Automating Google Workspace Incident Response

Megan Roddie
Automating Google Workspace Incident Response

GIAC GCIA Gold Certification

Author: Megan Roddie, megansroddie@gmail.com
Advisor: Bryan Simon

Accepted: November 1st, 2020

Abstract

Incident responders require a toolset and resources that allow them to efficiently investigate malicious activity. In the case of Google Workspace, there are an increasing number of subscribers, but resources to assist in the analysis of security incidents are lacking. The goal of this research is to develop a tool that expands on Google’s default administrative capabilities with the intent of providing value to incident responders. Through providing both additional context and purposeful views, incident responders can more quickly identify malicious activity and respond accordingly.
1. Introduction

The number of businesses seeking cloud-based solutions for file storage, application hosting, email, and more is increasing. One of the companies offering a collection of services for these businesses is Google, with Google Workspace (formerly G Suite). Google Workspace provides its customers with a suite of tools that include communication applications, email services, document collaboration capabilities, and cloud storage. While not as prevalent as one of its major competitor’s solutions, Microsoft 365, as of the end of 2019, Google had 6 million businesses paying for Google Workspace (Novet, 2020). The significance of this number warrants further assessment of the cybersecurity implications tied to the product.

Cloud services are at risk of cybersecurity threats, just like any other environment. One of the most significant risks to discuss when it comes to services like Google Workspace is business email compromises. With access to a victim’s Google Workspace account, an attacker gains access to their emails, files, calendar, and more. From this point, they can compromise other users via social engineering and exfiltrate sensitive data. If they are able to access an administrative account, they may establish persistence to ensure that they can maintain their access as long as possible.

With these risks in mind, one such consideration related to cloud services is the ability for incident responders to investigate a cybersecurity incident that occurs within the confines of the platform. This paper will analyze Google’s built-in logging tools and a custom open-source tool that has been built as part of the research. This custom tool aims to provide incident responders with a view of activity that is more efficient and useful than that which is provided by Google.

To perform this analysis, simulated activity that mimics a business email compromise and associated post-compromise actions will be generated. A cross-comparison between the logs from the Google Workspace administrator portal and those collected by the custom tool will be performed. The analysis will demonstrate a) how to find evidence of business email compromise activity in Google Workspace and b) the value of Google’s built-in logging versus a custom incident response tool.

Megan Roddie, megansroddie@gmail.com
The hypothesis is that activity associated with the incident will be more efficiently identified with the custom logging tool than Google’s built-in capabilities. This is because the different types of logs in the Google Workspace admin portal are separated without a complete view. Additionally, the API potentially provides greater detail associated with the logs than the details provided via the UI.

2. Research Method

The analysis performed involves the use of a Google Workspace instance subscribed to the lowest subscription tier that contains a history of legitimate activity. Additionally, simulated malicious activity was generated to show that the logging tool can bring such threats to light. After developing and executing the logging tool, a review of the logs is performed via both the logging tool’s output and Google Workspace’s built-in logging to show the tool’s effectiveness.

2.1. Overview of Google Workspace

2.1.1. Google Workspace Subscription Tiers

As mentioned, the analysis in this paper leverages the lowest subscription tier offered by Google. Several options are available to subscribers: basic, business, and enterprise (Google, 2020). While there is significant overlap in the applications provided across the tiers, higher subscription tiers provide additional administrative and security features. Some of these additional features include e-Discovery and other capabilities that may improve the incident response workflow. However, to ensure that this tool can assist Google Workspace customers regardless of their subscription level, the lowest tier is used for testing purposes.

It is worth noting that the built-in retention of logs and emails is also significantly lower for the lowest subscription tier. Most audit logs are stored for six months, while email search capabilities are restricted to 30 days (Google, 2020). Higher subscription tiers offer log archiving capabilities via Vault. In terms of challenges to incident responders, the limited log retention means that it is harder for them to identify an initial
infection vector and activity if the attacker compromised the system outside of the time frame that logs are available.

### 2.1.2. Google Workspace Applications

Google Workspace provides a range of products and services to its subscribers mainly focused on business productivity. First and foremost, especially when it comes to cybersecurity, is Gmail business email. Google Drive, their file storage service, is also of significant interest to attackers. With access to both of these applications, attackers would be able to obtain a treasure trove of sensitive business data and potentially even personally identifiable information (PII).

Further applications offered by the platform that may contain data valuable to attackers include their communication platforms, Chat and Meet, and business-wide shared calendars.

### 2.2. Attack Simulation

The Google Workspace environment used for simulating malicious activity for this research is a pre-existing legitimate organization’s account, so it contains a large amount of legitimate activity to make the simulation more realistic. For the attack activity, one of the existing user accounts, specifically an admin account, will represent the victim of the compromise.

First, the victim account will receive a fake phishing email with a link purporting to be a Google Workspace login page. Using a VPN configured to tunnel traffic through a foreign country, the victim account will then be logged into to represent suspicious login activity from a location not previously associated with the victim. This series of activities will be investigated in the logs as the source of the compromise.

With access to the account, the simulated attacker will change various settings in an attempt to gain persistence as well as interact with files and other data. The ideal outcome is that all this activity will be discoverable in the outputted logs so that responders can identify the impact and perform clean up.

Megan Roddie, megansroddie@gmail.com
While this simulation focuses on the aforementioned attack chain, various methods could be used by an attacker for initial compromise, persistence, and other post-compromise activity. Research into this topic has been published by Black Hills Information Security researchers, allowing responders to understand how their environments could be attacked so that they may be prepared to respond (Felch & Bullock, 2018).

2.3. Logging Tool Development

The tool developed to improve the efficiency of analysis of incidents in Google Workspace is written in Python. It leverages both the Google Workspace API and Pandas, a Python data analysis library, for the bulk of its operation. Additionally, integration with the MaxMind geolocation database will provide further context to the data. No user input is required outside of the initial configuration that provides the API credentials needed to authenticate to Google Workspace. The initial version of the tool focuses on pulling logs associated with login, Google Drive, admin, and user activity. The output provided by the tool will be presented in a spreadsheet containing the relevant data.

2.3.1. Program Structure

The code for the program instantiates a class that, at the time of this writing, contains seven functions related to obtaining logs from Google Workspace, generating additional context, and displaying the data in an efficient view to incident responders. The first function creates a Google Workspace API session, specifically with the Reports V1 API that is a part of the Google Workspace Admin SDK. Following this function is four similar functions, one for each of the categories of logs to be gathered: login, admin, user, and Google Drive logs. Via the API, the associated logs are pulled, and then the Pandas Python library is used to consolidate, organize, and format the returned data. Each function outputs the results to a separate sheet in an Excel workbook. Another one of the functions is used for geolocation and executed within each of these log collection functions to provide additional context to the data. Lastly, the “timeline” function adds a
final sheet to the Excel file that creates a single view containing all of the logs, regardless of type.

3. Findings and Discussion

3.1. Program Execution

Based on the format of the file, users of the tool can easily execute the program from the command line. Running the program will leverage the credentials stored in the JSON file created during setup. The results will be output to an XLSX file.

As mentioned in Section 2.3, the main focus of activity collection was logins, Google Drive, user account activity, and admin activity. Figure 1 shows the list of worksheets within the outputted XLSX file after program execution.

![Figure 1. Worksheets in Program Output](image)

The first four sheets include logs specific to each of the categories, while the “All” sheet combines all of the logs to a single pane of view, making it easier for users to receive an overview of all activity as opposed to clicking through multiple views and cross-reference data. This is one of the significant benefits of leveraging this custom tool as opposed to using Google Workspace’s built-in logging as Google does not provide all audit logs in a singular view, but rather they are only broken out by category.

3.2. Data Enrichment

One of the key aims of this research and the developed tool is to determine whether the tool could gather more data and context than shown in Google Workspaces’ admin interface. Reviewing both the Google Workspace admin view and the output Excel files reveals that the provided data is almost identical but presented differently. A significant piece of information not provided by Google is the geolocation of the IP address.
address. While only a single data point, later analysis of the logs will show the immense value that this provides to responders.

Geolocation is performed using a MaxMind database, and currently, the code returns the associated country’s ISO code and the city. Figure 2 shows the piece of code leveraging Python’s geoip2 library in combination with the MaxMind GeoLite2 City Database to provide location information for a given IP Address.

```
def get_geoip(self, ipAddress):
    reader = geoip2.database.Reader(
        self.geolocate_db)
    response = reader.city(ipAddress)
    return [response.country.iso_code, response.city.name]
```

Figure 2. Geolocation Function

Each IP Address in the collection logs is passed to this function and the result is provided in the output under the “loginCountry” and “loginCity” columns.

The API also provided some metadata related to each log, such as an etag, customerID, and more. However, during the development of the program it was determined that there was no significant value provided by this data to incident responders, so it was not included in the final output.

### 3.3. **Log Views and Analysis**

Based on the simulated activity, incident responders should observe a few relevant events across the logs, allowing for analysis of whether Google Workspace’s built-in logging or the output of the logging tool are more useful to a responder.

First, login audit logs should be analyzed in an attempt to find any abnormal login behavior. In the Google Workspace admin panel, this can be found by going to “Reports > Audit log > Login.” Figure 3 shows a screenshot of this panel and with all available...
fields displayed.

<table>
<thead>
<tr>
<th>Event Description</th>
<th>IP Address</th>
<th>Date</th>
<th>Login Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Megan Roddie logged in</td>
<td>89.206</td>
<td>Sep 27, 2020, 2:56:12 PM CDT</td>
<td>Google Password</td>
</tr>
<tr>
<td>Megan Roddie logged in</td>
<td>217.76</td>
<td>Sep 27, 2020, 1:27:16 PM CDT</td>
<td>Google Password</td>
</tr>
<tr>
<td>Megan Roddie logged in</td>
<td>217.76</td>
<td>Sep 27, 2020, 1:26:21 PM CDT</td>
<td>Google Password</td>
</tr>
<tr>
<td>Megan Roddie logged in</td>
<td>89.206</td>
<td>Sep 27, 2020, 1:25:38 PM CDT</td>
<td>Google Password</td>
</tr>
<tr>
<td>Megan Roddie logged in</td>
<td>89.206</td>
<td>Sep 27, 2020, 1:24:08 PM CDT</td>
<td>Google Password</td>
</tr>
<tr>
<td>bob logged in</td>
<td>33df:4a8:2307</td>
<td>Sep 25, 2020, 1:41:40 AM CDT</td>
<td>Google Password</td>
</tr>
<tr>
<td>bob was presented with login verification</td>
<td>33df:4a8:2307</td>
<td>Sep 25, 2020, 1:41:40 AM CDT</td>
<td>Google Password</td>
</tr>
</tbody>
</table>

Figure 3. Google Workspace Audit Logs – Login

Figures 4 and 5 show the same log type within the “Login Activity” tab of the tool’s output.

<table>
<thead>
<tr>
<th>userEmail</th>
<th>ipAddress</th>
<th>loginCountry</th>
<th>loginCity</th>
<th>timestamp</th>
<th>applicationName</th>
</tr>
</thead>
<tbody>
<tr>
<td>megan</td>
<td>99.12</td>
<td>HK Central</td>
<td></td>
<td>2020-09-24T02:21:55.573Z</td>
<td>login</td>
</tr>
<tr>
<td>megan</td>
<td>99.12</td>
<td>HK Central</td>
<td></td>
<td>2020-09-24T02:21:55.573Z</td>
<td>login</td>
</tr>
<tr>
<td>megan</td>
<td>99.12</td>
<td>HK Central</td>
<td></td>
<td>2020-09-24T02:21:55.573Z</td>
<td>login</td>
</tr>
<tr>
<td>megan</td>
<td>209.2</td>
<td>HK Central</td>
<td></td>
<td>2020-09-24T02:21:55.573Z</td>
<td>login</td>
</tr>
<tr>
<td>megan</td>
<td>209.2</td>
<td>HK Central</td>
<td></td>
<td>2020-09-24T02:21:55.573Z</td>
<td>login</td>
</tr>
<tr>
<td>megan</td>
<td>209.2</td>
<td>HK Central</td>
<td></td>
<td>2020-09-24T02:21:55.573Z</td>
<td>login</td>
</tr>
<tr>
<td>megan</td>
<td>99.12</td>
<td>HK Central</td>
<td></td>
<td>2020-09-24T02:21:55.573Z</td>
<td>login</td>
</tr>
<tr>
<td>megan</td>
<td>99.12</td>
<td>HK Central</td>
<td></td>
<td>2020-09-24T02:21:55.573Z</td>
<td>login</td>
</tr>
<tr>
<td>sadie</td>
<td>107.12</td>
<td>US Port Clinton</td>
<td>2020-09-24T02:21:55.573Z</td>
<td>login</td>
<td></td>
</tr>
<tr>
<td>megan</td>
<td>209.2</td>
<td>HK Central</td>
<td></td>
<td>2020-09-24T02:21:55.573Z</td>
<td>login</td>
</tr>
<tr>
<td>megan</td>
<td>99.12</td>
<td>HK Central</td>
<td></td>
<td>2020-09-24T02:21:55.573Z</td>
<td>login</td>
</tr>
<tr>
<td>megan</td>
<td>209.2</td>
<td>HK Central</td>
<td></td>
<td>2020-09-18T21:47:33.134Z</td>
<td>login</td>
</tr>
</tbody>
</table>

Figure 4. Tool Output - Login Activity
There are a few benefits to the custom tool being used to obtain and view logs. First and foremost, the presence of the loginCountry and loginCity columns provides additional context that may highlight illegitimate activity. As mentioned before, Google provides no context of geolocation in their logs; this was a function included in the tool. How this data is interpreted will vary based on the geographic demographic of a company’s staff. However, in this case, it can be seen how geolocation information may indicate abnormal patterns of activity. As highlighted in Figure 4, the user “megan@***” is recorded as logging in from Round Rock, United States but shortly after can be seen logging in again from Central Hong Kong. Regardless of whether the user has a legitimate reason for operating in China or not, the timeline in which the two distant locations appear should be a sign to incident responders that further investigation should occur.

A secondary benefit is that while filters can be applied in the Google Workspace UI, sorting capabilities are not built in. The default sorting is presented from the most recent to the oldest activity, which, while potentially the most valuable method, does not allow the flexibility that may be desirable to an incident responder attempting to find anomalies.

Figures 6 through 8 provide a comparison of the Google Workspace Audit Logs for Google Drive versus the tool’s output of logs related to Google Drive.
Once again, the most value appears from obtaining the geolocation context, which can be seen in Figure 7 where activity is taking place from Hong Kong right before activity based in the United States. At this point in the simulation activity, a responder would have used the login activity to discover at least one of the compromised accounts. A filter
could be applied in either the spreadsheet or the Google Workspace UI to discover what the victim account was used for post-compromise.

The last set of figures show user and admin audit logs related to user and admin activities or setting changes performed in Google Workspace. These logs are essential to review as they provide context as to what an attacker may have done post-compromise, especially in regard to establishing persistence.

<table>
<thead>
<tr>
<th>Event Description</th>
<th>Date</th>
<th>IP Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Megan Roddie has enabled 2-step verification</td>
<td>Sep 27, 2020, 1:28:26 PM CDT</td>
<td>193.102...</td>
</tr>
<tr>
<td>Megan Roddie has changed Account recovery email</td>
<td>Sep 27, 2020, 1:26:38 PM CDT</td>
<td>193.102...</td>
</tr>
<tr>
<td>Megan Roddie has disabled 2-step verification</td>
<td>Sep 23, 2020, 9:05:50 PM CDT</td>
<td>98.125...</td>
</tr>
<tr>
<td>mber has changed Account password</td>
<td>May 7, 2020, 9:19:17 PM CDT</td>
<td>260.5...</td>
</tr>
<tr>
<td>erci changed Account recovery email</td>
<td>Apr 8, 2020, 4:43:27 PM CDT</td>
<td>159.128...</td>
</tr>
</tbody>
</table>

Figure 9. Google Workspace Audit Logs - User

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>userEmail</td>
<td>appAddress</td>
<td>loginCountry</td>
<td>loginCity</td>
<td>timestamp</td>
<td>applicationName</td>
<td>email</td>
<td>callerType</td>
<td>type</td>
</tr>
</tbody>
</table>
| megan | | HK | Central | 2020-09-27 | 13:26:39 | 26.0.0.2 | USER | 2 | megan
| megan | | HK | Central | 2020-09-27 | 13:26:39 | user_accounts | USER | recovery_info_change | recovery_email_edit
| megan | | US | Round Rock | 2020-09-27 | 13:50:37 | 50.0.0.2 | USER | 2 | megan
| mber | | US | Port Clinton | 2020-09-03 | 13:17:30 | 65.0.0.2 | USER | password_change | password_edit
| erci | | US | Kansas City | 2020-04-08 | 13:27:77 | 77.0.0.2 | USER | recovery_info_change | recovery_email_edit

Figure 10. Tool Output - User Activity

<table>
<thead>
<tr>
<th>Event Name</th>
<th>Event Description</th>
<th>Admin</th>
<th>Date</th>
<th>IP Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Email Log Search</td>
<td>An email log search is performed for logs from 2020/09/28 09:00 UTC to 2020/09/28 09:09 UTC with a sender of megan, a recipient of megan, and an email message subject of megan_email_search_email, and an email log search email recipient ip ([1])</td>
<td>Megan Roddie</td>
<td>Sep 27, 2020, 2:57:27 PM CDT</td>
<td>99.125...</td>
</tr>
<tr>
<td>Email Log Search</td>
<td>An email log search is performed for logs from 2020/09/28 09:00 UTC to 2020/09/28 09:09 UTC with a sender of megan, a recipient of megan, and an email message subject of megan_email_search_email, and an email log search email recipient ip ([1])</td>
<td>Megan Roddie</td>
<td>Sep 27, 2020, 2:57:06 PM CDT</td>
<td>99.125...</td>
</tr>
<tr>
<td>Email Log Search</td>
<td>An email log search is performed for logs from 2020/09/28 09:00 UTC to 2020/09/28 09:09 UTC with a sender of megan, a recipient of megan, and an email message subject of megan_email_search_email, and an email log search email recipient ip ([1])</td>
<td>Megan Roddie</td>
<td>Sep 27, 2020, 2:56:59 PM CDT</td>
<td>99.125...</td>
</tr>
<tr>
<td>Email Log Search</td>
<td>An email log search is performed for logs from 2020/09/28 09:00 UTC to 2020/09/28 09:09 UTC with a sender of megan, a recipient of megan, and an email message subject of megan_email_search_email, and an email log search email recipient ip ([1])</td>
<td>Megan Roddie</td>
<td>Sep 27, 2020, 2:56:48 PM CDT</td>
<td>99.125...</td>
</tr>
</tbody>
</table>

Figure 11. Google Workspace Audit Logs - Admin

Megan Roddie, megansroddie@gmail.com
An example of the post-compromise evidence in these logs is the attacker changing recovery emails and two-step verification (2SV) settings on the victim account from Hong Kong, as seen in Figure 10. This can be leveraged for persistence so that even if the user changes the password, the recovery settings chosen by the attacker will allow them to recover access. Additionally, setting up two-step verification makes it harder for the user to regain access to their account with administrative changes. Furthermore, admin audit logs can provide context as to what kind of data was viewed during the compromise. An example of this can be seen in Figure 11-13, which show what it would look like if the attacker was searching through email logs, potentially indicating a leak of sensitive data and reconnaissance attempts. Although this activity should be suspicious based on the affected user and timeline, the geolocation context highlights it more immediately for responders.

The final section to highlight is the consolidated timeline, labeled “All” in the spreadsheet. For responders, the lack of a “single pane of glass” view is unfortunate since reviewing all activities that occurred across the environment can give a better picture of the attacker’s actions. With the current admin portal, the analyst would have to make a note of each log of interest and cross-reference between different pages or download each set of logs and try to merge the content manually. Therefore, the output of all gathered logs on a single page that can then be filtered and sorted through as needed provides

Megan Roddie, megansroddie@gmail.com
functionality to the incident responder not offered by Google. Figure 14 shows a view of the consolidated logs.

Figure 14. Tool Output - All Logs

The view of the activity details (stored in the “param” fields) is not easily readable in its current state. However, output still provides the intended purpose of allowing an incident responder to view, sort, and filter all logs as opposed to reviewing one application’s logs at a time in the browser.

3.4. Additional Findings

Unfortunately, some of the intended goals of this tool were unable to be met due to limitations of the Google API. Specifically, there is no API that can be used to gather email logs and activity. This presents a massive gap in knowledge for incident responders as the mail application is a significant aspect of Google Workspace that many companies leverage. It also is a common infection vector for business email compromises, which often use phishing emails. In order to analyze email logs, responders are still required to run searches via the admin portal.

Furthermore, the log retention limitations still apply when gathering data via the API, so this does not resolve that issue. If malicious activity occurred outside the retention period associated with a business’ subscription, there would be no evidence of it in the logs. To avoid this limitation, companies must backup their logs to a security information and event management (SIEM) or backup server regularly.

Megan Roddie, megansroddie@gmail.com
4. Recommendations and Implications

The custom tool presented in this paper shows some promising capabilities that could be of value to incident responders. While some aspects need to be improved on in order to provide the level of value intended, it provides a good starting point for providing incident responders with better data related to Google Workspace activity.

4.1. Recommendations for Practice

The hypothesis that the tool would provide better support to incident responders was partially correct in that some of additional context and alternate views were useful, but the extent of the additional data obtained was not as thorough as expected. Part of the intent of this project was to see whether the API provided more data than what was available to an admin via the Reports section of the Google Workspace admin portal. Upon further investigation, it appears the data provided via API is almost identical. What additional data was found would not be relevant to an investigator and was removed by the script as part of the data processing. Although this aspect of the hypothesis was not correct, the additional context provided by geolocation enrichment, along with the consolidated view of all the logs, did have the intended effect and proves the tool has some benefits over the built-in logging.

4.2. Implications for Future Research

To increase the value of the tool in a way that makes it worth using over reviewing the logs in the admin panel, there is some additional functionality that should be implemented.

First, this tool currently grabs login, Drive, user, and admin audit logs. However, there are additional audit logs available that could provide additional context, such as token, Chat, Calendar, LDAP, and other logs. Functions should be added for each of these log types in order to provide as much data as is available in the Google Admin portal.

Second, additional research should be done to identify whether there is any way to include email logs in the output. It was determined this is not available via the API, but
there may be methods such as merging a manual download of logs with the tool’s output so that the details can be found in the full timeline.

Lastly, the output can be improved upon to make it easier for the incident responder to review. The content of the “param” fields in the current output would be better viewed if the dictionary keys in each column were converted to the column name with the dictionary values remaining as the column values.

Outside of improvement to this custom tool, there seems to be a need for research related to exporting logs to a backup source, such as a SIEM, to avoid losing relevant logs due to short retention periods. While it appears many SIEM and logging companies have proprietary solutions for this, there are limited resources for how this can be done with open source tools for companies that don’t have the budget for a SIEM.

5. Conclusion

When performing incident response, it is important responders have the tools to efficiently investigate. In the case of Google Workspace, minimal tools exist to expand on the capabilities built-in to the admin portal. The tool developed during the course of the research aims to improve incident responders’ ability to handle an incident. In its initial state, it successfully provided additional context that could be used to highlight malicious activity. Additionally, it provided a single pane view not currently offered to Google Workspace admins, allowing for a bigger picture during analysis. While much work is still required to provide the intended value, a starting point is given with this initial development. With further development, this tool could allow for significant capabilities not provided by Google directly.

The tool developed as part of this research has been made publicly available via GitHub at the following link: https://github.com/megan201296/gsuite-dfir.

Megan Roddie, megansroddie@gmail.com
References


Megan Roddie, megansroddie@gmail.com