IP Agnostic Bot Detection

Michael Groskop
Director, Web Application Security
OWASP January 2017
About me

Director of Web Application Security @ Radware

Over 12 years in the Security Space:
• Web Applications Security
• Authentication & SSO
• Cloud solutions
• Database Security

michaelgro@radware.com

https://www.linkedin.com/in/michaelgroskop
AGENDA

- Scoping the Bot Problem
- Bots Behavioral Attributes
- IP Agnostic Bot Detection
Bots Generate \(\sim \frac{1}{2}\) of the Internet Traffic
~30% of the Web traffic is generated by Bad Bots
### DATA BREACHES

#### Breach Subtotal

<table>
<thead>
<tr>
<th>Breach Type</th>
<th>HACK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organization Type</td>
<td>BSF, BSO, BSR, EDU, GOV, MED, NGO</td>
</tr>
<tr>
<td>Company or Organization</td>
<td>all</td>
</tr>
<tr>
<td>Breaches made public fitting this criteria</td>
<td>1,063</td>
</tr>
<tr>
<td>Records total</td>
<td>330,240,018</td>
</tr>
</tbody>
</table>

#### US population projected to Jan 2017 is **325,400,000**

**Sensitive data** records breached by **hacking** in the **United States 2011 - 2016**
Scraping services are “Just a Google Away”

Data as a Service
Need web data? We can extract it for you, hassle-free.

Get data and images from any website
In a “reverse” Turing test, a computer is to determine whether it is interacting with a human or another computer.
Google reCAPTCHA?

Google reCAPTCHA Cracked in New Automated Attack
Facebook's CAPTCHA system too, over 70% accuracy achieved

RISK ASSESSMENT --

How a trio of hackers brought Google's reCAPTCHA to its knees
Hackers exploit weaknesses in Google's bot-detection system with 99% accuracy.

$0.5 for 1000 CAPTCHA

2Captcha

Risks devised a reCaptcha breaking system effective against Google and Facebook
~25% of the Web traffic is generated by **Good Bots**
Travel Industry Good Bots ?!
Protecting the Most valuable Asset - Data

• Competitors’ bots are extracting your data:
  – Price comparison to beat your prices
  – Content Theft & Data Aggregation
  – Faux buyers: 
    continuously creating but never completing reservations
AGENDA

- Scoping the Bot Problem
- Bots Behavioral Attributes
- IP Agnostic Bot Detection
Single Request vs. Continuous Attacks

**Single Request Attack**
- SQL Injection
- XSS
- CSRF
- ...

**Continuous Attack**
- Application DDoS
- Password Cracking / Brute Force
- Site Scraping / Data Harvesting
- Account Lockdown
- ...

radware
Every second counts
Commonly Used Frameworks

- SlimerJS
- CasperJS
- PhantomJS
Single Source with multiple IPs

The Problem:
- Single Attack source
- Attacker **dynamically changes** its IP
- DHCP reset, Anonymous proxies etc.
AGENDA

Scoping the Bot Problem
Bots Behavioral Attributes
• IP Agnostic Bot Detection
Solution Requirements

- IP Agnostic
- Unique
- Cross Platform
- Correlation
Device Fingerprint

- **Identify a browser/bot** through info collection.
- Dozens of browser attributes can be collected on the client side.
- JavaScript allows collecting detailed browser info.
- The power of the fingerprint is in the **consolidated information**.
How distinct does a fingerprint need to be?

- The current estimated world population: 7,477,780,179.

- How many bits of information are required to uniquely identify an individual from the entire population?

\[ \log_2 (7,477,780,179) < 33 \]
Entropy of Browser Fingerprint

To differentiate between 1,000,000 unique users, who access the secured environment requires 20 bits of information:

$$\log_2 (1,000,000) < 20$$
JavaScript Variables

**Navigator variable**
- UserAgent
- App Name
- App Code Name
- App Version
- Build ID
- Platform
- CPU Class
- OS CPU
- Product
- Product Sub
- Vendor
- Vendor Sub
- Language
- User Language
- Browser Language
- System Language

**Screen variable**
- Screen W x H
- Available W x H
- Color Depth
- Pixel Depth
- Device DPI (X, Y)
- Logical DPI (X, Y)
- Update Interval
- System DPI (X, Y)
Other Fingerprinting Approaches

HTTP Headers
- User Agent
- Accept
- Accept-Language
- Accept-Charset
- Accept-Encoding
- X-FORWARDED-FOR
- TRUE-CLIENT-IP
- Via
- DNT (Do not track)

TCP Packet Parameters
- Initial packet size
- IP Initial TTL
- TCP Window size
- Window scaling value
- Max segment size
- TCP Options
- IP flags
- IP Type of service
- IP Total Length
- "don't fragment" flag
- "sackOK" flag
- "nop" flag
for (plugin of navigator.plugins) { console.log(plugin.name); }
Mobile App Support

- **Native** vs. **Hybrid**

- **No** vs. **Yes**
Summary
What have we talked about?

- Scoping the Bot Problem
- Bots Behavioral Attributes
- IP Agnostic Bot Detection
- Device Fingerprint
michaelgro@radware.com
https://il.linkedin.com/in/michaelgroskop