

SENTINEL AI™ V4.0 — АРХИТЕКТУРНЫЙ ПАСПОРТ
10-КОНТЕЙНЕРНОЙ ЭКОСИСТЕМЫ

Инженерный паспорт «Великолепной Десятки» микросервисов: сквозные потоки данных, голосовой пульт Voice C2 (Asterisk 20) и корпоративный мониторинг Zabbix SLA 99.99%.

< 15 мс	~728 МБ	99.99%	0% Износа
Акустический AED Детекция выстрелов и взрывов	RAM Footprint Всех 10 контейнеров на старте	Доступность SLA Без единой точки отказа	Защита SSD дисков 30-сек RAM RingBuffer в Go

АРХИТЕКТУРНЫЙ МАНИФЕСТ И СТРАТЕГИЧЕСКИЕ ЦЕЛИ V4.0:
В релизе 4.0 суверенная платформа SENTINEL AI полностью завершила переход к модульной микросервисной экосистеме, объединив стек технологий 21 века в 10 изолированных контейнерах под управлением Docker Compose. Комплекс устранил критические уязвимости устаревшего легаси (потреблявшего до 32 ГБ VRAM и уничтожавшего SSD непрерывной видеозаписью 24/7). С интеграцией двух стратегических подсистем — Asterisk Dual-Stack Voice C2 и Zabbix Enterprise Telemetry — платформа гарантирует замкнутый автономный цикл безопасности: от ультразвуковой детекции инцидента за <15 мс до мгновенного электромеханического локдауна дверей, голосового обзвона руководства по PSTN/GSM и вывода 60 FPS телеметрии в ситуационный центр.

Сквозная Топология Потокow Данных (Real-Time Pipeline):

Этап 1: Захват и Буфер	Этап 2: Интеллект ИИ	Этап 3: Реакция СКУД & Связь	Этап 4: Мониторинг NOC
[7] Go Stream Recorder • Захват RTSP/ONVIF 24/7 • 30с циклический RAM-буфер • Спектральный скремблер речи • Нулевой износ дисков SSD • SHA-256 хэширование WORM	[6] FastConformer AI • ONNX AED детекция (<15 мс) • Билингвальный STT (KZ/RU) • Tier-1 сито: отсеив 85% шума • Tier-2 LLM верификация • RTF >= 1.0 (быстрее звука)	[3, 4] Backend & [8] Asterisk • Локдаун Modbus / Shelly Pro • Автодозвон Казахтелеком PSTN • PJSIP пейджинг в динамике • Пульт Voice C2 IVR (PIN 7788) • Celery изоляция очередей	[9] Cockpit & [10] Zabbix • 60 FPS Canvas спектрограмма • WebSocket-лента алертов • Zabbix LLD авто-дискавери • Контроль VRAM GPU и RTF • 4 языка интерфейса

Сетевая Изоляция Docker и Политика Доступа (Network Security Mesh):

Шлюз / Контейнер	Внешний Порт	Протокол & Назначение	Политика Защиты & Изоляция
Backend API Gateway	8080:8000	REST API, ASGI WebSockets, СКУД реле	Ingress шлюз, Rate-limit, CORS, JWT аутентификация.
Mission Cockpit	3002:3000	Next.js 14 Web UI, SOC консоль, Портал	HTTPS / WSS проксирование через Nginx / Cloudflare.
Asterisk Dual SIP	5060, 5080	chan_sip (Провайдер), res_pjsip (IP-динамики)	CPS 1.5с рейтлимит, изоляция RTP 10000-10050, PIN-защита.
Zabbix Agent 2	10050:10050	Пассивные/активные проверки Zabbix Server	Доступ только с IP Zabbix Server / Proxy (PSK шифрование).
PostgreSQL & Redis	ИЗОЛИРОВАНЫ	Внутренняя шина данных и брокер задач	Строгий запрет внешнего доступа! Доступны только внутри сети sentinel-net.

ИНЖЕНЕРНЫЙ ПАСПОРТ ЭКОСИСТЕМЫ • ЧАСТЬ 1

Контур Базы Данных, Шины Событий и Ядра Управления
(Сервисы 1–5)

Детальные технические спецификации системных сервисов хранения, маршрутизации и асинхронного исполнения задач.

Сервис / Контейнер	Базовый образ & Порты	Архитектурное Назначение & Роль	Механизмы Отказоустойчивости & RAM
[1] postgres Реляционное ядро	postgres:16-alpine Порт: 5434:5432 Вольюм: postgres-data	- Мультитенантная изоляция через FacilityScopedQuerySet. - Хранение инцидентов, пользователей, картотеки СКУД и правил. - Неизменяемые цифровые отпечатки аудиозаписей (SHA-256). - Оптимизированные B-Tree индексы для мгновенного поиска.	- Healthcheck pg_isready раз в 5 сек. - Изоляция от стандартного 5432 порта хоста. - WAL журналирование с защитой от сбоев питания. - Потребление RAM: ~35 МБ в покое.
[2] redis Шина событий & кеш	redis:7-alpine Порт: 6381:6379 Вольюм: redis-data	- Брокер очередей Celery (изолированные БД DB1 и DB2). - Канальный слой Django Channels для WebSockets (DB0). - Кеш синтезированных аудиофайлов TTS для повтора <10 мс. - Rate-limiting сессий и хранилище одноразовых токенов.	- Персистентный снапшоттинг --save 60 1. - Автоматический реконнект WebSocket каналов. - Политика вытеснения volatile-lru при переполнении. - Потребление RAM: ~18 МБ.
[3] backend Django 5.x ASGI Ядро	Custom Python 3.12 Порт: 8080:8000 Uvicorn / Gunicorn	- Центральный шлюз REST API и WebSocket-поток. - Диспетчер протоколов СКУД: Modbus TCP (ADAM-6060), Shelly Pro, Hikvision ISAPI, Dahua CGI. - Zabbix Metrics Collector: сбор LLD и метрик ИИ без нагрузки на БД. - Webhook интеграция экстренного пульта Asterisk Voice C2.	- Строгая мультитенантность без утечек данных. - Защита от перегрузки пула соединений БД. - Неблокирующий ввод-вывод на базе ASGI Uvicorn. - Потребление RAM: ~85 МБ.
[4] celery-worker Асинхронные воркеры	Custom Python 3.12 Внутренний сервис Concurency: 4 воркера	- Обработка 5 изолированных очередей задач: — alerts: доставка в Telegram, Email, Asterisk. — lockdown: экстренная блокировка замков СКУД. — voice_dispatch: синтез речи и SIP-пейджинг. — ai_callbacks, audio_ingestion.	- Гарантированная доставка критических алертов без зависания в очереди тяжелого аудио. - Автоматический перезапуск зависших воркеров. - Таймаут выполнения задач: 30 минут. - Потребление RAM: ~75 МБ.
[5] celery-beat Планировщик & Watchdog	Custom Python 3.12 Внутренний сервис Периодический шедулер	- Мониторинг застрявших или прерванных RTSP-поток. камер. - Контроль дисковых квот WORM-хранилища (порог 50 ГБ). - Синхронизация расписаний дежурств и графиков охраны. - Периодический аудит журнала доступа и прав операторов.	- Изолированный процесс, независимый от нагрузки воркеров. - Автоматический перезапуск при программных сбоях. - Точность шедулинга с шагом до 1 секунды. - Потребление RAM: ~40 МБ.

КЛЮЧЕВЫЕ ПРЕИМУЩЕСТВА КОНТУРА ЯДРА И АСИНХРОННОСТИ:

1. Разделение вычислительных контуров: Тяжелые вычисления (синтез речи, рассылка push-уведомлений, экстренный автодозвон Asterisk, генерация PDF) изолированы в воркерах Celery. Время отклика REST API стабильно <20 мс.

2. Мгновенная реакция СКУД: Блокировка замков (lockdown) выполняется по выделенной очереди за время <100 мс.

3. Защита улик: Каждая запись снабжается цифровым отпечатком SHA-256 в PostgreSQL сразу при фиксации, обеспечивая юридическую чистоту для следствия.

ИНЖЕНЕРНЫЙ ПАСПОРТ ЭКОСИСТЕМЫ • ЧАСТЬ 2

Контур ИИ, Потокowego Захвата, Телефонии и Мониторинга (Сервисы 6–10)

Спецификации передовых компонентов нейросетевого распознавания речи, телефонии Voice C2 и телеметрии Zabbix.

Сервис / Контейнер	Базовый образ & Порты	Архитектурное Назначение & Роль	Механизмы Отказоустойчивости & RAM
[6] ai Интеллект ИИ	FastConformer + ONNX Порт: 8081:8000 Профиль: local-laptop	- Акустический детектор AED: выстрелы, взрывы, паника, звон стекла (<15 мс). - Билингвальный STT FastConformer (смешанный КК/РУ). - Tier-1 Triage: Aho-Corasick сито отсекает 85% шума за 1.2 мс. - Эндпоинт /telemetry: передача VRAM, RTF и задержек в Zabbix.	- Защита от CUDA OOM при акустических всплесках. - Real-Time Factor (RTF) >= 1.0 (обгоняет звук). - Поддержка FP8 / INT8 квантования. - RAM: ~320 МБ / VRAM: 3.5 ГБ.
[7] recorder Go Stream Recorder	Custom Go 1.22 Alpine Порт: 8091:8090 Статический бинарник	- Непрерывный захват RTSP/ONVIF аудио и видео 24/7. 30s RAM RingBuffer: запись в память (0% износа SSD). - Скремблер приватности: обфускация речи невиновных граждан. - WORM Incident Vault: крипто-дамп при тревоге для суда.	- Статическая компиляция Go (нулевой overhead). - Сторожевой таймер dead-stream watchdog. - Авто-реконнект камер за 200 мс. - Потребление RAM: всего ~28 МБ!
[8] asterisk [NEW V4] Voice C2 Gateway	Asterisk 20 LTS Alpine Порты: 5060, 5080, 5061, 5038, 8088, 10000-10050	- Внешний транк chan_sip 5060 (Казахтелеком / Beeline) с CPS-рейтлимитом 1.5 сек (защита от бана 503). - Внутренний res_pjsip 5080/5061: пейджинг в IP-динамики зон. - Voice C2 IVR с PIN 7788: удаленный локдаун СКУД и эвакуация по звонку дежурного офицера.	- Режим Simulation Fallback (бекенд не падает при выключенном Asterisk). - Прямой сокет AMI без утечек памяти. - Ротация аудиозаписей звонков. - Потребление RAM: ~42 МБ.
[9] dashboard Cockpit & Портал	Next.js 14 Standalone Порт: 3002:3000 Node 20 Alpine	- SOC Mission Control Cockpit (/live, /doors, /speakers, /cases). - Публичный Sales Portal продаж на главной странице. 60 FPS Canvas спектрограмма звука с фильтрами частот. - Полная 4-язычность интерфейса (KZ, RU, EN, ZH).	- Нулевой Layout Shift (CLS = 0). - Standalone сборка (минус 80% размера node_modules). - Оптимизация кэширования статики Next.js. - Потребление RAM: ~70 МБ.
[10] zabbix-agent [NEW V4] Мониторинг NOC	zabbix-agent2:alpine Порт: 10050:10050 Zabbix 6.0/7.0 LTS	- Low-Level Discovery (LLD) всех устройств СКУД и камер. - Непрерывный опрос /unified/ (1 запрос раз в 15 сек). - Мониторинг VRAM GPU, задержек нейросети и показателя RTF. - Контроль транка Казахтелекома и очереди Celery.	7 критических триггеров аварий (Disaster, High, Warning). - Официальный шаблон template_sentinel_ai.yaml. - Сверхнизкая нагрузка на сеть. - Потребление RAM: ~15 МБ.

ИНЖЕНЕРНЫЙ ТРИУМФ ОПТИМИЗАЦИИ РЕСУРСОВ:
Суммарное потребление RAM всех 10 контейнеров в состоянии ожидания составляет всего ~728 МБ (менее 1 ГБ!).
Для сравнения: монолитная версия V2 требовала более 6 ГБ RAM и падала при нагрузке свыше 8 потоков. Архитектура SENTINEL AI V4.0 позволяет запустить полноценный автономный охранный комплекс даже на компактном защищенном ноутбуке или edge-контроллере с видеокартой RTX 5060 (8 ГБ VRAM), обеспечивая одновременный контроль до 16 видеокamer и аудиопотоков без потери кадров.

ЭВОЛЮЦИЯ, ОТКАЗОУСТОЙЧИВОСТЬ И ВНЕДРЕНИЕ

Сравнительная Матрица V2 vs V3 vs V4 и Гарантии SLA 99.99%

Анализ надежности комплекса, сценарии аварийного восстановления и варианты аппаратного базиса.

1. Эволюционная Матрица Развития Архитектуры:

Параметр Сравнения	Версия 2.0 (Монолит)	Версия 3.0 (8 Контейнеров)	Версия 4.0 (10 Контейнеров — Текущая)
Архитектурная модель	Монолит (PHP / Laravel)	Микросервисы (8 шт)	10 изолированных сервисов (Full Enterprise)
Потребление RAM (Idle)	4.8 – 6.5 ГБ	~560 МБ	~728 МБ (Оптимизированный микростек)
Требования к VRAM GPU	32 ГБ (A100 / RTX 6000)	8 ГБ (RTX 5060)	6–8 ГБ (Аппаратный инференс FP8 / INT8)
Износ дисков (SSD Wear)	Критический (запись 24/7)	0% (30s RAM RingBuffer)	0% (30s RAM RingBuffer + WORM Vault)
Экстренная телефония	Отсутствовала	Только Telegram / Email	Dual-Stack Asterisk (PSTN + PJSIP Voice C2)
Мониторинг периферии	Ручной (слепота оператора)	Базовый Docker healthcheck	Zabbix 6.0/7.0 LLD + Deep AI Telemetry (RTF)
Доступность (SLA)	95.0% (риск падения)	99.5%	99.99% (High Availability Enterprise)

2. Модель Отказоустойчивости (Graceful Degradation):

Сценарий Аварии / Сбоя	Реакция Платформы SENTINEL AI V4.0	Операционный Результат
Обрыв проводного интернета	Celery немедленно перенаправляет вызов через Asterisk в резервную аналоговую линию или GSM-шлюз Казахтелекома.	Дежурный офицер получает голосовой звонок на мобильный и активирует локдаун через PIN 7788.
Отказ / зависание Asterisk	Клиент AMI активирует симуляционный fallback; все критические алерты дублируются в Telegram и Email.	Бекенд не зависает, дежурные операторы продолжают видеть тревоги в Web Cockpit.
Всплеск нагрузки на GPU	Tier-1 Aho-Corasick отсекает 85% шума; Zabbix сигнализирует при превышении порога VRAM > 85%.	Предотвращается падение CUDA OOM; критические аудиосигналы обрабатываются с приоритетом.
Отвал камеры или замкаСКУД	Zabbix LLD фиксирует превышение таймаута Heartbeat (>120 сек) и генерирует триггер High/Disaster.	Красная индикация на интерактивном плане объекта задолго до возникновения ЧС в зоне.

3. Варианты Развертывания и Правовой Статус:

Профиль «Edge Laptop / Пост Охраны» - Железо: Ноутбук / Mini-PC (RTX 5060 8GB VRAM), 16GB RAM. - Емкость: До 16 камер, локальный Asterisk, Zabbix Agent. - Применение: Школы, колледжи, локальные КПП безопасности. - Автономия: Полная работа без внешнего интернета.	Профиль «Enterprise Rack / Спецобъект» - Железо: Сервер 19" 2U, GPU NVIDIA L4 / A10 / RTX 4090 24GB. - Емкость: До 128 камер, транки Казахтелеком, Zabbix Proxy. - Применение: Международные аэропорты, вокзалы, ТРЦ, заводы. - Резервирование: Двойной ввод питания, RAID 10 WORM.	ПРАВОВОЙ СТАТУС & ПРАВООБЛАДАТЕЛЬ ALLAYAROV ENGINEERING Автор и Главный Архитектор: Аллаяров Артур ИП Аллаяров Артур Алматы, Казахстан Официальный портал: sentinel-ai-1ru.pages.dev Служебный Telegram: @AllayarovcomBot
---	--	--