

Evolution of Web Bots and How They Are Detected

Robert Beane

Division of Science and Mathematics
University of Minnesota Morris
Morris, Minnesota, USA

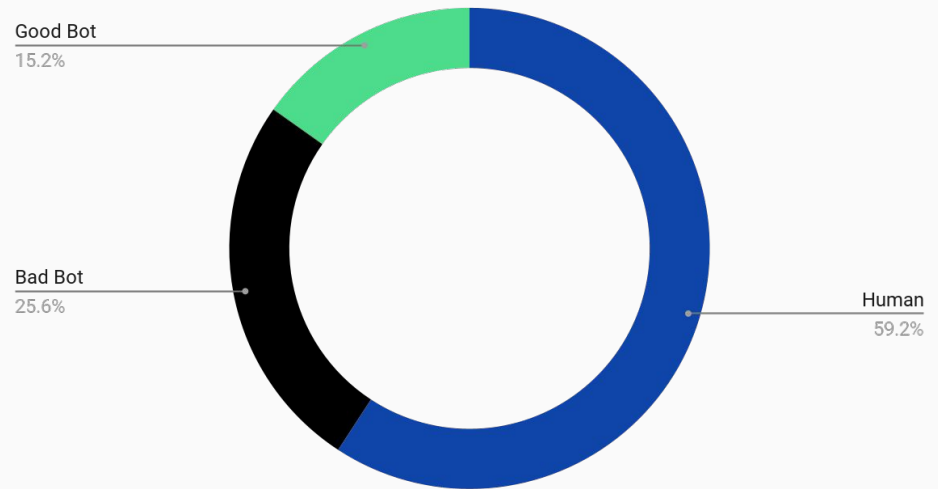
April 2022



Introduction

- What are “Web Bots”
- 40.8% of internet traffic is bots
- Good vs Bad
- Simple vs Advanced

Bad Bot v Good Bot v Human Traffic 2020



Outline

- **Background Information**
 - What Problems Can Bots Cause?
 - How Have Bots Evolved?
 - CAPTCHAs
- **Detection Techniques**
 - Using Web Logs to Detect Bots
 - Using Mouse Behavior to Detect Bots
- **Conclusion**

Problems Bots Cause

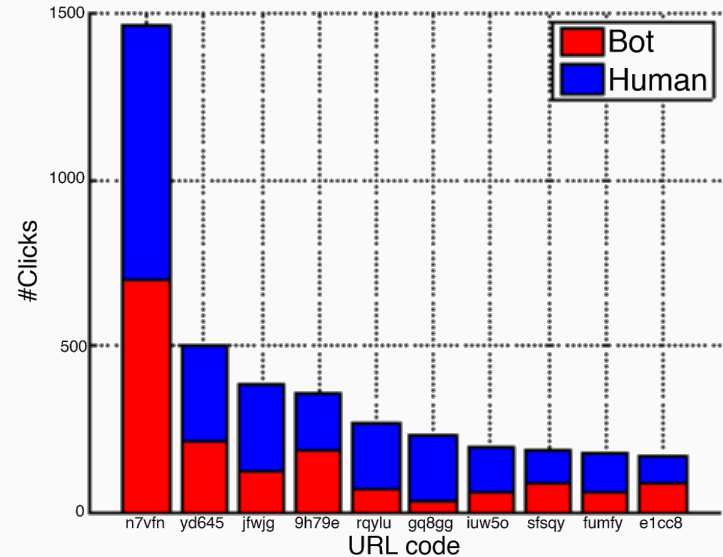
- Not all bots are bad
- 10% more bad bots than good (Imperva)
- Problems affect many areas of the internet

Examples of Problems Caused by Bots

- Vulnerability scanning
- Spamming
- Denial of Service attacks (DOS)
- Price scalping

Bot Problems and Social Media

- Important
 - Could promote artificial news
- Two year Twitter data logging
 - 44% clicks were bots
 - 4% were recurring bots



Source: Gilani (2017)

Social Media Response

- Twitter's new automated account label
 - Given to accounts utilizing Twitter API
- Spam/scam bots still plague social media
- May never know how companies detect bots
 - Harder for developers to avoid detection



Bot Evolution

- Do more besides simple commands
- More accessible
 - Selenium and python



Evasion Evolution

- Avoid new detection frameworks
 - “Arms race”
- Use new techniques to avoid detection
 - Simulate mouse behavior
 - Speech to text

CAPTCHAs

- CAPTCHA - Completely Automated Public Turing test to tell Computers and Humans Apart
- Became the go to solution

Common CAPTCHAs

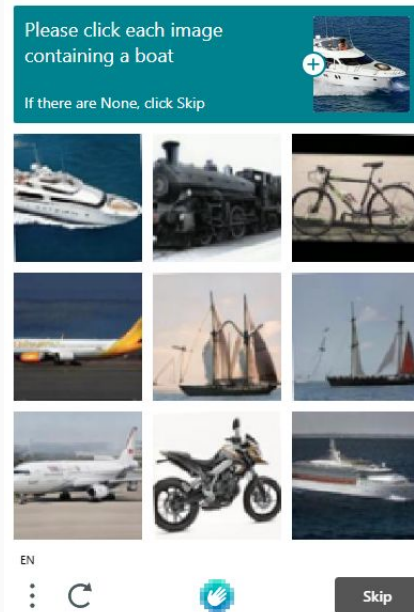
- Mainstream suppliers
 - Google reCAPTCHA
 - hCaptcha
- Different tests
 - Image selection
 - Text input
 - Simple click



reCAPTCHA



hCaptcha



CAPTCHA Disadvantages

Disadvantages

- Interrupts user experience
- Some lack accessibility settings
- Cases where they are bypassed
 - Speech to text generators

Outline

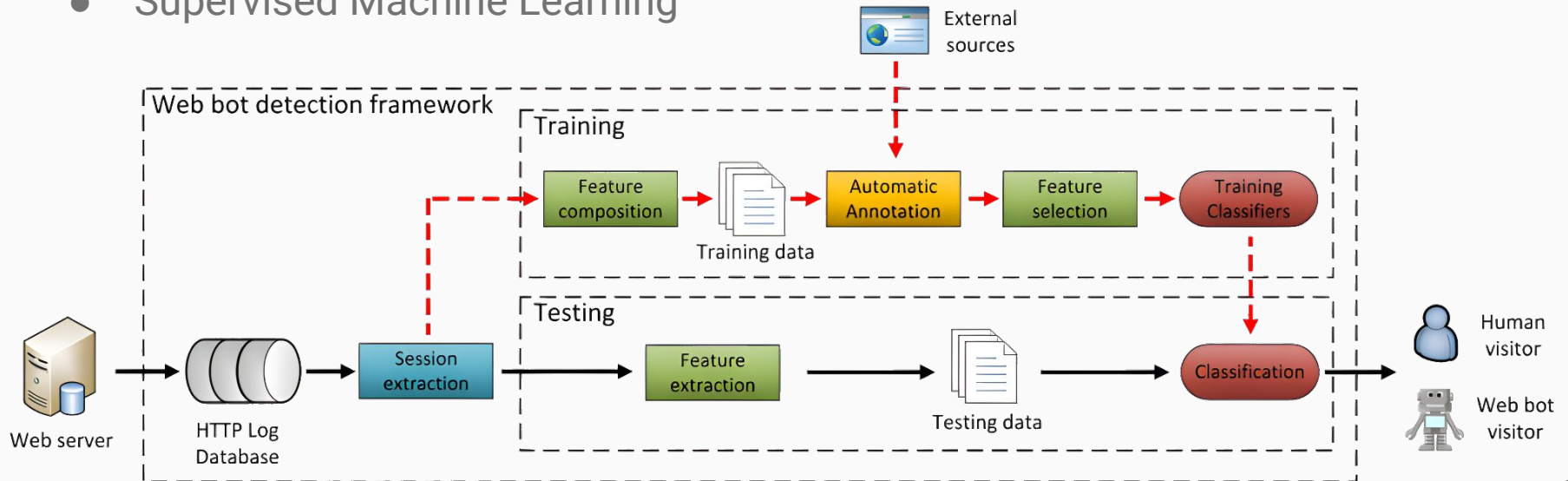
- Background Information
 - What Problems Can Bots Cause?
 - How Have Bots Evolved?
 - CAPTCHAs
- Detection Techniques
 - Using Web Logs to Detect Bots
 - Using Mouse Behavior to Detect Bots
- Conclusion

Why are these Frameworks Necessary?

- Alternative to CAPTCHAs
- Detection works offline
- Addresses advanced bots
 - Evade detection

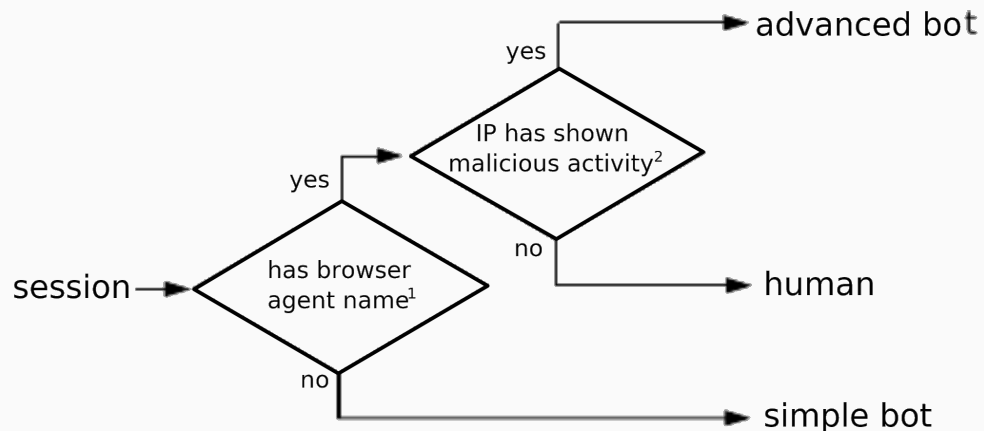
Detection Techniques - Web Logs

- Uses web logs
- Supervised Machine Learning



Session Automatic Annotation

- IP and Browser Agent name used for annotation
 - Browser agent displays system and browser information



[1] useragentstring.com

[2] greynoise.io

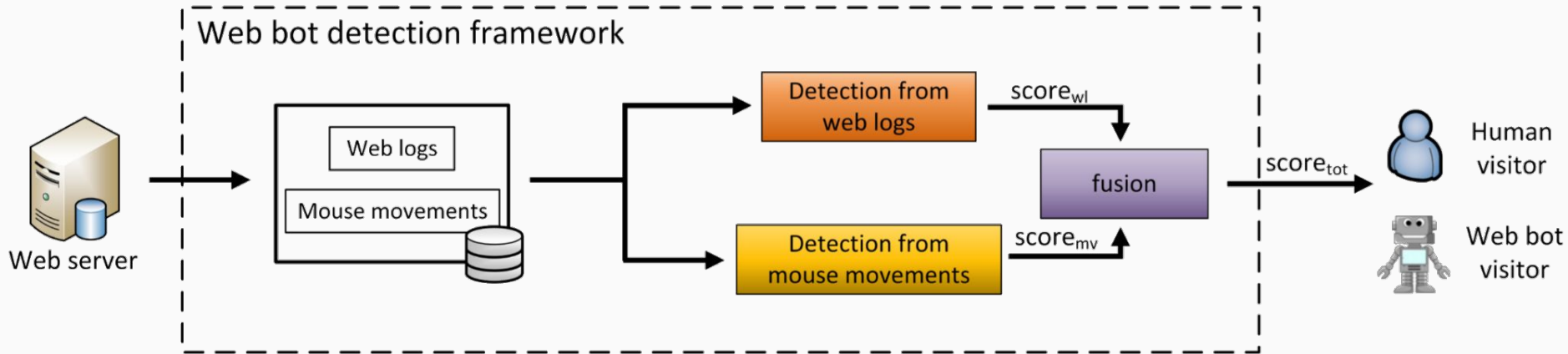
Source: Iliou (2019)

Web Log Framework Results

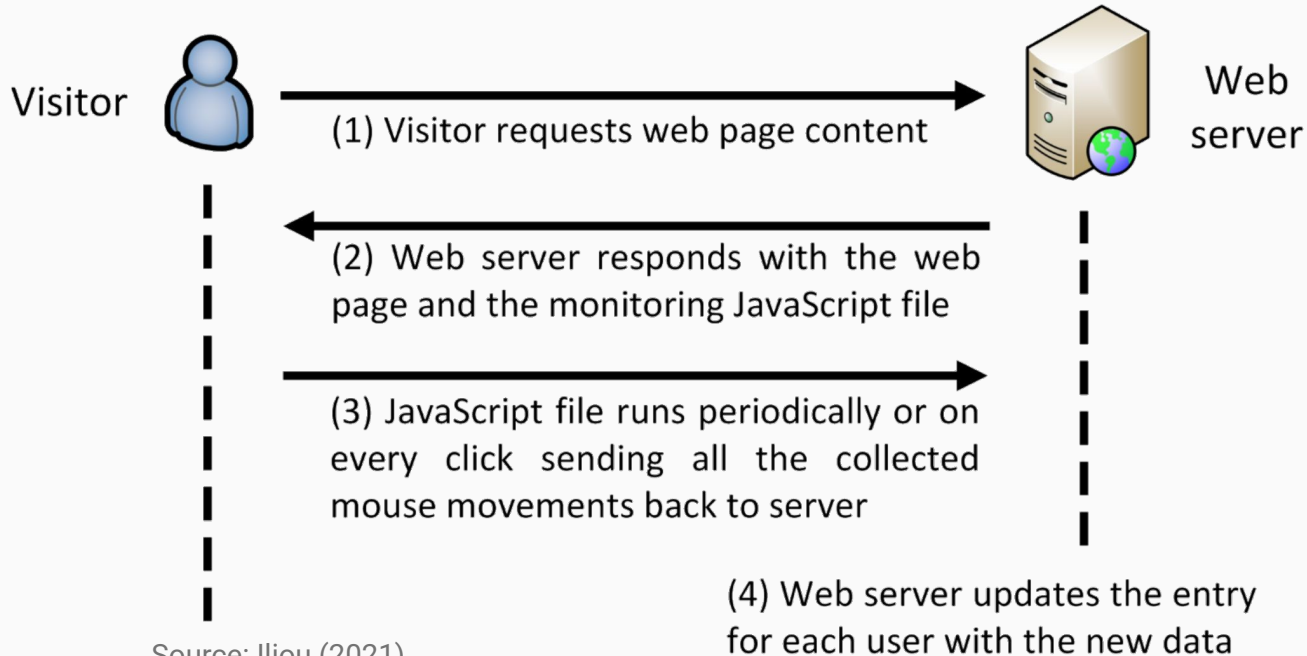
- “Web bot detection problem is a multifaceted one”
- Detecting simple bots easy
- Detecting advanced bots more difficult

Detection Techniques - Mouse Behavior







- Using Mouse Behavior
- Supervised Machine Learning



Mouse Behavior Logging



Logged Mouse Behavior

	Human	Moderate bot	Advanced bot
Characteristics	✓ Sessions made by human visitors	✓ Random hyperlink selection ✓ Direct mouse movements	✓ Heuristic hyperlink selection ✓ Advanced mouse movements
Example image 1			
Example image 2			

Source: Iliou (2021)

Mouse Behavior Bot Determination

- Scores given to each module
 - $Score_{wl}$
 - $Score_{mv}$
- Scores combined to evaluate session as human or bot
 - $Score_{tot}$
- Uses $Score_{mv}$ determination if score is super high or super low

Is Mouse Behavior Useful?

- Yes
- Mouse module more accurate than web log module
- Efficient

Outline

- Background Information
 - What Problems Can Bots Cause?
 - How Have Bots Evolved?
 - CAPTCHAs
- Detection Techniques
 - Using Web Logs to Detect Bots
 - Using Mouse Behavior to Detect Bots
- Conclusion

Conclusion

- Increase of advanced bots becoming a problem
- CAPTCHAs are useful but have downsides
- Web logs and mouse behavior are effective tools
- Mouse behavior shown to be more effective

Questions?

References

- Christos Iliou, Theodoros Kostoulas, Theodora Tsikrika, Vasilis Katos, Stefanos Vrochidis and Ioannis Kompatsiaris. *Detection of Advanced Web Bots by Combining Web Logs with Mouse Behavioural Biometrics*. 2021.
- Christos Iliou, Theodoros Kostoulas, Theodora Tsikrika, Vasilis Katos, Stefanos Vrochidis, and Yiannis Kompatsiaris. *Towards a Framework for Detecting Advanced Web Bots*. 2019.
- Imperva. *Bad Bot Report 2021: The Pandemic of the Internet*. 2021.
- Zafar Gilani, Reza Farahbakhsh and Jon Crowcroft. *Do Bots Impact Twitter Activity?* 2017.