Building Mitigation for the Next-Generation of Account Takeovers (ATOs) at Cloudflare

Dhruv Kuchhal

February 12, 2024

1 Introduction

Account takeover (ATO) is a form of identity theft in which attackers gain unauthorized access to legitimate user accounts on Relying Parties (RPs) and use them for malicious purposes. In the case of online retail stores as RPs, ATO attacks can cause financial losses, data breaches, reputation damage, and customer dissatisfaction. In 2023, ATO attacks increased by 354% year-over-year, resulting in over $6 billion in total losses [1]. Content delivery network (CDN) providers are intermediaries that deliver web content to users from geographically distributed servers. They are in a unique position to act as a first line of defense against a large variety of ATO attempts, as they can intercept and filter malicious requests before they reach the RP's server. Ultimately, this reduces the attack pressure on RPs, contributing towards a high-performance and cost-efficient defense-in-depth strategy.

Towards that goal, in this proposal we will explore how Cloudflare, as a CDN, can leverage its visibility over the vast internet landscape to provide an effective and efficient defense against the next generation of ATO attacks.

2 Background on Account Takeover (ATO)

Many authentication systems on the web still rely on shared secrets, such as passwords [2], despite the community's efforts to find secure alternatives [3]. These secrets can be stolen via phishing, data breaches, or malware [4, 5] and used for credential stuffing attacks, leading to large-scale ATOs. The community has developed several stopgap solutions to reduce ATO attack pressure at different stages of the authentication flow, such as credential breach alerting [6], risk-backed authentication [7], step-up login challenges [8], and multi-factor authentication (MFA) [9]. FIDO2 is a promising alternative protocol to passwords, as it eliminates shared secrets, but recent work has revealed that FIDO2's security guarantees often don't hold in today's deployments [10]. Apart from credentials, compromised accounts can also stem from vulnerabilities in web applications or misconfigurations in their deployment [11, 12, 13].

The features indicative of an ATO attack can vary depending on the modus operandi (MO) of the attacker. Some ATO attacks use automated bots to spray credentials in bulk, while others use targeted spearphishing, blended with legitimate traffic. These features manifest at different stages of the authentication process, from the network edge (e.g., when a request reaches a CDN like Cloudflare), to the Identity and Risk checkpoints within the RP. In this proposal, we will examine the features of the various ATO MOs that manifest in network traffic, and how Cloudflare can use them to enable a risk-aware environment for its customers.

3 Our Threat Model

ATOs can be sophisticated and complex attacks. For the scope of this proposal, we assume a secure implementation and deployment of the RP and its assets. Specifically, we assume:

(i) Traditional Man-In-The-Middle (MITM) attacks (e.g., SSL Stripping) and cookie hijacking attacks are out of scope, as Cloudflare's deployment would follow best practices (e.g., HSTS), which provides effective defense [14, 15].

(ii) Vulnerabilities in the RP's web application (e.g., predictable session tokens) [16] that could allow an attacker to bypass authentication are out of scope, as Cloudflare's present offerings, such as WAF, Page Shield, and API Shield, should provide defense [17, 18, 19].

(iii) Vulnerabilities in SSO/federated authentication [11, 12], if used, are out of scope, as Cloudflare might not be able to monitor or control the federated RP.

(iv) Weaknesses in account recovery (e.g., downgrade attacks) [20], and out-of-band MFA (e.g., SIM swapping attacks) [21, 22, 23] are out of scope. Such attacks are often targeted, and often require higher visibility/context than Cloudflare would have (e.g., account privileges, recovery methods). Hence, they are better mitigated at the RP's internal checkpoints.

Based on these assumptions, we limit the attack surface area for our consideration to authentication endpoints where *credentials (i.e., username and password)* and/or *FIDO2-based security keys* are used to login. Together they constitute an overwhelming majority of authentication systems on the Web.

4 Cloudflare's Existing Capabilities

Here, we synthesize Cloudflare's capabilities to defend against advanced ATOs, based on publicly available documentation:

(i) **DDoS Protection**: Large-scale bot-based brute force ATOs can be mitigated by DDoS protection. In fact, DDoS mitigation accounts for 52% of all mitigated traffic at Cloudflare [24].

(ii) **WAF**: For password-based authentication, all login attempts for any application protected by Cloudflare are routed through the WAF, where an "on-path" exposed credential check notifies RPs of accounts that can potentially be compromised. This allows RPs to resecure the accounts before ATOs occur [25, 26].

(iii) **Bot Management**: Bot-based attacks which are not identified at either DDoS or WAF are identified with a multi-pronged approach in this module:

- **Super Bot Fight Mode [27]**: Attempts to isolate automated traffic (from verified benign bot and human traffic) using heuristics (e.g., known malicious TLS fingerprints [28, 29, 30]), machine learning (GBDTs based on request attributes and inter-request features) [31, 32, 33], and JS detections (e.g., computing browser fingerprints) [30, 34, 35]. This prevents attacks such as credential stuffing from compromised clients.
- **Open Proxy Managed List [36]:** Cloudflare proactively searches for open proxy endpoints by analysing threat intelligence at its edge. Bots often use proxies to hide their identity and blend in with legitimate traffic. RPs can decrease the risk of malicious login attempts by challenging login attempts from such low reputation IP addresses.
- **Behavioral Analytics [31]:** This is an unsupervised ML approach that models the normal visitor behavior over an extended period of time, based on features such as previous devices, locations, network, activity time etc. – and detects anomalous traffic. Such an approach is harder to evade and allows novel (i.e., previously unseen) malicious behavior to be identified.

Bot traffic can be mitigated by employing challenges such as CAPTCHAs/Turnstile [37], or its alternatives such as CAP [38, 39] or PrivacyPass [40] which offer lesser friction to a legitimate user. However, it is important to note that this remains an arms race, and hence these systems need to continuously evolve to remain effective. For instance, browser fingerprints have been shown to be ineffective [41, 42] or can even be phished to evade detection [43].

As shown in Figure 1, every request is evaluated and adjudicated (i.e., inference) based on all of the components described above, by Cloudflare's Edge Datacenters in real-time. The logs, containing details about each applied detection, used features and flags, are asynchronously (via Kafka

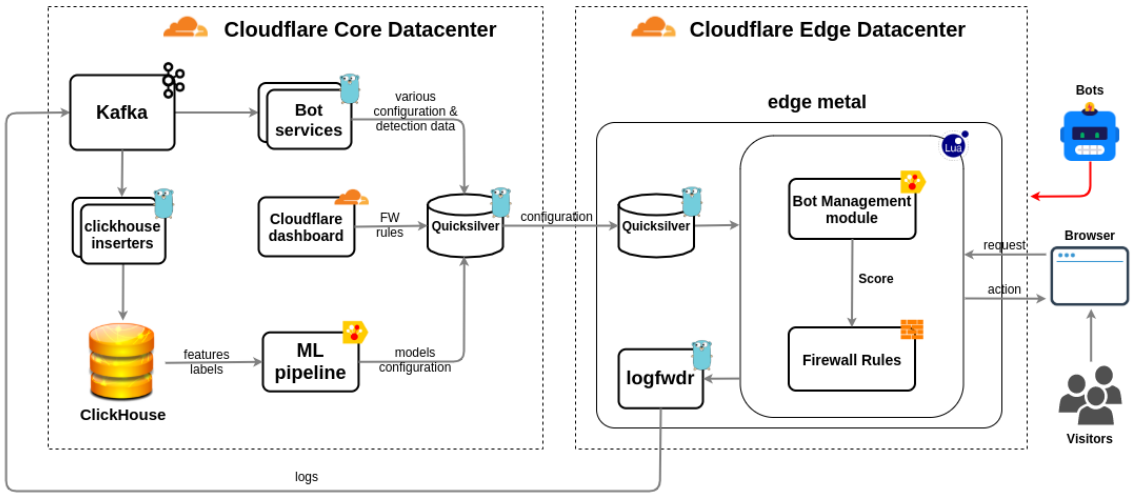

Figure 1: Cloudflare's Bot Management Architecture [31].

brokers) forwarded to ClickHouse for aggregate analysis and improving the ML models (e.g., adding new features, accounting for concept drift).

Mitigations for ATOs for password-based authentication have matured over time. We could continue improving these systems to keep attackers at bay. For instance, we could implement the GoldenHour framework at Cloudflare to proactively detect phishing sites targeting Cloudflare's customers, which will enable RPs to resecure phishing victims' accounts before they are attacked [44].

However, password-based authentication is gradually being replaced by FIDO2, passkeys being its most popular offering. As web authentication systems adopt FIDO2 passkeys [45], attackers and their methods will also evolve to exploit its weaknesses. Recent research has shown that today's FIDO2 deployments are not configured to defend against ATO attacks, especially when the client devices are compromised [10]. Therefore, we propose to study the ATO attempts on the next-generation of authentication, i.e., FIDO2-based authentication, and develop real-time detection and mitigation techniques for them.

5 Building Capability to Mitigate Next-Generation ATOs

Risk-backed authentication is an effective way to bridge the gap between security and usability in password-only and MFA authentication [46]. It leverages JS detection not only based on device attributes we previously discussed, but also signals inherent to password authentication, such as autofill/typing behavior, incorrect password attempts, and mouse movements [7]. With the wide variety of FIDO2 authenticators available on the Web today, ranging from secure hardware-based security keys to insecure browser extension based passkey wallets [10, 45], it is crucial to collect risk telemetry and build models to proactively detect vulnerable/compromised authenticators which can be abused in bot-based attacks. WebAuthn is the FIDO2 protocol that enables the client to communicate with the RP over HTTP. Figure 2 illustrates the authenticator's attestation information that Cloudflare can extract at an Edge datacenter (from the HTTP payload to the registration endpoint) and convey a risk score to the RP (based on internal threat intelligence) via a cookie in real time [47], during the authenticator's registration. At the Core datacenter, we can analyze authenticator registrations in aggregate to detect behavioral patterns, such as the time taken to complete registration, which can help in identifying bot-based registrations. Risk models can also evaluate logins in real time. For example, an anomalous round-trip communication time for an authentication may suggest that an authenticator was compromised after registration.

Regarding scalability, we expect that some additional operations, such as cryptographic verification of attestation signatures and decoding the various fields in the WebAuthn Attestation object, could introduce latency when computed in real-time. As part of this study, we will measure it and explore

strategies to optimize the process. The accuracy of our approach might differ based on the authenticator's trust anchor (i.e., attestation type) – for example, our risk score for a registration credential containing an Attestation-CA certificate might be much more reliable than one that is self-signed. The scalability and accuracy of this solution (i.e., our models) will also depend on the amount and diversity in risk telemetry we collect for authenticators. Lastly, since this approach only involves non-PII data processing and no data storage, it adheres to privacy laws such as GDPR. By combining this approach with existing defenses, Cloudflare can significantly enhance mitigation against tomorrow's ATOs.

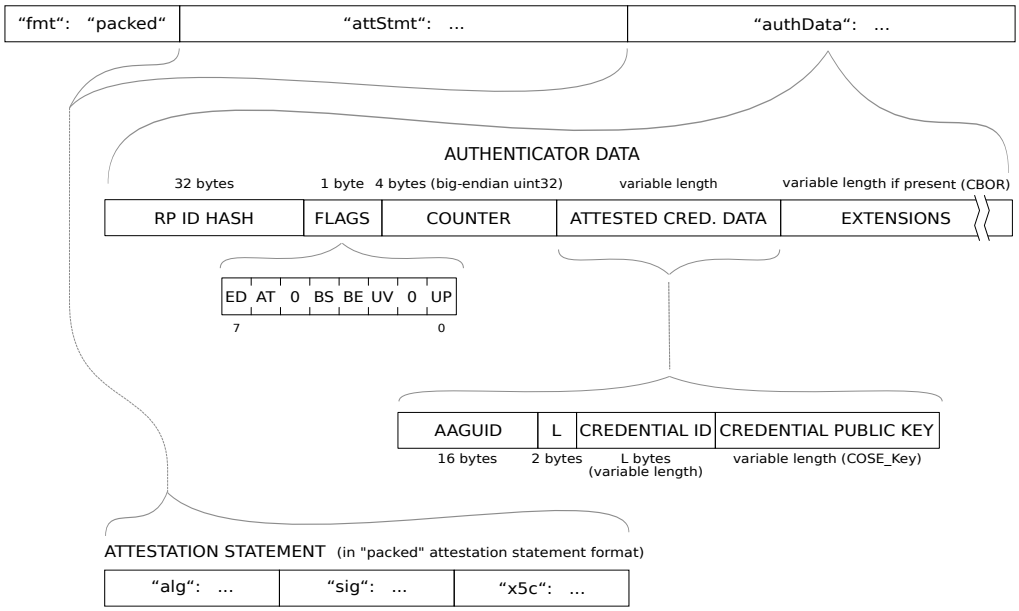

Figure 2: Layout of the Attestation object in a WebAuthn registration credential [48].

References

[1] Sift. Account Takeover Data, Consumer Insights, and Emerging Trends, Q3 2023. https://pages.sift.com/rs/526-PCC-974/images/Sift-2023-Q3-Index-Report_ATO.pdf.

[2] Suood Al Roomi and Frank Li. A Large-Scale Measurement of Website Login Policies. In *USENIX Security Symposium*, 2023.

[3] Joseph Bonneau, Cormac Herley, Paul C Van Oorschot, and Frank Stajano. The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. In *IEEE Symposium on Security and Privacy (S&P)*, 2012.

[4] Kurt Thomas, Frank Li, Ali Zand, Jacob Barrett, Juri Ranieri, Luca Invernizzi, Yarik Markov, Oxana Comanescu, Vijay Eranti, Angelika Moscicki, et al. Data breaches, phishing, or malware? understanding the risks of stolen credentials. In *ACM Conference on Computer and Communications Security (CCS)*, 2017.

[5] Athanasios Avgetidis, Omar Alrawi, Kevin Valakuzhy, Charles Lever, Paul Burbage, Angelos Keromytis, Fabian Monrose, and Manos Antonakakis. Beyond the gates: An empirical analysis of HTTP-managed password stealers and operators. In *USENIX Security Symposium*, 2023.

[6] Kurt Thomas, Jennifer Pullman, Kevin Yeo, Ananth Raghunathan, Patrick Gage Kelley, Luca Invernizzi, Borbala Benko, Tadek Pietraszek, Sarvar Patel, Dan Boneh, et al. Protecting accounts from credential stuffing with password breach alerting. In *USENIX Security Symposium*, 2019.

[7] David Freeman, Sakshi Jain, Markus Dürmuth, Battista Biggio, and Giorgio Giacinto. Who Are You? A Statistical Approach to Measuring User Authenticity. In *Network and Distributed System Security Symposium (NDSS)*, 2016.

[8] Periwinkle Doerfler, Kurt Thomas, Maija Marincenko, Juri Ranieri, Yu Jiang, Angelika Moscicki, and Damon McCoy. Evaluating Login Challenges as a Defense Against Account Takeover. In *The World Wide Web Conference (WWW)*, 2019.

[9] Anthony Gavazzi, Ryan Williams, Engin Kirda, Long Lu, Andre King, Andy Davis, and Tim Leek. A Study of Multi-Factor and Risk-Based Authentication Availability. In *USENIX Security Symposium (USENIX Security)*, 2023.

[10] Dhruv Kuchhal, Muhammad Saad, Adam Oest, and Frank Li. Evaluating the Security Posture of Real-World FIDO2 Deployments. In *ACM Conference on Computer and Communications Security (CCS)*, 2023.

[11] Avinash Sudhodanan and Andrew Paverd. Pre-hijacked accounts: An Empirical Study of Security Failures in User Account Creation on the Web. In *USENIX Security Symposium*, 2022.

[12] Mohammad Ghasemisharif, Amrutha Ramesh, Stephen Checkoway, Chris Kanich, and Jason Polakis. O Single Sign-Off, Where Art Thou? An Empirical Analysis of Single Sign-On Account Hijacking and Session Management on the Web. In *USENIX Security Symposium*, 2018.

[13] Kostas Drakonakis, Sotiris Ioannidis, and Jason Polakis. The cookie hunter: Automated black-box auditing for web authentication and authorization flaws. In *ACM Conference on Computer and Communications Security (CCS)*, 2020.

[14] The Cloudflare Blog. Performing & Preventing SSL Stripping: A Plain-English Primer. https://blog.cloudflare.com/performing-preventing-ssl-stripping-a-plain-english-primer.

[15] The Cloudflare Blog. Inside Cloudflare: Preventing Account Takeovers. https://blog.cloudflare.com/account-compromise-security-overview.

[16] Rohan Pagey, Mohammad Mannan, and Amr Youssef. All Your Shops Are Belong to Us: Security Weaknesses in E-commerce Platforms. In *ACM Web Conference (WWW)*, 2023.

[17] Cloudflare. WAF. https://www.cloudflare.com/application-services/products/waf.

[18] Cloudflare Docs. Page Shield. https://developers.cloudflare.com/page-shield.

[19] Cloudflare Docs. API Shield. https://developers.cloudflare.com/api-shield.

[20] Enis Ulqinaku, Hala Assal, AbdelRahman Abdou, Sonia Chiasson, and Srdjan Capkun. Is Real-time Phishing Eliminated with FIDO? Social Engineering Downgrade Attacks against FIDO Protocols. In *USENIX Security Symposium*, 2021.

[21] Kevin Lee, Benjamin Kaiser, Jonathan Mayer, and Arvind Narayanan. An Empirical Study of Wireless Carrier Authentication for SIM Swaps. In *Symposium on Usable Privacy and Security (SOUPS)*, 2020.

[22] Sabrina Amft, Sandra Höltervennhoff, Nicolas Huaman, Alexander Krause, Lucy Simko, Yasemin Acar, and Sascha Fahl. "We've Disabled MFA for You": An Evaluation of the Security and Usability of Multi-Factor Authentication Recovery Deployments. In *ACM Conference on Computer and Communications Security (CCS)*, 2023.

[23] Christian Peeters, Christopher Patton, Imani NS Munyaka, Daniel Olszewski, Thomas Shrimpton, and Patrick Traynor. SMS OTP Security (SOS) Hardening SMS-Based Two Factor Authentication. In *ACM Asia Conference on Computer and Communications Security (ASIA CCS)*, 2022.

[24] The Cloudflare Blog. The state of application security in 2023. https://blog.cloudflare.com/application-security-2023/.

[25] Cloudflare Docs. Check for exposed credentials. https://developers.cloudflare.com/waf/managed-rules/check-for-exposed-credentials.

[26] The Cloudflare Blog. Account Takeover Protection and WAF mitigations to help stop Global Brute Force Campaigns. https://blog.cloudflare.com/patching-the-internet-against-global-brute-force-campaigns/.

[27] The Cloudflare Blog. Introducing Super Bot Fight Mode. https://blog.cloudflare.com/super-bot-fight-mode.

[28] Cloudflare Docs. JA3 Fingerprint. https://developers.cloudflare.com/bots/concepts/ja3-fingerprint.

[29] Brian Kondracki, Babak Amin Azad, Oleksii Starov, and Nick Nikiforakis. Catching transparent phish: Analyzing and detecting MITM phishing toolkits. In *ACM Conference on Computer and Communications Security (CCS)*, 2021.

[30] Xigao Li, Babak Amin Azad, Amir Rahmati, and Nick Nikiforakis. Good bot, bad bot: Characterizing automated browsing activity. In *IEEE Symposium on Security and Privacy (S&P)*, 2021.

[31] The Cloudflare Blog. Cloudflare Bot Management: machine learning and more. https://blog.cloudflare.com/cloudflare-bot-management-machine-learning-and-more.

[32] The Cloudflare Blog. Evolving our machine learning to stop mobile bots. https://blog.cloudflare.com/machine-learning-mobile-traffic-bots/.

[33] Cloudflare Docs. Machine Learning Models. https://developers.cloudflare.com/bots/reference/machine-learning-models/.

[34] Cloudflare Docs. JavaScript detections. https://developers.cloudflare.com/bots/reference/javascript-detections/.

[35] Cloudflare Docs. Bot detection engines. https://developers.cloudflare.com/bots/plans/biz-and-ent/#bot-detection-engines.

[36] The Cloudflare Blog. Protecting your APIs from abuse and data exfiltration. https://blog.cloudflare.com/protecting-apis-from-abuse-and-data-exfiltration.

[37] Cloudflare Docs. Turnstile. https://developers.cloudflare.com/turnstile/.

[38] The Cloudflare Blog. Humanity wastes about 500 years per day on CAPTCHAs. It's time to end this madness. https://blog.cloudflare.com/introducing-cryptographic-attestation-of-personhood.

[39] Tara Whalen, Thibault Meunier, Mrudula Kodali, Alex Davidson, Marwan Fayed, Armando Faz-Hernández, Watson Ladd, Deepak Maram, Nick Sullivan, Benedikt Christoph Wolters, et al. Let The Right One In: Attestation as a Usable CAPTCHA Alternative. In *Symposium on Usable Privacy and Security (SOUPS)*, 2022.

[40] The Cloudflare Blog. Privacy Pass: upgrading to the latest protocol version. https://blog.cloudflare.com/privacy-pass-standard.

[41] Babak Amin Azad, Oleksii Starov, Pierre Laperdrix, and Nick Nikiforakis. Web Runner 2049: Evaluating Third-Party Anti-bot Services. In *Conference on Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA)*, 2020.

[42] Babak Amin Azad, Oleksii Starov, Pierre Laperdrix, and Nick Nikiforakis. Taming The Shape Shifter: Detecting Anti-fingerprinting Browsers. In *Conference on Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA)*, 2020.

[43] Xu Lin, Panagiotis Ilia, Saumya Solanki, and Jason Polakis. Phish in Sheep's Clothing: Exploring the Authentication Pitfalls of Browser Fingerprinting. In *USENIX Security Symposium*, 2022.

[44] Adam Oest, Penghui Zhang, Brad Wardman, Eric Nunes, Jakub Burgis, Ali Zand, Kurt Thomas, Adam Doupé, and Gail-Joon Ahn. Sunrise to Sunset: Analyzing the End-to-end Life Cycle and Effectiveness of Phishing Attacks at Scale. In *USENIX Security Symposium*, 2020.

[45] Biometrics Research Group. Passkey adoption lags great expectations but opportunities still growing. https://www.biometricupdate.com/202401/passkey-adoption-lags-great-expectations-but-opportunities-still-growing.

[46] Grzergor Milka. Anatomy of Account Takeover. In *Enigma*, 2018.

[47] Cloudflare Docs. Cloudflare Cookies. https://developers.cloudflare.com/fundamentals/reference/policies-compliances/cloudflare-cookies.

[48] W3C. Web Authentication: An API for accessing Public Key Credentials. https://www.w3.org/TR/webauthn-2/.