radware

September **2018**

Radware Research

Deconstructing Large-Scale Distributed Scraping Attacks

A Stepwise Analysis of Real-time Sophisticated Attacks On E-commerce Businesses

Table of Contents

Key Findings

Scraping - A Tool To Gain Competitive Advantage

Today, many online businesses either employ an in-house team or leverage the expertise of professional web scrapers to gain a competitive advantage over their competitors. Scrapers plan attacks in various stages to evade the vulnerabilities of existing systems such as WAFs, Intrusion Detection Systems/Intrusion Prevention Systems (IPS/IDS), and other in-house measures that lack the historical look-back, deep learning capabilities, and the ability to sniff automated behavior in syntactically-correct HTTP requests.

Usage Of Custom-built Exploit Kits

Attackers build an exploit kit that comprises a combination of tools (such as proxy IPs, multiple UAs, programmatic/sequential requests) to evade detection and perform large-scale and sophisticated scraping attacks. Websites are then hit by bots from tens of thousands of new IPs that are used once, and never again. For instance, in the case that we examined, attackers scraped product information and pricing details of 651,999 products from 11,795 categories using a combination of exploit tools and fake user accounts.

Systematic Attacks To Continuously Gather Market Intelligence

Our research shows that such organized and sophisticated attacks are fueled by the growing demand for data, price, and market intelligence. All large e-commerce firms track their competitors, hence large firms are more likely to be targeted by scrapers than small and mid-size e-tailers.

Real-world Case of A Large-Scale Scraping Attack On An E-tailer

Snapshot Of The Scraping Attack

Attack Overview

Industry:	E-commerce
Total Hits:	690,015
Duration of Study:	15 days
Fake Accounts Created:	2,345
Categories Scraped:	11,791
Products Scraped:	651,999
Scale Of The Attack:	Large-scale and distributed from thousands of locations using various evasion techniques

A popular e-commerce portal was inundated with scraping attacks and faced 690,015 hits on its category and product pages during our 15 day-long analysis.

Attackers created 2,345 fake user accounts, scraped 11,791 category results, and managed to get away with details of 651,999 products, including pricing information.

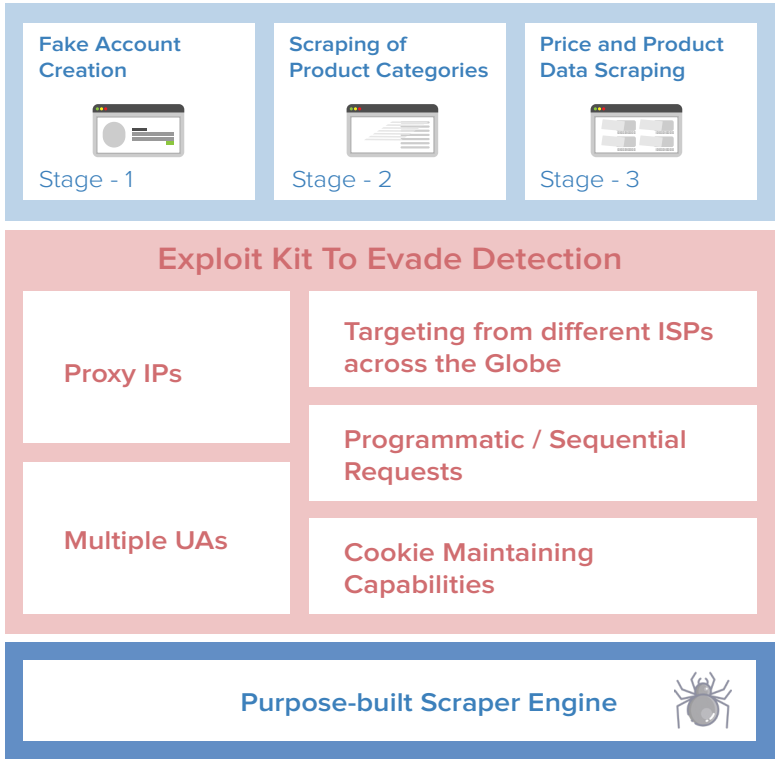
Business of Bots

The founders of Diapers.com, which Amazon acquired in 2010, have accused Amazon of using bots to automatically adjust its prices. - Brad Stone's book 'The Everything Store'

Real-world Case of A Large-Scale Scraping Attack On An E-tailer

Snapshot Of The Scraping Attack

Stages of the Attack



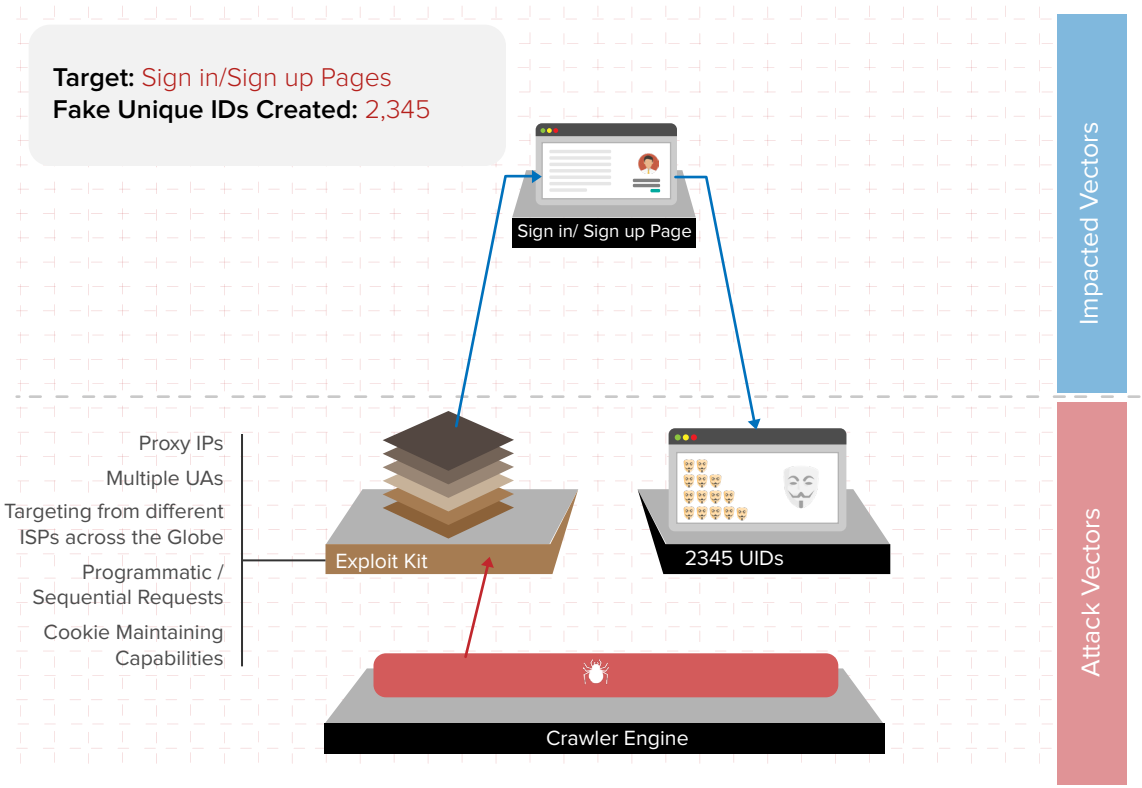
Attackers deployed a purpose-built scraper engine to execute attacks. They deployed an ‘exploit kit’ with different ready-to-use combinations of hardware and software to bypass web defense systems.

At first, they created fake accounts to register their bots as genuine users. Then they used those fake accounts to scrape category pages in the second phase. Once category pages were crawled, attackers regularly followed product pages to keep up with the latest pricing information and product updates.

Real-world Case of A Large-Scale Scraping Attack On An E-tailer

Stage 1: Fake Account Creation

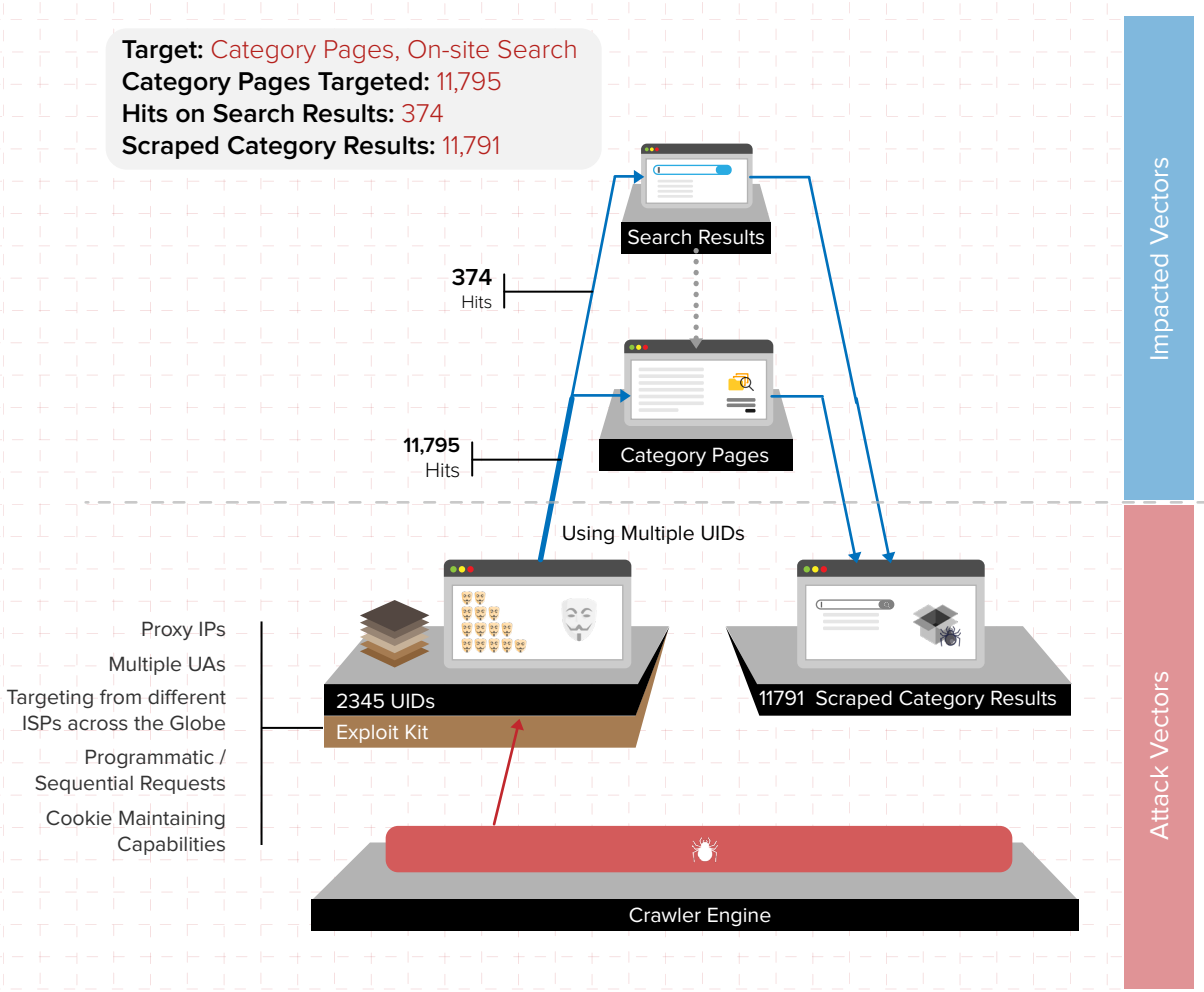
Attackers targeted the sign-up page using different attack vectors. They created 2,345 fake UIDs (User IDs) to register bots as legitimate users on the website. They used these fake accounts in combination with different device IDs, cookies, and UAs to masquerade as genuine users and generate perfectly-valid HTTP requests to easily circumvent rule-based conventional security measures.



Real-world Case of A Large-Scale Scraping Attack On An E-tailer

Stage 2: Scraping of Product Categories

Using fake UIDs, attackers logged into the website and made 11,795 hits on category pages. They managed to scrape 11,791 category results. Scrapers also performed 374 searches.



Business of Bots

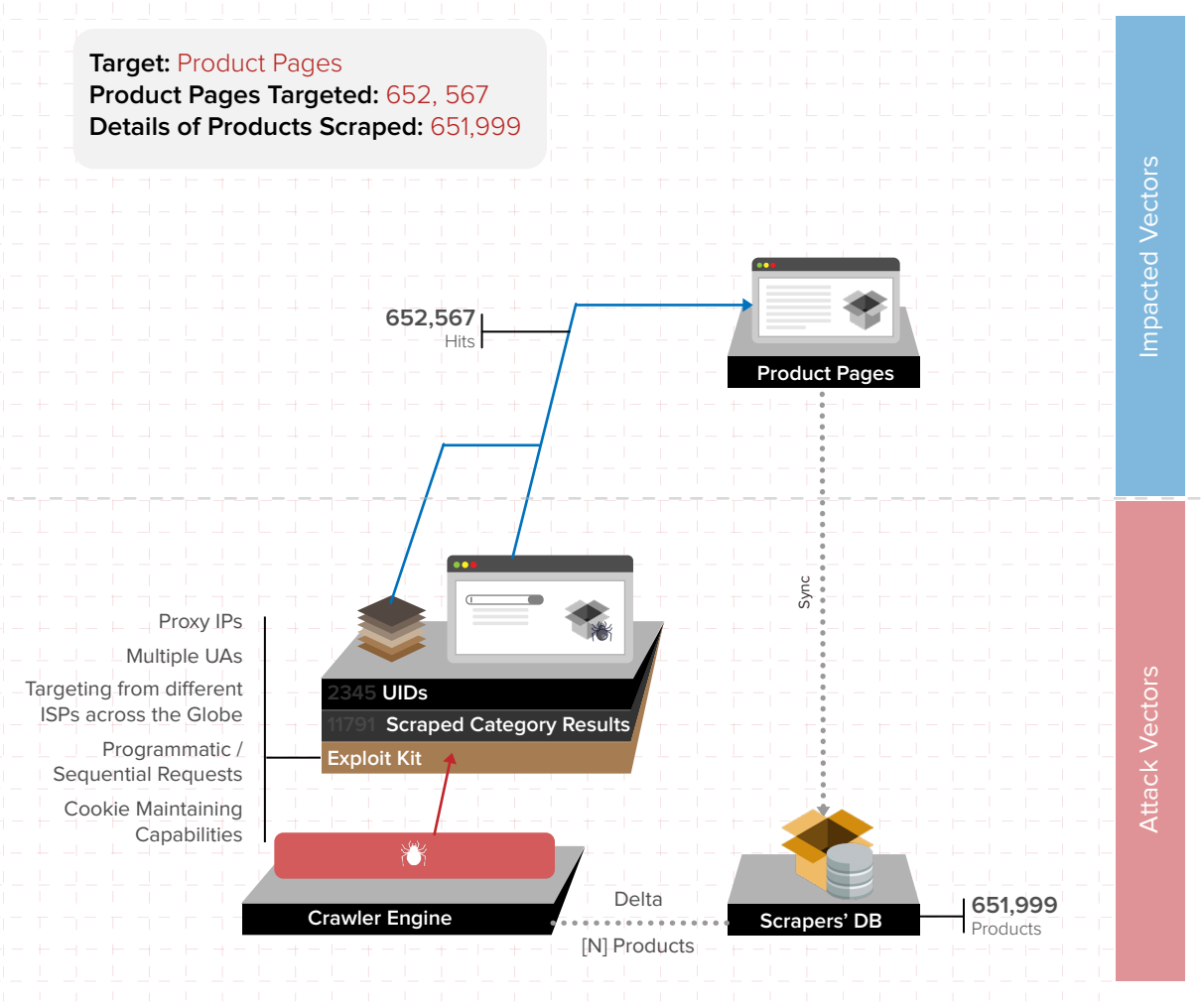
Thousands of scrapers are listed on sites such as upwork, guru.com, and freelancers.com

Real-world Case of A Large-Scale Scraping Attack On An E-tailer

Stage 3: Price and Product Data Scraping

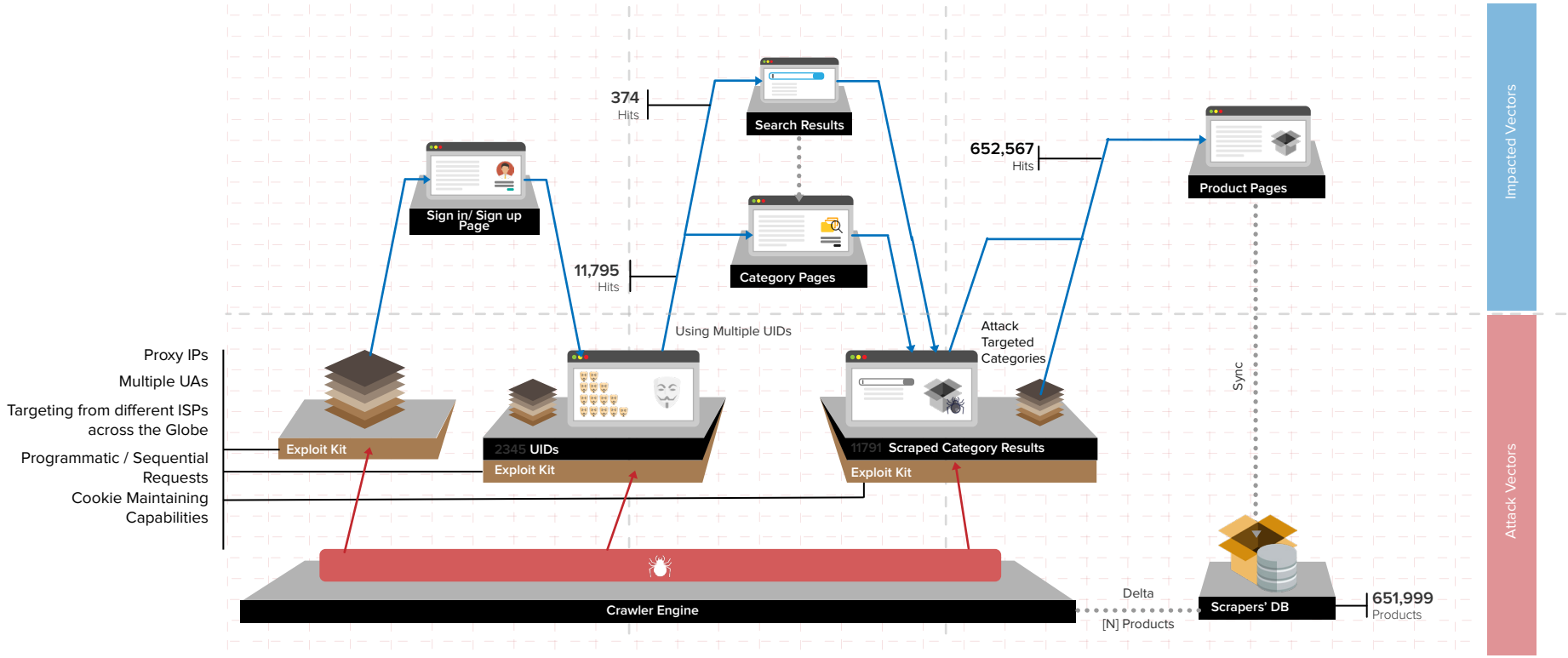
After scraping the category pages, attackers carried out 652,567 hits on specific product pages and managed to store the prices and product details of 651,999 products in their own database.

The attackers maintained a real-time repository of the entire product catalog on the e-commerce portal. They also regularly tracked the price changes to keep their database updated with the latest pricing information.



Real-world Case of A Large-Scale Scraping Attack On An E-tailer

Topology of The Attack — How Three Stages Work in Unison



All the three stages were part of a single large-scale scraping attack and worked together to perform real-time monitoring of product pages. During our analysis, we observed that rule-based systems are incapable of detecting such scraping attacks that are organized in different stages and executed using perfectly legitimate user activities.

Recommendations

Action Plan for E-commerce Businesses to Combat Scraping

All large e-commerce platforms have sophisticated bot activity on their website, mobile apps, and APIs that can expose them to scraping and loss of Gross Merchandise Value (GMV). E-tailers must be diligent in their approach to find and mitigate malicious sources of bot activity.

Spot highly active new or existing user accounts that don't buy

E-commerce portals must track old or newly-created accounts that are highly active on the platform but haven't made any purchase in a long time. Such accounts may be handled by bots which mimic real users to scrape product details and pricing information.

Don't overlook unusual traffic on selected product pages

E-tailers should monitor unusual spikes in page views of certain products. These spikes can be periodic in nature. A sudden surge in engagement on selected product pages can be a symptom of non-human activity on your website.

Watch out for competitive price tracking and monitoring

Many e-commerce firms deploy bots or hire professionals to scrape product details and pricing information from their rival portals. You must regularly track competitors for signs of price and product catalog matching.

Build capabilities to identify automated activity in seemingly legitimate user behaviors

Sophisticated bots simulate mouse movements, perform random clicks, and navigate pages in a human-like manner. Preventing such attacks require deep behavioral models, device/browser fingerprinting, and closed-loop feedback systems to ensure that you don't block genuine users. Purpose-built bot mitigation solutions can identify such sophisticated automated activities and can help you take action against them. In comparison, traditional solutions such as WAFs are limited to tracking spoofed cookies, user agents, and IP reputation.

About Radware

Radware® (NASDAQ: RDWR), a leading provider of cyber security and application delivery solutions, acquired ShieldSquare in March 2019. ShieldSquare is now Radware Bot Manager.

Radware® (NASDAQ: RDWR) is a global leader of cybersecurity and application delivery solutions for physical, cloud and software defined data centers. Its award-winning solutions portfolio secures the digital experience by providing infrastructure, application and corporate IT protection and availability services to enterprises globally. Radware's solutions empower more than 12,500 enterprise and carrier customers worldwide to adapt quickly to market challenges, maintain business continuity and achieve maximum productivity while keeping costs down. For more information, please visit www.radware.com

Radware encourages you to join our community and follow us on: Radware Blog, LinkedIn, Facebook, Twitter, SlideShare, YouTube, Radware Connect app for iPhone® and our security center DDoSWarriors.com that provides a comprehensive analysis of DDoS attack tools, trends and threats.

www.radware.com

www.shieldsquare.com