Detecting data exfiltration activities

When attackers are looking to identify and exfiltrate data from a target organization, they carry out attacks which contain three main activities: identification, collection, and staging data for exfiltration.

- Identification includes scanning systems and observing user activity.
- Collection includes the transfer of large amounts of data from various repositories.
- Staging, or preparation, includes moving data to a central location and compressing it, also optionally encoding or encrypting it.

These searches allow you to detect and monitor suspicious behavior related to these activities.

Required data

- Endpoint data
- Network protocol data

Some commands, parameters, and field names in the searches below may need to be adjusted to match your environment. In addition, to optimize the searches shown below, you should specify an index and a time range when appropriate.

How to use Splunk software for this use case

► Email files written outside of the Outlook directory

To complete this process, your deployment needs to ingest information on filesystem activity from your hosts. This is typically populated via endpoint detection-and-response products, such as Carbon Black, or by other endpoint data sources, such as Sysmon. You should also ensure you are ingesting normalized endpoint data, populating the Filesystem node of the Endpoint data model in the Common Information Model (CIM). For information on installing and using the CIM, see the Common Information Model documentation.

This search detects email files created outside of the normal Outlook directory.

False positives from this search might occur since administrators and users sometimes prefer backing up their email data by moving email files into a different folder. These attempts will be detected by this search.

```
| tstats summariesonly=false allow_old_summaries=true count, values("Filesystem.file_path") AS file_path, min(_time) AS firstTime, max(_time) AS lastTime FROM datamodel=Endpoint.Filesystem WHERE (("Filesystem.file_name"=*.pst OR "Filesystem.file_name"=*.ost) "Filesystem.file_path"!="C:\\Users\\\My
```

The information provided in Splunk Lantern is intended for informational and educational purposes only. All information is provided in good faith, however, Splunk disclaims any and all representations and warranties, express and implied, regarding the information provided, including without limitation any warranties and representations regarding the completeness, adequacy or accuracy of the information. You agree to take full responsibility for the results arising from the use of the information provided.
Hosts receiving a high volume of network traffic from an email server

To complete this process, your deployment needs to ingest network traffic data. Your email servers must be categorized as "email_server" for the search to work. You should also ensure you are ingesting normalized endpoint data, populating the Network_Traffic node of the Endpoint data model in the Common Information Model (CIM). For information on installing and using the CIM, see the Common Information Model documentation.

This search looks for an increase of data transfers from your email server to your clients. This could be indicative of an attacker collecting data using your email server.

False positives will vary based on how you set the deviation_threshold and data_samples values. You should adjust these values based on your network traffic to and from your email servers. The deviation_threshold field is a multiplying factor to control how much variation you're willing to tolerate. The minimum_data_samples field is the minimum number of connections of data samples required for the statistic to be valid.

Email servers sending a high volume of traffic to hosts

To complete this process, your deployment needs to ingest network traffic data. Your email servers must be categorized as "email_server" for the search to work. You should also ensure you are ingesting normalized endpoint data, populating the Network_Traffic node of the Endpoint data model in the Common Information Model (CIM). For information on installing and using the CIM, see the Common Information Model documentation.
This search looks for an increase of data transfers from your email server to your clients. This could be indicative of a malicious actor collecting data using your email server.

False positives will vary based on how you set the deviation_threshold and data_samples values. You should adjust these values based on your network traffic to and from your email servers. The deviation_threshold field is a multiplying factor to control how much variation you're willing to tolerate. The minimum_data_samples field is the minimum number of connections of data samples required for the statistic to be valid.

```
| tstats summariesonly=false allow_old_summaries=true sum("All_Traffic.bytes_out") AS bytes_out FROM datamodel=Network_Traffic WHERE "All_Traffic.src_category"=email_server BY "All_Traffic.dest_ip", _time span=1d |
| rename "All_Traffic.**" AS *** |
| eventstats avg(bytes_out) AS avg_bytes_out stdev(bytes_out) AS stdev_bytes_out |
| eventstats count AS num_data_samples avg(eval(if(_time < relative_time(now(), "@d"), bytes_out, null))) AS per_source_avg_bytes_out stdev(eval(if(_time < relative_time(now(), "@d"), bytes_out, null))) AS per_source_stdev_bytes_out BY dest_ip |
| WHERE ("_time" >= relative_time(now(),"@d")) |
| eval minimum_data_samples=4, deviation_threshold=3 |
| WHERE (((bytes_out > ((deviation_threshold * stdev_bytes_out) + avg_bytes_out)) AND (bytes_out > ((deviation_threshold * per_source_stdev_bytes_out) + per_source_avg_bytes_out))) AND (num_data_samples >= minimum_data_samples)) |
| eval num_standard_deviations_away_from_server_average=round((abs((bytes_out - avg_bytes_out)) / stdev_bytes_out,2), num_standard_deviations_away_from_client_average=round((abs((bytes_out - per_source_avg_bytes_out)) / per_source_stdev_bytes_out),2) |
| table dest_ip, _time, bytes_out, avg_bytes_out, per_source_avg_bytes_out, num_standard_deviations_away_from_server_average, num_standard_deviations_away_from_client_average |
```

**Next steps**

The content in this article comes from Splunk Enterprise Security (ES). As a Splunk premium security solution, ES solves a wide range of security analytics and operations use cases including continuous security monitoring, advanced threat detection, compliance, incident investigation, forensics and incident response. Splunk ES delivers an end-to-end view of an organization's security posture with flexible investigations, unmatched performance, and the most flexible deployment options offered in the cloud, on-premises, or hybrid deployment models. If you have questions about this use case, see the Security Research team’s support options on GitHub.

In addition, these Splunk resources might help you understand and implement this use case:

- **Use case:** Detecting data exfiltration through email
- **Docs:** Using Enterprise Security to find data exfiltration