Web Bot Authentication Use Cases

Chris Needham, Neil Craig

IETF Web Both Auth BoF | 21 July 2025

BBC

Current situation

- BBC publishes 80,000 news stories per year, history going back to early 1990s
- 100s of millions – billions of web pages served per day (total)
- About 23% of requests to www.bbc.com are bots (despite robots.txt “blocking”)
  - Hard to assess scale of bot usage currently due to inaccuracy of identification
- 10 – 100 million pages served to bots every day
- Some bots overwhelm smaller-scale origins (high number of requests per second)
- Not all bots obey robots.txt

BBC

Content usage

- Some bots are reducing our traffic by presenting our content directly, low click-through rates
  - See Cloudflare blog post https://blog.cloudflare.com/ai-search-crawl-refer-ratio-on-radar/
- We run B2B & B2C syndication. Uncontrolled scraping undermines our ability to license content
- Need to be able to firmly enforce contract duration (including hard blocking if necessary)
- Better access control means potential route to new customer contacts

BBC

Generative AI and misinformation

BBC research* found AI assistants misrepresent its news output:

- 51% of all AI answers to questions about the news were judged to have significant issues of some form
- 19% of AI answers which cited BBC content introduced factual errors – incorrect factual statements, numbers and dates
- 13% of the quotes sourced from BBC articles were either altered from the original source or not present in the article cited

* https://www.bbc.co.uk/aboutthebbc/documents/bbc-research-into-ai-assistants.pdf

BBC

Bot identification

Difficult to accurately identify bots, lots of manual work

- User-Agent is spoofable, != robots.txt token
- Look for documentation for what the bot actually does
  - What's their intent? We want to distinguish purposes (AI training, agentic queries, etc)
- Few bot authors publish source network info
  - Most bots run on hyperscalers
- If we need to hard block:
  - Identify the User-Agent, traffic source IP range(s) / ASN(s) etc.
- Keep up with changes in bot identity, purposes, and retirements over time

BBC

Thank you

BBC