



Internetworking Challenges for AI

Dirk KUTSCHER



香港科技大学 (广州)
THE HONG KONG
UNIVERSITY OF SCIENCE AND
TECHNOLOGY (GUANGZHOU)

Outline

Impact of AI Crawlers

AI Preferences

Agentic AI

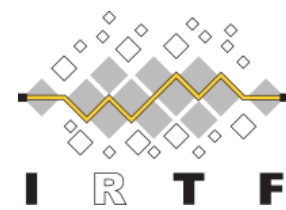
AI Infrastructure



Impact of AI Crawlers on the Internet

- Growing number of LLM services
- Constantly crawling the web
- Growing network traffic and server load
- Increasing interest to develop technical solutions and standards to control AI web crawler behavior

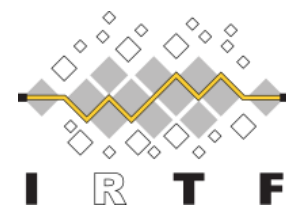




Measurement and Analysis for Protocols Research Group (MAPRG)

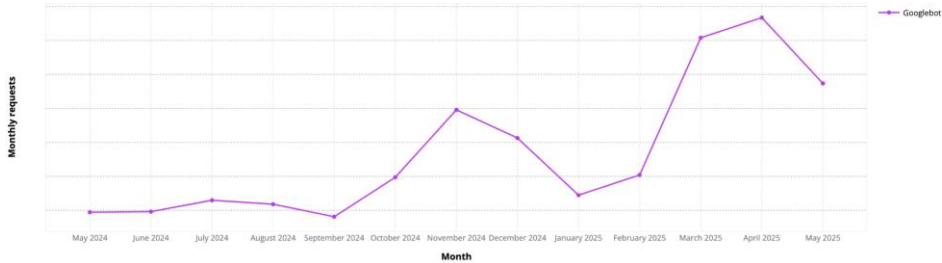
- **Special focus meeting on impact of AI traffic crawler impacts at IETF-123 in July 2025**
 - Assessing the impact on network, servers, CDN
 - Discussing technical ideas and current developments
- <https://datatracker.ietf.org/meeting/123/session/maprg>





AI Crawler Traffic

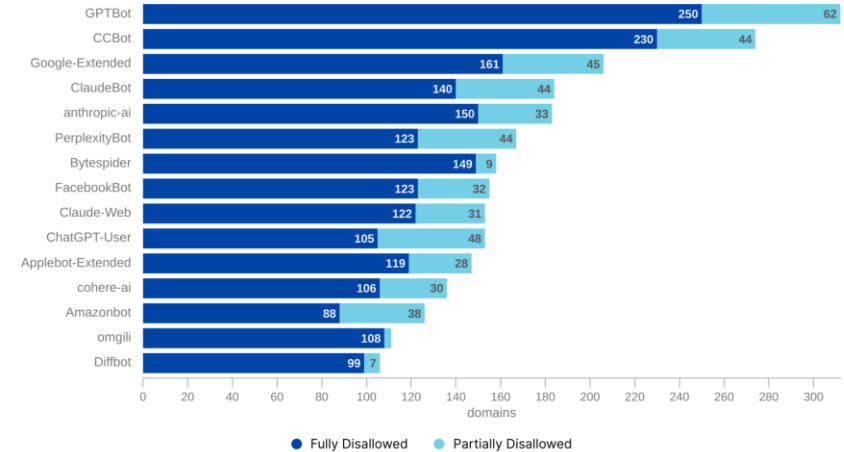
AI and search crawlers over time



AI user agents found in robots.txt

AI user agents found in 3816 robots.txt files across the top 10,000 domains

Directive: Disallow



Updated: Jun 2, 2025

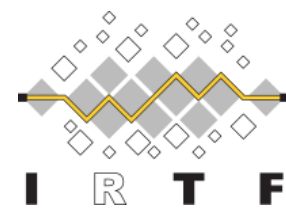


Last 7 days | Jun 6, 2025, 17:15 UTC

More traffic per crawler and more crawlers overall

<https://blog.cloudflare.com/from-googlebot-to-gptbot-whos-crawling-your-site-in-2025/>





Measurement and Analysis for Protocols Research Group (MAPRG)

• Some observations

- Automated traffic now makes up 30–50% of all web traffic
- AI crawlers are a major driver behind the recent surge
- Previously, crawlers indexed content and referred users: a fair exchange
- Now, their scraping imposes heavy infrastructure load, but offers little in return
- Bots now seen as exploitative and unwelcome

Anti-Crawler Protection is checking your browser and IP
[198.245.53.182](#) for spam bots

You will be automatically redirected to the
requested page after 3 seconds.
Don't close this page. Please, wait for 3 seconds
to pass to the page.

•••
3

The page was generated at Mon, 26 Feb 2024 10:17:16

Browser time Mon, 26 Feb 2024 10:17:27 GMT

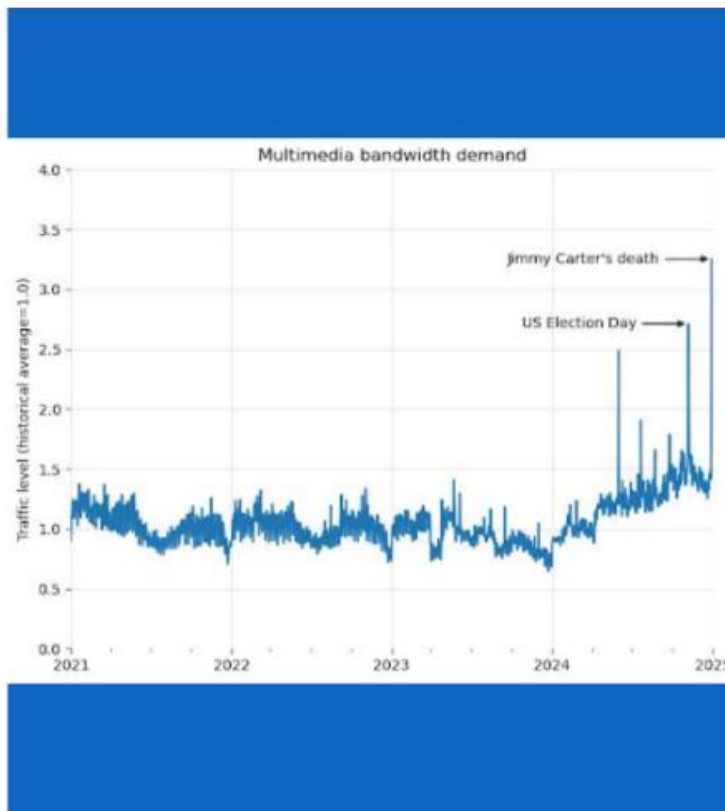
1210528, 15605, <https://myfso.org>, 0.27
[Anti-Spam by CleanTalk](#)



Example: Wikimedia

The Problem with Bots

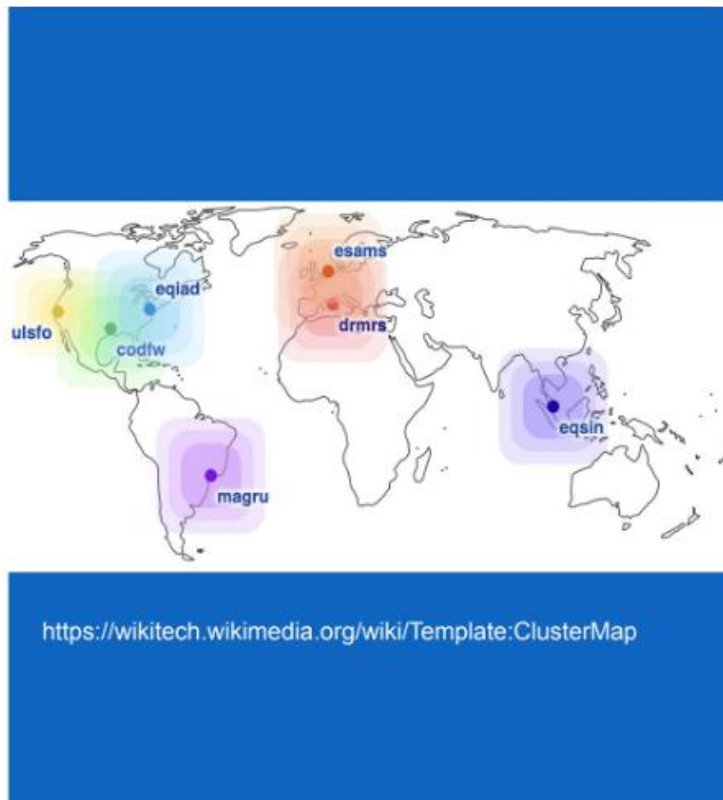
- Since January 2024, the Wikimedia Foundation has seen a 50% increase in bandwidth usage, largely caused by bots.
- Crawlers target not only the Wikimedia projects, but any URL within our infrastructure (e.g.: bug tracker, code review platform).
- Resources to host this demand have increased, especially given the newfound interest in high volumes of multimedia content.



Example: Wikimedia

Content delivery network: Optimized for humans

- 2 primary data centers, 5 caching centers
- Location-based DC-routing
- About 25B page views monthly
- Can sustain sudden, huge traffic spikes
- Requests by humans frequently terminate in the cache
- Requests by bots are more likely to get passed on to the main Data Center -> crawlers visit any URL; including less popular pages



Crawling Inefficiencies

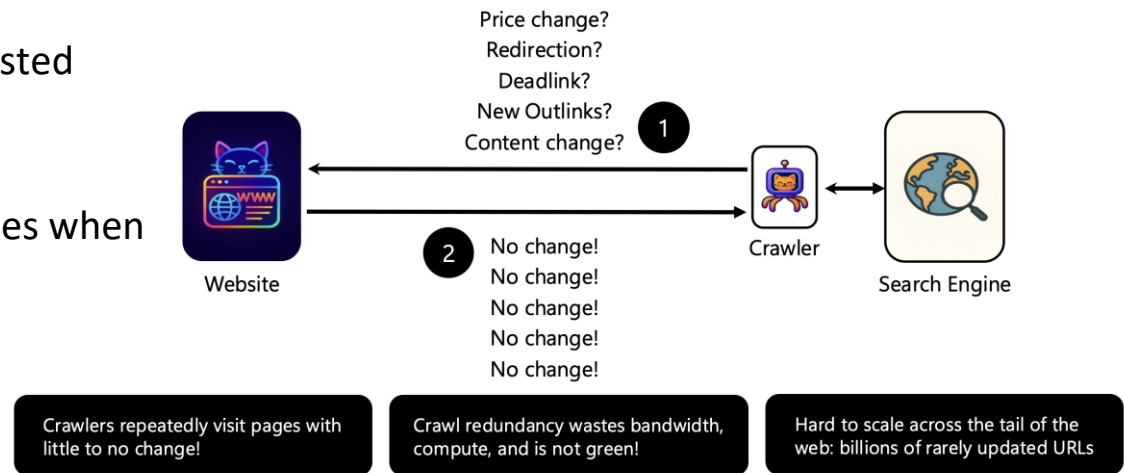
- Ratio of useful, fresh content crawled (new, updated, or deleted) relative to total crawl activity

- Not great today
- Many resources get requested repeatedly

- **New developments**

- Website to notify index sites when there are updates
- Crawlers would then start fetching the newer resources
- Example: IndexNow

Legacy crawl model is in bad need of innovation



Outline

Impact of AI Crawlers

AI Preferences

Agentic AI

AI Infrastructure



IETF AI Preferences Working Group

- **Rapid growth of AI systems that train on massive web datasets has created tension between:**
 - Publishers, authors, and rights holders who want to express whether (and how) their content can be used for AI training, inference, or related purposes.
 - AI developers and crawlers who currently lack standardized ways to discover and interpret those preferences at scale.
- **Today: no interoperable, machine-readable mechanism to communicate such preferences**
 - In a consistent way
 - Across different web resources



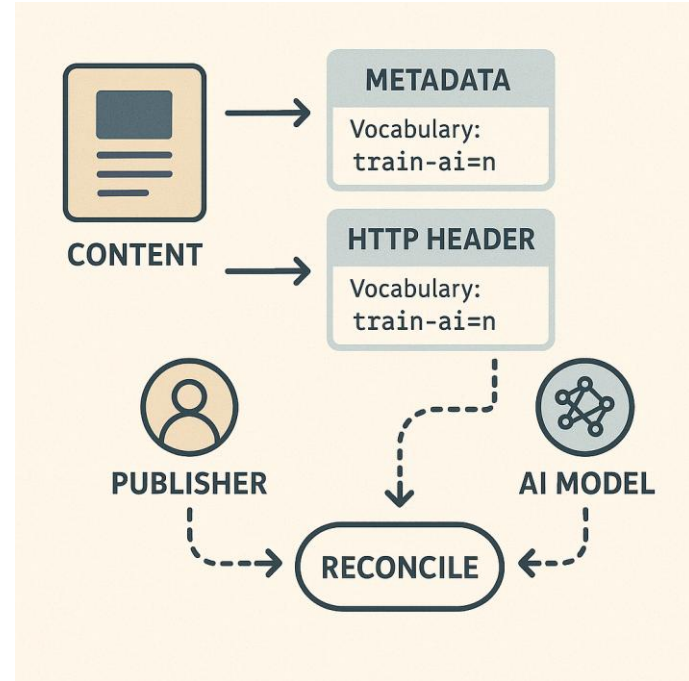
AIPREF

- **Mechanisms like robots.txt or metadata tags were designed for search engine crawling**
 - Not for nuanced AI-specific uses.
- **Different websites are experimenting with custom headers or formats, leading to fragmentation.**
- **AI crawlers have no common vocabulary for interpreting preferences**
 - “allow indexing but not training,”
 - “allow summarization but not reuse in inference”
- **Without standardization, content owners have little control**
 - AI crawlers face uncertainty about compliance



AIPREF Goals

- **AIPREF creating common framework**
 - Defines a clear vocabulary for AI-related content usage preferences.
 - Specifies standardized attachment methods (e.g., HTTP headers, robots.txt, metadata) for expressing those preferences.
 - Provides a reconciliation mechanism for resolving conflicts when multiple preference expressions exist.



Outline

Impact of AI Crawlers

AI Preferences

Agentic AI

AI Infrastructure



Agentic AI Challenges

Discovery of agents / APIs

Authentication & Identity

Authorization & Delegation

Lifecycle / Session Management

Safety, Security, Accountability

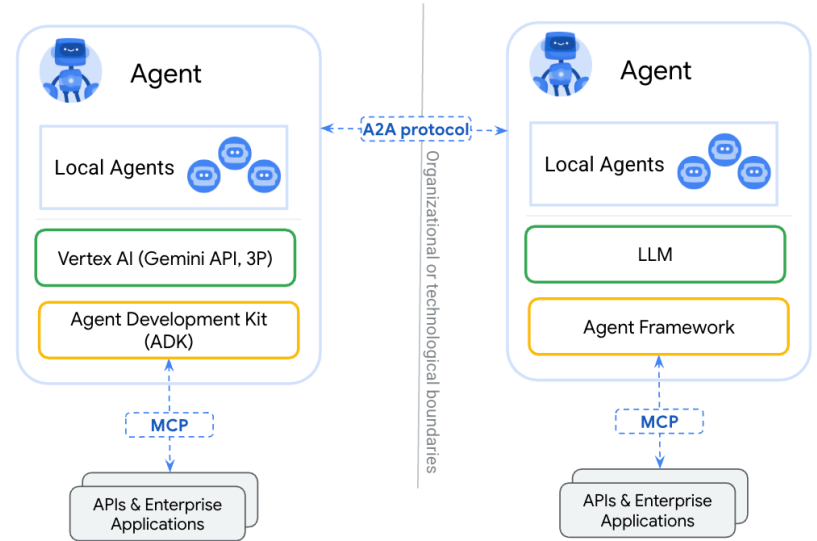
Protocol semantics / interoperability

Modality & capability negotiation



Selected Current Work in the Industry

- **Model Context Protocol (MCP)**
 - Connecting applications to agents for context sharing
 - <https://modelcontextprotocol.io/>
- **Agent2Agent (A2A) Protocol**
 - Interconnecting agents in the Internet
 - <https://github.com/a2aproject/A2A>
- **AGNTCY**
 - Infrastructure stack for agent collaboration
 - <https://agntcy.org/>



Status in the IETF

- Exploration field
- Side meeting on agent communication at IETF-123 in July 2025
- No consensus yet on WG establishment
- First individual Internet Drafts
 - <https://www.ietf.org/id/draft-rosenberg-ai-protocols-00>
 - <https://datatracker.ietf.org/doc/draft-stephan-ai-agent-6g/>



Outline

Impact of AI Crawlers

AI Preferences

Agentic AI

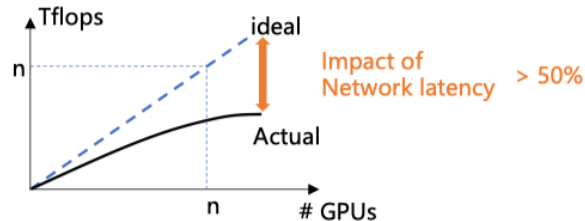
AI Infrastructure



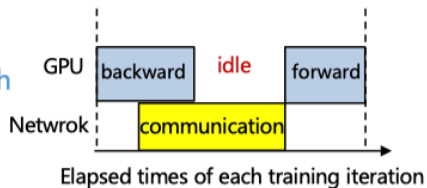
Distributed Machine Learning and the Network

Large Cluster Size \neq Effective Flops

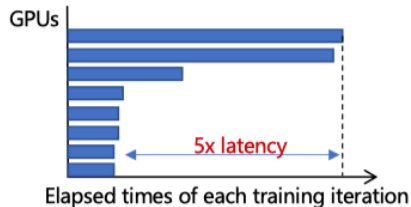
Network Performance constrains effective flops of GPU clusters



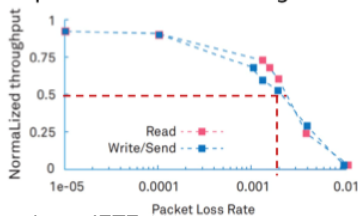
Impacts of Bandwidth (TB magnitude)



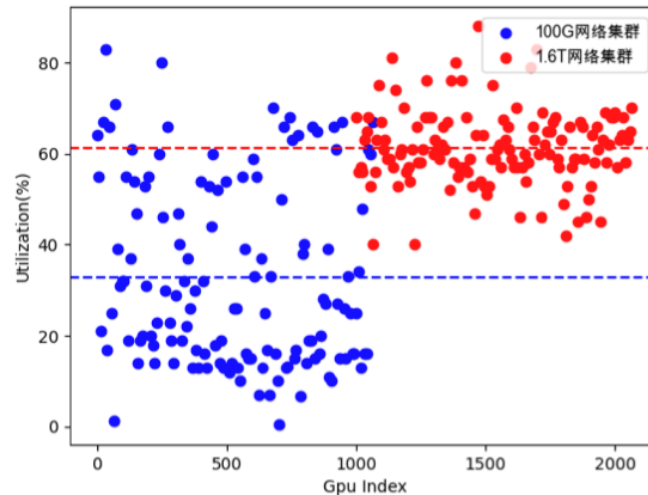
Impacts of Latency (5x)



Impacts of packet loss (0.1% packet loss incur 50%)



GPU usage **32%→61%**



Research Challenges

- **Transport over the Internet and in data centers**
 - Reliability: underlying network lacks communication reliability
 - Application data units instead of packets
 - What is the best response to congestion?
- **In-Network Aggregation**
 - Offload servers and the network
- **Computing in the network at line rate**
 - Machine learning training, data aggregation have max delay expectations
 - Stringent performance requirement

Dirk Kutscher

Collective Communication: Better Network Abstractions for AI

without comments

We have submitted two new Internet Drafts on Collective Communication:

1. Kehan Yao , Xu Shiping , Yizhou Li , Hongyi Huang , Dirk Kutscher; **Collective Communication Optimization: Problem Statement and Use cases**; Internet Draft [draft-yao-tsvwg-cco-problem-statement-and-usecases-00](#); work in progress; October 2023
2. Kehan Yao , Xu Shiping , Yizhou Li , Hongyi Huang , Dirk Kutscher; **Collective Communication Optimization: Requirement and Analysis**; Internet Draft [draft-yao-tsvwg-cco-requirement-and-analysis-00](#); work in progress; October 2023

Collective Communication refers to communication between a group of processes in distributed computing contexts, for example involving interaction types such as broadcast, reduce, all-reduce. This data-oriented communication model is employed by distributed machine learning and other data processing systems, such as stream processing. Current Internet network and transport protocols (and corresponding transport layer security) make it difficult to support these interactions in the network, e.g., for aggregating data on topologically optimal nodes for performance enhancements. These two drafts discuss use cases, problems, and initial ideas for requirements for future system and protocol design for Collective Communication. They will be discussed at [IETF-118](#).

Related

ICNRG @ IETF-118
2023-10-30
In "Events"

Towards a Unified Transport
Protocol for In-Network
Computing in Support of RPC-
based Applications
2024-01-25
In "IETF"

Privacy, Performance,
Protocols: ICN Researchers
meet in Prague
2015-07-23
In "Posts"



Summary

Impact of AI Crawlers

AI Preferences

Agentic AI

AI Infrastructure

Trends

- AI Crawlers causing traffic increase and server load
- Websites try to control traffic and crawler behavior

- Framework for specifying and transmitting website preferences
- More powerful than robots.txt

- Increasing trend of agent-to-agent communication
- First industry consortia and protocol specifications

- Challenges in making distributed machine learning scale well
- In data centers and over the Internet
- Development of protocols and infrastructure support

IETF/IRTF

Research in IRTF

Standards Development in IETF

Exploration in IETF

Exploration in IETF and IRTF





Dirk KUTSCHER
德克·库彻

Thank you!



香港科技大学(广州)
THE HONG KONG
UNIVERSITY OF SCIENCE AND
TECHNOLOGY (GUANGZHOU)

dirk-kutscher.info
dku@hkust-gz.edu.cn