Poster: Etrigan: Large Scale Adversarial Analysis of Elusive Bots

Hari Venugopalan
hvenugopalan@ucdavis.edu
UC Davis

Shaoor Munir
smunir@ucdavis.edu
UC Davis

Samuel T. King
kingst@ucdavis.edu
UC Davis

Zubair Shafiq
zubair@ucdavis.edu
UC Davis

The proportion of malicious bots on the web is on the rise [7]. As of 2023, bots constitute around 47.5% of online traffic [10], with 63.6% of those being malicious bots. Fraudsters employ such bots to launch a multitude of cyber-attacks including financial fraud [6], click fraud [5], account takeover [13], unauthorized scraping [3] etc., resulting in billions of dollars of loss to the industry.

To counter this threat, commercial anti-bot services detect and block requests generated by bots. Researchers have shown that such services employ browser fingerprinting to detect bots without disrupting the user experience of legitmate users [15], [2]. These fingerprints capture attributes of the device/browser of the user sending the request and can be used to differentiate between real users and bots.

Despite the presence of such services, industry reports a how a rise in the number of advanced bots that can alter their identities to evade detection [10]. One way in which bots can alter their identities is to alter their browser attributes to change their browser fingerprints [9], [11]. We refer to the fingerprints of bots that are able to evade detection as *elusive* fingerprints. It is imperative to characterize elusive fingerprints and understand the extent to which bots can modify their fingerprints in order to devise countermeasures to bolster bot detection.

Researchers have studied bot fingerprints by employing their own bots [1], [15] or by focusing on bots that naturally discover their honey sites [12]. Thus, their work does not capture the elusive fingerprints employed by bots seeking to evade detection in the wild. Recent research [17] by Wu et al., performs a large scale characterization of the differences between human and bot fingerprints in the wild. However, they use the decisions from their bot detection system to distinguish between human and bot fingerprints. Thus, their research cannot identify the fingerprints employed by bots that can evade their system.

In this paper, we perform the first large-scale measurement of elusive fingerprints. Concretely, we purchase traffic from different sellers and drive them to different versions of our honey site. We designed these versions to only differ in terms of the presence of random strings in their URL. These random strings ensure that the requests recorded at each version originated from their corresponding sellers and no one else. These sellers advertise their traffic as being realistic and natural, indicating that they likely employ elusive fingerprints to ensure they don't get detected as bots. We also integrate two commercial bot detection services (DataDome and BotD) on all versions of our site that predict if each request originated from a bot or a human. Since all requests originate from bots, true positives in their predictions correspond to bot requests that were detected while false negatives correspond to requests that evaded detection. Our honey sites, thus, provide strong ground truth to analyze bots with elusive fingerprints.

We received over 500,000 requests from 20 different sellers over a period of 6 months. DataDome had an accuracy of 55% and BotD had an accuracy of 44% on these requests. We identified browser attributes that aid in evasion by analyzing differences in the distribution of browser attributes across requests that were detected as bots against those that evaded detection. We used results from this analysis to alter browser attributes of our own bots to have them evade detection.

To further push the boundary of elusive fingerprints, we adopted techniques from adversarial machine learning [8], [16], [4] to automatically alter the fingerprints of requests in our dataset that were previously detected to have them evade detection. These techniques have predominantly been applied in domains such as images where there are no restrictions on adversarial perturbations since inputs can take any real value. Browser attributes include a combination of real, discrete, categorical and binary values. Thus, in order to remain practical, adversarial perturbations have to respect these domain specific constraints when applied to browser fingerprints. Taking inspiration from prior research [18], we enforce these constraints on perturbations produced by the Fast Gradient Sign Method [8]. We practically demonstrate these perturbations being successful in converting previously detected fingerprints into elusive fingerprints by evading BotD and DataDome.

We attribute the threat of elusive fingerprints (and their adversarial perturbations) to the unrestricted freedom that bots have towards altering browser attributes. However, we crucially observe that the freedom to alter any attribute without restrictions introduces avenues for inconsistent combination of attributes that cannot exist in the real world. We hypothesize that it is difficult for bots to guarantee that the fingerprints that emerge as a result of their alterations are consistent and will exist in the real world. In contrast, fingerprints of benign users can never be inconsistent by virtue of not having any alterations made to their browser attributes. We

thus believe that inconsistencies provide an interesting avenue towards detecting elusive fingerprints. While prior research has proposed the use of inconsistencies for bot detection [14], [15], they rely on anecdotes to define inconsistencies. Coming up with such anecdotes is a manual process which is time consuming. We propose a semi-automated, data-driven approach to discovering inconsistencies. Our approach uses our dataset of bot requests to identify combinations of features that occur rarely to propose them as potential inconsistencies. We then manually vet these combinations to ensure that we do not consider a rare occurrence in the real-world as an inconsistency. Our inconsistency analysis revealed inconsistent combination of features in the elusive fingerprints from the wild as well as our adversarially generated fingerprints.

Our contributions include:

- The first large-scale measurement and characterization of "elusive fingerprints" - the fingerprints of bots that are able to evade commercial bot detection services.
- The development of techniques to automatically alter bot fingerprints that help with evasion to push the boundary of elusive fingerprints.
- A novel, semi-automated approach to discovering inconsistencies in browser fingerprints as a means of detecting evasive bot behavior.

REFERENCES

[1] Babak Amin Azad, Oleksii Starov, Pierre Laperdrix, and Nick Nikiforakis. Web Runner 2049: Evaluating Third-Party Anti-bot Services. In *DIMVA 2020 - 17th Conference on Detection of Intrusions and Malware & Vulnerability Assessment*, Lisboa / Virtual, Portugal, June 2020.

[2] Dylan Cutler Asuman Senol, Alisha Ukani and Igor Bilogrevic. The double edged sword: Identifying authentication pages and their fingerprinting behavior. 2024.

[3] Elisa Chiapponi, Marc Dacier, Olivier Thonnard, Mohamed Fangar, Mattias Mattsson, and Vincent Rigal. An industrial perspective on web scraping characteristics and open issues. In *2022 52nd Annual IEEE/IFIP International Conference on Dependable Systems and Networks - Supplemental Volume (DSN-S)*, pages 5–8, 2022.

[4] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A. Bharath. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 35(1):53–65, 2018.

[5] Vacha Dave, Saikat Guha, and Yin Zhang. Viceroi: Catching click-spam in search ad networks. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer Communications Security*, CCS '13, page 765–776, New York, NY, USA, 2013. Association for Computing Machinery.

[6] Zainul Abi Din, Hari Venugopalan, Jaime Park, Andy Li, Weisu Yin, HaoHui Mai, Yong Jae Lee, Steven Liu, and Samuel T. King. Boxer: Preventing fraud by scanning credit cards. In *29th USENIX Security Symposium (USENIX Security 20)*, pages 1571–1588. USENIX Association, August 2020.

[7] Erez Hasson. Evasive Bots Drive Online Fraud. https://www.imperva.com/blog/evasive-bots-drive-online-fraud-2022-imperva-bad-bot-report/.

[8] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples, 2015.

[9] Daniel Goßen, Hugo Jonker, Stefan Karsch, Benjamin Krumnow, and David Roefs. Hlisa: towards a more reliable measurement tool. In *Proceedings of the 21st ACM Internet Measurement Conference*, IMC '21, page 380–389, New York, NY, USA, 2021. Association for Computing Machinery.

[10] imperva.com. 2023 Imperva Bad Bot Report. https://www.imperva.com/resources/resource-library/reports/2023-imperva-bad-bot-report/.

[11] Jordan Jueckstock, Shaown Sarker, Peter Snyder, Aidan Beggs, Panagiotis Papadopoulos, Matteo Varvello, Benjamin Livshits, and Alexandros Kapravelos. Towards realistic and reproducibleweb crawl measurements. In *Proceedings of the Web Conference 2021*, WWW '21, page 80–91, New York, NY, USA, 2021. Association for Computing Machinery.

[12] Xigao Li, Babak Amin Azad, Amir Rahmati, and Nick Nikiforakis. Good bot, bad bot: Characterizing automated browsing activity. In *2021 IEEE Symposium on Security and Privacy (SP)*, pages 1589–1605, 2021.

[13] Minh Hieu Nguyen Ba, Jacob Bennett, Michael Gallagher, and Suman Bhunia. A case study of credential stuffing attack: Canva data breach. In *2021 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 735–740, 2021.

[14] Antoine Vastel, Pierre Laperdrix, Walter Rudametkin, and Romain Rouvoy. Fp-stalker: Tracking browser fingerprint evolutions. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 728–741, 2018.

[15] Antoine Vastel, Walter Rudametkin, Romain Rouvoy, and Xavier Blanc. FP-Crawlers: Studying the Resilience of Browser Fingerprinting to Block Crawlers. In Oleksii Starov, Alexandros Kapravelos, and Nick Nikiforakis, editors, *MADWeb'20 - NDSS Workshop on Measurements, Attacks, and Defenses for the Web*, San Diego, United States, February 2020.

[16] Rey Wiyatno and Anqi Xu. Maximal jacobian-based saliency map attack, 2018.

[17] Shujiang Wu, Pengfei Sun, Yao Zhao, and Yinzhi Cao. Him of many faces: Characterizing billion-scale adversarial and benign browser fingerprints on commercial websites. In *30th Annual Network and Distributed System Security Symposium, NDSS 2023, San Diego, California, USA, February 27 - March 3, 2023*. The Internet Society, 2023.

[18] Shitong Zhu, Zhongjie Wang, Xun Chen, Shasha Li, Keyu Man, Umar Iqbal, Zhiyun Qian, Kevin S. Chan, Srikanth V. Krishnamurthy, Zubair Shafiq, Yu Hao, Guoren Li, Zheng Zhang, and Xiaochen Zou. Eluding ml-based adblockers with actionable adversarial examples. In *Proceedings of the 37th Annual Computer Security Applications Conference*, ACSAC '21, page 541–553, New York, NY, USA, 2021. Association for Computing Machinery.

Etrigan: Large Scale Adversarial Analysis of Elusive Bots

Hari Venugopalan, Shaoor Munir, Samuel T King, Zubair Shafiq
University of California, Davis

IEEE S&P

Problem Statement

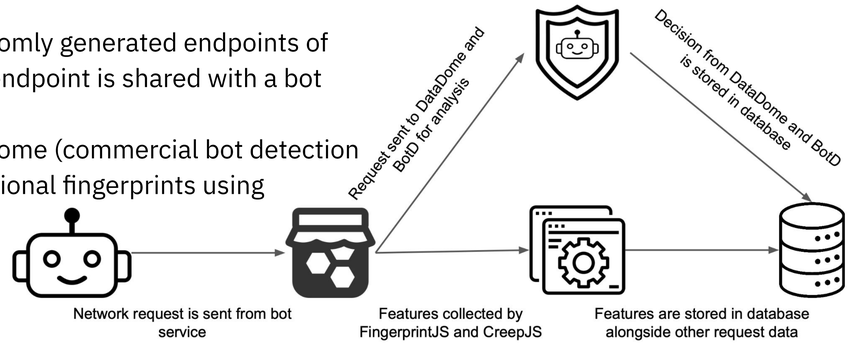
- Proportion of malicious bots on the web is on the rise
  - Bots constitute 47.5% of online traffic
  - 63.6% of bots are malicious
- Commercial bot detection services make use of browser fingerprinting to detect such malicious bots
- Advanced bots evade detection by changing their browser attributes and altering their fingerprints (creating elusive fingerprints)
- It is imperative to understand how evasive these elusive fingerprints can be and create defenses against these advanced bots

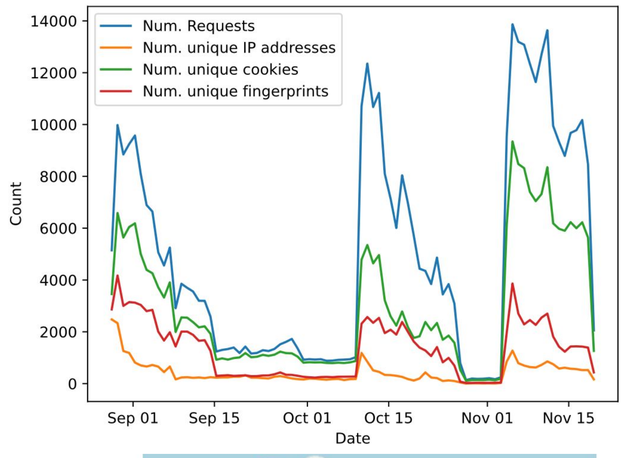
Project Goals

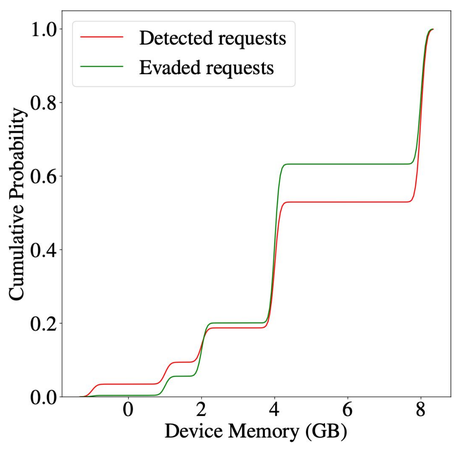
- Perform large-scale measurement of advanced bots in the wild
- Analyze evasive performance of advanced bots against commercial bot detection services
- Analyze inconsistencies in elusive fingerprints as a defence
- Demonstrate it's possible to adversarially create fingerprints to evade bot detection

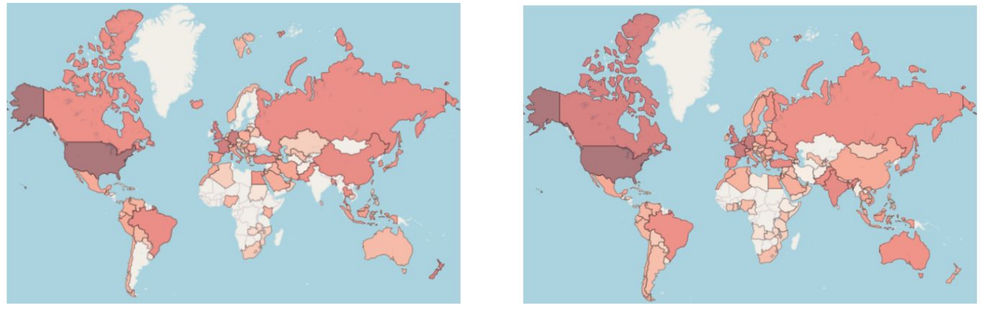
Measurement

- We deploy honey sites on randomly generated endpoints of our master domain, and each endpoint is shared with a bot traffic source
- We integrated BotD and DataDome (commercial bot detection services) and also collect additional fingerprints using FingerprintJS and CreepJS









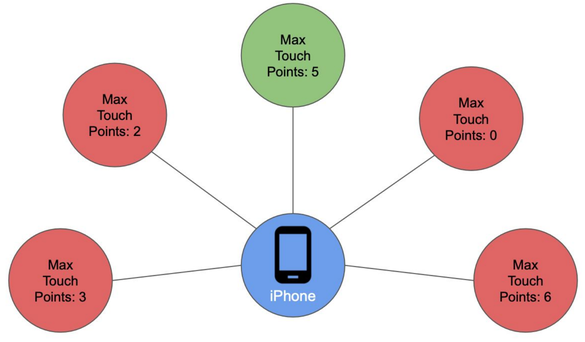
(a) Heatmap of geographical locations inferred using the browser's timezone APIs

(b) Heatmap of geographical locations inferred from IP address

Adversarial Fingerprints

- We adopt techniques from adversarial machine learning to automatically alter fingerprints of network requests
- We prove efficacy of our adversarial model by altering fingerprints of detected bot requests
- Our experiment on BotD shows that it is possible to evade detection 100% of times when using our adversarial model to perturb attributes of detect bots

Inconsistency Analysis



- We perform large scale analysis to show that altering fingerprints result in inconsistencies between different browser attributes
- Changing attributes obtained through one JavaScript API creates inconsistencies with other APIs which relay similar information
- We report more than 50 inconsistencies across different features in our dataset