Digital Forensics of Data Theft on the Google Cloud Platform

TJEERD SLOKKER | FRANK WIERSMAN
SUPERVISOR: KORSTIAAN STAM

Monday February 3th
### Introduction

<table>
<thead>
<tr>
<th>Initial Access</th>
<th>Persistence</th>
<th>Privilege Escalation</th>
<th>Defense Evasion</th>
<th>Credential Access</th>
<th>Discovery</th>
<th>Collection</th>
<th>Exfiltration</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exploit Public-Facing Application</td>
<td>Account Manipulation</td>
<td>Valid Accounts</td>
<td>Redundant Access</td>
<td>Account Manipulation</td>
<td>Cloud Service Dashboard</td>
<td>Data from Cloud Storage Object</td>
<td>Transfer Date to Cloud Account</td>
<td>Resource Hijacking</td>
</tr>
<tr>
<td>Trusted Relationship</td>
<td>Create Account</td>
<td></td>
<td>Revert Cloud Instance</td>
<td>Cloud Instance Metadata API</td>
<td>Cloud Service Discovery</td>
<td>Data from Information Repositories</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid Accounts</td>
<td>Implant Container Image</td>
<td>Unused/Unsupported Cloud Regions</td>
<td>Credentials in Files</td>
<td>Network Service Scanning</td>
<td>Network Share Discovery</td>
<td>Data Staged</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Redundant Access</td>
<td>Valid Accounts</td>
<td></td>
<td></td>
<td>Remote System Discovery</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>System Information Discovery</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>System Network Connections Discovery</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**MITRE ATT&CK Matrix**
Research questions

What design, utilizing exclusively GCP native tooling, is required to establish digital forensic readiness on the Google Cloud Platform to investigate the Data from Cloud Storage Object and Data from Local System techniques from the MITRE ATT&CK Matrix?

1. What evidence needs to be acquired for investigation on the Data from Cloud Storage Object and Data from Local System techniques?
2. What are the sources for the evidence using exclusively GCP native tooling?
3. What evidence can be acquired with different GCP configurations?
Related work

• Haag, Leuenberger and van Ginkel, described the basics of digital forensics

• Zawoad and Hasan, proposed a log management solution

• Baryamureeba and Tushabe, defined the Abstract Digital Forensics Model (ADFM)
# Evidence

## Data from Cloud Storage Object
- IP addresses
- Usernames
- Time of access
- What is accessed
- What operations
- Authentication attempts

## Data from Local System
- + Network connections
- + Temp folders
- + Caches
- + Recycle bin
- + OS Event logs
Sources for evidence

Storage locations
- BigQuery (data warehouse)
- Google Cloud Storage bucket

- Disk Forensics
  - Snapshots

- Live (OS) Forensics
  - Logs
    - Virtual Private Cloud Network
    - Data Access
    - Identity Access Management
    - Admin Activity
Methodology

• Forensic readiness

• Experiments
  • Data exfiltration from a virtual machine
  • Privilege escalation on a storage bucket
  • Integrity on storage location
Test environment

- VPC Flow logs
- Data Access Logs
- IAM Logs
- Admin Activity Logs

Diagram:
- Virtual Machine
  - Stackdriver agent
- Stackdriver
  - GCS Bucket
  - BigQuery
- Snapshot
  - Plaso
  - Splunk
Experiment I – Data exfiltration from a VM

.pdf  .xls  .xlsx  .doc  .docx  .pptx
Experiment I – VM data exfiltration

Generated Logs
Experiment I – VM data exfiltration

Disk Forensic Investigation

- Firewall change
- Creation of temporary folder
- File copy operations
- Tracks of a temporary ftp connection file
- Deletion of the zip afterwards
## Experiment I – VM data exfiltration

### Evidence collection

<table>
<thead>
<tr>
<th>Potential evidence</th>
<th>Stackdriver logging-agent</th>
<th>Stackdriver logging-agent</th>
<th>Network flow logs</th>
<th>Network flow logs</th>
<th>Disk forensics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OFF</td>
<td>ON</td>
<td>OFF</td>
<td>ON</td>
<td></td>
</tr>
<tr>
<td>IP addresses</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Usernames</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Time of access</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>What is accessed</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>What file operations</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Authentication attempts</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Network connections</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Temporary folders</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Caches</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Recycle bin</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>OS event logs</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Yes** = did provide evidence  
**No** = did not provide evidence
Experiment II – Storage Bucket Privilege escalation

UNAUTHENTICATED ACCESS ALLOWED: data-collection-simulation
- VULNERABLE TO PRIVILEGE ESCALATION (storage.buckets.setIamPolicy)
- ALL PERMISSIONS:
  [ "storage.buckets.getIamPolicy", "storage.buckets.setIamPolicy" ]
Experiment II – Storage Bucket Privilege escalation

<table>
<thead>
<tr>
<th>Type</th>
<th>Members</th>
<th>Role(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><a href="mailto:448704278198@cloudbuild.gserviceaccount.com">448704278198@cloudbuild.gserviceaccount.com</a></td>
<td>Cloud Build Service Account - Inherited</td>
</tr>
<tr>
<td>allAuthenticatedUsers</td>
<td></td>
<td>Storage Admin</td>
</tr>
<tr>
<td>allUsers</td>
<td></td>
<td>Storage Admin</td>
</tr>
<tr>
<td><a href="mailto:frankwiersma1947@gmail.com">frankwiersma1947@gmail.com</a></td>
<td></td>
<td>Security Reviewer - Inherited</td>
</tr>
<tr>
<td><a href="mailto:private-storage-access@public-cloud-forensics.iam.gserviceaccount.com">private-storage-access@public-cloud-forensics.iam.gserviceaccount.com</a></td>
<td></td>
<td>Storage Admin - Inherited</td>
</tr>
</tbody>
</table>
Experiment II – Storage Bucket Privilege escalation Success!
### Experiment II – Storage Bucket Privilege escalation

**Evidence collection**

<table>
<thead>
<tr>
<th>Potential evidence</th>
<th>GCS data access audit logs OFF</th>
<th>GCS data access audit logs ON</th>
<th>IAM audit logs OFF</th>
<th>IAM audit logs ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP addresses</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Usernames</td>
<td>No</td>
<td>Yes, if authenticated</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Time of access</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>What is accessed</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>What file operations</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Authentication attempts</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Unusual API requests</td>
<td>No</td>
<td>Partially</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

**Yes** = did provide evidence  
**No** = did not provide evidence
## Experiment III – Integrity

<table>
<thead>
<tr>
<th>Storage Location</th>
<th>Mutation prevention</th>
<th>Security options evidence</th>
<th>Retrievability evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>BigQuery</td>
<td>Permissions</td>
<td>-</td>
<td>Querying Downloading</td>
</tr>
<tr>
<td>Google Cloud Storage bucket</td>
<td>Permissions</td>
<td>Customer-managed key</td>
<td>Downloading</td>
</tr>
</tbody>
</table>
What design, utilizing exclusively GCP native tooling, is required to establish digital forensic readiness on the Google Cloud Platform, to investigate the Data from Cloud Storage Object and Data from Local System techniques from the MITRE ATT&CK Matrix?
Conclusion

• GCP native tooling not sufficient for live forensics
• Combine logs & disk forensics

Key findings:
• Stackdriver agent collects minimal OS event logs
• No traces of the intentional privilege escalation
• Hard to check integrity during the preservation and collection phase
• Disk forensics provided the most evidence
Future work

- More tests within MITRE matrix
- Try to get Google’s help with evidence collection
- Research on Chain of Custody
- Third party agents