Technische Universität Braunschweig

Mehrstufiges Vorgehen gegen Überlastung durch KI-Bots an der UB der TU Braunschweig

Michael Kotzyba, Jannis Ohms & Robert Strötgen

Agenda

KEY TO DIGITAL OBJECTS

- Problemstellung
- Erkennung von Bots
- Maßnahmen
  - 1. Geoblocking
  - 2. Rate Limiting
  - 3. Proof of Work Firewall

Problemstellung

KITODO
KEY TO DIGITAL OBJECTS

- KI Anbieter verwenden sogenannte Bots zum sammeln großer Textmengen aus verfügbaren Quellen, z. B. Repositorien
- Die Flut der Anfragen überlasten die angefragten Systeme
- Wie kann die Überbeanspruchung durch Bots verhindert werden?

Erkennung von Bots

KEY TO DIGITAL OBJECTS

- Auswerten des Surfverhaltens
  - Anzahl an Suchen pro IP
  - Verhältnis von Suchen zu Detail Ansichten

- IP Adressen

- User-Agent
  - Wird in der HTTP Anfrage versendet; Inhalt wird durch den Bot kontrolliert; nicht immer korrekt

Maßnahme/Schritt 1

KITODO
KEY TO DIGITAL OBJECTS

→Geoblocking←

- Anfragen durch Bots kommen Großteiles aus dem Ausland z. B. China, USA, Vietnam und teils Russland
- IP Adressen können über eine Datenbank ihrem Ursprungsland zugeordnet werden und anschließend abgewiesen werden
- Apache Module wie https://github.com/maxmind/mod_maxminddb ermöglichen es den Prozess zu automatisieren

Maßnahme/Schritt 2

KITODO
KEY TO DIGITAL OBJECTS

→Rate Limiting←

- Die Anzahl der HTTP Anfragen pro IP Adresse limitieren
  - Die Wahl eines passenden Limits ist kompliziert und hängt von der Anzahl der Elemente der Website ab
  - Hilft nicht, wenn die Bot Anfragen über verschiedene IPs verteilt werden
  - Problematisch für Proxys oder NAT

- Wir nutzen: https://wiki.ubuntuusers.de/Archiv/Apache/mod_evasive/

Maßnahme/Schritt 3

KEY TO DIGITAL OBJECTS

→Proof of Work Firewall←

- Eine Proof-of-Work-Firewall verlangt von jedem, der eine Anfrage an einen Server oder Dienst stellt (z.B. per HTTP-Anfrage), dass er eine kleine Rechenaufgabe löst, bevor seine Anfrage akzeptiert wird.
  - Die Berechnungen werden im Browser durchgeführt
  - Der erhöhte Aufwand macht den Dienst für Bots unattraktiv
  - Die Komplexität und Häufigkeit der Berechnung können eingestellt werden

- Wir nutzen: https://github.com/TecharoHQ/anubis
  - Alternative: https://github.com/altcha-org/altcha

Danke für die Aufmerksamkeit

Mehrstufiges Vorgehen gegen Überlastung durch KI-Bots an der UB der TU Braunschweig

Michael Kotzyba, Jannis Ohms & Robert Strötgen