РФЯЦ-ВНИИТФ
РОСАТОМ

Оценочное тестирование распределенной системы хранения данных Ceph поверх высокоскоростной сети Альфа

28 сентября 2026

АО «НПО «КИС»

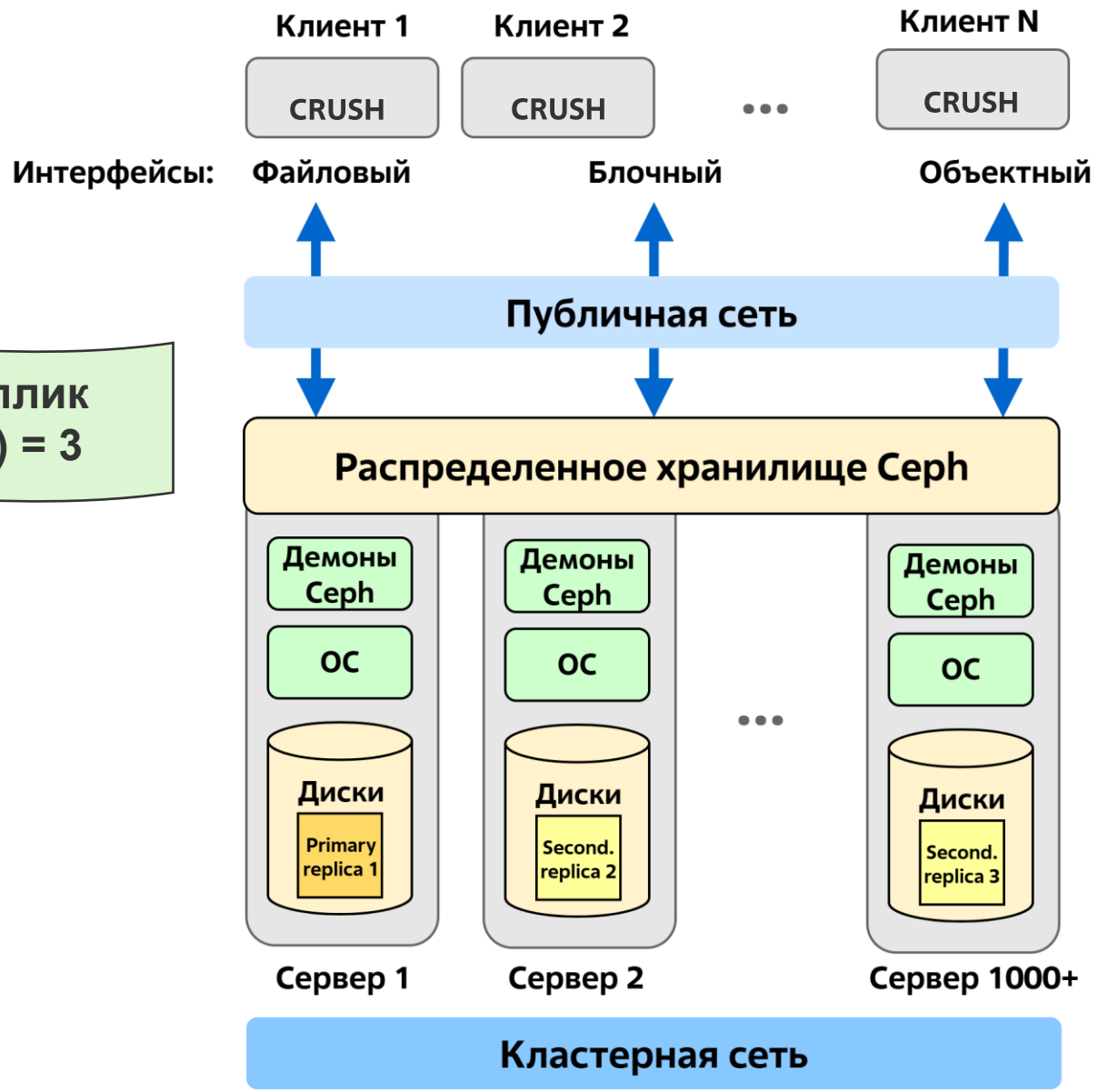
Семенов Александр Сергеевич

Главный инженер проекта, к.т.н.

Архитектура Ceph (2004–2026)

РФЯЦ-ВНИИТФ
РОСАТОМ

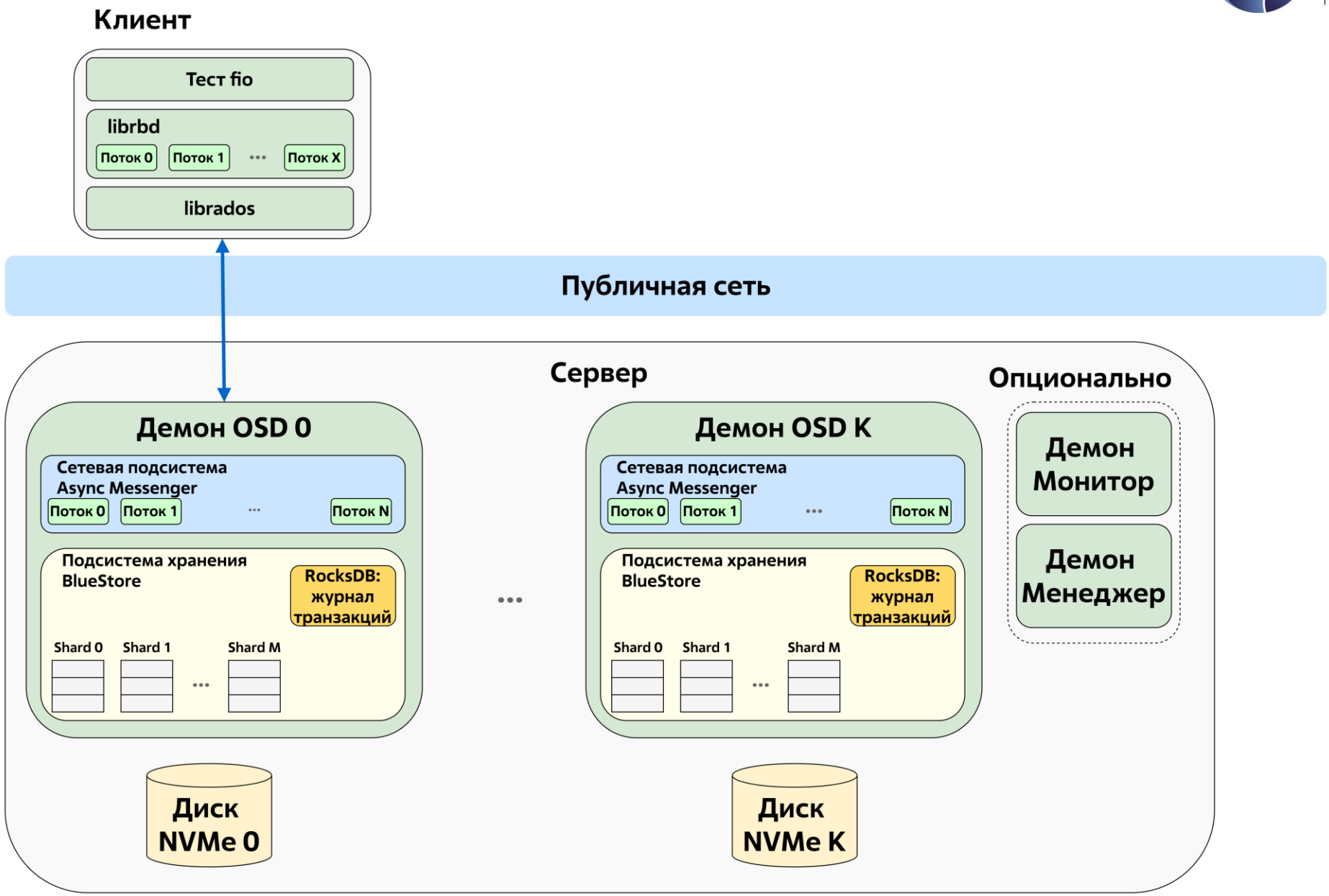
1. Количество реплик (replication factor) = 3



2. Нет метаданных: алгоритм CRUSH

3. Клиенты и сервера с данными могут совпадать!

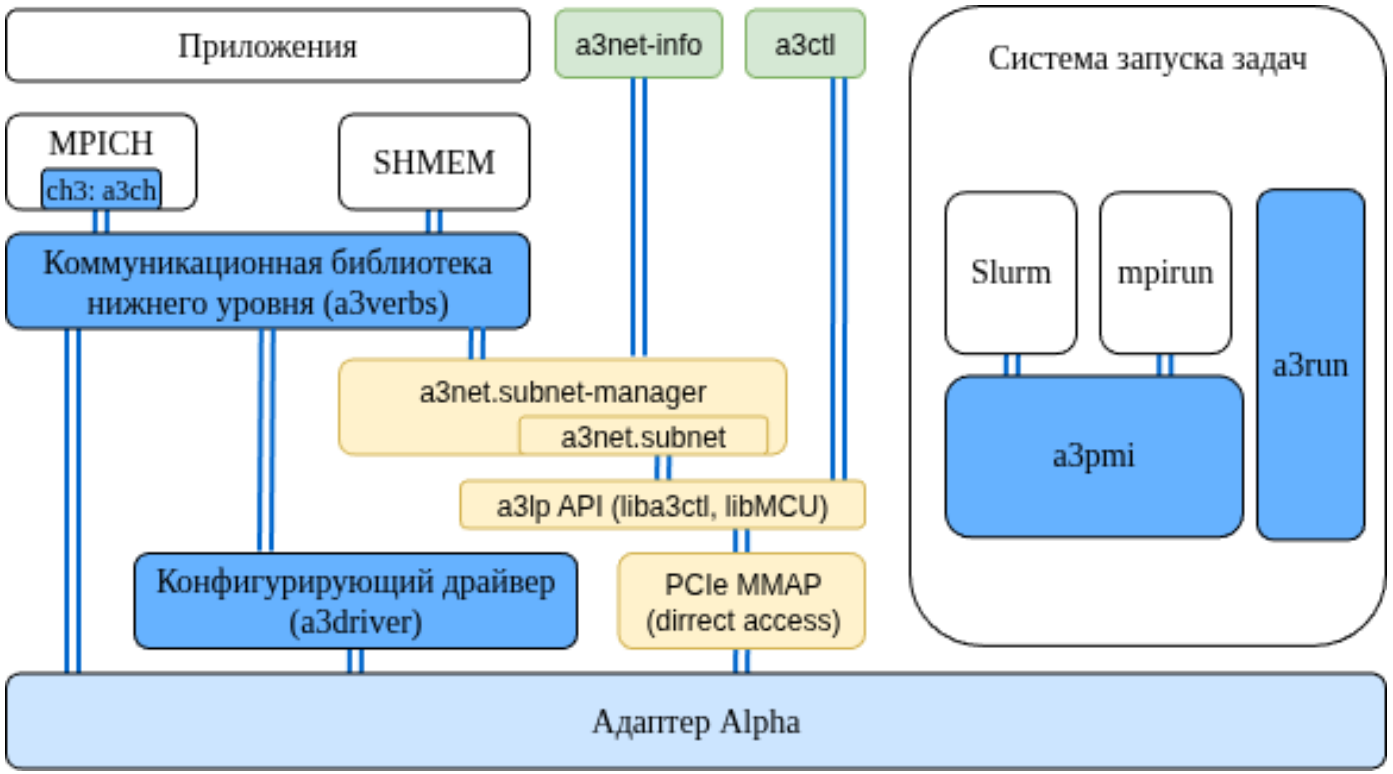
Взаимодействие, версия Ceph Tentacle 20.2.1



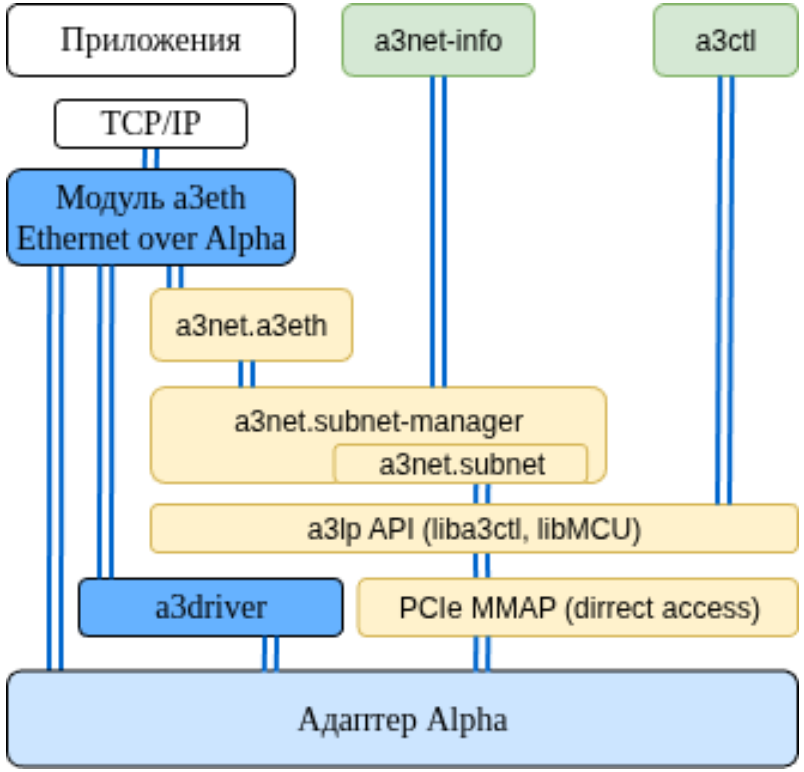
Архитектура программного обеспечения сети Альфа

РФЯЦ-ВНИИТФ
РОСАТОМ

Стек суперЭВМ



Универсальный стек



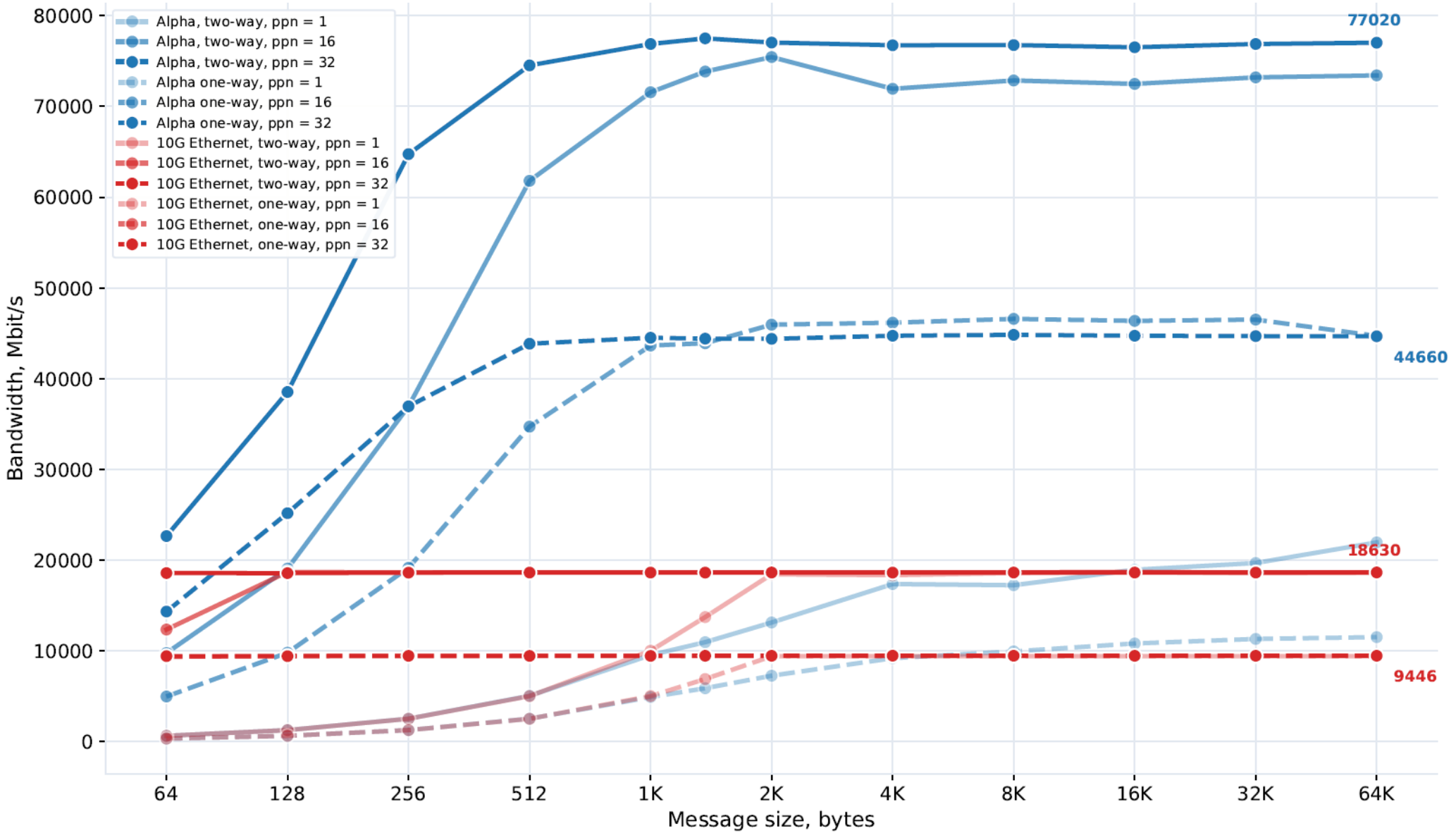
Вычислительный 3-узловой стенд

РФЯЦ-ВНИИТФ
РОСАТОМ

Компонент узла	Характеристика
Платформа	Asus RS700-E10-RS12U
CPU (Процессор)	2x Intel Xeon Gold 6330 2.00 ГГц
Ядра / Потоки (в узле)	56 ядер / 112 потоков
Память узла	256 ГБ
Диски	7x Intel SSDPF2KX960HZ 960 ГБ
Сеть 1: **Альфа**	Адаптер Альфа (подключен через коммутатор Альфа)
Сеть 2: **10G Ethernet**	Intel X710-AT2 10 Gbit/s Ethernet (подключен через коммутатор D-Link DXS-3400-24TC)
ОС, ядро	Debian Linux 12, 6.1.0-41-amd64

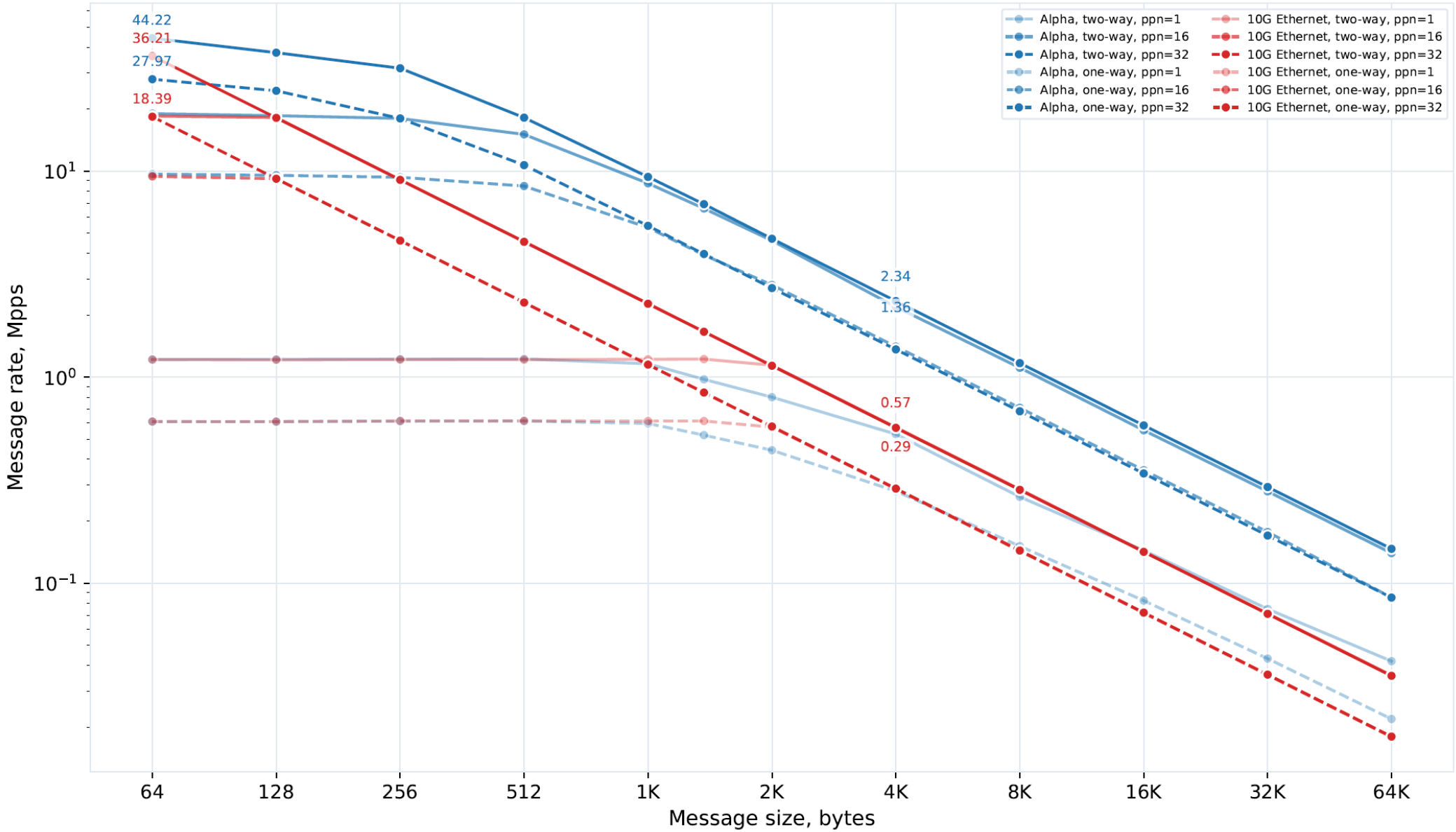
Модуль a3eth, iperf: пропускная способность

РФЯЦ-ВНИИТФ
РОСАТОМ



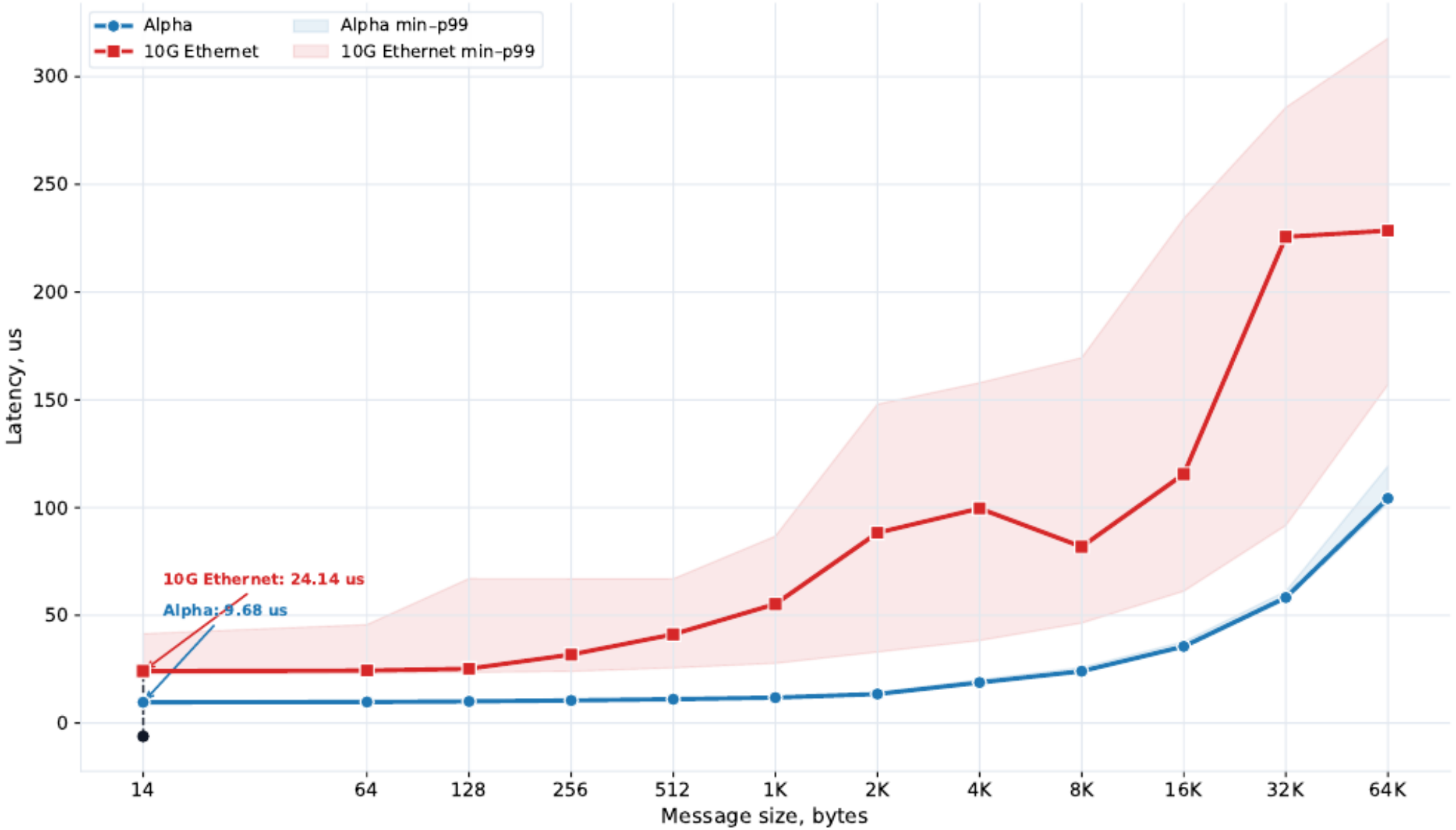
Модуль a3eth, iperf: темп выдачи сообщений

РФЯЦ-ВНИИТФ
РОСАТОМ



Модуль a3eth, sockperf: задержка

РФЯЦ-ВНИИТФ
РОСАТОМ



Настройка хранилища и параметры тестов fio



РФЯЦ-ВНИИТФ
РОСАТОМ

- Используется **блочный интерфейс RBD** (Rados Block Device) – виртуальные блочные устройства Ceph
- В **распределенном на 3 узла** хранилище данных создано 2 RBD диска по 400 ГБ
- Общий объем распределенного хранилища = 2 RDB диска * 400 ГБ * фактор репликации 3 = 2.4 ТБ

Тест	Размер блока	Описание	Размер очереди (iodepth)	numjobs	Количество RBD дисков
read	4 МБ	Последовательное чтение	32	4	2
write	4 МБ	Последовательная запись	4	6	2
randread	4 КБ	Случайное чтение	140	4	2
randwrite	4 КБ	Случайная запись	128	4	2

Стабильность результатов

РФЯЦ-ВНИИТФ
РОСАТОМ

$$\mathrm{CV} = s/\bar{x} \times 100\%$$

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

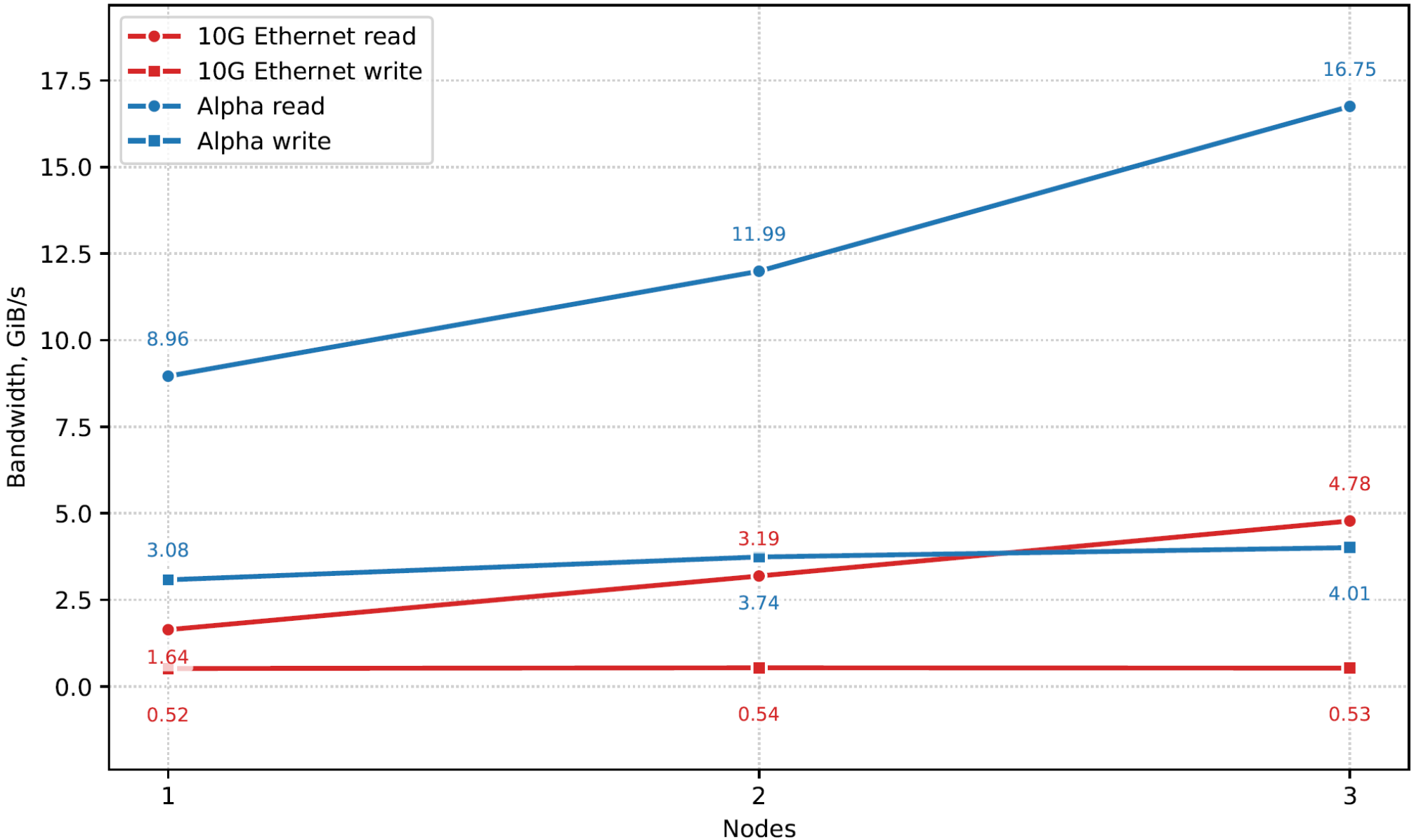
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$n = 5$ итераций запуска каждого теста по 5 минут

	Тест	Альфа	10G Ethernet
CV	read	5.4%	1.9%
	write	6.9%	2.5%
	randread	1.2%	0.9%
	randwrite	1.5%	1.3%

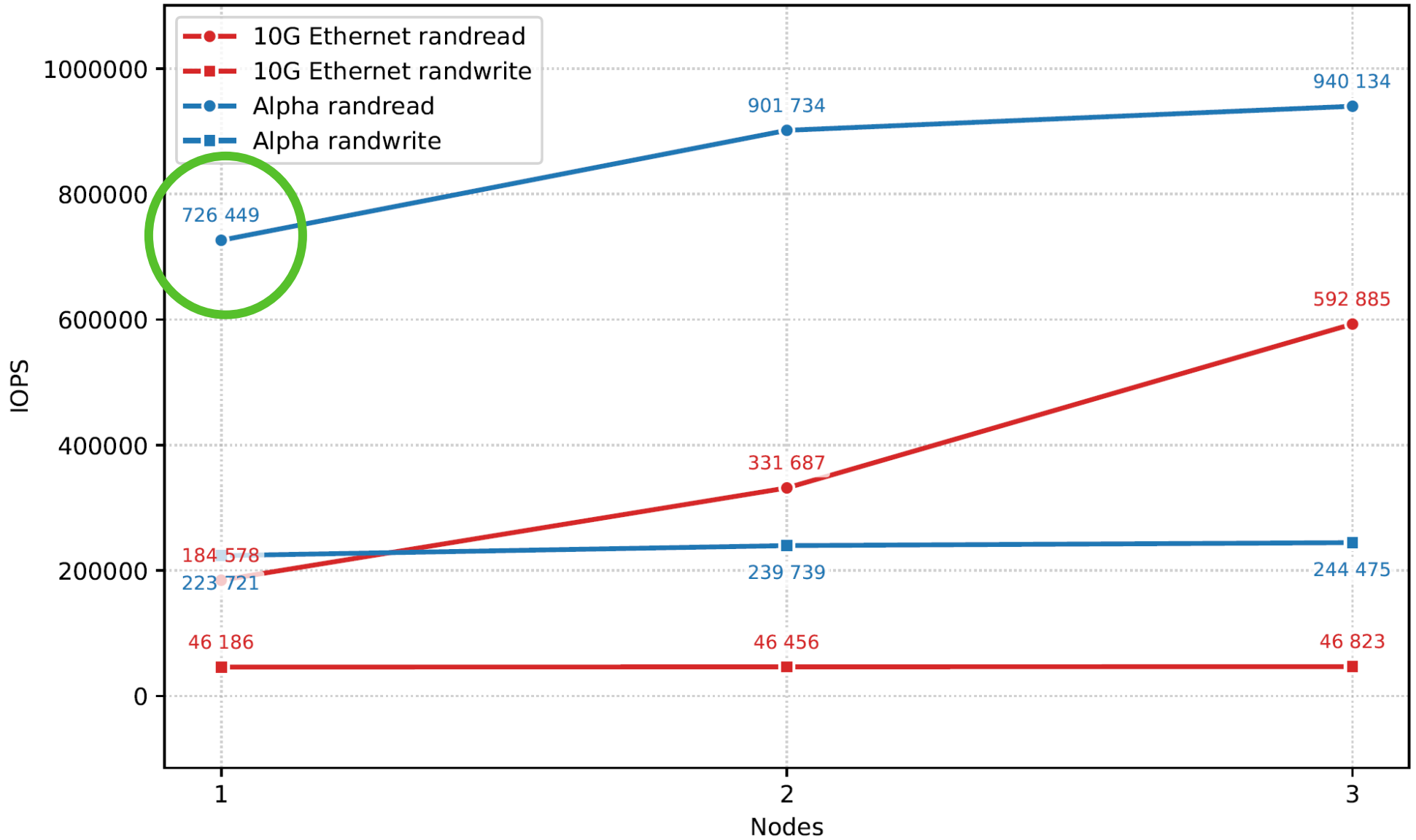
Производительность
Тесты read и write

РФЯЦ-ВНИИТФ
РОСАТОМ



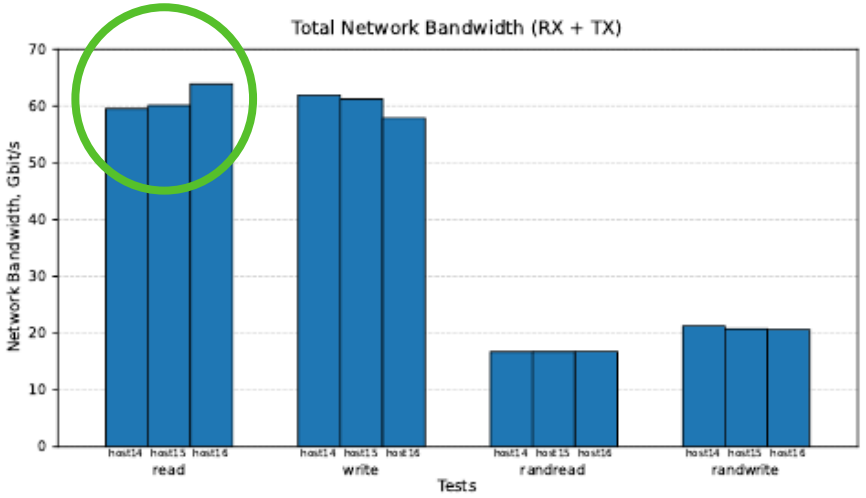
Производительность
Тесты randread и randwrite

РФЯЦ-ВНИИТФ
РОСАТОМ

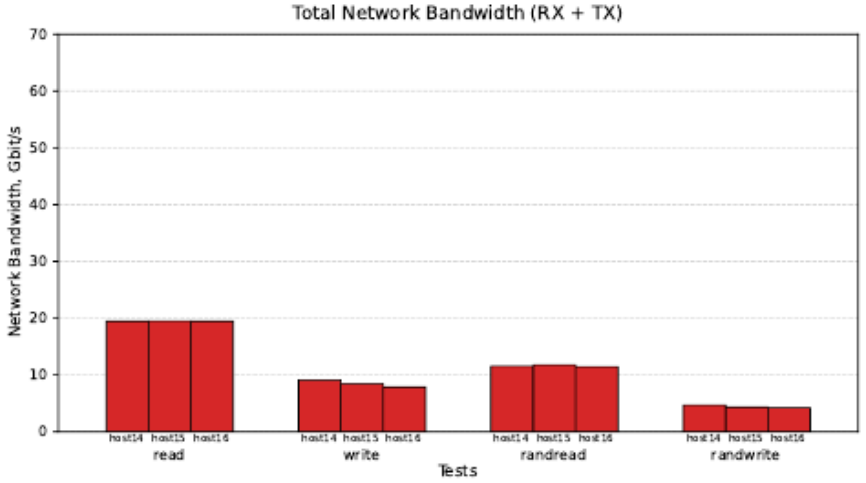


Профилирование: sar, iostat

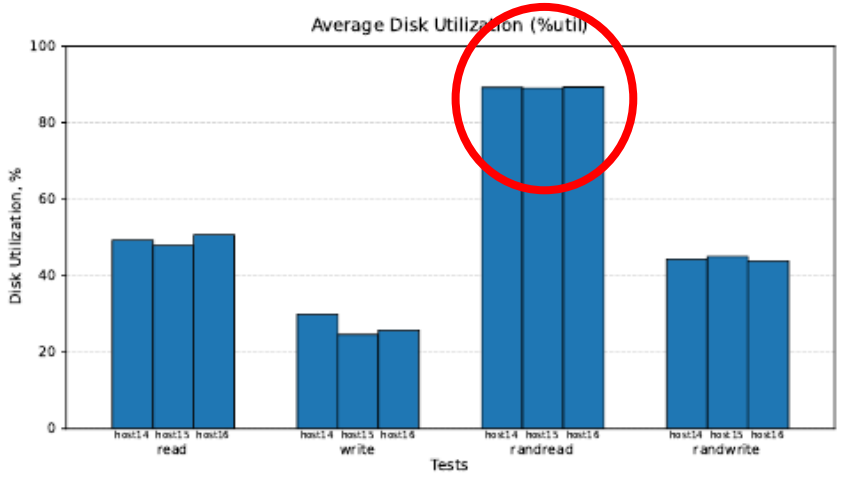
РФЯЦ-ВНИИТФ
РОСАТОМ



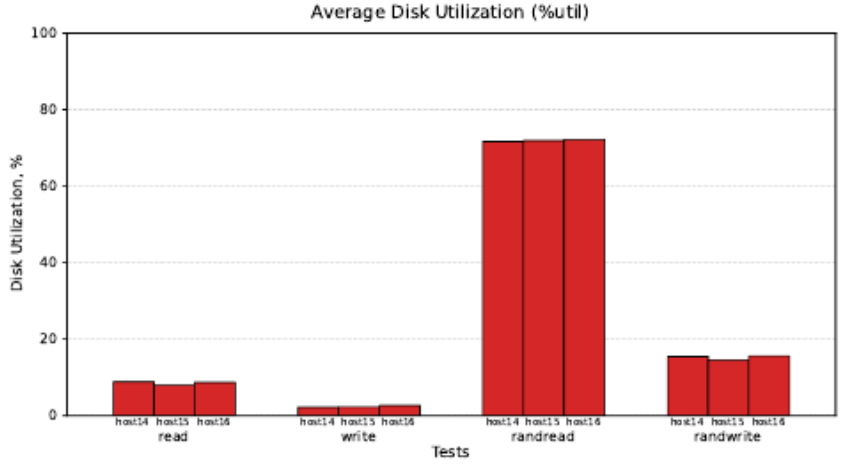
(c) Сеть: Ceph через сеть Alpha



(d) Сеть: Ceph через сеть 10G Ethernet



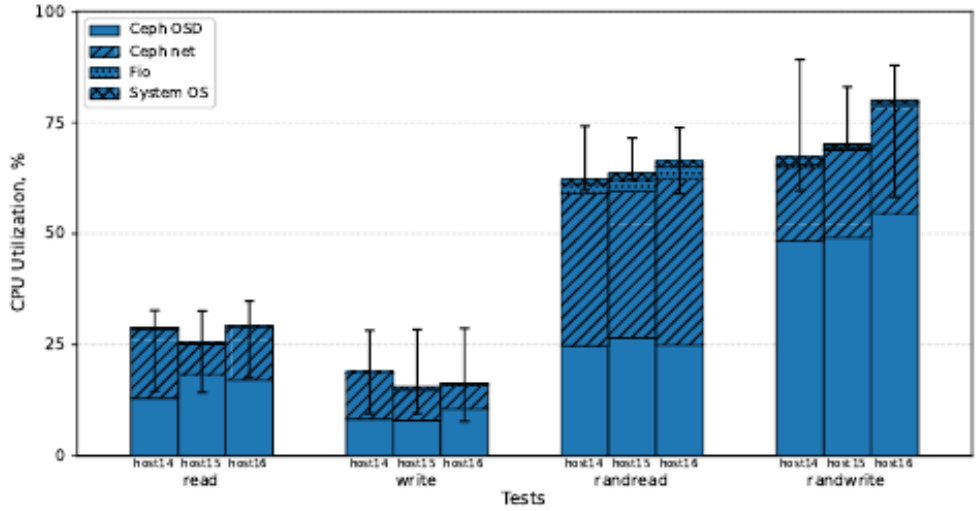
(e) Диски: Ceph через сеть Alpha



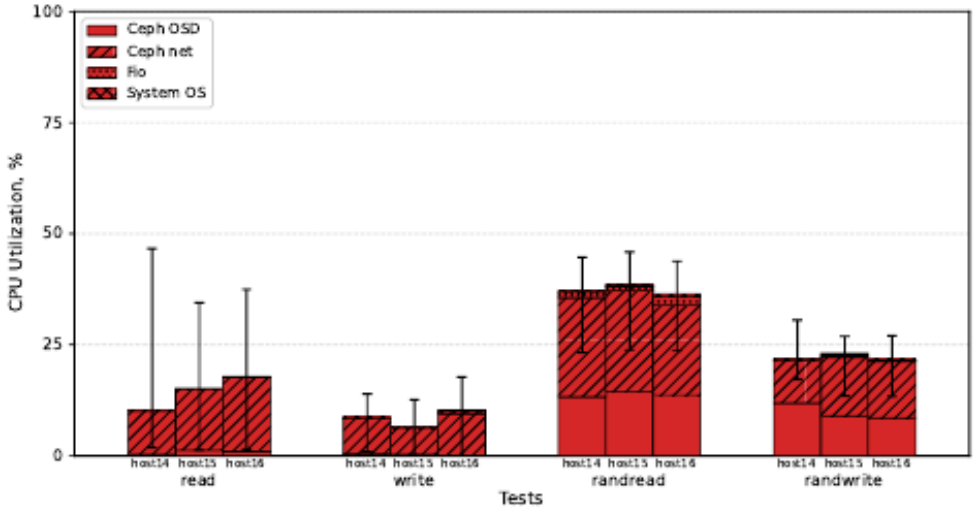
(f) Диски: Ceph через сеть 10G Ethernet

Профилирование: perf sched, mpstat

РФЯЦ-ВНИИТФ
РОСАТОМ



(a) CPU: Ceph через сеть Альфа

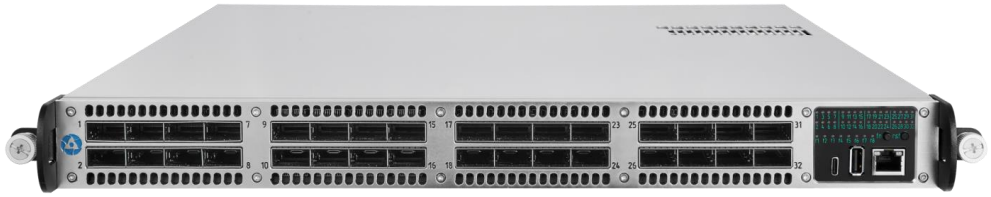


(b) CPU: Ceph через сеть 10G Ethernet

Заключение и дальнейшие планы

РФЯЦ-ВНИИТФ
РОСАТОМ

- Ceph стабильно функционирует на сети Альфа



- Улучшение программного модуля поддержки Ethernet поверх сети Альфа
- Использование SPDK
- Необходимо сравнение на большом кластере
- Сравнение с более мощной сетью

Спасибо за внимание!

Семенов Александр Сергеевич

AleksandrSSemenov@rosatom.ru