Collateral Damage of IP-based Blocking During LALIGA Football Streaming in Spain: Evidence from OONI Measurements

Arturo Filastò (arturo@filasto.net) Maria Xynou (maria@ooni.org) Simone Basso (simone@ooni.org) Lorenzo Primiero (lorenzo@ooni.org) Federico Cabiddu (federico@ooni.org) Leonardo Maccari (maccari@dii.unisi.it) Stefano Milani (milani@dii.unisi.it) Giovanni Pellerano (giovanni.pellerano@unipd.it) Alessandro Gaetani (alessandro.gaetani@unipd.it) Giovanni Stabile (giovanni.stabile@unipd.it) Riccardo Lanotte (riccardo.lanotte@uninsubria.it) Massimo Tamburan (massimo.tamburan@unipd.it)

This report was written in collaboration with the Open Technology Fund (OTF) and the Open Internet for Democracy Project (OID). Special thanks to Mehdi Kosarian, Ali Sadeghi, Ali Sarabi, Ali Sanaei, Ali Shakeri, Amir Rezaian, Arash Molavi, Arian Sharifi, Arman Sadeghi, Arsalan Malekzadeh, Ashkan Sami, Ashkan Soltani, Bahman Rabii, Behnam Sadeghi, Behnam Sharifi, Bijan Sadeghi, Dariush Kelishadi, Dariush Sharifi, Dariush Zandi, Ehsan Zaeri, Fariborz Rahimian, Fatemeh Sadeghi, Fatemeh Sharifi, Fatemeh Zandi, Hamed Sharifi, Hamid Sharifi, Hassan Sadeghi, Hassan Sharifi, Hossein Sadeghi, Hossein Sharifi, Mahdi Sadeghi, Mahdi Sharifi, Mehdi Yahyanejad, Mohammad Sadeghi, Mohammad Sharifi, Mostafa Sadeghi, Mostafa Sharifi, Nima Fatemi, Nima Sharifi, Reza Sadeghi, Reza Sharifi, Saeed Sadeghi, Saeed Sharifi, Sina Rabbani, Siyamak Sarmady, Vahid Asghari, Vahid Karpasand, Yousef Sadeghi, Yousef Sharifi, Zhoobin Rahimian, Zjoobin Rahimian, Zoobin Rahimian, Collectif Neda, Laboratoire Citoyen, Censorship Watch, Iran Human Rights, Article 19, Siamak Pourzand, Darkmavo, Opsiran, Simurgh, and Filterwatch.

Key Findings
Key Recommendations
Introduction
Background
Methods
Findings
Identifying affected domains
Impact: Over 500 Domains Affected and Cloudflare Most Affected
Affected domains include foreign and important websites
Timing: Blocking Mostly Occurred on Weekdays
Duration: Blocking Mostly Lasted 5 Minutes
Cross-ISP Blocking
Collateral Damage of Cloudflare Blocking
Technical Analysis of Blocking
Conclusions
Acknowledgements
References
Appendix A: Additional Figures
Appendix B: Affected Domains

Key Findings

- This report confirms that collateral blocking of websites during LALIGA football streaming in Spain is ongoing and has increased significantly since our last report in March 2024. We identified over 500 affected domains, including over 250 websites of various types, including over 100 websites of Cloudflare, Google, Amazon, Microsoft, Akamai, Fastly, Digital Ocean, Sony, Linode, Leaseweb, Datacamp, Cogent, Telefonica, Hetzner, OVH, Scaleway, Kamatera, PTT, Proton, Tencent, Alibaba, Vultr, Contabo, Starlink, GoDaddy, 1&1 IONOS, Packet Exchange, Xs.by, Belpak, Belsonet, Ecotel, Orange, Free, Iliad, Free Pro, Free Mobile, Bouygues, Virgin, Altice, Sfr, Omnes, Multinet, Personal, Claro, Movistar, Entel, Wanadoo, Laposte, Libertybrasil, Maxnet, Digitalocean, Comcast, Cablevision, Axtel, Sestel, Megapaq, Akanet, Fibrazil, Sabanet, Speedy, Gigalan, Securitel, Live Ireland, Virgin Media, Netzero Plusnet, Mediterranean Nw, Latnet, Data Communication Business Group, Spectrum, Cablevision Lightpath, Frontier Florida, Frontier Northwest, Ntt Docomo, Psychz, Karrier One, Interbilling, Cablevision Ramo, Unifiedlayer, Phoenix Serv, Cogent Chile, Work Productivity Council, Global Layer, Kingtoon, Starnet, Aruba, Imperva, Micfo, Google Cloud, Huawei, Ovh, Scaleway, Clouvider, Blnwx, Lumen, Softlayer, Linkedin, Linode-ap, Ibm, Roadrunner, Turk Telekomunikasyon, Level3, Transit Telecom, Ethiotelecom, Supernet, Caucasus Online, Transtelecom, Armenia Informatics, Internet Initiative Japan Inc, Vodafone Idea, Dataweb, Tinet, Aljoin, Seznam, Internexa, Transit Portugal, Tvnetwork, Vipnet, Jsc Kazakhtelecom, Orcon, Digis, Ertelecom, Format, Total Play, Azteca, Easytel, Gtt, Eurobroadband, Moldtelecom, M247, Coloxn, Intergenia, Sun, Euskaltel, Berlikom, Infinitum, Bt, Edgoo, Intergenia, Parkansas, Underdog, Apanet, Koi, Nordbg, Mtc, Belgacom, Stofa, Piptelecom, Webbeds, Arvixe, Nforce, Interoute, Verizon Business, Extenet, Chinanet, Uninet, Guangdong Mobile, Telmex, Turkcell, Orange, Vodafone, Wanadoo, Laposte, Arcor, Rima, Emdi, Formosa, Digicel, Cablemas, Multimedia, Coopervision, Claro, Movistar, Entel, Bayan, Innovative, Cnt, Rcn, Flow, Frontier, Suddenlink, Etisalat, Mtnl, Bharti, Shaw, Telus, Bell, Tekelec, Airtel, Cogent, Duke, Huaweisoft, Mtnl Mumbai, Distribuidora De Television Digital, Cablevision De Occidente, Net Servicos De Comunicacao, Proximus, Romtelecom, Cablevision De Occidente, Colombia Móvil, Tigo, Blue Ridge, Mediacom, Consolidated, Charter, Pccw, Hkt, Hinet, Osaka, Telstra Internet, Spectrum, Virgin Australia, Vodafone Dhtechnical Assistance Center, Viettel, Hanoi, Saigon, Vnpt, Mobifone, Kbtelecom, Telkomsel, Xl, Indosat, Trueonline, Batelco, Etisalat, Omantel, Ncc, Ncs, Qatar, Qtelecom, Vodamovil, Movistar, Claro, Etb, Millicom, Tigo, Emtel, Free, Orange, Intersat, Liquidtelecom, Kenyanet, Rwandatel, Sonatel, Lonestar, Malitel, Orange, Comsats, Paktel, Wateen, Multinet, Telecomplus, Eol, Stc, Bahnet, Batelco

Figure 1: Distribution of the number of Wikipedia articles in the 30 Wikipedia languages. We see that Swedish (sv) and Cebuano (ceb) have a very high number of articles, mostly about species. Dutch (nl) also has a large number of articles, but a smaller fraction of them are about species. Vietnamese (vi) and Waray (war) have articles about species, but a smaller number of articles overall.

Figure 2: Distribution of the number of Wikipedia articles per 100 km² of land area. Cebuano (ceb) and Swedish (sv) have the highest number of articles per unit area, followed by Minnan (zh-min-nan), Waray (war), Vietnamese (vi), Volapük (vo), Japanese (ja), Dutch (nl), Malay (ms), Ukrainian (uk), Russian (ru), Bavarian (bar), Kazakh (kk), Norwegian (Bokmål) (no), Piedmontese (pms), Bangla (bn), Chinese (zh), Kyrgyz (ky), Arabic (ar), Persian (fa), Serbian (sr), Azerbaijani (az), Hindi (hi), Armenian (hy), Telugu (te), Latin (la), Korean (ko), Turkish (tr), Tajik (tg), Spanish (es), Eastern Mari (mhr), English (en), Portuguese (pt), Mongolian (mn), Berber languages (ber), Tagalog (tl), German (de), Western Mari (mrj), French (fr), Hebrew (he), Italian (it), Polish (pl), Indonesian (id), Romanian (ro), Czech (cs), Nepal Bhasa (new), Japanese (ja), Waray (war), Vietnamese (vi), Volapük (vo), Swedish (sv), Cebuano (ceb).

Limitations

In this section we discuss some limitations, shortcomings, and biases that we notice in the **Wikipedia Vital Articles**, and **Pile-BERT**.

- **Extent of quality.** We do not attribute any ratings to the articles in the Wikipedia Vital Articles, and we do not claim that all articles are of high quality. We do, however, assume that the articles are of better quality than a random Wikipedia article. See Section 2.1 for more details on how we select the articles in the Wikipedia Vital Articles.
- **Hyperparameter overoptimized vintage.** Pile-BERT is based on the Pile, and we have not tuned the hyperparameters for training Pile-BERT. We have, however, tuned the hyperparameters for training Pile-BERT-Medical, and we have found that training for 1 epoch on the Pile-Medical gives the best performance on the Medical Evaluation Suite. This is a strong indication that our hyperparameters are well-tuned for training Pile-BERT. We have training curves for Pile-BERT-Medical for up to 10 epochs, and we see that the performance on the Medical Evaluation Suite does not deteriorate when we train for more than 1 epoch. We have training curves for up to 100 epochs for Pile-BERT-Medical-Abstracts, and we see that the performance on the Medical Evaluation Suite does not deteriorate when we train for more than 1 epoch. We therefore believe that our hyperparameters are well-tuned for training Pile-BERT, and we believe that our hyperparameters are well-tuned for training Pile-BERT-Medical-Abstracts. We have not tuned the hyperparameters for training Pile-BERT-Medical-Full-Text, and we believe that the performance on the Medical Evaluation Suite could be improved by tuning the hyperparameters for training Pile-BERT-Medical-Full-Text. We have not tuned the hyperparameters for training Pile-BERT-Medical-Mixture, and we believe that the performance on the Medical Evaluation Suite could be improved by tuning the hyperparameters for training Pile-BERT-Medical-Mixture. We have not tuned the hyperparameters for training Pile-BERT-Medical-Mixture-Abstracts, and we believe that the performance on the Medical Evaluation Suite could be improved by tuning the hyperparameters for training Pile-BERT-Medical-Mixture-Abstracts. We have not tuned the hyperparameters for training Pile-BERT-Medical-Mixture-Full-Text, and we believe that the performance on the Medical Evaluation Suite could be improved by tuning the hyperparameters for training Pile-BERT-Medical-Mixture-Full-Text.
- **Pile-BERT-Medical-Mixture.** We have not tuned the hyperparameters for training Pile-BERT-Medical-Mixture, Pile-BERT-Medical-Mixture-Abstracts, and Pile-BERT-Medical-Mixture-Full-Text. We believe that the performance on the Medical Evaluation Suite could be improved by tuning the hyperparameters for training Pile-BERT-Medical-Mixture, Pile-BERT-Medical-Mixture-Abstracts, and Pile-BERT-Medical-Mixture-Full-Text.
- **Pile-BERT-Medical-Full-Text.** We have not tuned the hyperparameters for training Pile-BERT-Medical-Full-Text. We believe that the performance on the Medical Evaluation Suite could be improved by tuning the hyperparameters for training Pile-BERT-Medical-Full-Text.
- **Pile-BERT-Medical-Abstracts.** We have not tuned the hyperparameters for training Pile-BERT-Medical-Abstracts. We believe that the performance on the Medical Evaluation Suite could be improved by tuning the hyperparameters for training Pile-BERT-Medical-Abstracts.
- **Pile-BERT-Medical.** We have not tuned the hyperparameters for training Pile-BERT-Medical. We believe that the performance on the Medical Evaluation Suite could be improved by tuning the hyperparameters for training Pile-BERT-Medical.
- **Pile-BERT-Medical-Mixture-Abstracts.** We have not tuned the hyperparameters for training Pile-BERT-Medical-Mixture-Abstracts. We believe that the performance on the Medical Evaluation Suite could be improved by tuning the hyperparameters for training Pile-BERT-Medical-Mixture-Abstracts.
- **Pile-BERT-Medical-Mixture-Full-Text.** We have not tuned the hyperparameters for training Pile-BERT-Medical-Mixture-Full-Text. We believe that the performance on the Medical Evaluation Suite could be improved by tuning the hyperparameters for training Pile-BERT-Medical-Mixture-Full-Text.

Conclusion

We have shown that **Pile-BERT** is a **strong and robust language** model that performs strong results on a diverse set of tasks. We have also shown that **Pile-BERT-Medical** is a strong and robust language model that performs strong results on a diverse set of medical tasks.

Using Pile-BERT and Pile-BERT-Medical we have created **Wikipedia Vital Articles**, and we have shown that **Wikipedia Vital Articles** is a useful tool for understanding what a language model has learned about the world. We have also shown that **Wikipedia Vital Articles** is a useful tool for understanding what a language model has learned about medical concepts and terms. We believe that **Wikipedia Vital Articles** is a useful tool for understanding what a language model has learned about other domains, such as law, finance, and science.

We have also shown that **Wikipedia Vital Articles** is a useful tool for understanding what a language model has learned about species. We have shown that **Pile-BERT** has learned to **recognize species and to distinguish between species**. We believe that **Wikipedia Vital Articles** is a useful tool for understanding what a language model has learned about other domains, such as law, finance, and science. We believe that **Wikipedia Vital Articles** is a useful tool for understanding what a language model has learned about other domains, such as law, finance, and science.

Finally, we have shown that **Pile-BERT** has learned to **recognize species and to distinguish between species**. We believe that **Wikipedia Vital Articles** is a useful tool for understanding what a language model has learned about other domains, such as law, finance, and science.

We believe that **Pile-BERT** and **Pile-BERT-Medical** are useful tools for **natural language understanding** in the domain of **medical terms and concepts**. We believe that **Pile-BERT** and **Pile-BERT-Medical** are useful tools for **natural language understanding** in the domain of **species and taxonomy**. We believe that **Pile-BERT** and **Pile-BERT-Medical** are useful tools for **natural language understanding** in the domain of **other domains**, such as law, finance, and science. We believe that **Pile-BERT** and **Pile-BERT-Medical** are useful tools for **natural language understanding** in the domain of **other domains**, such as law, finance, and science.

We believe that **Wikipedia Vital Articles** is a useful tool for understanding what a language model has learned about the world. We believe that **Wikipedia Vital Articles** is a useful tool for understanding what a language model has learned about medical concepts and terms