ZEROFOX INTELLIGENCE

Flash

Data Set Features Billions of Leaked User Credentials

F-2025-07-16a

Classification: TLP:CLEAR

Criticality: HIGH

Intelligence Requirements: Data Leak, Personally Identifiable Information

July 16, 2025

Scope Note

ZeroFox Intelligence is derived from a variety of sources, including—but not limited to—curated open-source accesses, vetted social media, proprietary data sources, and direct access to threat actors and groups through covert communication channels. Information relied upon to complete any report cannot always be independently verified. As such, ZeroFox applies rigorous analytic standards and tradecraft in accordance with best practices and includes caveat language and source citations to clearly identify the veracity of our Intelligence reporting and substantiate our assessments and recommendations. All sources used in this particular Intelligence product were identified prior to 9:00 AM (EDT) on July 16, 2025; per cyber hygiene best practices, caution is advised when clicking on any third-party links.

| Flash | Data Set Features Billions of Leaked User Credentials

| Key Findings

- ZeroFox has procured a significant quantity of leaked data, the existence of which was first announced by cybersecurity researchers on June 18, 2025.
- Among approximately 16 billion compromised data points, ZeroFox's preliminary analysis identified at least 2.7 billion lines of URL, login, and password (ULP) data, alluding to at least this many separate victims.
- Although initial media reporting suggested that this data was associated with a singular data breach, it is significantly more likely that it was compiled from various stealer logs.
- As of this writing, ZeroFox is unable to determine the usability of this data or the extent to which it is associated with contemporary accounts that may be vulnerable to compromise.
- Initial analysis suggests that much of the data is dated between January and June 2025, indicating there is a likely chance that a high proportion of the identified data would provide malicious utility to those in possession of it.

Details

ZeroFox has procured a significant quantity of leaked data, the existence of which was first announced by cybersecurity researchers on June 18, 2025. The approximately 14 terabyte raw data set consists of login credentials associated with technology organizations Apple and Google, social media platforms Facebook and Telegram, and many others.

- Among the approximately 16 billion compromised data points, ZeroFox's preliminary analysis identified at least 2.7 billion lines of ULP data, alluding to at least this many separate victims.

Although initial media reporting suggested that this data was associated with a singular data breach, it is significantly more likely that it comprises multiple datasets compiled from various stealer logs. As of this writing, ZeroFox has observed evidence that the RedLine malware family was leveraged to obtain some of the data, though there is a likely chance that other prominent stealers such as StealC, Vidar, or Lumma were also used.

- Numerous data-stealing techniques are leveraged by infostealers, depending on the strain deployed. Form grabbing, keylogging, credential dumping, and screen scraping are all commonly observed in such attacks, which usually seek to uncover and extract information from elements of the targeted endpoint device.

The data originated from a misconfigured server, and the identity of the actor or group that obtained and compiled it is unknown at this time—as is how they planned to leverage it. It is very likely that the actor behind the dataset had intended to sell the data, either raw or parsed, within deep and dark web (DDW) marketplaces such as Russian Market or Exodus or Telegram-based marketplaces.

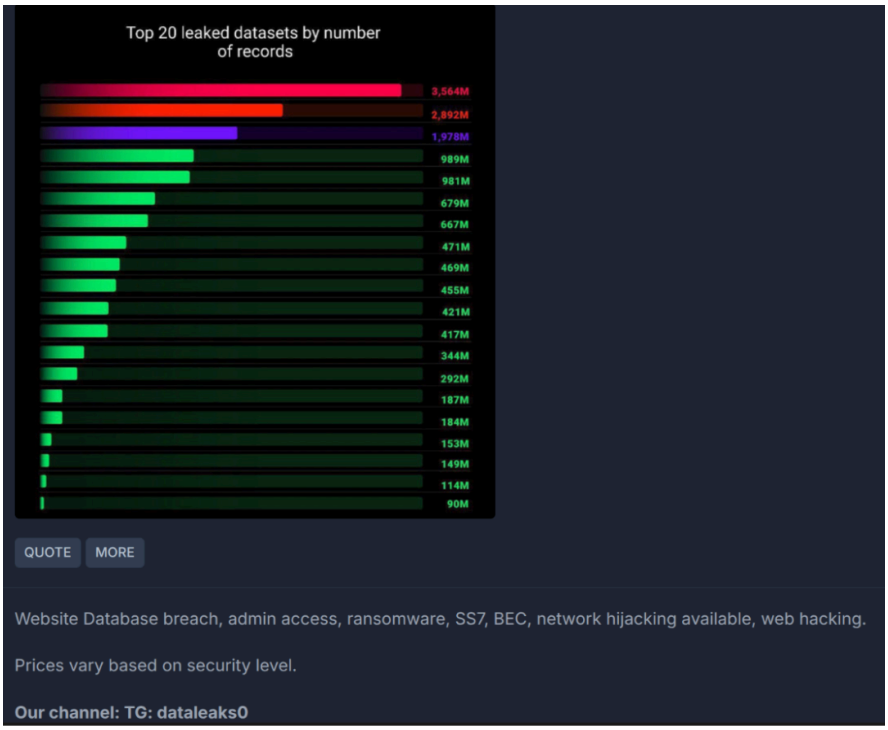
- These marketplaces often categorize and group victim data together by region, sector, or type, enabling buyers to easily obtain the information they seek.

Obtaining personally identifiable information (PII) such as this enables threat actors to conduct subsequent exploitation, such as the deployment of further malware, digital extortion, or targeted social engineering campaigns. Further, other cybercriminals use

such data to establish initial network access that can then be sold, providing a supply to the array of threat actors seeking to conduct malicious cyber activity.

- As of this writing, ZeroFox is unable to determine the usability of this data or the extent to which it is associated with contemporary accounts that may be vulnerable to compromise.
- However, initial analysis suggests that much of the data is dated between January and June 2025, indicating there is a likely chance that a high proportion of the identified data would provide malicious utility to those in possession of it.

ZeroFox has observed several unverified advertisements by low-reputation actors claiming to offer the data for sale, primarily in the Cracked, Leakbase, and NulledBB forums. These posts were all created by low-credibility actors with no meaningful samples offered, as is common within the community of low-credibility actors trying to capitalize on high-visibility compromised data sets.



NulledBB actor claiming to sell dataset, including imagery obtained from security researchers

Source: ZeroFox Collections

What Are Some Common Misconceptions Surrounding Stealer Logs?

"Data contained in stealer logs is new, relevant, valuable, and exploitable."

- In some cases this is true, especially if the data has not been recycled or previously sold. However, many stealer logs for sale within DDW circles often contain old, recycled, or outdated credentials and information that have appeared in numerous previous data breaches. By the time exploitation is attempted, victim users may have changed passwords or closed accounts, rendering the data obsolete. Stealer logs also often contain fake or manipulated information to bolster the file size and attract potential buyers.
- While some logs contain valuable corporate credentials, many are from personal devices with limited value to threat actors. Research indicates that about 1 percent of logs pertain to corporate victims.[1]
- Not all data in stealer logs is immediately exploitable. Factors such as multi-factor authentication (MFA), account lockouts, and changed passwords can limit the utility of stolen credentials.

"Organizational data in a stealer log must have come from a recent data breach."

- Threat actors often compile stealer logs from multiple sources and timeframes. This means that a single compilation may contain information collected over an extended period, not just data from recent infections or breaches. Additionally, threat actors often resell and repackage stealer logs in various combinations. This can lead to fresher data sitting alongside older data, creating a misleading perception of the data's recency.

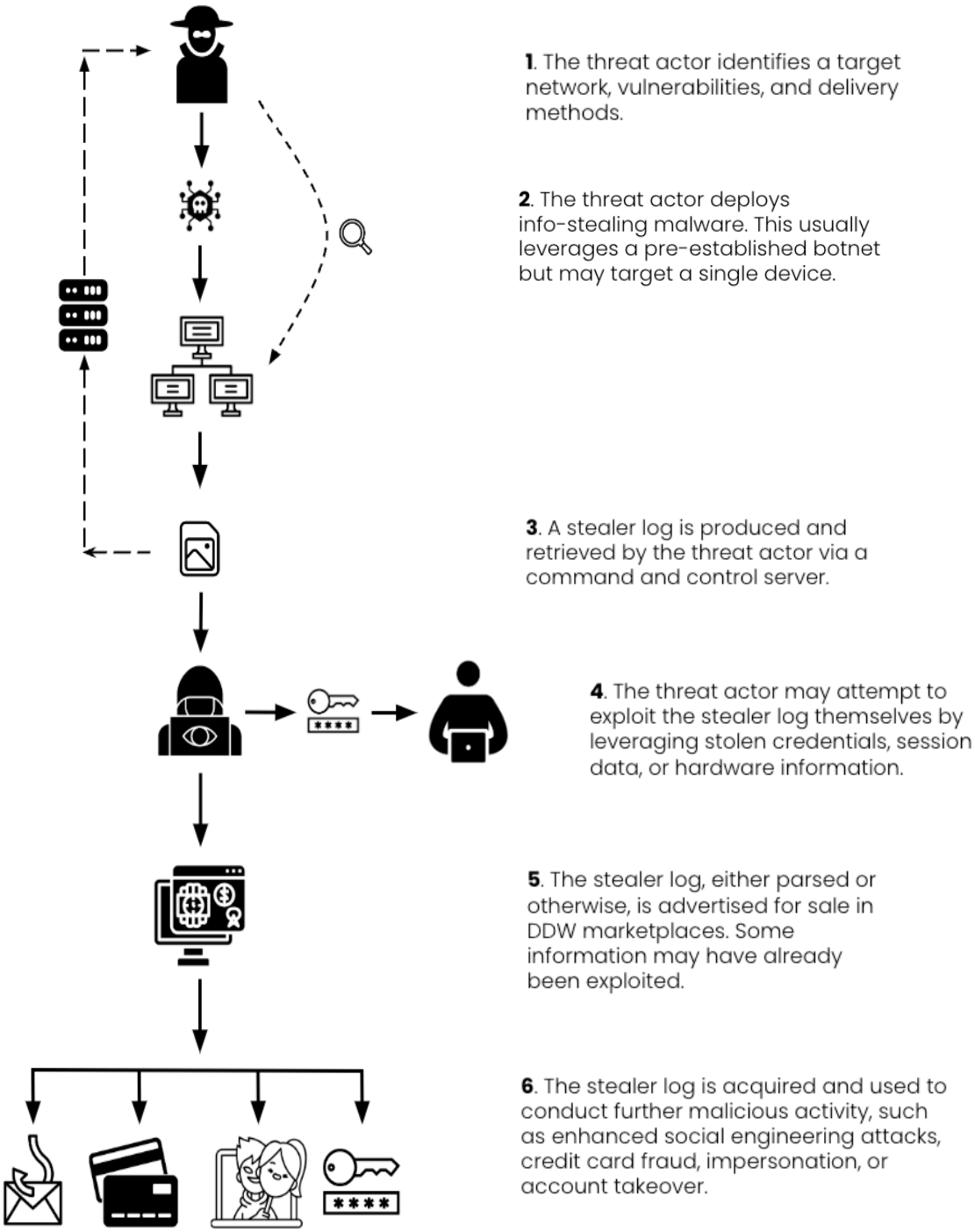
[1] hXXps://sosintel[.]co[.]uk/stealer-logs-what-you-need-to-know/

"My organization or network was targeted specifically."

- While stealer logs may contain corporate information, this does not necessarily imply that the organization was specifically targeted. The presence of such data is often a result of broader, non-targeted malware campaigns that affect individual users who happen to have access to corporate resources. This information is often obtained via breaches of third parties when employees leverage corporate credentials on their personal devices or use corporate devices for personal use.

"My information in a stealer log indicates that a broad compromise of my network has taken place."

- The presence of corporate credentials in stealer logs does not automatically mean the organization itself has been breached. The presence of corporate credentials in stealer logs is often a result of individual user compromises rather than a wider organizational breach.



Overview displaying the production, sale, and exploitation of a stealer log

Source: ZeroFox Collections

How Can the Threat Be Mitigated?

The diverse information found in stealer logs is leveraged for malicious activities by a range of threat actors of different capabilities and motivations. The mitigation strategies of individuals and organizations must therefore be equally holistic, addressing the full scope of network access vectors that can be targeted by an attacker in possession of credentials, cookies and tokens, or other browser data. Ensuring that basic cyber hygiene measures are properly enacted and scrutinized can reduce the likelihood of both being targeted by stealer malware and any subsequent exploitation.

- Adopt a Zero-Trust cybersecurity architecture based upon a principle of least privilege.
- Implement network segmentation to separate resources by sensitivity and/or function.
- Implement secure password policies with phishing-resistant MFA, complex passwords, and unique credentials.
- Leverage cyber threat intelligence to inform detection of relevant cyber threats and associated tactics, techniques, and procedures (TTPs).
- Ensure critical, proprietary, or sensitive data is always backed up to secure, off-site, or cloud servers at least once per year—and ideally more frequently.
- Develop a comprehensive incident response strategy.
- Configure email servers to block emails with malicious indicators, and deploy authentication protocols to prevent spoofed emails.
- Deploy a holistic patch management process, and ensure all IT assets are updated with the latest software updates as quickly as possible.
- Proactively monitor for compromised accounts being brokered in DDW forums.
- Ensure employees are aware of contemporary cyber threats and educated in how to recognize and report suspicious activity.
- Configure ongoing monitoring for Compromised Account Credentials.

Appendix A: Traffic Light Protocol for Information Dissemination

	Red	Amber
WHEN SHOULD IT BE USED?	**Sources may use** TLP:RED when information cannot be effectively acted upon by additional parties and could lead to impacts on a party's privacy, reputation, or operations if misused.	**Sources may use** TLP:AMBER when information requires support to be effectively acted upon but carries risks to privacy, reputation, or operations if shared outside of the organizations involved.
HOW MAY IT BE SHARED?	**Recipients may NOT share** TLP:RED with any parties outside of the specific exchange, meeting, or conversation in which it is originally disclosed.	**Recipients may ONLY share** TLP:AMBER information with members of their own organization and its clients, but only on a need-to-know basis to protect their organization and its clients and prevent further harm. **Note that** TLP:AMBER+STRICT restricts sharing to the organization only.

	Green	Clear
WHEN SHOULD IT BE USED?	**Sources may use** TLP:GREEN when information is useful for the awareness of all participating organizations, as well as with peers within the broader community or sector.	**Sources may use** TLP:CLEAR when information carries minimal or no risk of misuse in accordance with applicable rules and procedures for public release.
HOW MAY IT BE SHARED?	**Recipients may share** TLP:GREEN information with peers and partner organizations within their sector or community but not via publicly accessible channels.	**Recipients may share** TLP:CLEAR information without restriction, subject to copyright controls.

Appendix B: ZeroFox Intelligence Probability Scale

All ZeroFox intelligence products leverage probabilistic assessment language in analytic judgments. Qualitative statements used in these judgments refer to associated probability ranges, which state the likelihood of occurrence of an event or development. Ranges are used to avoid a false impression of accuracy. This scale is a standard that aligns with how readers should interpret such terms.

Almost No Chance	Very Unlikely	Unlikely	Roughly Even Chance	Likely	Very Likely	Almost Certain
1-5%	5-20%	20-45%	45-55%	55-80%	80-95%	95-99%