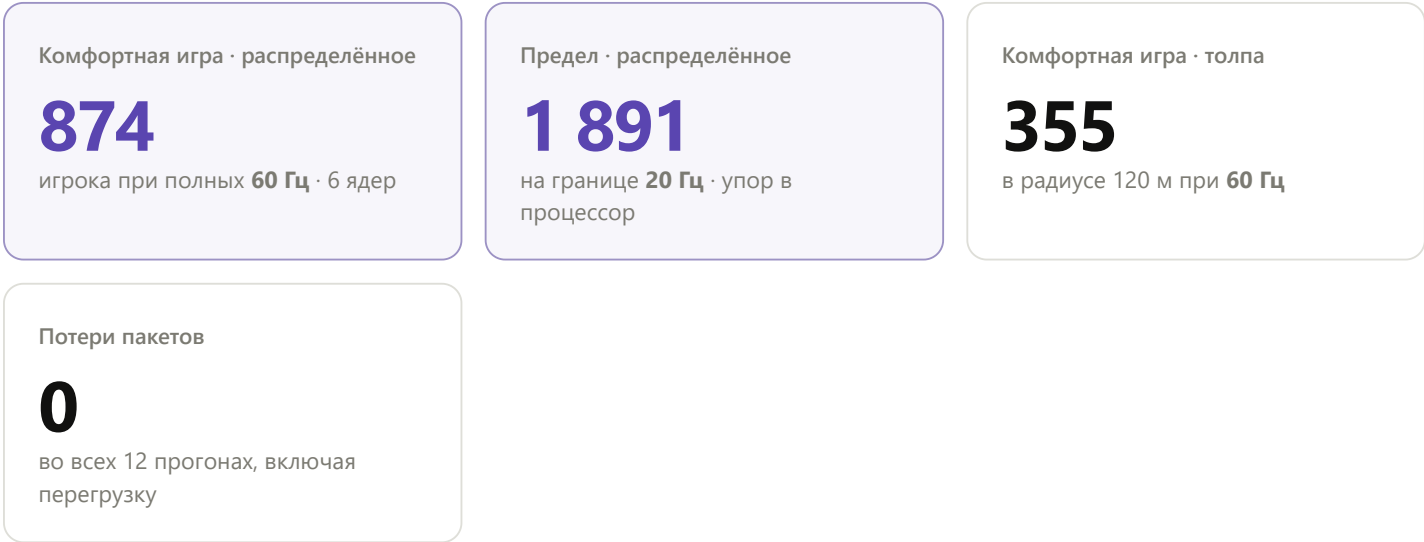


CloudV:MP — предельная ёмкость и профиль нагрузки

Двенадцать нагрузочных прогонов в единых условиях с имитацией реального поведения игроков. Два сценария размещения доведены до полного исчерпания процессорного ресурса на двух конфигурациях ядер. Приведены полные распределения времени такта, отзывчивости и сетевого профиля. Все значения получены прямым измерением.

Машина: AMD EPYC-Rome · 8 ядер @ 4,49 ГГц · 8 ГБ RAM · Linux Целевой темп: 60 Гц (16,67 мс/такт)
Окно замера: 45 с после стабилизации Метрик снято: 145



Ключевые значения приведены для конфигурации из 6 ядер сервера. Полные данные обеих конфигураций — в таблицах ниже.

01 Резюме

Сервер удерживает целевой темп 60 Гц при **874** распределённых по карте игроках и при **355** игроках, сосредоточенных в одной точке. Предел играбельности (20 Гц) достигается при **1 891** и **780** игроках соответственно. Различие вместимости в 2,4 раза определяется плотностью размещения: вычислительная стоимость растёт от числа взаимно видимых пар игроков, а не от их общего количества.

Ограничивающим ресурсом во всех режимах является центральный процессор — на пределе занято 4,2–4,7 ядра из шести. Оперативная память (0,3–1,6 ГБ) и пропускная способность канала (57–233 Мбит/с) не исчерпываются. Потери пакетов и ожидание на внутренних блокировках равны нулю во всех прогонах, включая режимы с трёхкратным превышением бюджета такта: под перегрузкой сервер снижает частоту

обновления, но сохраняет целостность и порядок доставки. Медианный интервал обновления состояния игрока в рабочем режиме составляет 48 мс, размер одной записи состояния — 18 байт.

02 Архитектура движка

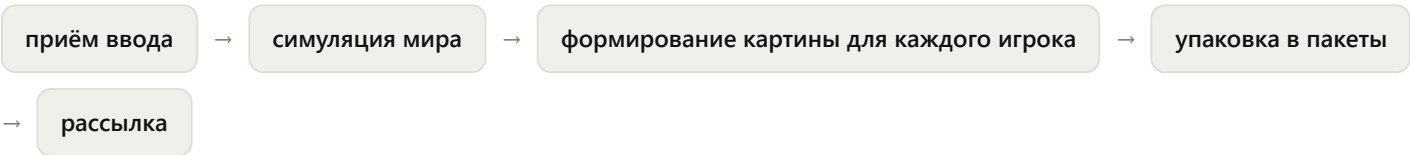
Раздел поясняет, как устроен серверный движок, за счёт чего достигаются приведённые показатели и где находятся ограничения. Изложение рассчитано на технического читателя без погружения в детали реализации.

Модель: авторитетный сервер

Сервер является единственным источником истины о состоянии игрового мира. Клиент передаёт лишь ввод и состояние принадлежащих ему объектов; решение о том, что видит и получает каждый игрок, принимает сервер. Такая модель (server-authoritative) противоположна одноранговой схеме, где клиенты доверяют друг другу: сервер не принимает на веру ни позицию, ни действия клиента, что является основой защиты от несанкционированного вмешательства. Серверная часть написана на языке Rust, клиентская построена на движке игры (сборка GTA V 1.0.3788).

Такт обновления

Мир пересчитывается фиксированное число раз в секунду (целевой темп — 60). Один такт состоит из последовательности этапов, каждый из которых измеряется отдельно:



Если сумма этапов укладывается в бюджет такта (16,67 мс при 60 Гц), сервер держит целевой темп; при превышении темп снижается. Формирование индивидуальной картины мира и упаковка составляют около двух третей такта — это и есть определяющие производительность этапы (раздел 07).

Ключевые механизмы

- **Область интереса (AOI).** Игроку передаётся не весь мир, а только объекты поблизости. Мир разбит на пространственную сетку; для каждого игрока обсчитываются лишь соседние ячейки. Поэтому стоимость определяется локальной плотностью, а не размером карты.
- **Работа с лимитом движка на 128 персонажей.** Игровой клиент по своей природе отображает ограниченное число персонажей одновременно — это ограничение самого движка GTA. У нас есть механизм, который позволяет держать в одной точке несколько сотен игроков, а на сервере — тысячи, при этом у каждого клиента остаётся плавная картинка ближайших. Как именно это устроено — часть нашего ноу-хау, детали не раскрываем.
- **Клапан плотности.** Под перегрузкой сервер ограничивает число одновременно транслируемых одному игроку объектов, отдавая приоритет ближайшим. Это

удерживает нагрузку управляемой на массовых скоплениях.

- **Параллельный такт.** Формирование картины мира для игроков распределяется по ядрам процессора, что позволяет наращивать вместимость увеличением числа ядер (раздел 06).
- **Сжатие состояния.** Состояние объектов квантуется и кодируется как разность от опорного кадра, что сокращает размер записи до 18 байт и удерживает сетевой расход низким.

■ **Сильные стороны**

- Авторитетный сервер — устойчивость к клиентскому вмешательству по построению.
- Параллельный такт — вместимость растёт с числом ядер.
- Низкое потребление памяти — тысячи игроков в пределах 0,3–1,6 ГБ.
- Под перегрузкой снижает темп, но не теряет пакеты и не рассинхронизируется.
- Компактный сетевой формат — 18 байт на запись состояния.

■ **Ограничения**

- Картина мира кодируется заново для каждого наблюдателя — в толпе это доминирующая стоимость.
- Один узел ограничен процессором; концентрация в одной точке — наиболее дорогой случай.
- Отображение 128 персонажей в кадре — предел движка игры, а не сервера.
- Реальная производительность арендованных машин расходится до 20 % при равных характеристиках.

■ **Направления развития**

- **Общее кодирование.** Для наблюдателей с совпадающей картиной упаковывать состояние один раз, а не отдельно для каждого — снимает основную долю стоимости упаковки.
- **Групповая обработка.** Обсчитывать близко стоящих наблюдателей совместно, поскольку их области интереса почти совпадают.
- **Распределение по узлам.** Разделение мира между несколькими серверными процессами даёт кратный рост общей вместимости (масштабирование по населению).

03 Методика измерения

Нагрузка формируется флотом программных клиентов, подключающихся к серверу по штатному сетевому протоколу. Синтетический поток пакетов не применяется: каждый клиент проходит полное рукопожатие, авторизацию и создание персонажа, после чего нагружает те же обработчики сервера, что и реальный игрок.

Модель поведения клиентов

Поведение назначается каждому клиенту индивидуально, в пропорциях типовой активности ролевого сервера. Однородный флот нагрузил бы единственный обработчик и зависил бы оценку.

ПОВЕДЕНИЕ	ДОЛЯ	НАГРУЖАЕМЫЙ ПУТЬ СЕРВЕРА
Передвижение пешком	42 %	приём и рассылка состояния персонажа

ПОВЕДЕНИЕ	ДОЛЯ	НАГРУЖАЕМЫЙ ПУТЬ СЕРВЕРА
Управление транспортом	28 %	приём состояния машины, экстраполяция скорости
Стрельба и урон	14 %	разбор боевых событий, компенсация задержки
Обмен сообщениями	11 %	гарантированный канал доставки
Ожидание	5 %	удержание сессии

Фактическая активность подтверждается счётчиками сервера: в установившемся режиме регистрируется **90–200 боевых событий/с**, до **250 подтверждённых попаданий/с** с компенсацией задержки, до **38 сообщений/с** и **330–740 тысяч записей состояния/с**.

Сценарии размещения

Сценарий А — распределённое население. Клиенты размещаются по карте с весами по населённым зонам: городской центр и прилегающие районы получают основную долю, пригороды — меньшую, трассы и незаселённые территории — фоновую. Соответствует типовой картине живого сервера и определяет показатель суммарной вместимости.

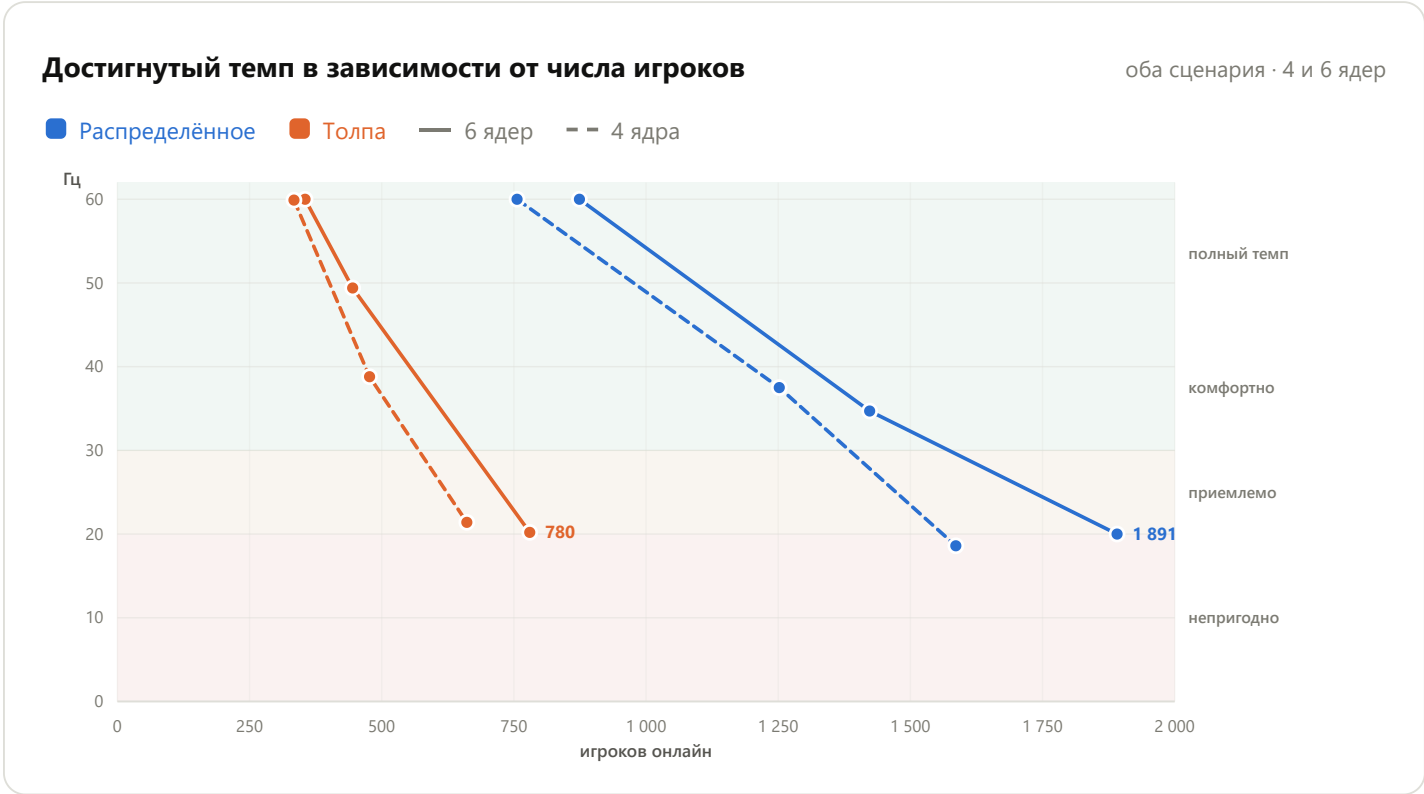
Сценарий В — концентрация в одной точке. Всё население в круге радиусом 120 метров. Предельный случай: стоимость определяется числом взаимно видимых пар. Соответствует массовым мероприятиям.

Разделение вычислительных ресурсов

Генератор нагрузки выполняется на той же машине, поэтому ядра закреплены раздельно: при совместном использовании он конкурировал бы с сервером за процессор и занижил бы результат. Измеренное потребление генератора не превышает одного ядра даже при флоте в 7 000 клиентов. Основная серия снята в конфигурации «6 ядер — серверу, 2 — генератору»; серия на 4 ядрах приведена для оценки масштабирования.

04 Кривая ёмкости

Горизонтальная ось — число подключённых игроков, вертикальная — достигнутый темп обновления. Цвет соответствует сценарию, тип линии — числу ядер. Фоновые полосы — зоны субъективного восприятия темпа игроком.



Все двенадцать прогонов. «Соседей» — медианное число других игроков, обсчитываемых для одного клиента; именно эта величина определяет вычислительную стоимость. Такт — интервал между тактами, р50 / р95 / р99 — процентиля за окно. ЦП — загруженных ядер, RAM — резидентная память процесса. Строки предела и полного темпа выделены.

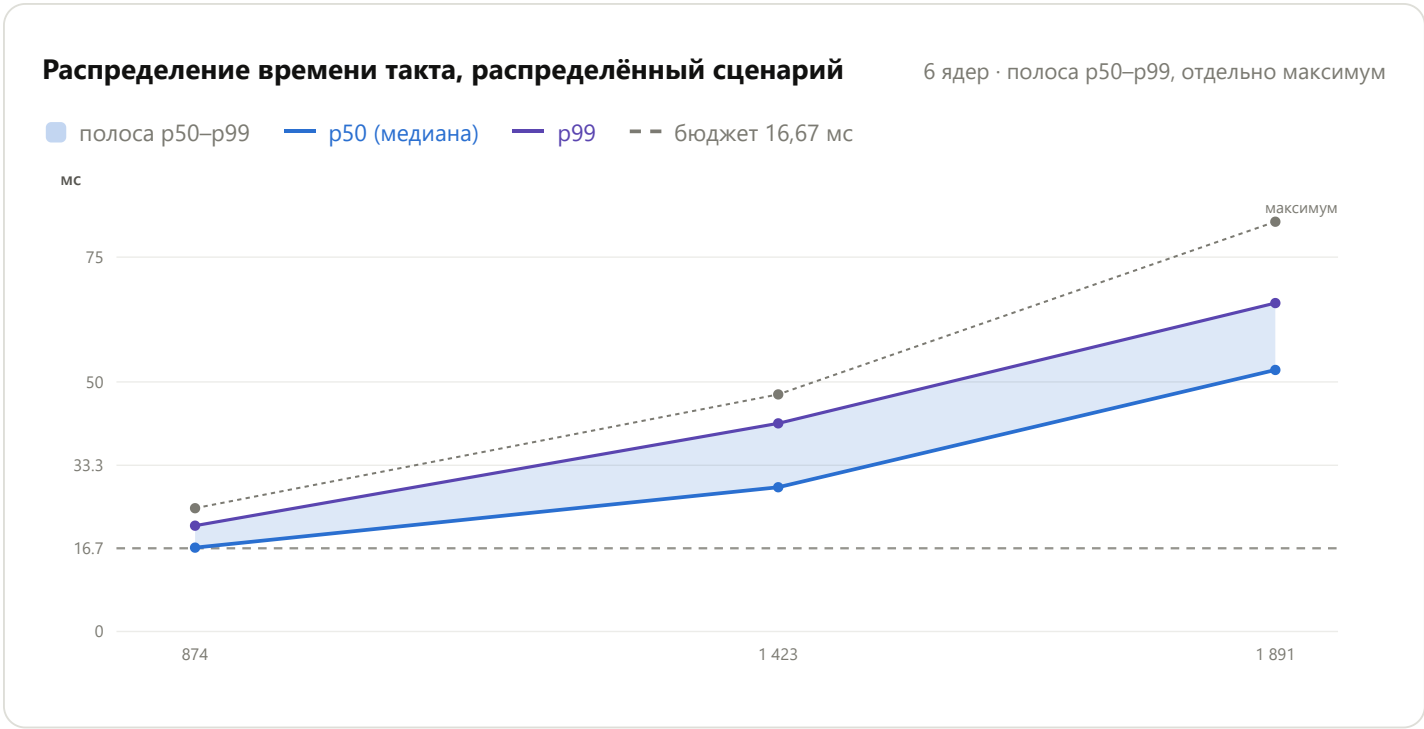
СЦЕНАРИЙ	ЯДЕР	ИГРОКОВ	ТЕМП	ТАКТ P50	P95	P99	МАКС	СОСЕДЕЙ	ЦП	RAM	РЕЖИМ
Толпа	4	334	59.9 Гц	16.6	18.9	21.2	27.2	333	3.0	410 МБ	полный темп
Толпа	4	477	38.8 Гц	25.6	30.4	32.6	38.8	476	3.1	588 МБ	комфортно
Толпа	4	661	21.4 Гц	45.7	55.1	59.7	68.2	660	3.4	870 МБ	приемлемо
Толпа	6	355	60.0 Гц	16.7	17.8	18.5	23.0	354	3.3	434 МБ	полный темп
Толпа	6	445	49.4 Гц	19.9	23.1	24.7	28.1	444	4.0	532 МБ	полный темп
Толпа	6	780	20.2 Гц	49.0	55.3	60.6	63.5	779	4.7	1.0 ГБ	приемлемо
Распределённое	4	756	60.0 Гц	16.8	19.4	21.4	25.4	8	2.4	334 МБ	полный темп
Распределённое	4	1 252	37.5 Гц	25.5	39.1	43.5	48.8	13	3.2	862 МБ	комфортно
Распределённое	4	1 586	18.6 Гц	55.7	66.6	71.4	81.8	16	3.5	1.5 ГБ	непригодно
Распределённое	6	874	60.0 Гц	16.8	19.0	21.2	24.7	8	3.2	373 МБ	полный темп
Распределённое	6	1 423	34.7 Гц	28.9	38.3	41.7	47.5	14	4.7	1.1 ГБ	комфортно
Распределённое	6	1 891	20.0 Гц	52.4	61.5	65.8	82.1	18	4.2	1.6 ГБ	приемлемо

Зависимость вместимости от плотности размещения

Один и тот же сервер вмещает **1 891** распределённого игрока и **780** при концентрации в одной точке. Различие в 2,4 раза объясняется столбцом «соседей»: в распределённом сценарии на одного клиента приходится порядка 18 взаимно видимых игроков, в концентрированном — до 779. Показатель вместимости корректен только в привязке к сценарию размещения.

05 Задержки и отзывчивость

Средний темп не отражает равномерности: игрок ощущает не среднее, а редкие затянутые такты. Ниже приведено полное распределение времени такта, а также интервал обновления состояния игрока — показатель, наиболее близкий к воспринимаемой отзывчивости.



Интервал обновления состояния игрока (owner update interval) — как часто сервер получает и ретранслирует новое состояние объекта игрока. p50 48 мс соответствует штатному темпу отправки клиента; рост p95 / p99 под нагрузкой отражает прореживание дальних объектов клапаном плотности, а не потерю данных. 6 ядер.

сценарий	игроков	темп	обновление p50	p95	p99	соседей p50	соседей p99	просадок такта >33 мс
распределённое население								
Распределённое	874	60.0 Гц	48 мс	192 мс	288 мс	8	111	528
Распределённое	1 423	34.7 Гц	144 мс	432 мс	672 мс	14	183	1 519
Распределённое	1 891	20.0 Гц	336 мс	1391 мс	2064 мс	18	272	902
концентрация в одной точке								
Толпа	355	60.0 Гц	48 мс	96 мс	145 мс	354	354	201
Толпа	445	49.4 Гц	48 мс	97 мс	144 мс	444	444	2 225
Толпа	780	20.2 Гц	48 мс	337 мс	720 мс	779	779	909

06 Удельная стоимость и масштабирование по ядрам

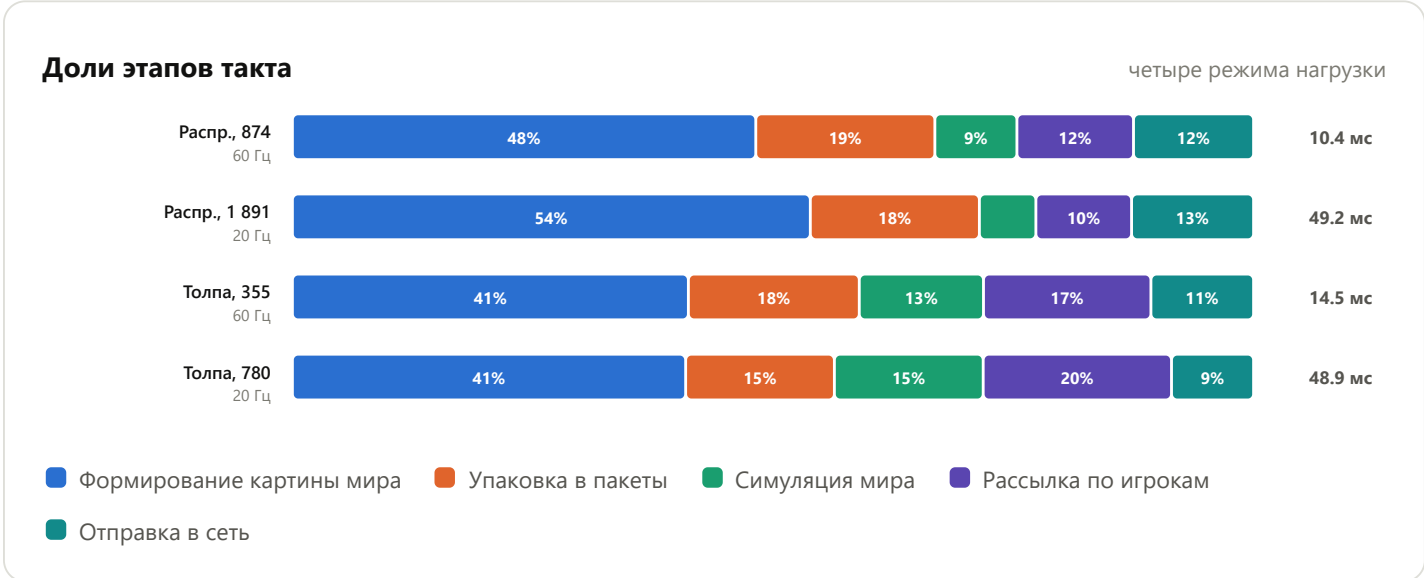
Слева — процессорное время на одного игрока за такт (при линейной зависимости линии были бы горизонтальными; наклон вверх означает сверхлинейный рост). Справа — изменение темпа при переходе с 4 ядер на 6 при сопоставимой нагрузке.



Удельная стоимость в распределённом сценарии возрастает с **12,0** до **26,3 мкс** при двукратном росте онлайн; в концентрированном она втрое выше и достигает **63,3 мкс**. Переход на 6 ядер повышает вместимость на границе 20 Гц с 1 586 до **1 891** игрока (+19 %). Прирост темпа умереннее прироста вместимости, поскольку часть работы выполняется последовательно: **наращивание ядер расширяет суммарный онлайн, но не повышает предел концентрации в одной точке пропорционально** — последний определяется структурой алгоритма.

07 Структура такта по этапам

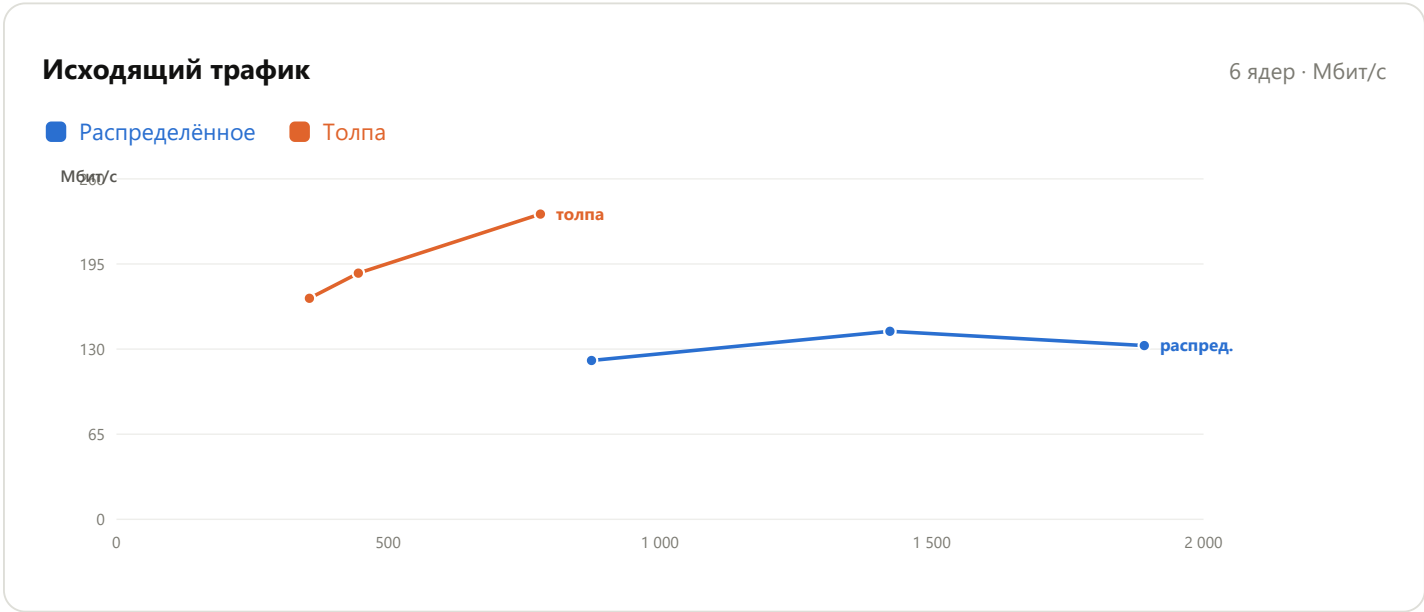
Каждый такт декомпозирован по этапам обработки; сумма измеренных этапов покрывает 99 % времени такта. Структура устойчива во всех режимах и определяет приоритет оптимизации.



Формирование индивидуальной картины мира занимает 41–54 % такта, упаковка в сетевые пакеты — ещё 15–23 %. Совокупно на эти два этапа приходится около двух третей вычислений; они и определяют предел производительности. Симуляция мира, рассылка и отправка в сеть в сумме не превышают трети такта.

08 Сеть, пакеты и репликация

Сетевой профиль и объём репликации. Исходящий трафик на игрока с ростом онлайна снижается, поскольку частота отправки убывает с расстоянием до объекта.



Сетевой и репликационный профиль, 6 ядер. «На игрока» — исходящий трафик на клиента. «Пакетов/с» — темп датаграмм. «Байт/запись» — размер одной записи состояния на проводе. «Записей/с» — темп авторинга сетевых записей.

сценарий	игроков	исходящий	входящий	на игрока	пакетов/с исх.	пакетов/с вх.	байт/запись	записей/с
распределённое население								
Распределённое	874	121 Мбит/с	31 Мбит/с	139 кбит/с	100 648	72 358	18.3 Б	377 928
Распределённое	1 423	144 Мбит/с	25 Мбит/с	101 кбит/с	96 668	58 492	18.3 Б	568 336
Распределённое	1 891	133 Мбит/с	11 Мбит/с	70 кбит/с	57 889	23 542	18.3 Б	728 264
концентрация в одной точке								
Толпа	355	169 Мбит/с	17 Мбит/с	475 кбит/с	110 279	36 497	18.2 Б	738 265
Толпа	445	188 Мбит/с	20 Мбит/с	422 кбит/с	132 332	44 296	18.3 Б	737 008
Толпа	780	233 Мбит/с	27 Мбит/с	299 кбит/с	225 354	64 356	18.3 Б	553 840

09 Полный справочник измеренных параметров

Все ключевые параметры для трёх характерных режимов: распределённое население при полном темпе, оно же на пределе, и толпа на пределе. Значения вычислены из сохранённых срезов метрик и проверяемы.

Сводка параметров по трём режимам. Каждая строка — измеренная величина.

ПАРАМЕТР	РАСПР. · 874 · 60 ГЦ	РАСПР. · 1 891 · ПРЕДЕЛ	ТОЛПА · 780 · ПРЕДЕЛ
НАГРУЗКА			
Игроков подключено	874	1 891	780
Достигнутый темп, Гц	60.0	20.0	20.2
Соседей видимо, медиана	8	18	779
Соседей видимо, р99	111	272	779
ВРЕМЯ ТАКТА, МС			
Медиана (р50)	16.8	52.4	49.0
р90	18.0	59.2	53.9
р95	19.0	61.5	55.3
р99	21.2	65.8	60.6
Максимум	24.7	82.1	63.5
Просадок >33 мс за окно	528	902	909
ОТЗЫВЧИВОСТЬ			
Обновление игрока р50, мс	48	336	48
Обновление игрока р95, мс	192	1 391	337
Обновление игрока р99, мс	288	2 064	720
ПРОЦЕССОР И ПАМЯТЬ			
Загружено ядер	3.2	4.2	4.7
Стоимость на игрока, мкс	12.0	26.3	63.3
Резидентная память, МБ	373	1 612	1 014
СЕТЬ			
Исходящий, Мбит/с	121	133	233
Входящий, Мбит/с	31	11	27
На игрока, кбит/с	139	70	299
Пакетов/с исходящих	100 648	57 889	225 354
Пакетов/с входящих	72 358	23 542	64 356
Байт на запись состояния	18.3	18.3	18.3
Записей состояния/с	377 928	728 264	553 840
АКТИВНОСТЬ (В СЕКУНДУ)			
Боевых событий	198	93	174
Попаданий (компенс.)	11	11	252
Сообщений чата	21	18	17
Изменений области интереса	21	9	12
ЦЕЛОСТНОСТЬ			
Потеряно пакетов	0	0	0
Ожидание блокировки, мс	0.000	0.000	0.000
Уровень перегрузки	0	0	0

■ Меры контроля корректности

- Все показатели — разности за окно замера (счётчики накопительные).
- Сырые срезы метрик сохранены до и после каждого окна; вычисления повторяемы.
- Сопоставление — при равном числе подключённых клиентов.
- Генератор нагрузки изолирован по ядрам, его потребление фиксируется.
- Размещение и поведение каждого клиента детерминированы — прогон воспроизводим.

■ Границы применимости

- Производительность арендованных машин расходится до 20 % при равных характеристиках; абсолютные значения привязаны к экземпляру.
- Игровой клиент отображает не более 128 персонажей — ограничение движка игры.
- Частота кадров у клиента — самостоятельное ограничение, не зависящее от сервера.
- Измерения на 4 и 6 ядрах; поведение на 32+ ядрах требует отдельной проверки.

Все значения получены прямым измерением на тестовой машине и воспроизводимы через нагрузочный стенд. Ни одна цифра не является расчётной оценкой; прогнозы отмечены явно.