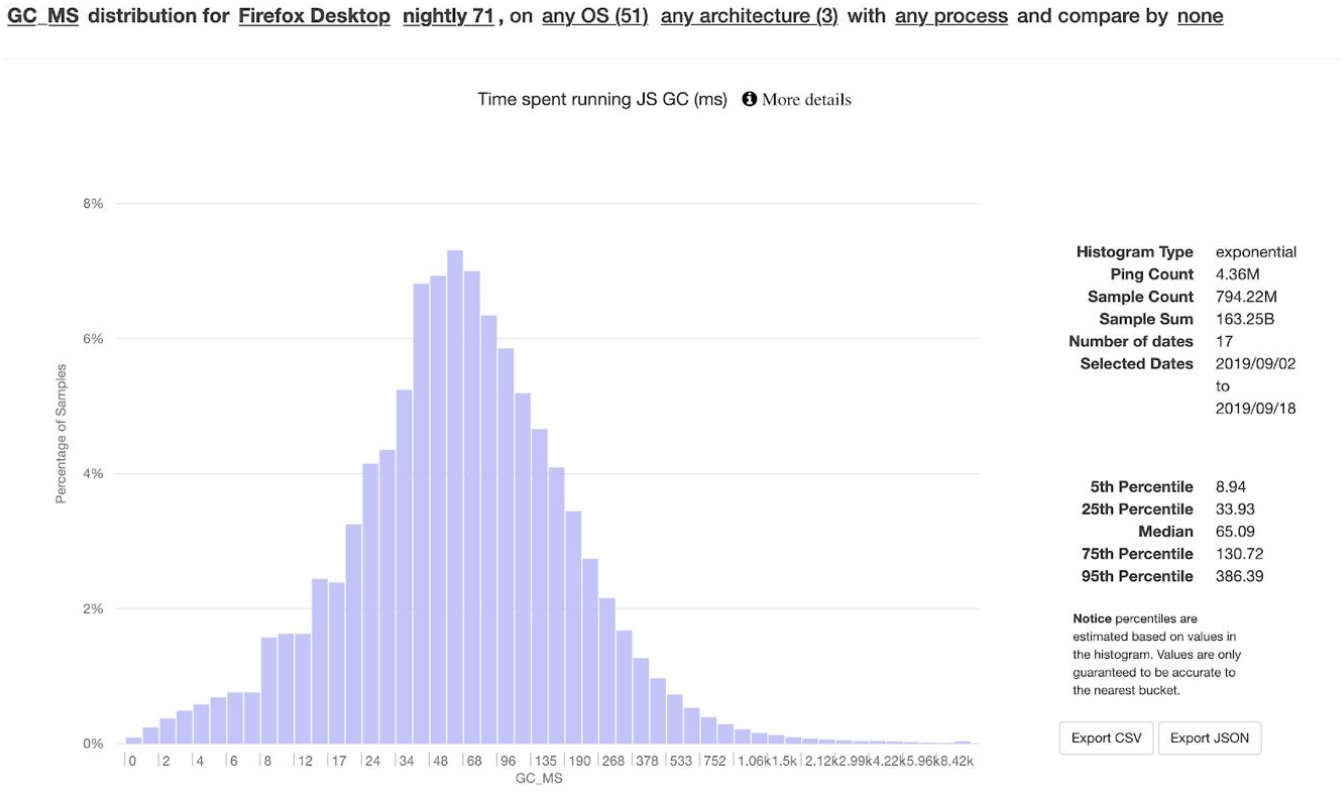
moz://a

Firefox Origin Telemetry with Prio

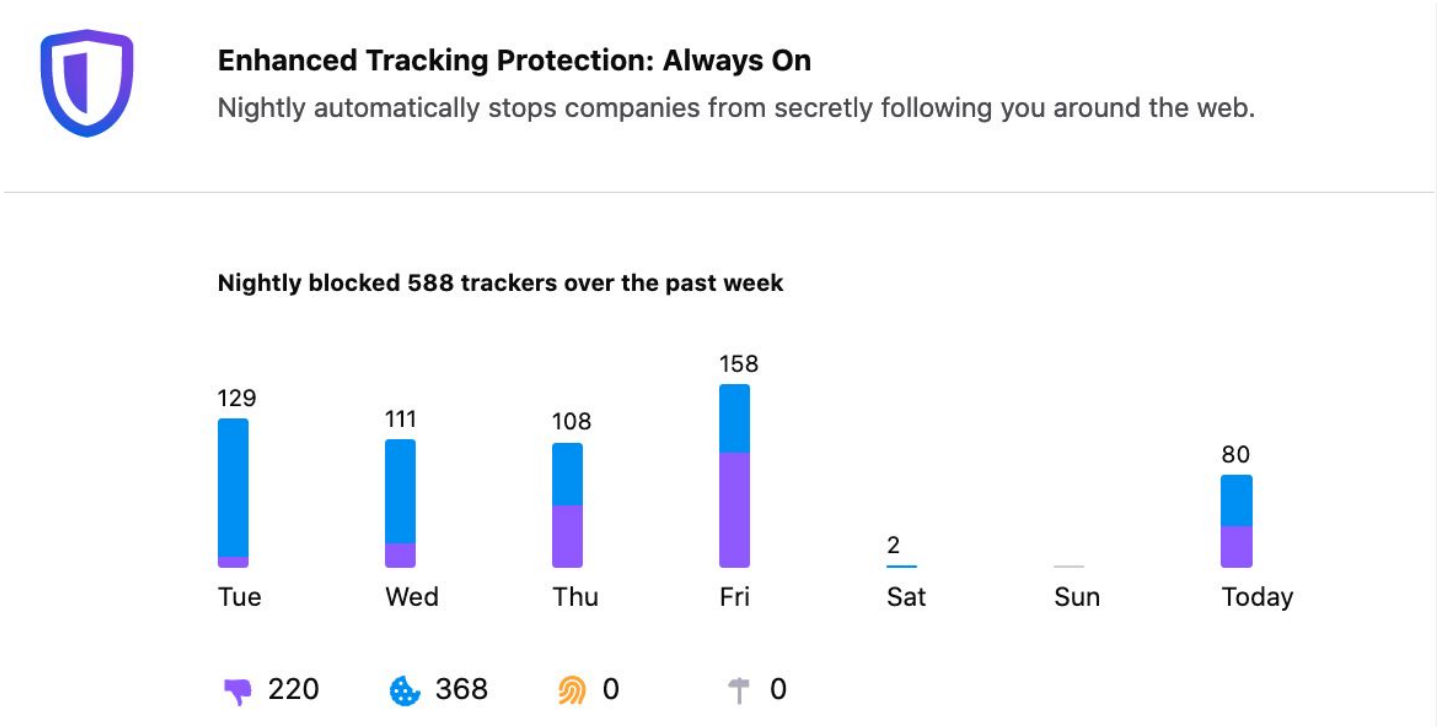
2020-08-15

Anthony Miyaguchi
Data Engineer

Data at Mozilla

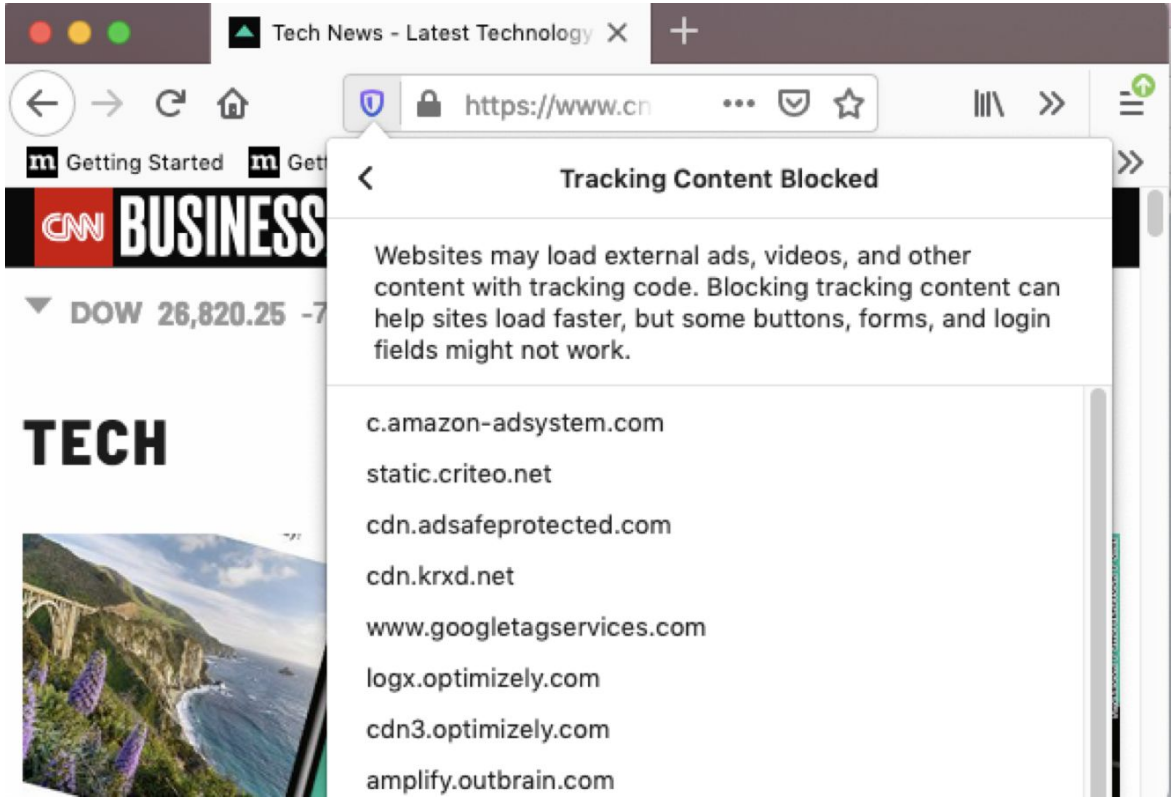


Content Blocking



about:protections

How many times has `example.com` been blocked?

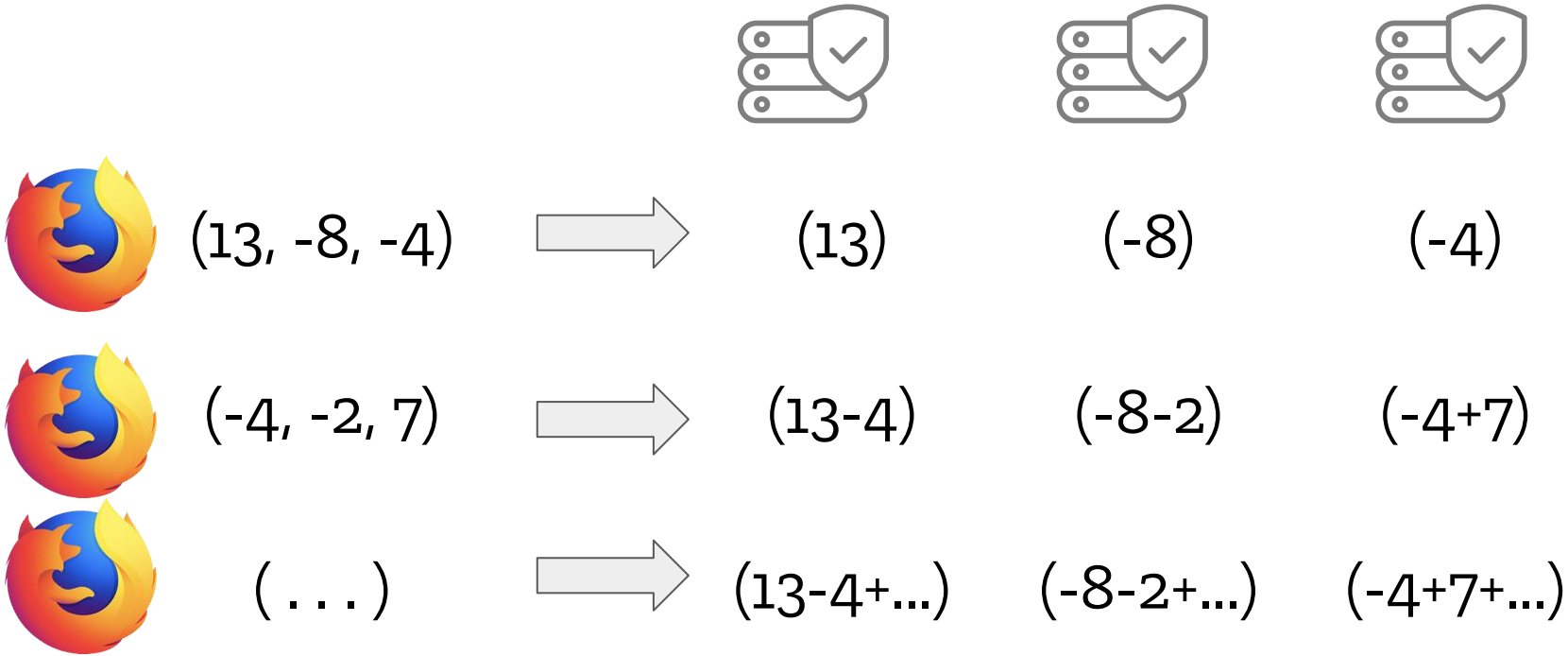


Prio

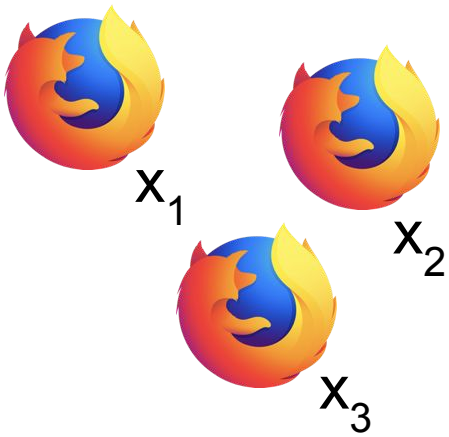
A system for privacy preserving aggregates

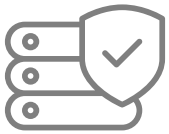
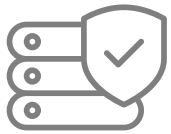
Private	Clients are anonymous if at least one server is honest
Robust	Malicious clients cannot corrupt the resulting output
Scalable	Secret-shared non-interactive proofs

Scheme for Private Sums



Scheme for Private Sums

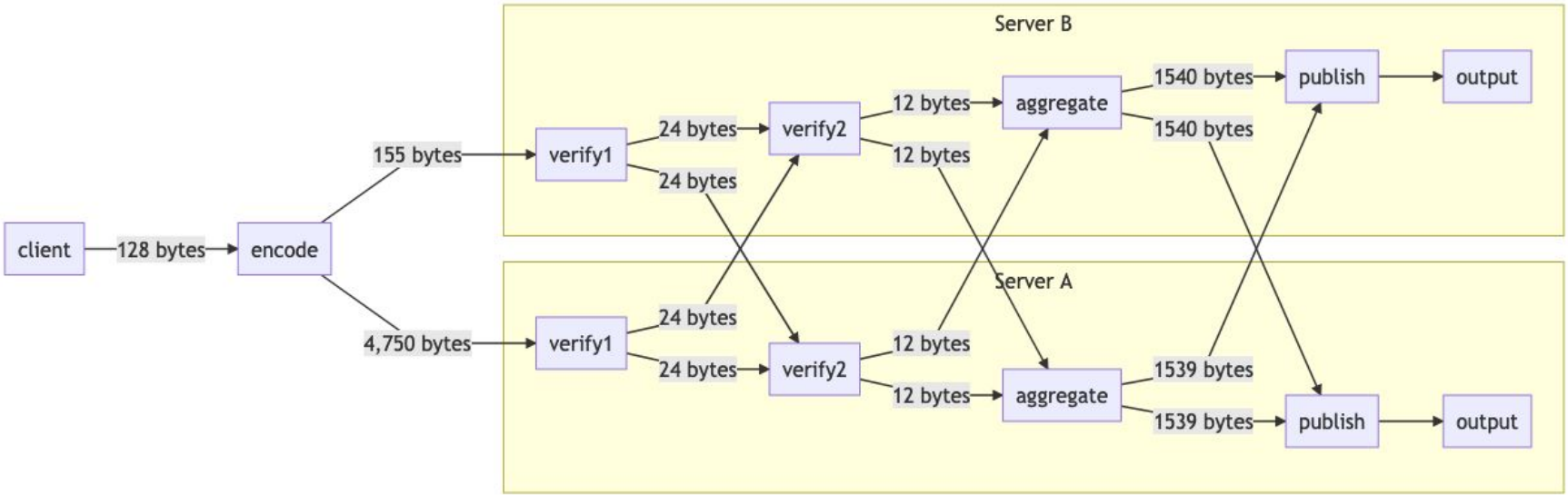



(13-4+...) (-8-2+...) (-4+7+...)

$= x_1+x_2+x_3$

Data Flow


Server B
Server A
client
128 bytes
encode
155 bytes
4,750 bytes
verify1
verify1
24 bytes
24 bytes
24 bytes
24 bytes
verify2
verify2
12 bytes
12 bytes
12 bytes
12 bytes
aggregate
aggregate
1540 bytes
1540 bytes
1539 bytes
1539 bytes
publish
publish
output
output

Pilot Program and Verification

libprio - A Prio library in C using NSS

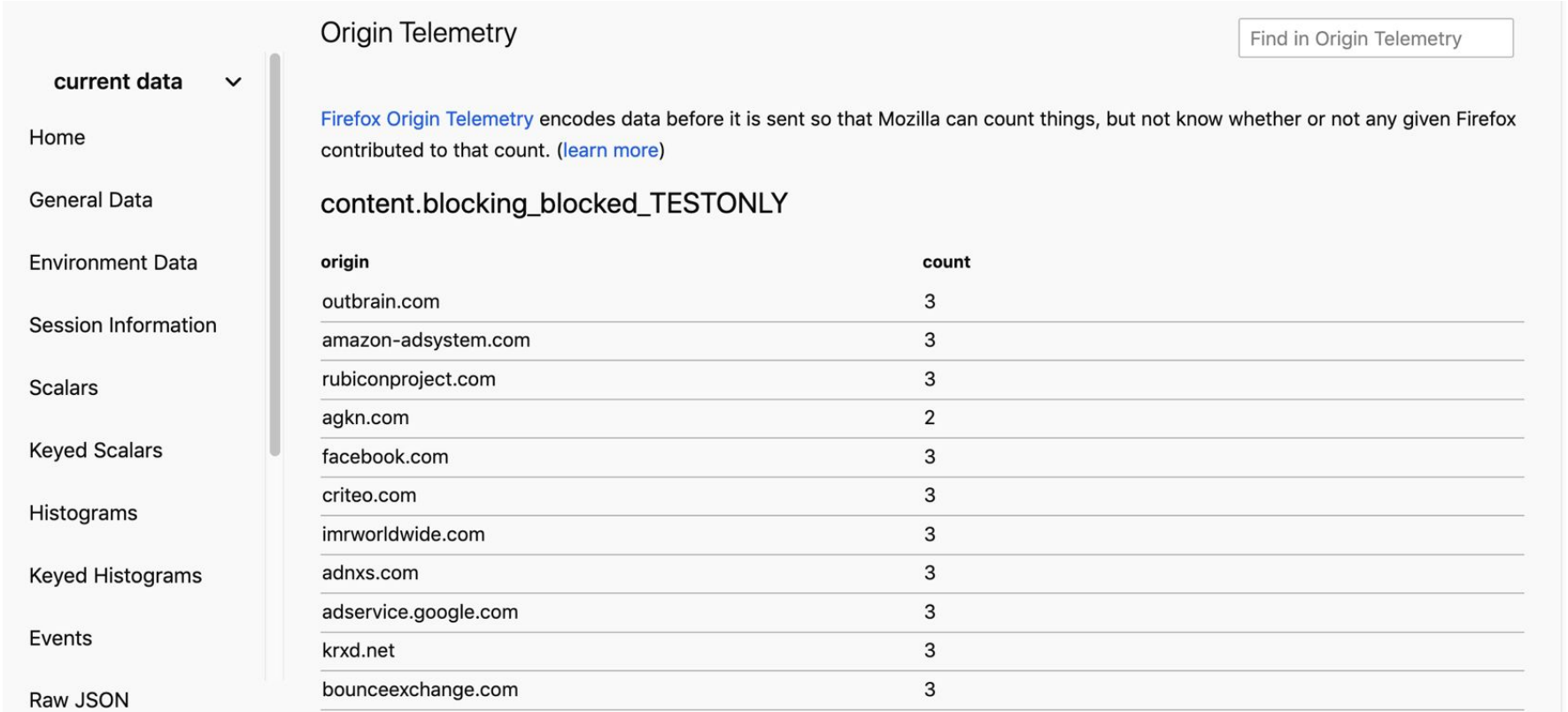
Warning: We do our best to write bug-free code, but I have no doubt that there are scary bugs, side-channel attacks, and memory leaks lurking herein.

Security bugs: If you find a security-critical bug in libprio, please report it to Mozilla using the contact information on this page.

Verifying Correctness

reason	prio_control	prio_observed	prio_diff
shutdown	[14891, 1578, 1]	[14891, 1578, 1]	[0, 0, 0]
shutdown	[10829, 3160, 1]	[10829, 3160, 1]	[0, 0, 0]
shutdown	[18333, 2939, 3]	[18333, 2939, 3]	[0, 0, 0]
aborted-session	[467, 155, 0]	[459, 150, 0]	[8, 5, 0]
aborted-session	[470, 129, 0]	[465, 129, 0]	[5, 0, 0]
aborted-session	[265, 49, 0]	[261, 49, 0]	[4, 0, 0]
daily	[1067, 471, 0]	[0, 0, 0]	[1067, 471, 0]
daily	[935, 325, 0]	[0, 0, 0]	[935, 325, 0]
daily	[1566, 934, 1]	[0, 0, 0]	[1566, 934, 1]
environment-change	[1923, 500, 0]	[0, 0, 0]	[1923, 500, 0]
environment-change	[1411, 697, 0]	[0, 0, 0]	[1411, 697, 0]
environment-change	[2100, 684, 0]	[0, 0, 0]	[2100, 684, 0]

Firefox Origin Telemetry and Validation

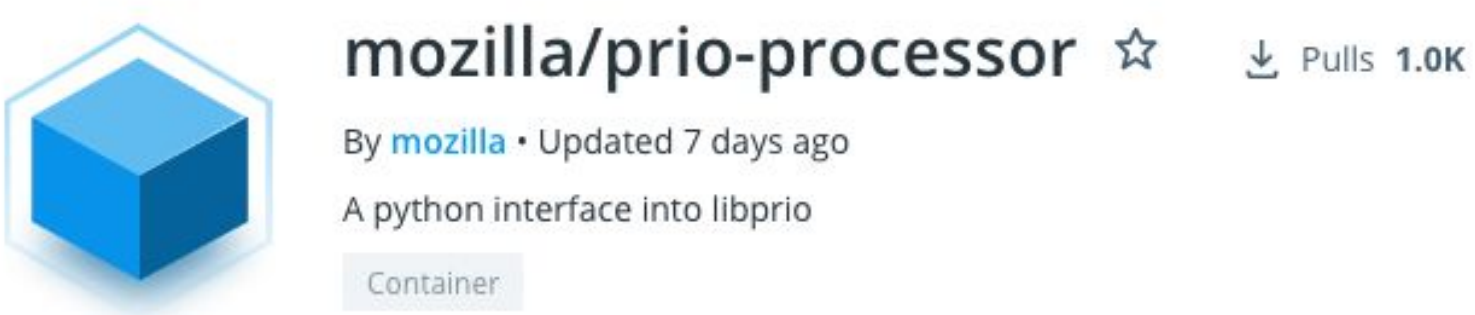


“prio” ping

This ping transmits Origin Telemetry data.

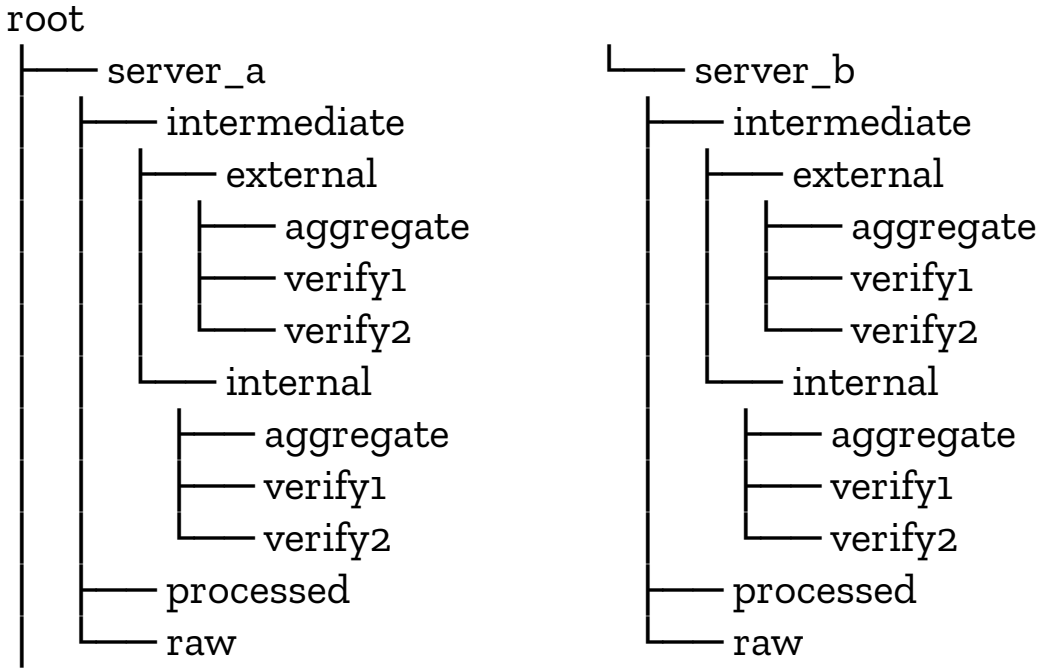
The client id is not submitted with this ping. The Telemetry Environment is not submitted in this ping.

```
{
  "type": "prio",
  ... common ping data
  "payload": {
    "reason": {periodic, max, shutdown}, // Why the ping was submitted
    "prioData": [{
      encoding: <encoding name>, // Name of App Encoding applied. e.g. "content-blocking-1"
      prio: {
        // opaque prio-specific payload. Like { a: <base64 string>, b: <base64 string> }
      },
    }, ... ],
  }
}
```



- Spark ETL to prepare Prio pings from Mozilla ingestion (BigQuery, GCS)
- Scripts for idempotent batch processing using S3-compatible object store
- Deployed as a docker image and configured via environment variables e.g. `docker run mozilla/prio-processor:latest`

Interoperability via Cloud Object Storage




On
DAG: prio_processor
schedule: 1 day, 0:00:00
Graph View
Tree View
Task Duration
Task Tries
Landing Times
Gantt
Details
Code
Refresh
Delete
success
Base date:
2019-09-30 00:02:56+0(
Number of runs:
25
Run:
scheduled__2019-09-30T00:02:55.575489+00:00
Layout:
Left->Right
Go
Search for...
GoogleCloudStorageToGoogleCloudStorageOperator
PythonOperator
SubDagOperator
success
running
failed
skipped
rescheduled
retry
queued
no status
bootstrap
staging
clean_processor_b
load_processor_b
trigger_processor_b
processor_b
clean_processor_a
load_processor_a
trigger_processor_a
processor_a
insert_into_bigquery

submission_date	origin	AGGREGATE	origin_rank	pageloads	normalized
2020-07-12	youtube.com	3	1	26	0.12
2020-07-12	__UNKNOWN__	2	2	26	0.08
2020-07-12	doubleclick.net	2	3	26	0.08
2020-07-12	quantserve.com	2	4	26	0.08
2020-07-12	accounts.google.com	1	5	26	0.04

Current State

- Collection is enabled on Firefox Nightly (opt-in) for ~2500 block list rules
- In the process of engaging with other organizations for running another server
- libprio Python bindings deployed to PyPi (pip install prio)
- Spark backend implementation in prio-processor

Links

[Paper] Prio: Private, Robust, and Scalable Computation of Aggregate Statistics

[Blog] Testing Privacy-Preserving Telemetry with Prio

[Blog] Firefox Origin Telemetry: Putting Prio in Practice

[Blog] Next steps in privacy-preserving Telemetry with Prio

[GitHub] mozilla/libprio

[GitHub] mozilla/prio-processor

[Firefox Source Docs] Origin Telemetry