Log Management and Metadata

Georgia Tech

Log Management and Metadata

Log Management Introduction

- Security Information and Event Management

Network Metadata

- HTTP, DNS, Wireless Network Logs, and Netflow

Final Thoughts

- Example Exercise

Georgia Tech

Log Management

Georgia Tech

Logs

Systems generate logs:

- Debug logs
- Application logs

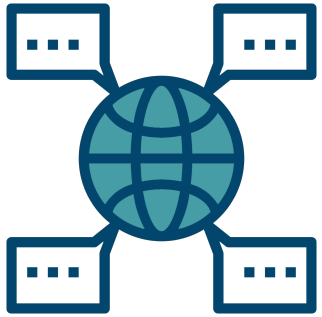
Useful for:

- Diagnosing issues
- Usage statistics
- Performance statistics
- **Security**

Georgia Tech

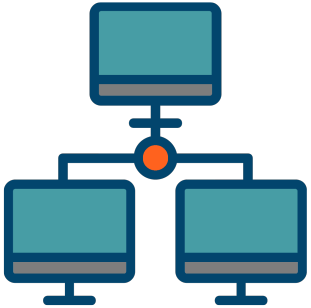
What types of logs are important for a security team?

Georgia Tech



Network

- Authentication logs
- Firewall logs
- Intrusion Detection and Prevention System logs
- Metadata



Host (Client and Server)

- Authentication logs
- Firewall logs
- Anti-malware and HIPS
- Application logs

Log management refers to a system that provides these functionalities:

Log Ingestion Log Augmentation Log Storage Log Searching

Security Information and Event Management (SIEM)

Georgia Tech

SIEMs usually specialize in these functionalities:

Log Aggregation and Retention

Correlation and Analysis

Alerting and Dashboards

Security Information and Event Management (SIEM)

Georgia Tech

Examples of Products

- Splunk
- Elastic (ELK)
- Exabeam
- SumoLogic
- Graylog

Georgia Tech

Network Metadata

Georgia Tech

What is Metadata?

- Information related to **connections that occurred on the network**.
- Consists of **useful properties about the network transaction**
- Allows for the analysis of the transaction **without having to store the entire contents of the data.**
  - Storing the entire content of the data would get **expensive.**

Georgia Tech

Types of Metadata?

- HTTP
- DNS
- SMTP
- TLS
- Netflow
- SSH
- Wireless access point logs
  - Detailed location at every moment

Georgia Tech

Extremely useful for **Forensics**

Extremely **Sensitive**

Metadata

Georgia Tech

HTTP Metadata

```
{
    "proto": "TCP",
    "vlan": [
        2200
    ],
    "timestamp": "2020-04-30T18:08:45.852437+0000",
    "tx_id": 0,
    "src_ip": "143.215.XX.XX",
    "event_type": "http",
    "flow_id": 327429317788103,
    "http": {
        "url": "/autodiscover/autodiscover.xml",
        "status": 302,
        "hostname": "autodiscover.gatech.edu",
        "http_user_agent": "MacOutlook/16.33.0.200113 (Intelx64 Mac OS X Version 10.14.6 (Build
18G103))",
        "length": 0,
        "http_method": "GET",
        "protocol": "HTTP/1.1",
        "redirect": "https://autodiscover-s.outlook.com/autodiscover/autodiscover.xml"
    },
    "src_port": 53128,
    "dest_ip": "52.96.88.120",
    "dest_port": 80
}
```


HTTP Metadata Example

Georgia Tech

What is it used for?

- Who **visited phishing pages**?
- **POST vs GET** – did the user submit credentials?

Georgia Tech

DNS Metadata

Georgia Tech

What is it used for?

Helps inform decision about blocking domains

- **How many users** have queried the domain?
- Is the **domain legitimate**?
- **New domains** can be indicative of **malware**

Georgia Tech

What is it used for?

Passive DNS is extremely useful

- What **domains has this IP had**?
- Not all domains have **reverse records**
  - **Forward**: domain -> IP
  - **Reverse**: IP -> domain
- **Passively records** queries and answers
  - Can **search forward and reverse**

Georgia Tech

```
{
    "vlan": [
        2198
    ],
    "dns": {
        "rrtype": "A",
        "id": 17437,
        "type": "answer",
        "rcode": "NOERROR",
        "answers": [
            {
                "rrname": "www.gatech.edu",
                "rdata": "d2mlopss7pe3bd.cloudfront.net",
                "rrtype": "CNAME",
                "ttl": 60
            },
            {
                "rrname": "d2mlopss7pe3bd.cloudfront.net",
                "rdata": "54.230.138.67",
                "rrtype": "A",
                "ttl": 4
            },
            {
                "rrname": "d2mlopss7pe3bd.cloudfront.net",
                "rdata": "54.230.138.77",
                "rrtype": "A",
                "ttl": 4
            },
            {
                "rrname": "d2mlopss7pe3bd.cloudfront.net",
                "rdata": "54.230.138.79",
                "rrtype": "A",
                "ttl": 4
            },
            {
                "rrname": "d2mlopss7pe3bd.cloudfront.net",
                "rdata": "54.230.138.53",
                "rrtype": "A",
                "ttl": 4
            }
        ],
        "qr": true,
        "flags": "8580",
        "rrname": "www.gatech.edu",
        "rd": true,
        "ra": true,
        "aa": true
    },
    "in_iface": "ens6",
    "src_ip": "10.2.XX.XX",
    "dest_ip": "130.207.244.244",
    "host": "detective6",
    "proto": "UDP",
    "dest_port": 53,
    "event_type": "dns",
    "timestamp": "2020-05-03T17:08:11.326476+0000",
    "src_port": 18927
}
```


DNS Metadata Example

Georgia Tech

Wireless Network Logs and Netflow

Georgia Tech

Wireless network logs can be useful for:

Diagnosing Issues

Improving Service

Stolen Property

Missing Persons

Think of **Netflow** like a **phone bill for network traffic.**

Layers 3 / 4 only.

- **Protocol**
- **Source Address**
- **Dest Address**
- **Source Port**
- **Dest Port**
- **All TCP flags used**
- **Number of bytes transferred**
- **Number of packets transferred**
- **Time of “connection”**
- **Potentially other stuff**

What is it used for?

- **“Call log”** for the network
- **Verifying connections occurred** (or didn’t occur)
- Has an **IP communicated with campus**?
- **How much data was transferred** between IPs?

Georgia Tech

Final Thoughts

Georgia Tech

What is GT doing about it?

- We **capture and store it.**
- We only retain metadata for the **minimum necessary time**.
- We created a **data privacy policy.**
- **Used for:**
  - **Forensics** in incident response
  - **Investigate alerts**

Georgia Tech

Example Exercise

You are a major university incident response team.

You receive notification from the financial aid department that several students are claiming that their accounts should have been payed in full, but now show a balance and they suspect something malicious has occurred.

Answer the following question:

- **What log data and metadata would you want to collect and preserve to determine if something malicious happened?**

You've confirmed that the students in question did initially receive they student loan payment, but then it appears that the students logged into their Student Information System and requested a refund

You interview all these students that reported this issue and each of them claim that they never requested the refund.

Answer the following question:

- **What log data and metadata would you want to review in order to corroborate the students' stories?**

You've determined that each of the logins where a refund was requested originated from a single IP address in South America.

Answer the following question:

- **What log data and metadata would you want in order to analyze the full extent of the issue?**

You suspect the hacker used a phishing attack to target these students and steal their credentials.

Answer the following questions:

- **What log data and metadata would you analyze in order to determine who else may have been compromised by this phishing attack?**
- **You've determined that there were approximately 500 other users that were compromised, how will you handle their accounts?**

Final Questions:

- How would you recover from this incident?
- What solutions would you propose to help ensure this type of incident doesn't occur again?
- Should the University cover the student's stolen tuition money?
- Should the University provide identity theft protection?