AI-Assisted Insider Threat Detection

Correlating Gemini Activity with Drive Exfiltration in Google Workspace

haasonsaas

BSides

October 16, 2025

Overview

The Insider Threat Landscape

- Traditional DLP focuses on **content inspection**
- Misses **behavioral patterns** that indicate intent
- New attack surface: **LLM-assisted reconnaissance**
- Insiders now use AI to rapidly understand sensitive documents

The New TTP

Use Gemini to analyze files $\rightarrow$ Immediately exfiltrate them

Why This Matters

Traditional Exfil:

1. Manually read documents
2. Identify sensitive content
3. Extract/share

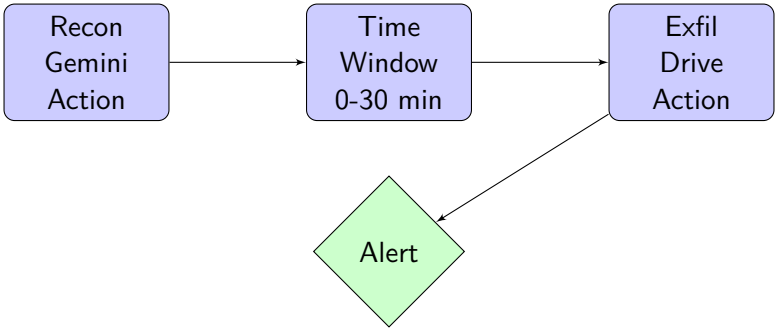
Time-consuming, leaves traces

LLM-Assisted Exfil:

1. Ask Gemini "summarize this"
2. Instantly understand value
3. Immediately exfiltrate

Fast, efficient, higher-value targets

Detection Logic



Correlation Key: Actor + Temporal Proximity

Recon Signals (Gemini Events)

Data Source: Admin SDK Reports API
Application: `gemini_in_workspace_apps`
Event: `feature_utilization`

High-Signal Actions:

- `ask_about_this_file` - Direct file query
- `summarize_file` - File summarization
- `analyze_documents` - Multi-file analysis
- `catch_me_up` - Bulk triage
- `report_unspecified_files` - Report generation

Available Since: 2025-06-20 (180-day retention)

Exfil Signals (Drive Events)

Data Source: Admin SDK Reports API
Application: `drive`

High-Risk Events:

- `change_visibility` - Made public/external
- `change_acl` - External principal added
- `export` - Export to PDF/DOCX/CSV
- `download` - File download
- `copy` - File duplication
- `add_to_folder` - Move to external folder

Key Parameters: `doc_id`, `visibility`, `new_value`, `old_value`

Detection Algorithm

```
for exfil_event in drive_events:
    for recon_event in gemini_events:
        if exfil_event.actor == recon_event.actor:
            delta = exfil_event.time - recon_event.time

            if 0 <= delta <= 30_minutes:
                severity = calculate_severity(
                    delta,
                    exfil_event.type,
                    exfil_event.visibility
                )

                emit_finding(severity, recon_event, exfil_event)
```

Severity Rubric

Severity	Criteria	Response
High	External share/export $\leq$ 10 min	Page on-call
Medium	External share/export 10-30 min	Next-day review
Low	Any permission change within 30 min	Log for analysis

Severity Overrides:

- Actor in high-risk OU (Exec, Finance, R&D) $\rightarrow +1$ level
- File labeled confidential/restricted $\rightarrow +1$ level

Architecture

1. **Authentication:** Service account with domain-wide delegation
2. **Data Collection:** Fetch Gemini + Drive events via Admin SDK
3. **Correlation Engine:** Temporal matching by actor
4. **Scoring:** Apply severity rules and suppressions
5. **Output:** JSON findings to SIEM/alerting

Deployment Options:

- Cron job (every 10 minutes)
- Systemd timer
- Cloud Function / Lambda

Example Finding

```
{
  "severity": "high",
  "actor": "john.doe@company.com",
  "exfil_event": "change_visibility",
  "exfil_time": "2025-01-15T14:23:45-08:00",
  "doc_id": "1abc123def456",
  "doc_title": "Q4 Financial Projections.xlsx",
  "recon_action": "summarize_file",
  "recon_time": "2025-01-15T14:18:12-08:00",
  "delta_minutes": 5.55,
  "visibility": "people_with_link",
  "reason": "External share within 10min of recon"
}
```

Tuning & Suppression

False Positive Reduction:

- Allowlist trusted external domains (partners)
- Suppress security/IT OUs investigating files
- Exclude service accounts
- Adjust time windows based on your environment

Calibration Process:

1. Week 1: Observation mode (no alerts)
2. Week 2: Tune suppressions, enable high severity
3. Week 3-4: Refinement
4. Month 2+: Ongoing review

Why This Works

Behavioral Sequence Detection:

- **Intent:** Gemini logs reveal *what* the user wanted to understand
- **Action:** Drive logs reveal *what* they did with it
- **Correlation:** Temporal proximity reveals insider TTP

No Content Inspection Required:

- Privacy-preserving detection
- No prompt/response content visible
- Metadata-only telemetry

Detection Advantages

High-Fidelity Signal:

- Precise timestamps
- Actor email
- File IDs
- Action types

Google-Recommended:

- Official audit logs
- Designed for security telemetry
- 180-day retention
- No extra cost

Common Patterns Detected

1. **Rapid Reconnaissance:** User asks Gemini about 10+ files, then downloads 3 with highest value
2. **Pre-Resignation Exfil:** Employee uses Gemini to triage entire project folder, then exports key documents
3. **Competitive Intelligence:** Sales rep summarizes customer contracts, immediately shares with personal email
4. **IP Theft:** Engineer asks Gemini to analyze codebase documentation, then copies to external Git repo

Deployment Checklist

1. Create service account with domain-wide delegation
2. Grant `admin.reports.audit.readonly` scope
3. Configure suppressions for your environment
4. Test with known benign patterns
5. Start with observation mode (no alerts)
6. Gradually enable alerting by severity
7. Integrate with SIEM/SOAR

Security Best Practices:

- Store service account key in secrets manager
- Rotate keys every 90 days
- Monitor the monitor (alert if detector stops)

Metrics to Track

Detection Quality:

- True positive rate
- False positive rate
- Time to detection
- Time to response

Operational:

- Alert volume (findings/day)
- API success rate
- Coverage (Gemini users / total users)
- Alert fatigue score

Limitations & Future Work

Current Limitations:

- Gemini events only since 2025-06-20
- 180-day retention window
- No file content analysis
- Requires Google Workspace Enterprise

Future Extensions:

- ML-based user risk scoring
- Automated response (revoke links, notify owner)
- Integration with HR data (resignations, PIPs)
- Cross-correlation with physical security (badge logs)

Key Takeaways

1. **New Attack Surface:** LLMs enable rapid, efficient insider reconnaissance
2. **Behavioral Detection:** Focus on sequences, not content
3. **High-Signal:** Temporal correlation of recon + exfil is highly indicative
4. **Privacy-Preserving:** No content inspection required
5. **Actionable:** Deploy today with Google Admin SDK

The Bottom Line

AI-assisted insider threats require AI-aware detection. This technique provides high-fidelity telemetry without compromising user privacy.

Resources

GitHub Repository:
`https://github.com/haasonsaas/gemini-exfil-detector`

Google Documentation:

- Admin SDK Reports API
- Gemini in Workspace Apps Events
- Drive Audit Events

Contact:
GitHub: @haasonsaas

Questions?