

Harvester Collection Automation API

Facilitating automated online collection



The Need for Automation

The internet is a rich source of valuable data, but sifting through copious amounts of ever-expanding content across the surface, deep, and dark web is a daunting task for organizations. The ability to automate data collection when a signal is received is critical to downstream analyst research and data analysis. Early enrichment through automation, including before the sources are updated and relevant information is erased, is a critical tool for timely and effective incident response.

The Challenge of Secure and Anonymous Collection

While several tools are now available to help automate data collection, most of them still require continuous input from analysts, and don't take into consideration the obfuscation requirements needed to keep organizations protected. Unfortunately, most organizations' automated collections are not able to properly disguise identity and intent, risking revealing the organizations collection activity to an adversary, getting blocked from important sites, and jeopardizing their missions. Organizations that rely on "dirty" networks to access questionable sites incur additional costs, and are still risking exposure, potential retaliation or malware infection.

The Solution: Authentic8 Harvester Automation API

Authentic8's Harvester Automation API enhances automated web-based collection through:

- The ability to integrate into existing content collection workflows via a simple API
- A range of collection options including full-page screenshots, video and other web assets
- Dedicated egress network and granular fingerprint control, allowing for unattributable collection
- Storage of collections using Authentic8 Secure Storage, with File API for file download automation
- API task management and audit logging for administrators

The Benefits

- Enhance obfuscation and security of all automated collections
- Automate and easily scale repetitive and time-sensitive content gathering activities
- Remove overhead of hosting, managing and maintaining homegrown "dirty" networks
- Improve analyst efficiency by taking collection off their task list while increasing data collection output