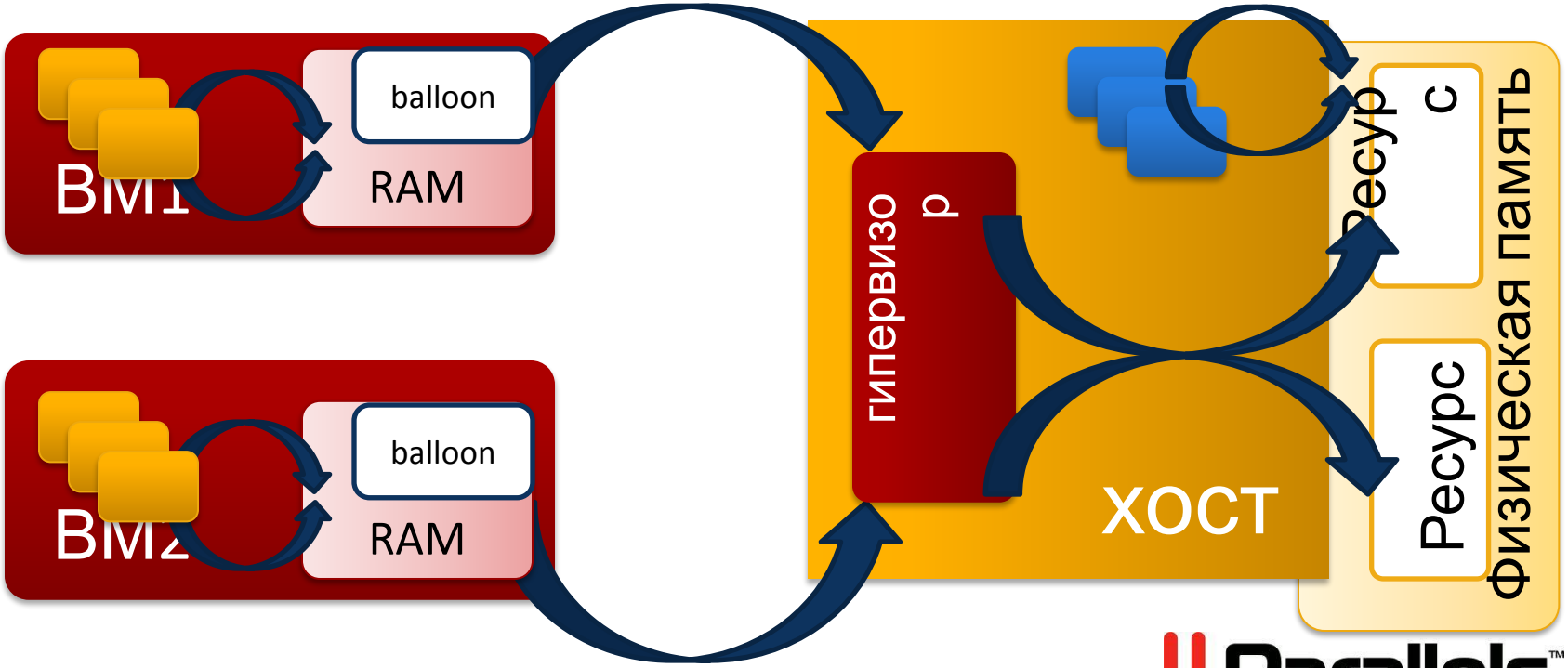
Динамические данные

- ✔ Используемая память
- ✔ Бездействующая память (idle)
- ✔ Статистические данные

Опасности конфигурируемых данных

- ✔ Избыток назначенной памяти (32 no-pae + 4GB RAM)
- ✔ Своп из-за низкой гарантии
- ✔ Незаслуженный дефицит при лимите меньше назначенной памяти
- ✔ Оптимистичный лимит для ноды

Распределение памяти: шаг 3



Backing storage

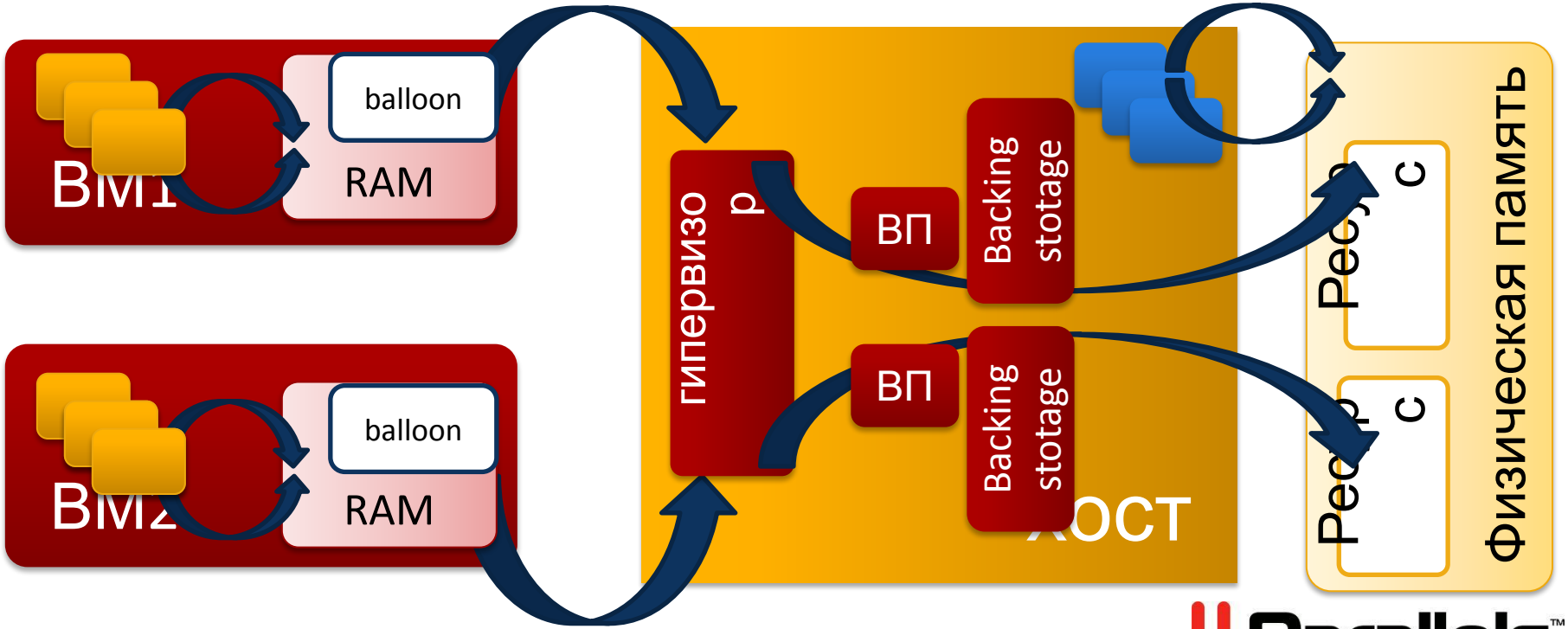
Влияет на

- ✔ Suspend/snapshot
- ✔ Resume/switch to snapshot
- ✔ Подкачка

Популярные решения

- ✔ File mapping
- ✔ Anonymous mapping
- ✔ HugeTlbFs

Распределение памяти: шаг 4



Parallels™

Меняем тики на данные

Page sharing

- ✔ Посчитать хэш
- ✔ Сравнить
- ✔ Защитить по COW
- ✔ По записи отвязать
  - Для Read-Only страниц

Compression

- ✔ Сжать
- ✔ Оставить в кэше либо записать на диск
- ✔ По требованию развернуть
  - Для редко используемых

Меняем тики на данные

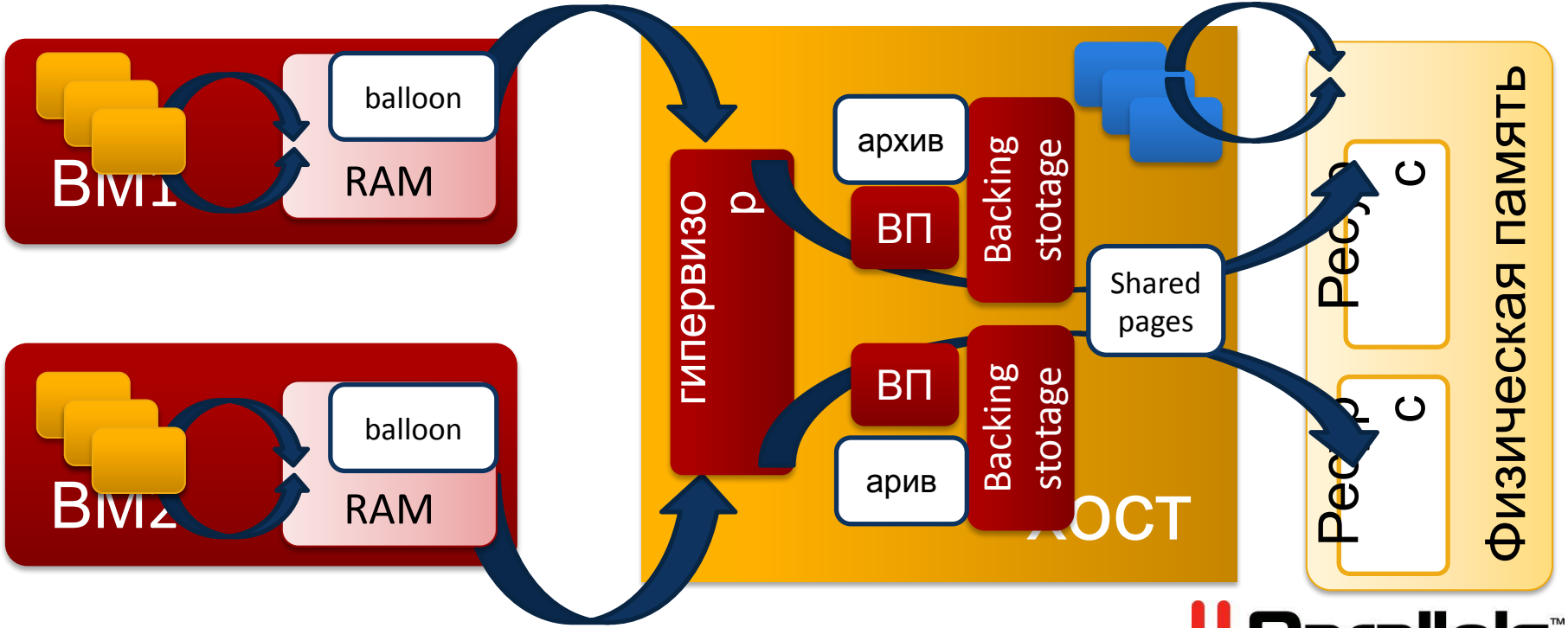
Page sharing

- ✔ Эффективность при одинотипной нагрузке на ноду
  - Накладные расходы могут быть напрасны
  - Запрет на большие страницы

Compression

- ✔ Интеграция с suspended image
- ✔ Уверенный compression вне зависимости от нагрузки

Распределение памяти: полная картина



Parallels™

Disclaimer

СРАВНЕНИЕ ПРОДУКТОВ, ПРЕДСТАВЛЕННОЕ ДАЛЕЕ, ЯВЛЯЕТСЯ ЛИШЬ МНЕНИЕМ.

	Balloon	Page sharing	Compres-sion	Swapout	Quota	Backing store
ESX	+	+	+	+	+	Huge/VSWP?
Xen Server	+	-	-	-	+	Own
PSBM	+	-	+	+	+	Huge/File/An
HyperV	+	-	-	-	+	File
KVM	+	+	-	+	+	Huge/File/An

Сравнение: Xen Server – осторожность превыше всего

- ✔ Исключительно ballooning
- ✔ Page-sharing & swapout присутствуют в xen hypervisor 4.0

Сравнение: VMWare ESX – сильнейшие со времен Waldspurger-а

- ✔ В статье 2002ого года они уже описывают balloon, квоту, page sharing, idle-memory tax swapout
- ✔ Некоторая инертность в новом, compression не интегрирован с suspend-ом

Сравнение: KVM – все блага Linux-а

- ✔ Balloon включен в дерево Linux
- ✔ Эффективнейший KSM достался бесплатно
- ✔ Блага надежного вытеснения
- ✔ Compression и алгоритмы, специфичные для виртуализации, могут идти с запозданием

Сравнение: HyperV – все что не от нас, то от лукавого

- ✔ Hot-plug memory + balloon
- ✔ Оверкоммит опасен и вреден

Сравнение: PSBM

- ✔ Свой алгоритм компрессии и его интеграция:
  - ✔ Эффективная реализация для разнотипной нагрузки
  - ✔ Быстрый suspend/resume/snapshot
- ✔ Для однотипной нагрузки – контейнеры

	Balloon	Page sharing	Cmprs	Swap	Quota	Backing store	Usage
ESX	+	+	+	+	+	Huge/VSWP?	Hi-end ENT
Xen Server	+	-	-	-	+	Own	Providers, middle Linux ENT
PSBM	+	-	+	+	+	Huge/File/An	Service-providers, ENT
HyperV	+	-	-	-	+	File	Windows middle ENT
KVM	+	+	-	+	+	Huge/File/An	Providers

Вопросы?

mailto: anyav@parallels.com

Parallels™