Swarming Postmortem: 2014-12-04

Status:final

Summary

- Description: Incapacity to create a Swarming task
- Component(s): Chromium CI, CQ, TS
- Date/time: 2014-12-04 14:15:57 UTC
- Duration of problem: 47:22
- User impact: Loss of CI coverage, False negative on CQs. **15518 tasks** were aborted.

Owner(s)

maruel@chromium.org

Timeline

`14:15:57` 1167-c87df13.2014 is pushed to chromium-swarm. Author was in a hurry.
`14:16:00.392764` First exception report, 3 seconds after rolling in the broken version.
`14:49` First user report about HTTP 500 sent to two internal MLs (c-i-t@ and s-e@).
`15:00` User report sent on irc on #chromium
`15:00` Automated report from both swarming and isolate server.
`15:01` Ack of failure and start of investigation.
`15:01` User report sent to 3rd internal ML (c-t@) alerting Swarming failures.
`15:03:21` chromium-swarm is rolled back to 1166-53a754b.
`15:03:22.693404` Last exception report, 1 second after the rollback was completed.

Root causes

1. Type of *365*24*60*60*1000*1000* is **int** on local server, **long** on prod.
2. Overzealous assert asserted that the variable type was int.

Action Items

Automate deployment

Action Item	Owner	Tracking bug	Notes
Write remote_smoke_test.py to automate remote smoke testing; similar to local_smoke_test.py	maruel	swarming:186	
Write deployment script	maruel	swarming:187	

This would have caught this issue immediately on the staging server and would still be fast enough (<5 minutes) to be usable. Doing this has been discussed for a long time but was not done yet.

Increase ereporter2 rate

Change it from 1h to 5m cycle and see if we can survive the flow.

What Worked

- ereporter2 reported the failure on its hourly schedule.
- ereporter2 could be used to know both the extent of the downtime and the exact number of task requests denied.
- Maintainer monitors irc and emails.

What didn't Work

- Local tests passed.
- Staging server was not used at all before pushing to prod.
- Alert send to s-e@ and c-i-t@ was ignored for 11 minutes.
- CI, TS and CQ were affected.

Lessons Learned

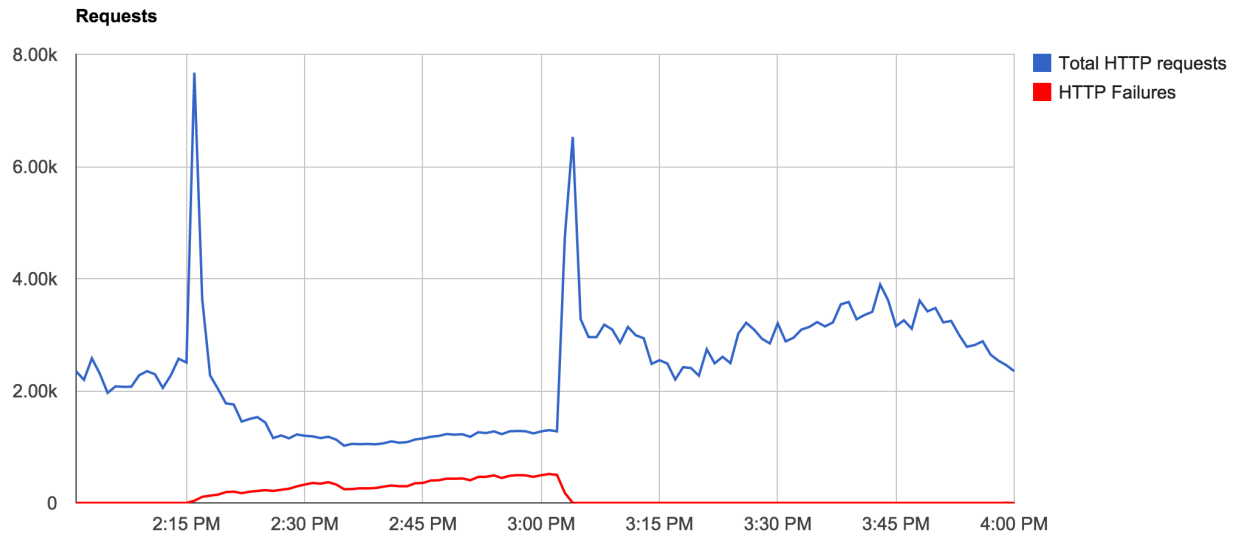
Don't bypass the staging server, damn!

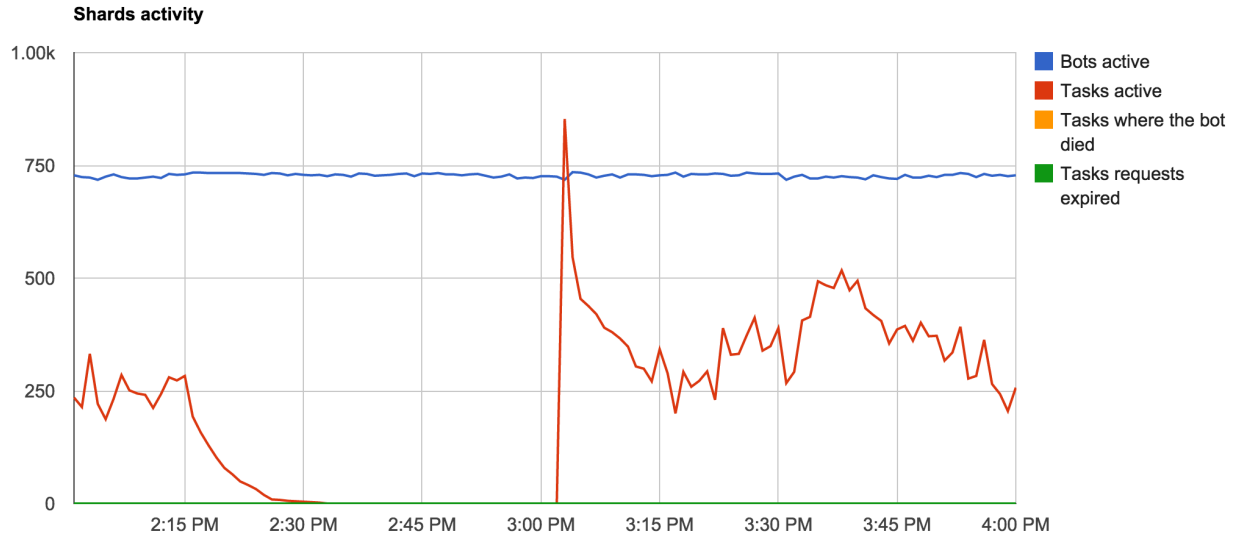
References

Public

IRC: http://echelog.com/logs/browse/chromium/1417647600

Swarming server stats:
https://chromium-swarm.appspot.com/stats?resolution=minutes&now=2014-12-04%2016:00





Private

https://appengine.google.com/adminlogs?tz=UTC&app_id=s~chromium-swarm
http://g/chrome-troopers/3MHstklzMf8
http://g/chrome-infrastructure-team/o7QxSzKsVbI