Summary

Some numbers to show what goes on in sites that do not share information about their visitors (unlike Windows-centric sites which target non-technical audiences)

The common perception of GNU/Linux is that it is scarcely used, based on statistics gathered from privacy-hostile Web sites that share (or sell) access log data, embed spyware in all of their pages, and so on. Our sites are inherently different because of a reasonable -- if not sometimes fanatic -- appreciation of privacy at both ends (server and client). People who read technical sites know how to block ads, impede spurious scripts etc. These sites also actively avoid anything which is privacy-infringing, such as interactive 'social' media buttons (these let third parties spy on all visitors in all pages).

Techrights and Tux Machines attract the lion's share our traffic (and server capacity). They both have dedicated servers. These are truly popular and some of the leaders in their respective areas. Techrights deals with threats to software freedom, whereas Tux Machines is about real-time news discovery and organisation (pertaining to Free software and GNU/Linux).

The Varnish layer, which protects both of these large sites (nearly 100,000 pages in each, necessitating a very large cache pool), handles somewhere between a gigabyte to 2.5 gigabytes of data per hour (depending on the time of day, usually somewhere in the middle of this range, on average).

The Apache layer, which now boasts 32 GB of RAM and sports many CPU cores, handled 1,324,232 hits for Techrights (ranked 6636th for traffic in Netcraft [2]) in this past week and 1,065,606 for Tux Machines (ranked 6214th for traffic in Netcraft [3]).

Based on VISITORS Web Log Analyzer [4], this is what we've had in Techrights:

Windows: (36.2%)
Linux: (31.8%)
Unknown: (e.g. bots/spiders): (23.0%)
Macintosh: (8.8%)
FreeBSD: (0.1%)

As a graph (charted with LibreOffice):
Tux Machines reveals a somewhat different pattern. Based on `grep`ing/filtering the of past month's log at the Apache back end (not Varnish, which would have been a more sensible but harder thing to do), presenting the top 3 only:

One month is as far as retention goes, so it's not possible to show long-term trends (as before, based on Susan's summary of data [5]). Logs older than that are automatically deleted, as promised, for both sites -- forever! We just need a small tail of data (temporarily) for DDOS prevention. ?

Site News

Source URL: http://www.tuxmachines.org/node/77503

Links: