RSS Podcast Feed Inefficiency brought to you by Apple, Amazon, and Peers

Damon Hart-Davis *

August 19, 2024

DRAFT

Abstract

Centralised social media systems are somewhat out of favour in 2024 for reasons from fake news and privacy to the actions of billionaire owners. Federated and more decentralised systems such as Mastodon and the Fediverse, plain old email, and RSS feeds including podcasts, are cool again. With much of the workings being out of sight for ordinary users, and in a system designed before intermittent renewable power generation was a thing, podcasting and RSS in particular are unnecessarily wasting an appreciable portion of their bandwidth and CPU time, and adding to climate change. Indeed, key players are consuming orders of magnitude more resources across their systems and others than necessary. There are already several simple and widely-used technical mechanisms that could help, but many participants are ignoring them. This paper suggests sustainability improvements for elements of the ecosystem that should be largely transparent to end users, including Cache-Control, conditional GET and skipHours, saving likely much more than 100kWh per day of electricity globally.

1 Introduction

In a world fighting climate change, one of our most powerful weapons is efficiency. There is no cheaper nor lower-carbon kWh of electricity for a participating client or server than the kWh not used at all. RSS feeds [2] are regaining popularity as a robust decentralised distribution for blogs and podcasts, with a comparatively ancient (in Internet timescales) and stable protocol but with a vibrant ecosystem and useful extensions regularly added and used It is problem that some of the biggest central participants such as directories are consuming bandwidth and CPU resources (their own, and

*Damon Hart-Davis is at the Centre for Environment and Sustainability, University of Surrey, Guildford, UK.

that of creators) at two or three orders of magnitude more than is useful. This at the very least is money wasted and unnecessary carbon pollution. A huge amount of electricity, not nearly all zero-carbon, is being consumed in data centres that host RSS servers and clients, such as in Ireland where those data centres use more than urban homes [1]. As of 2019 data centres accounted for approximately 3% of the electricity consumption and almost 4% greenhouse gas emissions globally [8] and in the UK alone will likely grow six-fold over the next decade [7]; the waste described here is one small piece of the puzzle to fix. A few days of engineering time across the top perpetrators would easily recover much of this, but they have indicated to the author that they see no problem in what they do and are unwilling to change their behaviour. They are ignoring multiple signals both RSS-specific and so generic that all mainline browsers already observe them Indeed the Internet would be unworkable slow without browsers doing so. Better news is that many of the niche and end-user client developers are very willing to engage and improve and almost disappear out of usage logs, supporting the notion that this is easily fixable. This paper both describes what is needed by existing participants to reduce the waste, and some defensive measures, stateless and stateful, that feed servers can take to mitigate load from bad actors.

It should be noted, however, that energy consumption in data centres and networks is not necessarily simply (positively) related to traffic [4–6]. Though fewer hits and bytes should allow at least some equipment to use less power dynamically, eg CPUs to spend more time asleep, or alternatively to postpone a little upgrades for otherwise growing traffic.

Once these very low-level issues are fixed, invisible to end users such as podcast listeners, there is another layer of more salient improvements that directories, aggregators and clients could make to reduce streaming bandwidth by an order of magnitude in some cases. There are large savings available in other (especially podcasting) aspects too, outside of the scope of this paper, such as handling of transcripts and cover art, that could be tackled in parallel. A future paper aims to address those.

1.1 Structure

The paper is structured as follows. Section refS-XXX presents XXX, ...

2 Literature review

3 Method

3.1 Case study podcast

3.2 Signals available

3.3 Technical defences erected

3.4 Sonification

Per [9]...

4 Results

4.1 Responses to signals given

4.2 Traffic hits and bytes

An analysis of feed traffic by 'hits' (number of HTTP accesses) and 'bytes' (nominally HTTP response bytes, excluding some overheads) is presented in Table 1. This is approximately week-long slices, exact boundaries determined by when the Web server logs happen to be rolled, show normalised to per-day and percentage-of-full-site values. Sample visualisaton are shown in Figure 1.

The overall story is that the defensive techniques deployed have reduced 'bytes' per day by about a factor of three, though 'hits' have go up a little because of the bad response of many clients to being asked to slow down with an HTTP error response (429 or 503) where they instead poll more and faster.

Figure 1: Visualisations of the RSS feed data by hour of day UTC to 2024-08-19 for each data chunk (approximately a week in each case) showing just bits/bytes per hour (top) and additionally HTTP response code (middle) and in an alternate presentation (bottom). An overall downward trend in load (left to right) over time is visible, though with a hump in the middle where Googlebot misbehaved badly until manually banned. A time-of-day pattern is also fairly evident within each chunk, amplified once the worst offenders were only allowed to download the feed at noon (in the middle of each chunk) and when low-carbon power is available. These sample visualisations are generated as side-effects of sonification.

Interval		Feed/day		% of EOU site traffic		Selected events
Ending	Days	Hits	MBytes	Hits	Bytes	
2024-04-01T06:25Z	8	1077	16.8	7.5	1.5	
2024-04-15T06:25Z	8	1205	21.9	8.3	2.3	
2024-04-21T06:25Z	6	1027	11.6	4.2	0.9	
2024-04-29T06:25Z	8	1204	12.8	7.6	0.7	2024-04-23: added Spotify with lite feed
2024-05-05T06:25Z	6	1285	12.6	8.9	1.4	
2024-05-13T06:24Z	8	1070	7.1	6.4	0.8	
2024-05-19T06:25Z	6	1144	6.1	7.5	0.6	
2024-05-27T06:25Z	8	1607	8.9	11.2	0.7	2024-05-24: Googlebot goes rogue
2024-06-02T06:25Z	6	2393	10.8	14.2	0.4	
2024-06-10T06:25Z	8	2345	11.6	13.6	0.5	2024-06-08: added 503 rejections for top-3 bots
2024-06-16T06:25Z	6	2282	11.5	11.6	0.5	
2024-06-23T06:25Z	7	2322	11.4	11.4	0.4	
2024-07-01T06:25Z	8	1460	7.0	8.4	0.2	2024-06-24: Googlebot reined in; blocked from feed in robots.txt 2024-06-28: added skipDays to podcast feed
2024-07-07T06:25Z	6	1562	7.7	8.7	0.5	2024-07-05: now using 503 error codes instead of 429s; more clients respond in some way to 503
2024-07-14T06:25Z	7	1666	8.0	8.3	0.4	
2024-07-22T06:25Z	8	1526	6.4	9.2	0.4	
2024-07-28T06:25Z	6	1557	6.1	8.8	0.2	
2024-08-05T06:25Z	8	1524	6.6	9.3	0.6	
2024-08-11T06:24Z	6	1510	6.5	8.6	0.4	2024-08-08: RSS feed lastBuildDate now timestamp of newest primary media file in feed after any filtering, so much older than previously
2024-08-19T06:25Z	8	1274	5.1	8.2	0.4	

Table 1: Summary of RSS feed fetch stats over time.

5 Conclusions

This paper makes two contributions to the important debate on efficiency in network protocol use in practice in climate mitigation. First, ...

Funding

This research received no external funding.

Data Availabilty

The data and code supporting this research is available at `https://www.earth.org.uk/img/research/RSS-efficiency/` and Zenodo [3].[1] Some obfuscation of IP address and User-Agent values has been applied for the privacy of human users of the podcast.

Acknowledgments

Thanks to James Cridland of Podnews and Michael Boelen amongst others — for support, ideas, and data!

Declaration of interests

The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

EOU	Earth.Org.Uk (website)
RSS	Really Simple Syndication

References

[1] National energy and climate plan (NECP) 2021-2030, 07 2024. URL `https://www.gov.ie/en/publication/a856a-national-energy-and-climate-plan-necp-2021-2030/`.

[2] RSS Advisory Board. RSS 2.0 specification, 03 2009. URL `https://www.rssboard.org/rss-specification`.

[3] Damon Hart-Davis. RSS feed efficiency study dataset, 08 2024. URL `https://zenodo.org/records/13292718`.

[1] 2024-08-10 note: still incomplete and a work in progress.

[4] Kerry Hinton, Jayant Baliga, Michael Feng, Robert Ayre, and Rodney S. Tucker. Power consumption and energy efficiency in the internet. *IEEE Network*, 25(2):6–12, 03 2011. ISSN 0890-8044. doi: 10.1109/mnet.2011.5730522. URL https://ieeexplore.ieee.org/abstract/document/5730522.

[5] Jonathan Koomey and Eric Masanet. Does not compute: Avoiding pitfalls assessing the internet's energy and carbon impacts. *Joule*, 5(7): 1625–1628, 07 2021. ISSN 2542-4351. doi: 10.1016/j.joule.2021.05.007. URL https://www.sciencedirect.com/science/article/pii/S2542435121002117.

[6] Louise Krug, Mark Shackleton, and Fabrice Saffre. Understanding the environmental costs of fixed line networking. In *Proceedings of the 5th international conference on Future energy systems*, e-Energy '14. ACM, 06 2014. doi: 10.1145/2602044.2602057. URL https://dl.acm.org/doi/abs/10.1145/2602044.2602057.

[7] John Pettigrew. Transforming the supergrid of the 1950s to a network built on an electrified future for generations to come., 03 2024. URL https://www.linkedin.com/pulse/transforming-supergrid-1950s-network-built-future-come-john-pettigrew-jpbcf/.

[8] Xiaolei Yuan, Yumin Liang, Xinyi Hu, Yizhe Xu, Yongbao Chen, and Risto Kosonen. Waste heat recoveries in data centers: A review. *Renewable and Sustainable Energy Reviews*, 188:113777, 12 2023. ISSN 1364-0321. doi: 10.1016/j.rser.2023.113777. URL https://www.sciencedirect.com/science/article/pii/S1364032123006342.

[9] Jonathan Zong, Isabella Pedraza Pineros, Mengzhu (Katie) Chen, Daniel Hajas, and Arvind Satyanarayan. Umwelt: Accessible structured editing of multi-modal data representations. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24. ACM, 05 2024. doi: 10.1145/3613904.3641996. URL https://dl.acm.org/doi/10.1145/3613904.3641996.