

Lab: Hunting SCATTERED SPIDER Phishing Infrastructure

Background

ACME Corp. processes tens of thousands of web transactions per hour through its Secure Web Gateway (SWG) / Zscaler Internet Access (ZIA) platform. The overwhelming majority of this traffic is legitimate business activity: SaaS usage, collaboration tools, document sharing, accounting platforms, and remote work tooling.

The ZIA CSV file provided for this lab contains 16,000+ web transactions (one hour window) representing normal, day-to-day enterprise user activity. This includes legitimate SaaS usage, browsing, analytics traffic, and background application noise.

Recently, multiple vendors released reports describing an increase in eCrime activity attributed to the SCATTERED SPIDER group. Unlike traditional phishing campaigns that rely on obviously malicious infrastructure, this group is known for:

- Targeting legitimate, widely used SaaS platforms
- Creating look-alike authentication infrastructure
- Leveraging allowed traffic to blend in with normal user behavior
- Focusing on identity compromise rather than malware delivery

Because the activity often occurs over HTTPS and uses trusted brands, signature-based detections are insufficient.

Threat Context

A recent vendor report highlights that SCATTERED SPIDER and similar eCrime groups are actively phishing users of the following legitimate platforms:

- QuickBooks
- AppSheet
- Canva
- SurveyMonkey
- Google Docs
- Zoom

The phishing infrastructure is intentionally designed to closely resemble legitimate authentication endpoints, often differing by only:

- A hyphen
 - Word order
 - Subdomain placement
-

Lab Objective

In this lab, you will analyze a large ZIA CSV [dataset](#) (~16,000+ rows) containing mostly benign web activity of a user over the course of an hour.

Your objectives are to:

- Reduce noise in a high-volume dataset
- Identify domain-level patterns associated with phishing
- Apply threat intelligence traits specific to SCATTERED SPIDER
- Determine whether the observed activity warrants incident response escalation

You are not looking for known bad malware domains. You are hunting for tradecraft.

CTI Traits to Guide the Hunt

Use the following threat intelligence as directional guidance, not hard filters:

1. **Infrastructure**
 - Frequent use of AS63949
 - Legitimate hosting providers mixed with short-lived domains
2. **Domain Naming Patterns**
 - Use of authentication-related keywords:
 - auth
 - login
 - identity
 - sso
 - Keywords combined with target brand names
 - Hyphenated or reordered strings (e.g., auth-`<brand>`, sso-`<brand>`, auth-us-`<brand>`, etc.)
3. **Phishing Content**
 - SSO and login workflows closely resembling legitimate SaaS platforms

- Allowed HTTPS POST traffic (credential harvesting)
 - No malware delivery required
-

How the Lab Data Is Structured

- CSV file ([lab_phishing.csv](#)) within the `/data/` directory contains ZIA web proxy logs
- Dataset includes:
 - Normal SaaS usage
 - Content delivery traffic
 - Advertising and analytics domains
 - Collaboration platforms
- Mixed into this noise are phishing lures attempting to blend in

NOTE: Some (or all) phishing domains are **Allowed**, not blocked.

Step 1: Reduce the Dataset to Unique Domains

Start by extracting and grouping domains from the URL field.

```
Shell
awk -F',' '{
    gsub(/^"|"$/ , "", $8)
    split($8, a, "/")
    print a[1]
}' ./data/lab_phishing.csv | sort | uniq -c | sort -nr
```

This reduces the dataset from 16,000+ rows to smaller grouped domains, making analysis feasible. With a dataset spanning more than an hour and a single user, further filtering will be needed.

Step 2: Focus on Targeted Platforms

Based on threat intelligence, begin pivoting on known targeted brands.

For example, QuickBooks, AppSheet, Canva, SurveyMonkey, Google Docs, and Zoom:

```
Shell
awk -F',' '{
    gsub(/^"|"$/ , "", $8)
    split($8, a, "/")
    print a[1]
}' ./data/lab_phishing.csv | sort | uniq -c | sort -nr | grep
<keyword>
```

Pro Tip: You can perform an OR condition in grep using `grep -E 'keyword1|keyword2|keyword3'`

At first glance, this looks like normal SaaS usage.

Step 3: Exclude Known Legitimate Domains

Since [surveymonkey.com](https://www.surveymonkey.com) is a legitimate domain, exclude it to surface look-alike (masquerading) infrastructure.

```
Shell
awk -F',' '{
    gsub(/^"|"$/ , "", $8)
    split($8, a, "/")
    print a[1]
}' ./data/lab_phishing.csv \
| sort | uniq -c | sort -nr \
| grep -i surveymonkey \
| grep -viE '^(|\.)surveymonkey\.com($|\.)'
```

Result:

```
None
6 auth-us-surveymonkey.com
2 surveymonkey.com
```

```
1 surveymonkey.chilipiper.com
1 auth-us-surveymonkey.com:443
```

This dramatically reduces the investigation scope and makes it easy to spot the doppelganger (auth-us-surveymonkey[.]com) likely used for credential harvesting .

Step 4: Apply Threat Intelligence Reasoning

Ask yourself:

- Why would a legitimate platform host credentials on a hyphenated apex domain?
- Is `POST` traffic allowed to this host?

```
awk -F',' 'NR>1 && /,POST,/ { split($8,a,"/"); print a[1] }'
./data/lab_phishing.csv | sort | uniq -c | sort -nr
```

- How closely does the domain naming pattern resemble known SCATTERED SPIDER tradecraft?
 - Does this align with recent vendor reporting?
-

Questions

Question	Answer
How many distinct domains were reduced from the original dataset?	
Which domains appeared visually similar to legitimate SaaS applications (e.g. Zoom, SurveyMonkey, etc.) infrastructure?	
What domain naming traits suggest phishing tradecraft?	
Why is allowed <code>POST</code> traffic significant in this context?	
How does this activity align with SCATTERED SPIDER TTPs?	
Would you recommend incident response escalation? Why or why not?	

Final Decision Exercise

You are briefing leadership.

Choose one:

- ☐ Document and monitor only
- ☐ Block domains and reset impacted credentials
- ☐ Initiate incident response and identity compromise investigation

Be prepared to justify your decision using:

- Dataset evidence
- Domain patterns
- Threat intelligence alignment