Poorly-coded bots can harm site performance and increase the load on web servers.

**WHAT THEY DO:**

**Competitive intelligence**
- Metasearch
- Mobile apps
- Location tracking
- Fraud
- Portfolio analysis
- Data mashups
- Statistics
- Financial statements

**GOOD BOTS & BAD BOTS**

![Diagram showing Good and Bad Bots with attributes and solutions.](image-url)

### GOOD BOTS
- Highly desired, low aggression
  - Helps users
  - Respects robots.txt
  - Not too many requests
- Solution
  - Offer an API
  - Dedicated data feed

### BAD BOTS
- Undesired, highly aggressive
  - Poor error handling
  - GET and POST floods
  - Thousands of requests/second
- Solution
  - Blocklists
  - Rate controls
  - Tar pits
  - Spider traps
  - Protect login pages

### UGLY BOTS
- Highly desired, high aggression
  - Access to emerging markets
  - Baidu-bots
  - Poor request throttling
- Solution
  - Rate controls with high threshold
  - User-prioritization application

### WEBSITE COPYING
- Low desirability, low aggression
  - Scrape content for reuse
  - Fraud & counterfeiting
  - Headless browsers
  - Scraping-as-a-service
- Solution
  - Java checkers
  - CAPTCHA
  - Rate controls
  - Validation on sensitive pages
  - Validation for suspicious IPs

### 3 ways to identify client reputation for bots, spiders and scrapers
1. Volume of requests
2. Type of scraped content
3. User agent info

Website: www.akamai.com/stateoftheinternet-security

**Statistics based on attack campaigns mitigated by Akamai**

**Performance vs. Load**

- **Performance:**
  - Website Copier
  - Location Scraper
  - Baidu-bot
  - Metasearch scraper
  - Product page scraper
  - Googlebot

- **Load:**
  - Low desirability, low aggression
    - Scrape content for reuse
    - Fraud & counterfeiting
    - Headless browsers
    - Scraping-as-a-service
  - Solution
    - Java checkers
    - CAPTCHA
    - Rate controls
    - Validation on sensitive pages
    - Validation for suspicious IPs

**DESIRABILITY vs. AGGRESSIVENESS**

- **HIGH DESEIRABILITY, LOW AGGRESSIVENESS**
  - Helps users
  - Respects robots.txt
  - Not too many requests
- **LOW DESEIRABILITY, HIGH AGGRESSIVENESS**
  - Scrape content for reuse
  - Fraud & counterfeiting
  - Headless browsers
  - Scraping-as-a-service
- **HIGH DESEIRABILITY, HIGH AGGRESSIVENESS**
  - Access to emerging markets
  - Baidu-bots
  - Poor request throttling
- **LOW DESIRABILITY, LOW AGGRESSIVENESS**
  - Scrape content for reuse
  - Fraud & counterfeiting
  - Headless browsers
  - Scraping-as-a-service

**3 ways to identify client reputation for bots, spiders and scrapers**

1. Volume of requests
2. Type of scraped content
3. User agent info

Statistics based on attack campaigns mitigated by Akamai.