Andrew Grathwohl

Cloud-native software and systems engineering professional with 15 years of extensive subject matter expertise in artificial intelligence, multimedia technology, digital signal processing, and real-time communications protocols.

andrew@grathwohl.me (917) 412-2617 Nashville, TN https://grathwohl.me

SKILLS

- PROGRAMMING LANGUAGES: Rust, C, C++, Swift, TypeScript, Golang, Python, Elixir.
- FRONTEND FRAMEWORKS: Electron, Tauri, Next.js, Vue 3, Remix, React Native.
- DATABASES: PostgreSQL, Qdrant, MongoDB, LevelDB, Sqlite, BigQuery, Clickhouse, Redis, Pinecone, Elasticache, DynamoDB
- NETWORKING TOOLS: bird2, FRR, Nginx, Varnish, Netbox, ntop-ng, ip-tools.
- STORAGE GlusterFS, Zfs/OpenZFS, HAMMER2, Btrfs, Ceph, IPFS, JuiceFS, XFS
- MEDIA FRAMEWORKS: FFmpeg, SoX, librosa, QCTools, GPAC, GStreamer.
- ML & NLP TOOLS: PyTorch, TorchAudio, SpaCy, SpeechBrain, ONNX
- SYSTEM INTERFACES: sysctl, cpupower, Syslogd, bcc/bpftrace, perf, tcpdump.
- INFRASTRUCTURE TOOLING: Terraform, Nix, kubectl, Jenkins, Ansible, Vagrant, Docker.

WORK EXPERIENCE

Horizon3 Venture Studio

Audio AI & Telephony Consultant [Contract] Oct 2024 - Present

- Led development of a LLM-driven AI voice agent implementation for Prudential Insurance's customer support phone numbers, to reduce call avoidance and incorrect routing KPIs at US call centers.
- Deployed Kubernetes infrastructure to support Kamailio and FreeSWITCH SIP and RTP proxy services. Enabled high-availability routing, transfer, conferencing, and SIPREC features as LLM tools.
- Produced custom Linux kernel builds with RT PREEMPT patches and sysctl tunings optimized for low-latency multimedia processing use cases
- Reduced call authorization mistakes by nearly 100%, by developing a novel SIP messaging protocol extension that passed encrypted customer metadata between call center servers and agent telephones, enabling continuous authorization of the caller delivered over phone call and WebRTC connections.
- Integrated Kubernetes network's telephony and presence services as LLM tools, enabling use within live phone calls directly. This led to a dramatic reduction in call avoidance metrics (between 10-20% drop each day since deployment.)

Rumble

Senior Systems Engineer - CDN July 2024 - Sept 2024

- Architected a KVM-based anycast CDN solution for Rumble’s 12 distributed data center locations throughout the United States, responsible for handling an average of 2Tbps traffic during peak usage times
- Generated FRR BGP routing configurations for pfSense and bird2 BGP routing configurations for Ubuntu hypervisor guests, optimizing for the greatest number of active connections per server
- Produced numerous kernel optimizations that enabled Ubuntu guests to serve cached assets with a 24ms reduction in round-trip travel times
- Implemented a custom Rust-based caching service to replace nginx and openresty caching services, enabling Rumble’s live streaming servers to exclusively leverage in-house hardware to route RTMP, WHIP, and SRT broadcasts to end users
- Developed inventory analysis scripts to assess current server memory footprints, which automatically produced purchase orders for additional memory DIMMs based on a provided capacity target

Tube.sh

Principal Engineer [Contract] Dec 2023 - Present

- Established and operate Tube.sh, a media technology consulting company that centralizes my software development projects for media companies, content creators, and broadcast networks.
- Developer and operator of PartOfTheProblem.com for the Dave Smith political podcast, "Part Of The Problem," which generates >$34,000 MRR for the client and conists of over 5,000 subscribing users.
- Developed the "patreon in a box" software suite utilized by clients to host their own content archives and subscription services for direct interaction with fans.
- Implemented payment processing services via Stripe, PayPal, and bank wire transfer, along with Klarna, Apple Pay, Google Pay, and Amazon Pay.
- Host and maintain numerous cloud and on-prem RTMP ingest endpoints and VOD archives across three CDN providers with fine-tuned failover specifications for optimal reliability and cache ratio.

- Deployed a private podcast feed generation service for subscribers, ensuring each paying customer gets their own feed URL that is constantly re-validated for content access permissions

Storyboard

Lead Engineer — Aug 2020 - Sept 2023

- Bootstrapped and mentored Storyboard's audio-first engineering team as the company's first engineering hire, growing headcount from 1 to 16 in one year.
- Oversaw and implemented CI/CD workflows, code quality standards, and cloud architecture for the world's largest private podcasting network for enterprises and its companion audio chat service for front-line and first responder work forces.
- Developed Storyboard's machine learning pipelines, responsible for producing transcripts from spoken word recordings at 98% accuracy. Devised AWS Sage-Maker pipeline to parse entities in transcripts and separate concepts. Utilized temporal convolution and chromaprint to detect authenticity of recorded voices.
- Invented a novel approach to achieving sub-30ms latency audio communication channels over a swarm WebRTC decentralized network, with STUN/TURN signaling at both the client and within a cluster of centralized publicly-available WebTorrent nodes.
- Implemented company's iOS, tvOS, and CarPlay audio frameworks as Swift Packages, providing support for Opus transcoding, real-time audio classification with CoreML, and live Whisper transcript features.
- Implemented AWS Lambda low-latency audio analysis and transcoding pipeline for validating uploaded content and generating RF64 masters and opus encodes.
- Developed recording, editing, and playback suite consisting of Web Audio API modules and Rust-WASM binaries, enabling generalist engineers to build audio creation tools that perform complex audio mixing and DSP operations.
- Led development of white label app for Amazon.com, enabling the distribution of private podcast episodes to Amazon's 10,000+ fleet of delivery vehicles

Littlstar

Director - Media Technology — Nov 2017 - June 2020

- Developed and implemented multi-CDN Node.js services responsible for ingesting, processing, tracking, and validating 360/3D video, VR experiences, traditional SVOD media, RTMP live streams, and volumetric holograms. Performed frequent benchmarks and performance analyses to ensure reliable low-latency peer-to-peer connectivity for all services.
- Co-creator of Ara, a decentralized blockchain-backed content distribution proof-of-stake cryptocurrency which rewards peer-to-peer sharing of media content.
- Launched, managed, and mentored the Media Services remote engineering team in pursuit of executing media technology strategy. Carried out hiring, budget, sprint planning, code review, and performance tracking responsibilities.
- Technical business development lead for enterprise contracts signed with AT&T 5G, Intel, Microsoft, ViacomCBS, Sony, MyndVR, WWE, and Live Nation.
- Produced novel implementations of new MPEG standards with the goal of improving streaming QoE, including CUDA transcode acceleration, tile-adaptive MPEG-DASH, real-time object detection neural nets, and peer-to-peer distributed encoding.

Lead Media Engineer — April 2016 - Nov 2017

- Led the integration of proprietary Littlstar video player technologies into business partners' products, including Alcatel, CNN, and Mattel.
- Liaised with content partners to promote success on the platform, providing documentation and hands-on technical support to ensure compliant media assets and strong content performance KPIs.
- Developed Littlstar's media library standards, including audiovisual QC workflows, livestream setup requirements, digital rights management, metadata internationalization, storage and permissions management, and content taxonomy.

CBS Corporation - Showtime Networks Inc.

Senior Business Analyst - Video — Jan 2016 - April 2016

- Evaluated business requirements, technical processes, and operations improvements necessary to bring Showtime Networks' digital asset management services in-house.
- Identified and tracked key performance indicators for content management operations, presenting weekly reports to CBS IT executive leadership.
- Spearheaded initiative to conduct regular transfer speed tests for hard drive, solid-state, and NAND storage media, increasing the accuracy of CBS IT's predictive models for storage requirements.
- Modernized Showtime Networks' content metadata taxonomy, optimizing for greater interoperability throughout the CBS organization.

Streamable

Backend Engineer - Video [Contract] — Sept 2015 - Dec 2015

- Assessed the company's video transcoding pipeline and identified key improvement opportunities to better target their growing mobile user base.
- Addressed bottlenecks in video processing services that prevented timely delivery of high-frame rate content (particularly sports and gaming content) to mobile devices.
- Updated chrome browser plugin that Streamable's users leveraged to make real-time clips of their video gaming sessions to share among their social media connections.

Overture.me

Backend Engineer - Video April 2015 - Sept 2015

- Developed the video editing backend for Overture.me, a video sharing social media app for iOS. The backend acquired video editing instructions from the mobile application and performed editing, concatenation, and final authoring of video content for rapid distribution to users' networks of followers.
- Implemented custom FFmpeg concatenation filters that enabled videos to be published and edited within 30 seconds of their initial submission.

SpokenLayer

Backend Engineer - Audio [Contract] Dec 2014 - April 2015

- Developed the SpokenLayer "cloud studio," consisting of a web app for recording and editing spoken word audio, a real-time bidding platform for new narration jobs, and a backend processing pipeline to convert and process audio recordings for distribution.
- Integrated microservices into a standardized workflow which enabled audio narrations for news articles to be published to client websites within 30 minutes of the text's publication.
- Head of production for the Stitcher Top 5 Podcast, TIME's The Brief, and Reuters World News Report, which yielded a combined 500k listeners per day.

Audible, Inc. (An Amazon Company)

ACX Production Coordinator Sept 2013 - Dec 2014

- Managed Audible's Audio QA team, consisting of four full-time audio engineers, responsible for validating all Audible titles prior to publication.
- Director of ACX audio operations, including ingestion, production guidelines, QA workflows, and transcoding.
- Scaled ACX's production volume from 200 audiobooks per month to over 3,000 audiobooks per month.
- Developed SoX audio processing recipes to auto-master all incoming audio to Audible's loudness specification, and to measure the dry and wet signals in real-time during the mastering process.
- Introduced backend audio quality metrics and QA validation results into Amazon's internal dashboard framework, leading to my team winning Audible's first ever Hackathon event.
- Inventor of Audible voice search technology, US Patent #9412395: Narrator selection by comparison to preferred recording features.
- Regularly blogged about and presented audio production techniques under the name "Andrew the Audio Scientist."

Post-Production Associate Oct 2011 - Sept 2013

- Edited and produced more than 150 premium Audible Studios audiobooks.
- Defined department-wide audio QA guidelines and encoding specifications.
- Developed an AWS web service which re-formats Kindle book files into screenplay style narration scripts for voice actors.

EDUCATION

Indiana University - Bloomington

Bachelor of Science in Recording Arts August 2011

- Concentration: Computer Music