FibRace Performance Report

Cairo M Client-side Proving Benchmark

Simon Malatrait <simon@kakarot.org>

October 1, 2025

1 FibRace Proof Performance Report

A comprehensive analysis of proof generation performance, device characteristics, and crash behaviour across the FibRace dataset.

1.1 Executive Summary

1.1.1 Global Stats

The proof generation frequency is computed as the number of Cairo M cycles divided by the proof generation duration, it weighs the duration with the computational workload.

Stat	Value
Total Proofs	**2,195,488**
Total Players	**6,047**
Total Players with at least one proof	**5,156**
Total Unique Devices Model	**1,420**
Total Cairo M cycles proven	**332,995,036,024**
Average Proof Generation Duration (s)	**7.506**
Top 10% Proof Generation Duration (s)	**1.548**
Top 25% Proof Generation Duration (s)	**3.673**
Median Proof Generation Duration (s)	**6.370**
Top 75% Proof Generation Duration (s)	**9.365**
Top 90% Proof Generation Duration (s)	**13.752**
Average Proof Generation Frequency (Hz)	**30,234.240**
Median Proof Generation Frequency (Hz)	**10,374.369**

1.1.2 RAM Impact on proof generation

Available RAM limits whether or not the proof generation is possible. The minimum RAM a device must have to generate Cairo M proof on FibRace is **3 GB**. When device is low on available memory, we notice a significative degradation of performance, usually ending up in a crash. On most devices, it is due to continuously generating proofs, which stresses the hardware. Hence, devices with less RAM available tends to prove smaller numbers which yield smaller proof frequencies. Also, devices with lower RAM capacity tends to have slower/older CPUs.

1.1.3 SoC Impact on proof generation

The proof generation speed is slightly coupled with the CPU architecture and its max core frequency.

1.2 Dataset Overview

Snapshot of the FibRace campaign (09/11/2025 16:00 UTC - 09/30/2025 23:59 UTC)

- `snapshot_09_30_2025/mintedItems/`: successful proof generation records (one row per minted proof).
- `snapshot_09_30_2025/players/`: player profiles, including historical best proof stats and minted counters.
- `snapshot_09_30_2025/devices/`: device registrations per player with hardware and OS details.
- `snapshot_09_30_2025/crashLogs/`: reported crashes during proof attempts, capturing player and device identifiers when available.

All timestamps are stored as Unix epoch milliseconds. Numerical performance fields are floating-point values (seconds for duration, Hz for frequency, cycle counts in `n`).

1.3 Environment Setup

This notebook uses `polars` for streaming-friendly analytics and `matplotlib` for visualisations.

1.4 Data Access Helpers

Helper utilities encapsulate file-system paths, enforce schemas derived from Convex exports, normalise textual fields, and provide reusable expressions for safe ratios and timestamp conversion.

1.5 Lazy Dataset Views

Create cached `LazyFrame` objects for each export. These preserve the streaming behaviour of JSONL sources while letting downstream queries reuse common transformations.

1.5.1 Record Counts & Coverage

Validate row counts and player/device coverage across the exported tables before diving into deeper analysis.

1.6 Global Activity & Reliability Trends

High-level KPIs describing throughput, performance, and reliability provide quick sanity checks and context for deeper slices.

1.7 Player-Level Performance

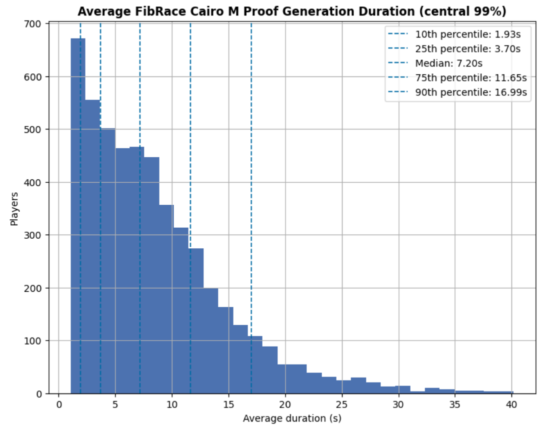
Aggregate successful proofs per player to benchmark throughput, proving speed, and frequency.

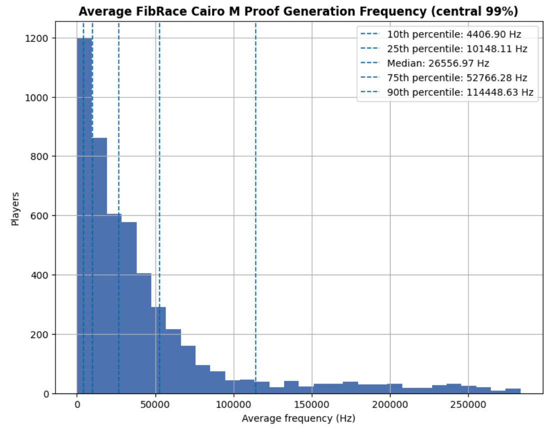
1.7.1 Fastest And Slowest Cohorts

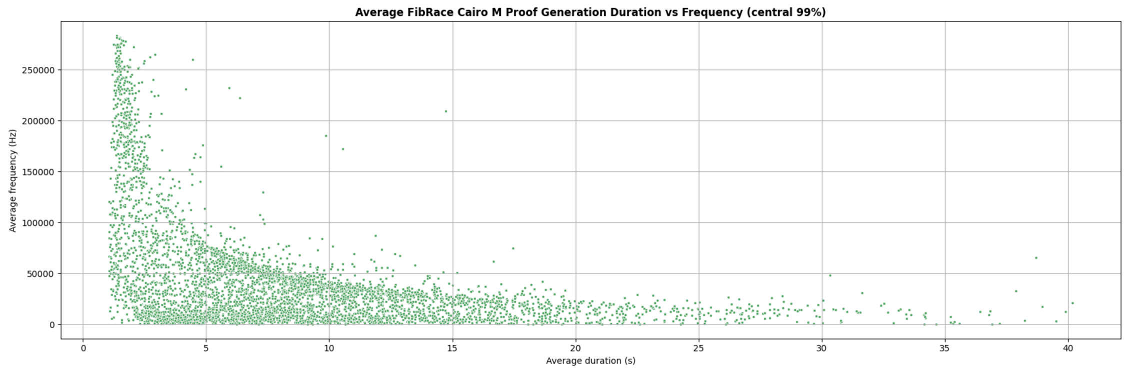
Compare players with at least five proofs to highlight best-in-class devices and outliers needing deeper investigation.

1.7.2 Performance Distributions & Frequency Relationship

Outlier-Clipped View Focusing on the central 99% of players (0.05th – 99.5th percentiles) highlights the bulk behaviour without extreme devices skewing the scale.





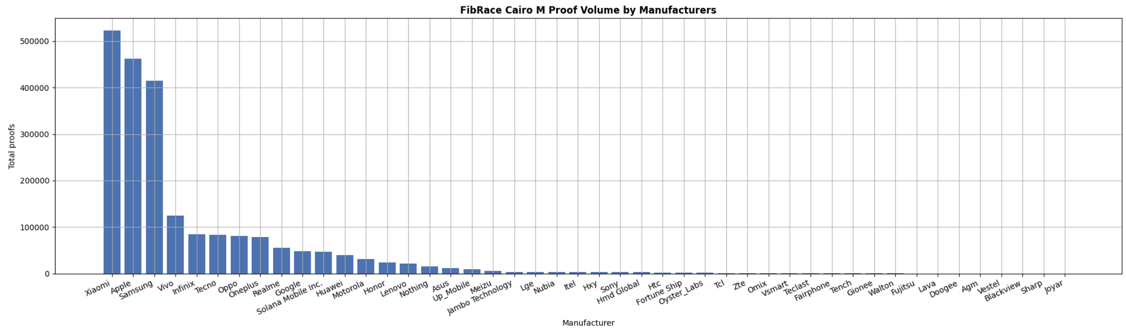


1.8 Device Specs and Proof Performance

Linking enriched device specs to proof performance highlights how hardware generation influences proving throughput.

1.8.1 Manufacturer Summaries

Most of the FibRace proofs were generated on Xiaomi, Apple and Samsung devices. This is largely due to their market share and large offer of devices.



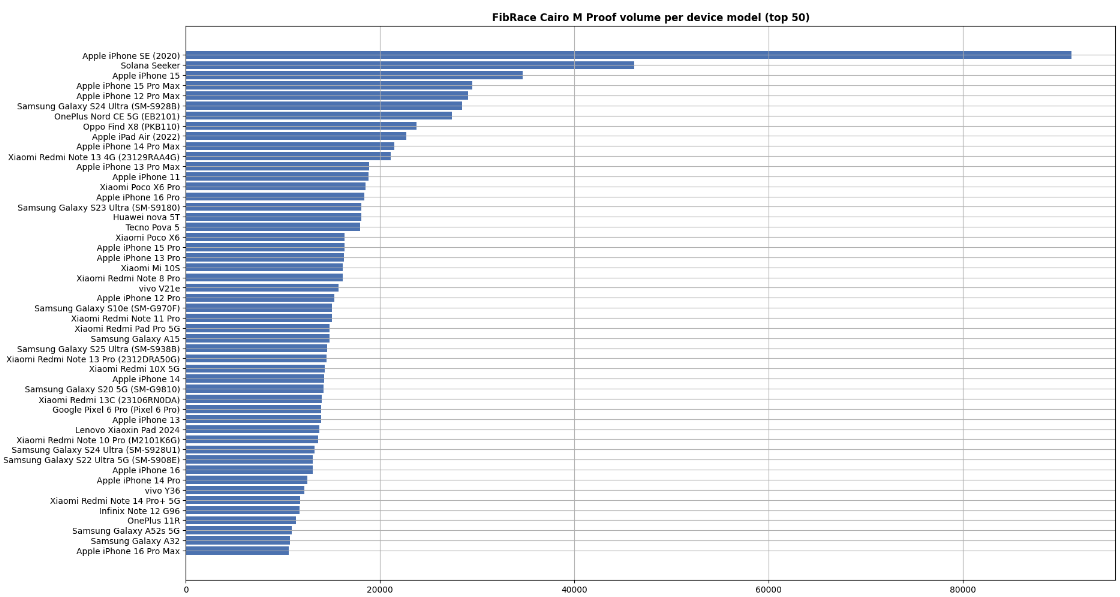
1.8.2 Device Top Dataset

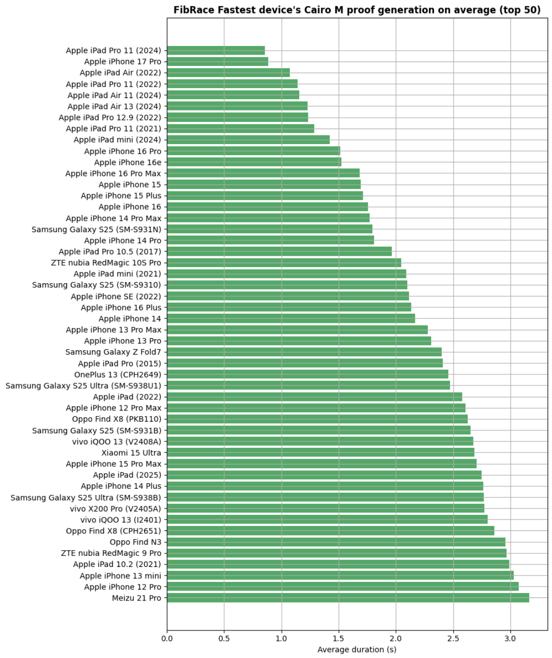
Create datasets sorting the proof generation performance of the various benchmarked devices, sorted by best and average proving speed and frequency.

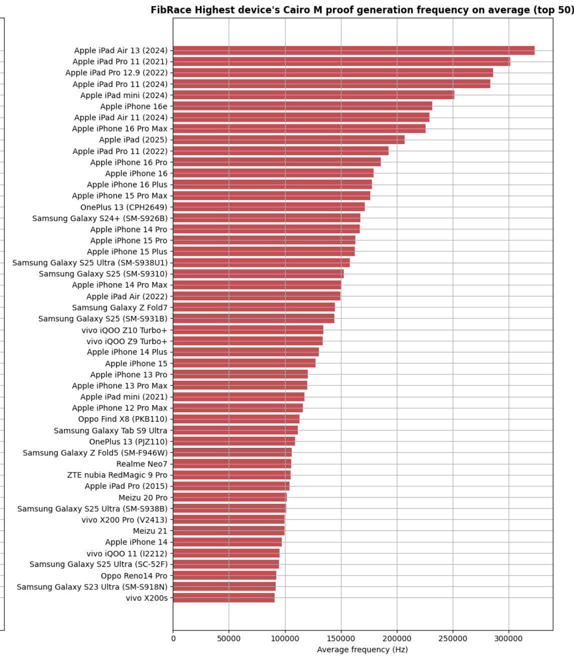
1.8.3 Proving Metrics by Device

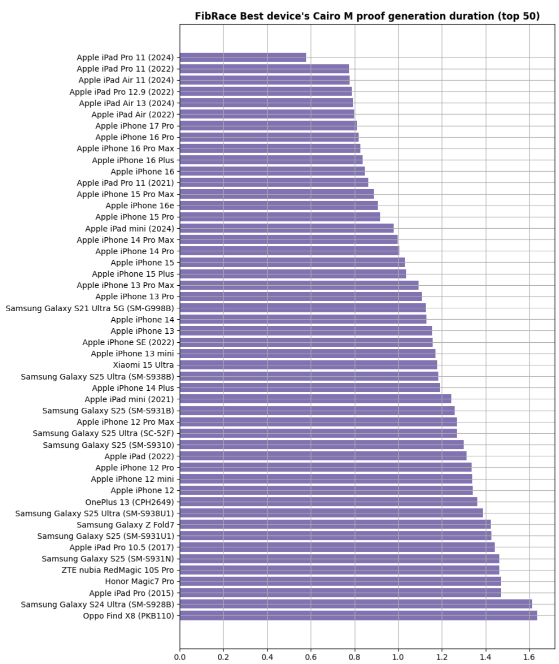
Aggregated view of how devices fare across proof generation frequency and duration.

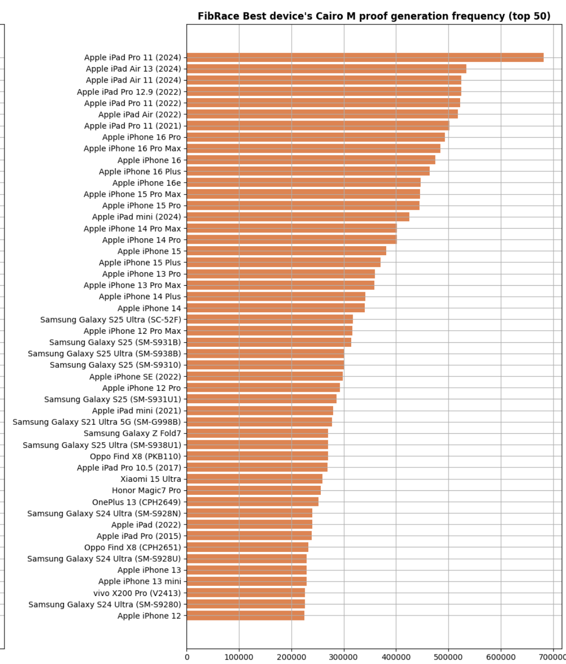
The device brands that constitue the top 50 are Apple, Samsung, ZTE, OnePlus, Xiaomi, Oppo, Vivo and Google. The best devices which generated Cairo M proofs in FibRace were iPads, mainly the latest generations of iPad Pro, with the Apple M4, Apple M2 and Apple M1 chips. The iPhone 17 Pro, with an A19 Pro chip is close to the performance of the latest iPad Pro. Best in class phones are not far away from iPad performances, with various iPhones models at the top of the ranking.









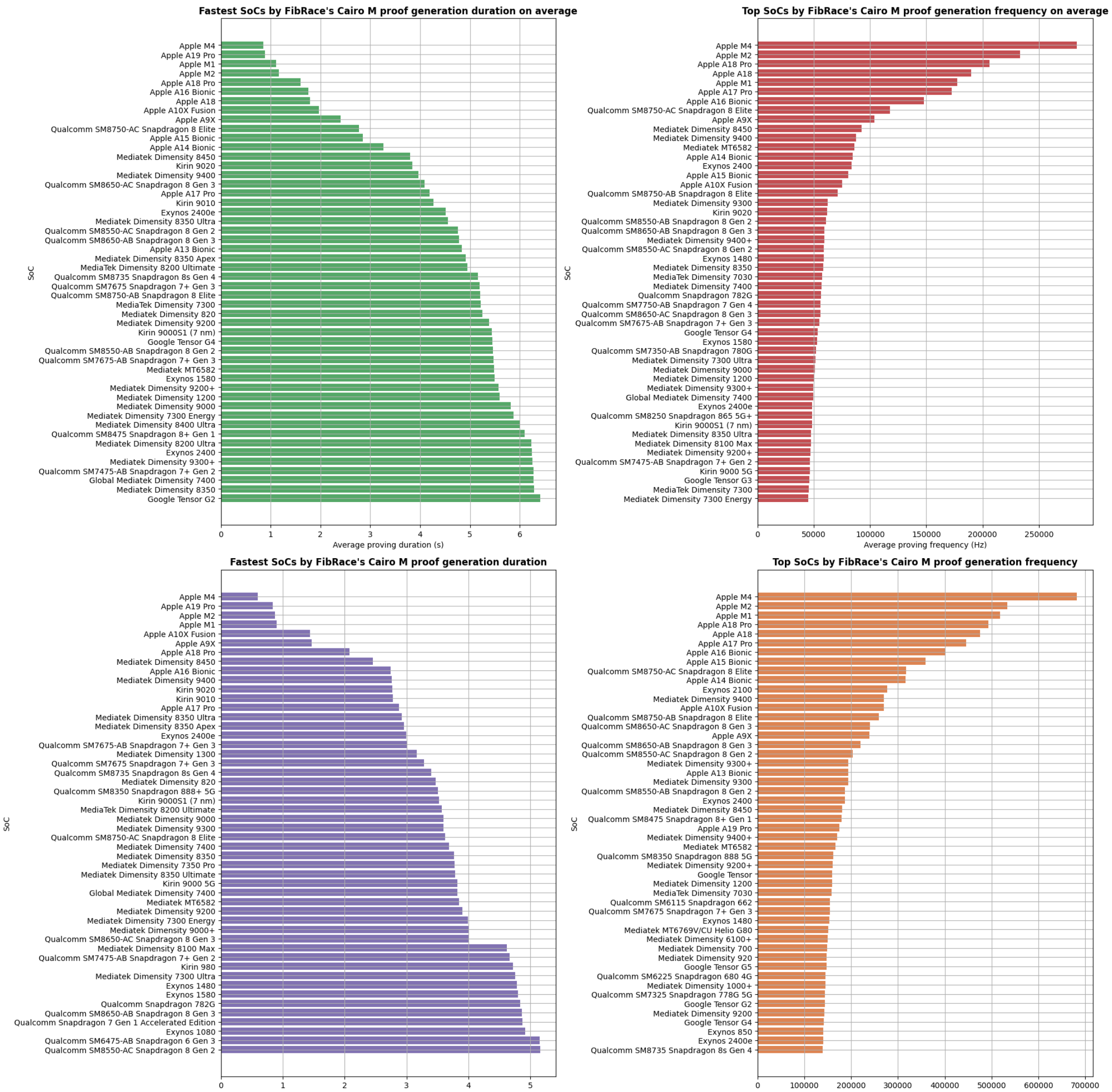


1.8.4 Proving Metrics by SoC

Aggregated view of how SoC fare across proof generation frequency and duration.

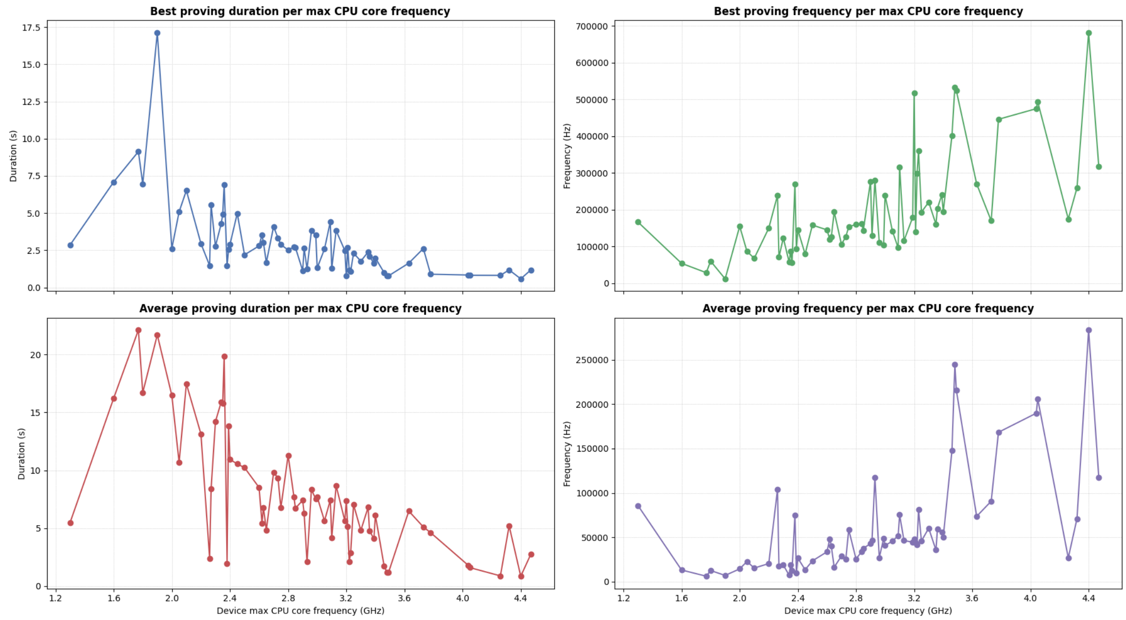
Apple SoCs are scoring the top of the leaderboards, introducing a gap between its latest chipsets.

The Apple A19 Pro SoC, available in the iPhone 17 Pro competes with the Apple M1, M2 and M4 ones. The best proving duration performance of the A19 Pro benchmarked are more than twice faster than the A18 Pro, the latter being on par with SoCs from other manufacturers (Mediatek, Qualcomm, Huawei and Samsung).

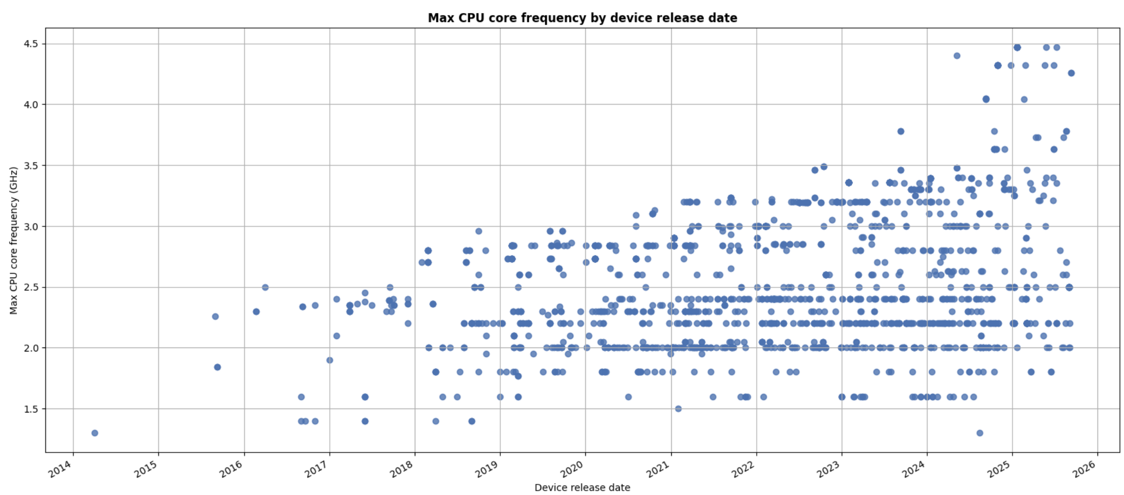


Proof Generation Performance Distribution over the SoC max core frequency These panels shows the distribution of the proving duration and frequency over the core frequency of the device's chipset they are executed on. We remark that the higher the core frequency is, the faster the proving is, while proving more (higher frequency on average).

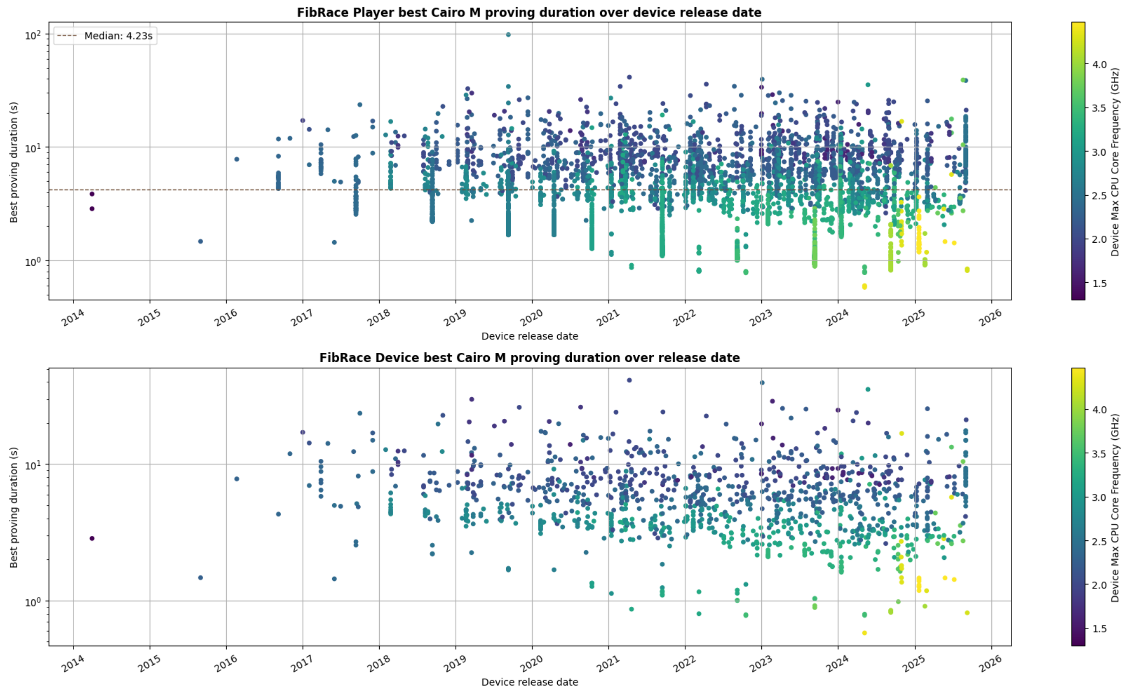
For a given max SoC core frequency, we still see notable variations between the devices, the architecture and manufacture of the overall chipset plays a significative roles which is not analyzed here (e.g.the other cores, as most recent CPUs are octa-core).



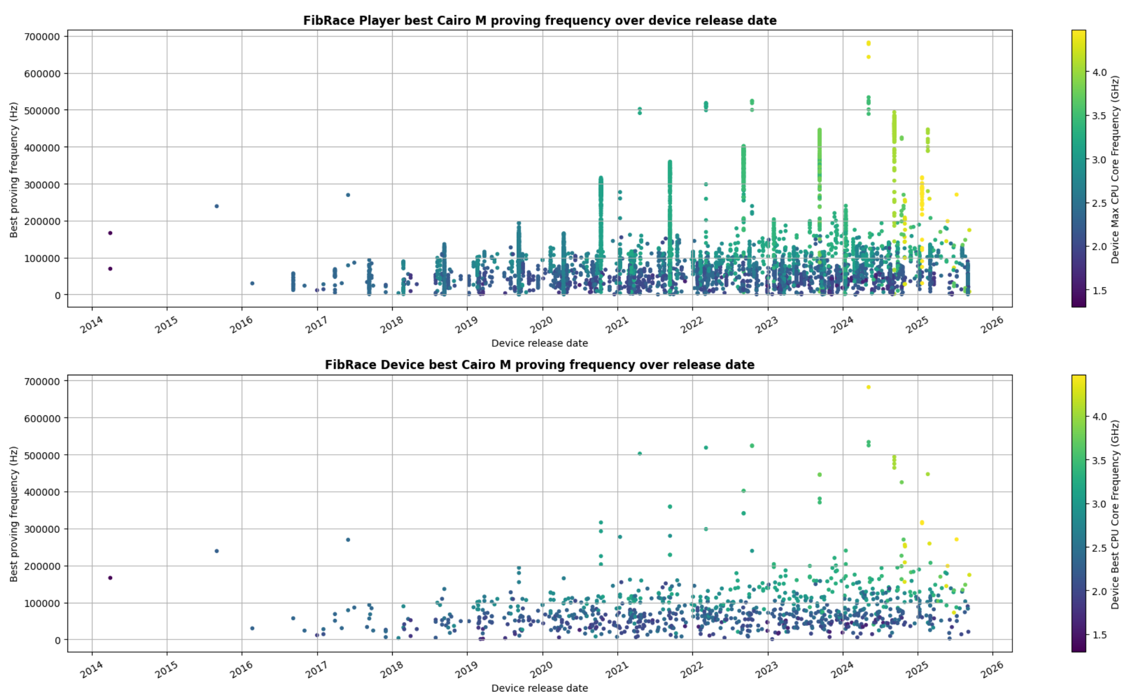
CPU Core Frequency Timeline Scatter of each device's fastest CPU core against its release date to show the max CPU core frequency of the device dataset. The release date used is the one of the device, not the one of the CPU chipset.



Player & Device proving duration over release date Panel A shows each player's best proof generation duration distributed across their device's release date ; Multiple players could have the same device: Panel B rolls that up to the top-performing player for the given device.



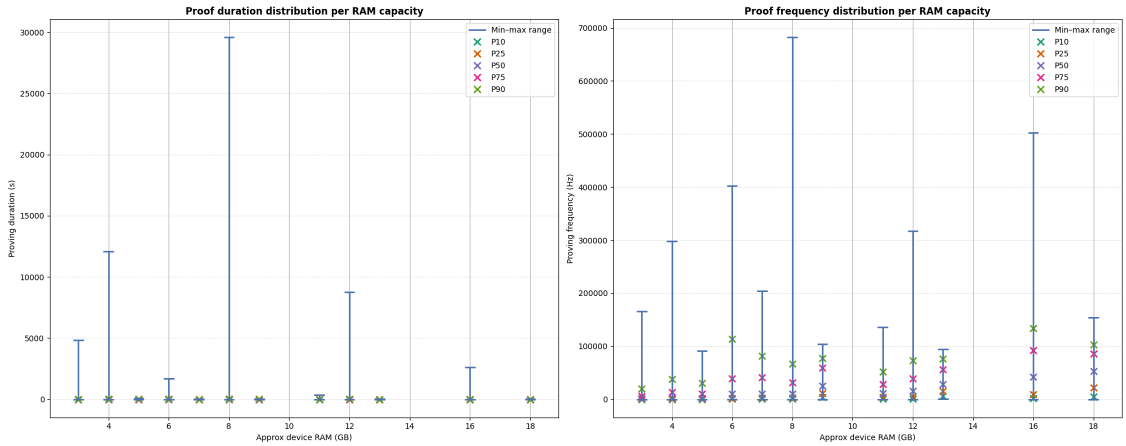
Player & Device proving frequency over release date Panels mirror the duration view, but focus on highest frequency proofs: fastest while wheighting the computational workload.

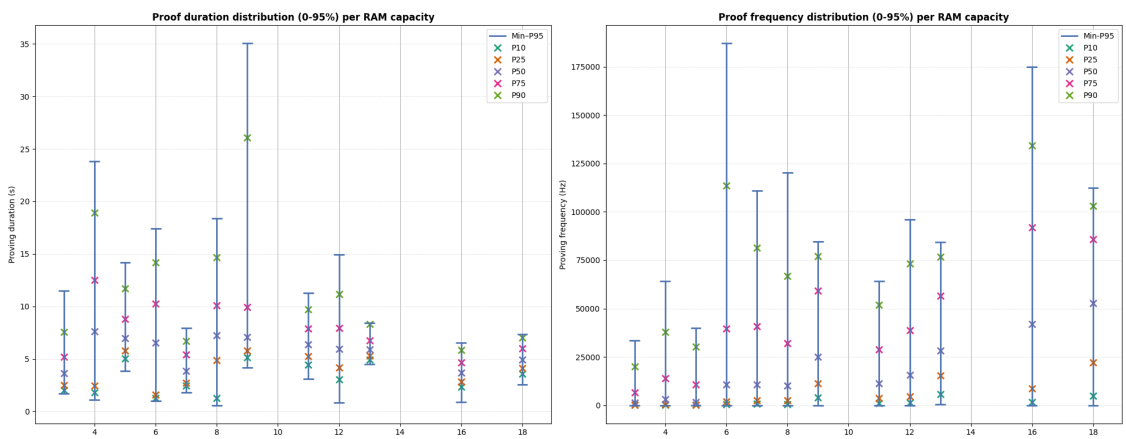


1.8.5 Proving Metrics by RAM

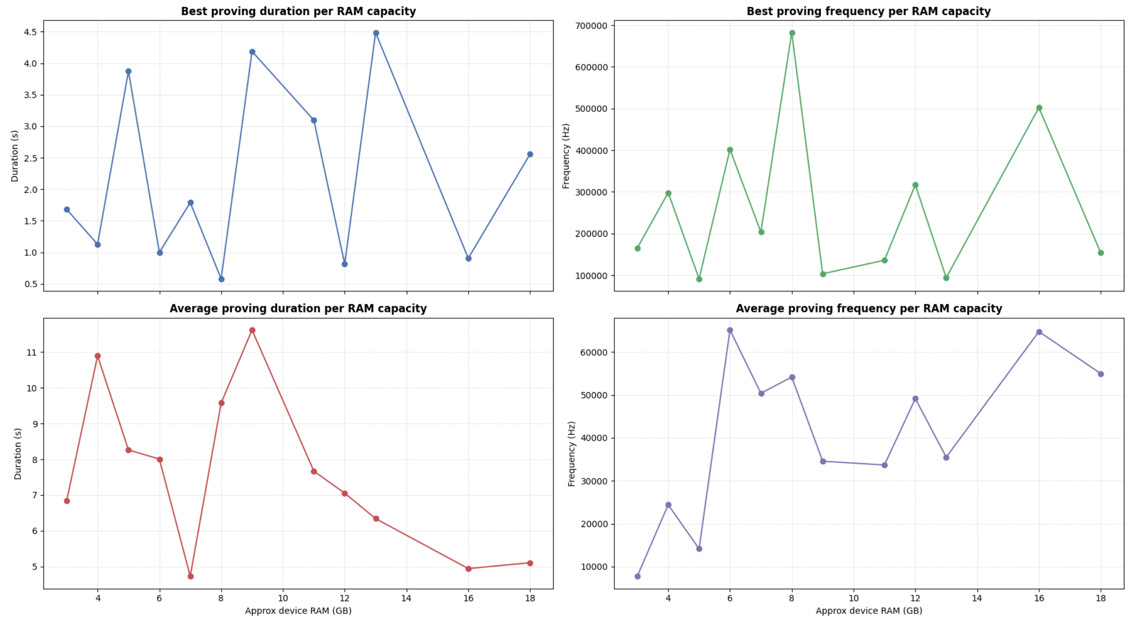
Aggregated view of how RAM fare across proof generation frequency and duration.

Distribution of Proof Generation Duration and Frequency per RAM capacity The first panel shows the complete distributions of the FibRace proofs duration and frequency per RAM capacity. The second panel shows the distribution of a trimmed view 0-95% of the proofs of each RAM capacity, to remove the few extreme outliers (worst duration, best frequencies), and better visualize the main frame.

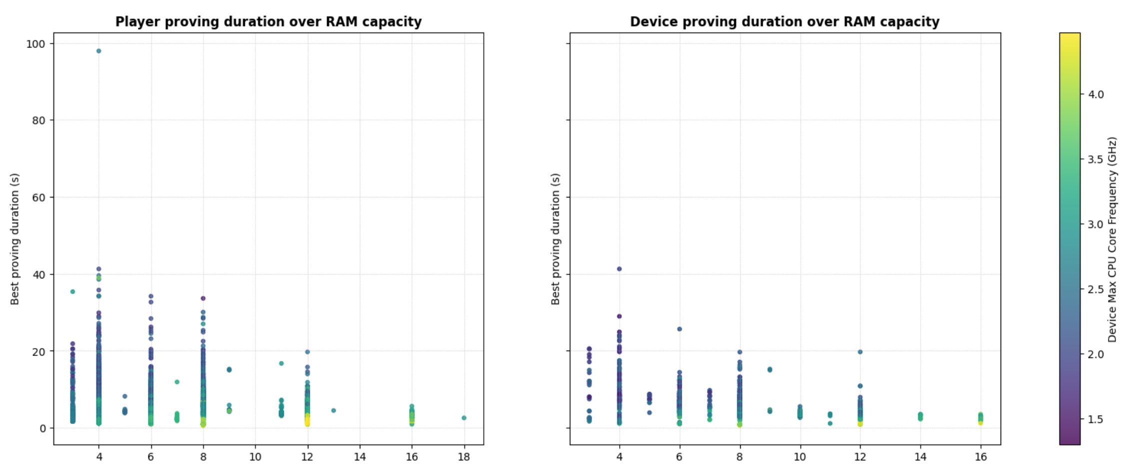




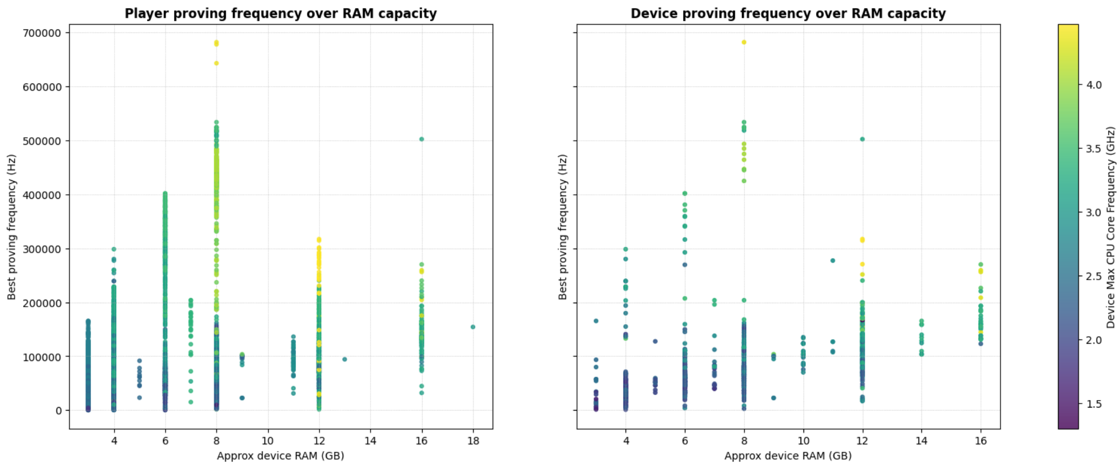
Best Proof Generation Duration & Frequency per RAM capacity Now we focus on the best proving performance for each RAM capacity. We can see that regardless of the device RAM, better performance are tied to better SoC (which is slightly tied to the max CPU core frequency).



Player & Device proving duration over RAM capacity Panel A shows each player's best proof generation duration distributed per device RAM capacity ; Multiple players could have the same device: Panel B rolls that up to the top-performing player for the given device.



Player & Device proving frequency over RAM capacity Panels mirror the duration view, but focus on highest frequency proofs: fastest while wheighting the computational workload.



1.9 Crash Analysis

Combine crash logs with performance metrics to estimate failure rates and identify unstable hardware.

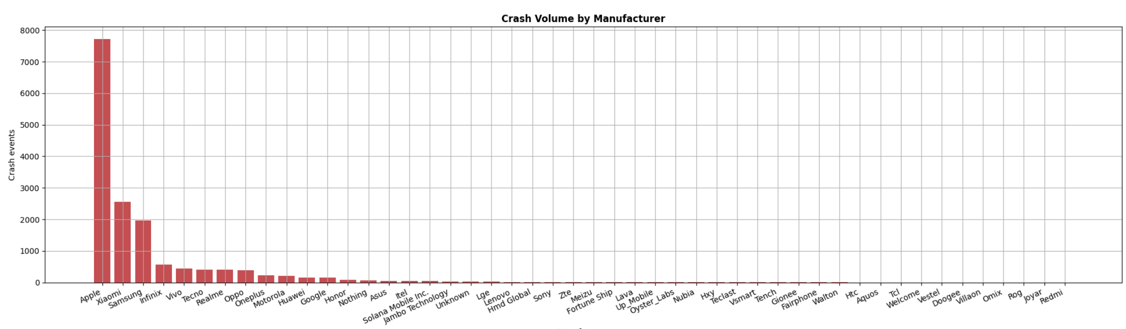
1.9.1 Players With Frequent Crashes

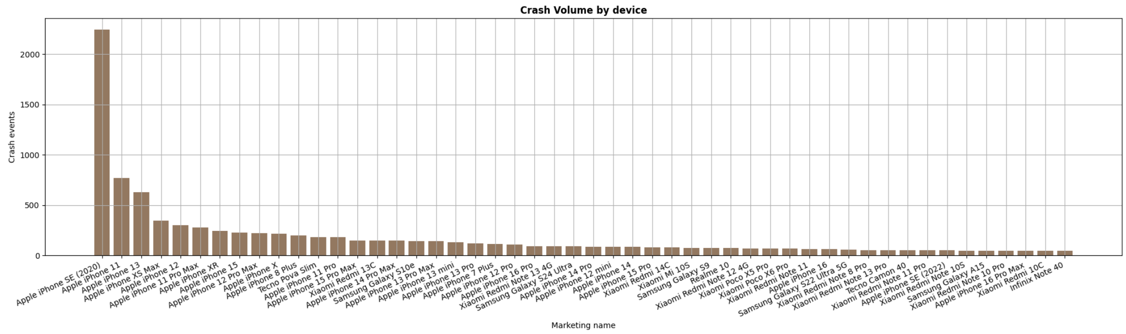
Focus on players with at least three incidents to surface high-risk accounts.

1.9.2 Device Stability Snapshot

Crash counts aggregated by device provide a hardware view of reliability issues.

Most crashes happened on devices being about 5 years old.

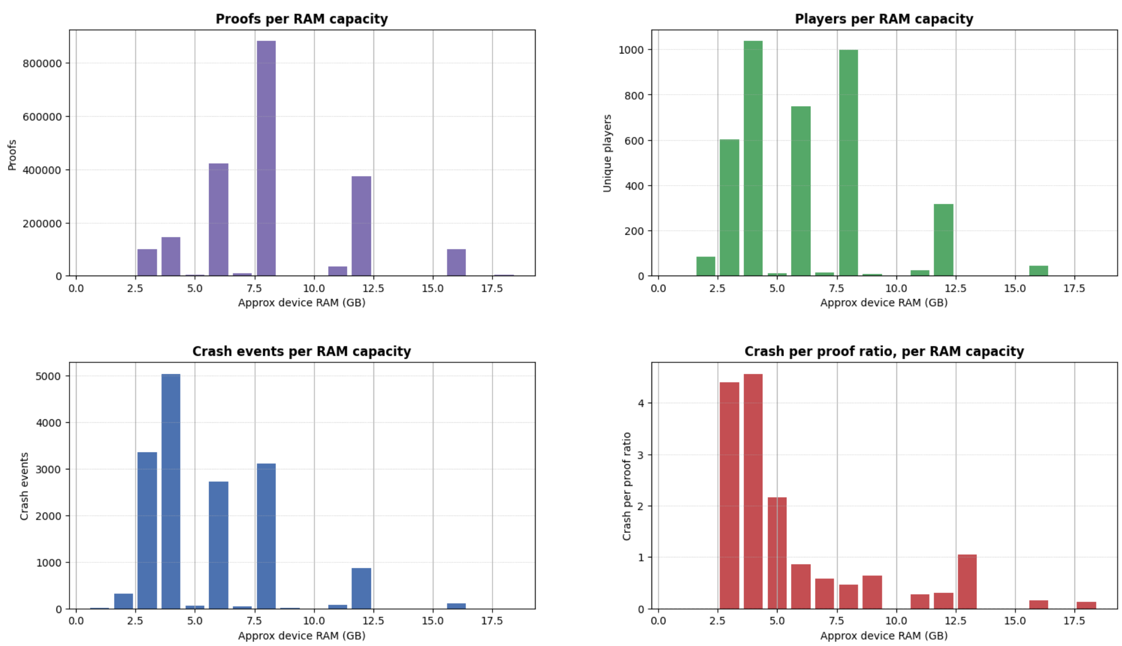




1.9.3 Crash Patterns by Device RAM

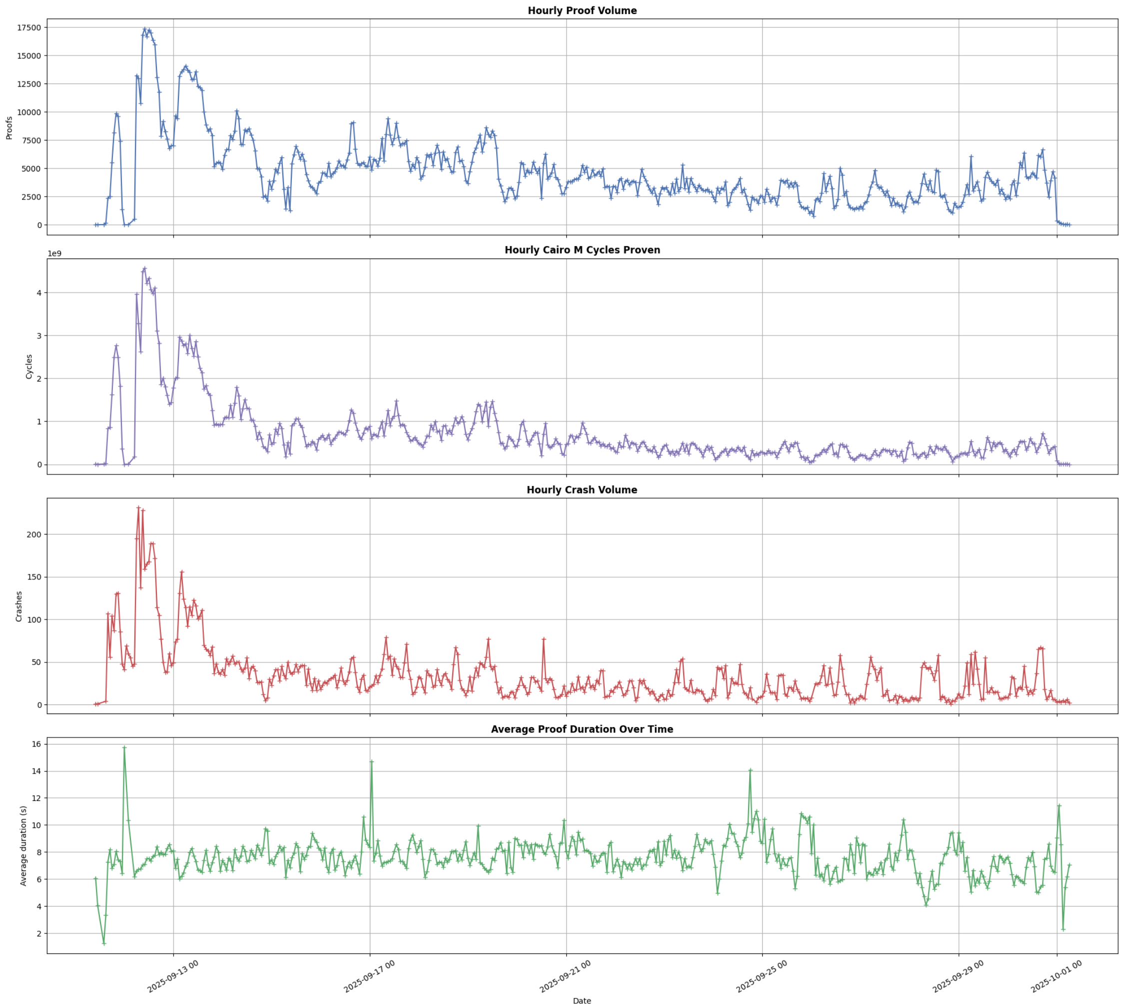
Crash counts bucketed by approximate device memory (GB).

Devices with less than 3 GB of RAM didn't succeed to generate a single proof. Devices with 3 and 4 GB of RAM were the most prone to crashes, due to the memory footprint of Cairo M and the underlying prover S-two when generating a proof.



1.10 Temporal Trends

Hourly aggregates highlight seasonality, onboarding spikes, and crash/throughput divergence over time.



Hourly Proof Volume
Proofs
Hourly Cairo M Cycles Proven
Cycles
1e9
Hourly Crash Volume
Crashes
Average Proof Duration Over Time
Average duration (s)
2025-09-13 00
2025-09-17 00
2025-09-21 00
2025-09-25 00
2025-09-29 00
2025-10-01 00
Date