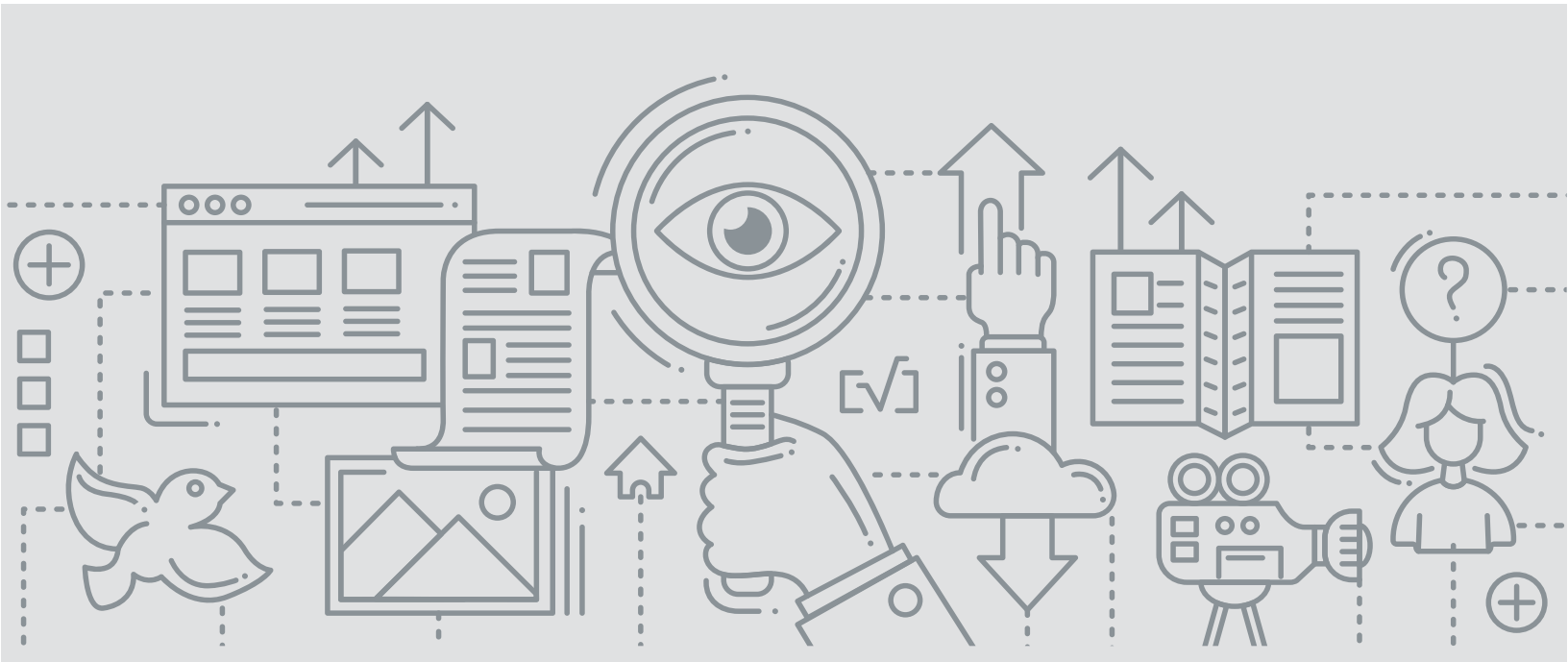
Redefining Security Analytics with Chronicle



DISCLAIMER: This whitepaper applies to Google Cloud products described at cloud.google.com. The content contained herein represents the status quo as of the time it was written. Google's security policies and systems may change going forward, as we continually improve protection for our customers.

Table of contents

Chronicle is a global security telemetry platform for detection, investigation and threat hunting within your enterprise network. Chronicle makes security analytics instant, easy, and cost-effective.

The cyber security world has changed

The threats and technology environments around us have changed radically. Security professionals lament that they must manage threats that originated in the 1980s, which means that old threats never leave; new threats simply pile on.

It is clear that the environments an organization must secure and monitor have also dramatically expanded. Much of this change is additive - in addition to mainframes and Windows servers, there are cloud platforms, mobile, and SaaS. As a result, there is more of everything - both threats and assets - to protect. As such, the tools that help detect threats, enable incident responders, and perform security monitoring must change as well.

SIEM reimagined: past to future

How does security monitoring and threat detection need to change? How can it accommodate the older and the newer systems, while the new (e.g., containers, cloud, and IoT) continue to grow?

Historically, most analysis and complex detection capabilities have been bundled together into a technology called a "SIEM." In time, the SIEM gathered more capabilities ranging from compliance reporting to security workflow management to machine learning algorithms for threat detection.

What would SIEM look like if it were invented today?

> First, some things never change. For example, every organization needs to collect security information, analyze it to find non-obvious problems (detection), investigate, look for suspicious activity beyond alerts (threat hunting), initiate and support a response, and prove that the security architecture is working (reporting).
>
> Second, some things change radically or have no historical equivalent. Threat hunting, for example, was not a core use case for a SIEM in 1999. Many IT environments, such as containers and microservices, didn't exist years ago. Practices have also changed - DevOps is different from a planned waterfall approach and it affected how system security is deployed and monitored. Cloud environments that auto-scale and don't rely on IP addresses for system identification present a known trouble spot for conventional security tools.
>
> Third, some things have grown in size and variety. Big data in 2012 meant "terabytes" but today equates to petabytes or more. There are more systems of different types producing telemetry just as there are more threats affecting both new and old systems. Finally, in this time of digital transformation, more organizations are using IT for more things, hence IT overall has grown dramatically as more areas of business are being digitized.

In light of this, it is easy to bet that if a new security analytics technology were born in 2020, as replacement for SIEM, it would be cloud-based (hence essentially maintenance-free), blazing fast even on petabytes of data, able to collect data without onerous (and risky) intake filtering, be effective for detection, investigation and hunting, use threat intelligence heavily and would present clean and enriched data to the users in the form of timelines and stories. On the other hand, it will focus less on compliance reports and auditor requirements.

In short, a modern security analytics solution born today would be:

- SaaS-based to avoid the burden of deployment, performance tuning and expansions
- Scalable for the world of petabytes - both physically and economically
- Providing data retention suitable for detecting and hunting modern stealthy threats
- Fast to search in order to accommodate modern hunting use cases; SIEM at Google speed
- Offering high quality data enriched from many sources automatically thus saving clients from most data enrichment tasks
- Natively integrating threat intelligence for detection, alert triage, hunting, investigation
- Enabling threat detection using multiple methods including rules and algorithms
- Offering a pricing model that does not punish usage or expansion, and does not break the bank for most clients, large and small
- Simple to deploy, administer and operate

How to make sense of all the security data

Security analytics promises to help analysts make sense of the security data, to find useful signals in the noise before it's too late. For most organizations, however, creating an effective security analytics solution is an expensive and complex exercise in systems integration, with heavy IT operations support required simply to keep the system up and running as it grows. As CIOs migrate corporate IT to the cloud, CISOs roll out advanced threat protection such as EDR and network traffic analyzers to protect the pieces that remain under their control.

New SIEM requirements

While it sounds counter-intuitive that we need another type of a tool, the world has changed dramatically and many existing security tools did not evolve fast enough to stay relevant.

Today, organizations still operate legacy systems, have vast on-premise IT presence, but also large cloud presence, often across multiple cloud providers. The type of security telemetry they collect expands, and the volumes grow.

The requirements are:

- Scale
- Speed
- Simplicity
- Cost
- Cloud and on-prem

Chronicle, a global security telemetry platform designed to help those same analysts understand threats that exist in their own corporate networks. With Chronicle, analysts can hunt for threats and investigate petabytes of their own data to get answers in milliseconds. And all that without paying for data volumes.

Scale

Data volumes have increased dramatically over the last decade. Using a popular SIEM metric, events per second (EPS), a large environment 15 years ago may have had “thousands” of EPS. Today, it is not uncommon to see environments with hundreds of thousands and more events per second, pushing into high terabyte daily telemetry volumes.

Admittedly, increased focus on security monitoring is one of the primary reasons. We have more detection tools. Some of the newer tools, such as EDR, are also more verbose.

In addition, more digitization and IT infrastructure have pushed the log volumes way up. This includes both cloud and on-premise environments’ product logs and other types of security telemetry, from cloud VPC flow logs to the classic Windows event logs.

Now, apart from higher volumes of data, we also have longer retention periods of data, and not merely for compliance. Breaches may be discovered 200-300 days after the initial compromise and this calls for investigative data to be retained.

As a result, there are many orders of magnitude greater security telemetry that needs to be ingested and analyzed. This either breaks the legacy tools or breaks their economic model. In the latter case, the challenge is as severe. Gartner in 2009 said that “security is becoming a big data analytics problem” but today customers are the ones paying for it. Economic scaling is as important since no one wants to buy $10m worth of hardware to run their security data search tools.

Speed

How hard is it to search the entire internet in under a second? Well, it has not been difficult since the day Google launched back in 1999. Then why do people tolerate waiting for minutes if not hours while searching their security telemetry?

Today, organizations deal with too many alerts, and confirming alerts is anything but fast for many. As organizations increasingly shift focus to threat detection and response, one issue seems to get worse over time: alert triage. Prioritizing security alerts has been a critical function for security teams since the early 1990s — so why does it remain such a challenge decades later? Today we have more logs to search and more alerts to confirm.

Worse, many of the alerts are false positives, and the recent advances in machine learning for threat detection has, sadly, contributed to the problem. False alarms and ambiguous alerts all need investigating - and fast.

Finally, there is also a real need for performance around incident investigations which are too complex and slow today. Search EDR data from 50,000 machines?

Simplicity

It is often said that SMBs need simple tools for everything in security. But guess who else does? Large enterprises! Extreme staff shortages, and in fact worse talent shortages in security has led to many security analysts and even threat hunters being hired without deep experience. They all need to simple tools to detect, confirm and investigate without stress - and without a 3 day class on a new search language.

Furthermore, using threat intelligence can be simple - but it almost never is. Threat intelligence feeds are supposed to add more context, but are often too noisy or redundant and cause more static than they eliminate. Over time, the indicators in those embedded feeds and signals may change. For example, a domain first seen a year earlier and judged as “good” in the threat feeds may begin hosting malware; the indicator turns “bad” in the feed.

Thus there is a need to automatically and instantly re-calculate any customer activity to that now-bad indicator and alerts analysts about all machines that have ever communicated with this domain. Note that the focus on simple and automatic here.

Similarly, having the threat data, security telemetry enriched and connected makes alert triage much simpler. Given such a system, alert triage should not require deep knowledge of threats and the environment.

Cost

Does a tool scale? This is a useful question to answer and security tools should cover the data needs of an organization. However, does the tool scale without an exponential increase in hardware cost and commensurate increase in complexity? Or, if deployed in public cloud, do they scale without you being stuck with a 7 digit cloud provider bill for collected and stored data, that keeps growing and growing? In fact, even some open source tools with a license cost of free has led to huge cloud bills if used for security telemetry collection without adequate planning.

Security analytics tool should not be priced based on collected data value. This license models create a disincentive to send all the data, and causes the customer to make decisions not based on security

Cost is ultimately is not only about low price, but about economic scalability and predictability - can the tool grow with you without breaking the bank. Predictable pricing is often more important than low pricing for some scenarios.

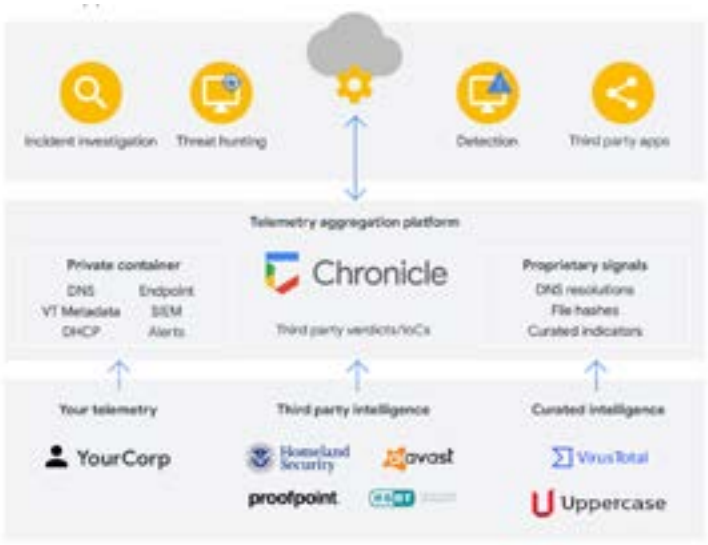
Cloud and on-premises

It is very clear that a modern security analytics tool should use the cloud backend. But it cannot be only supporting the assets deployed in the cloud since this is simply not reality at most organizations today. Hence the ideal solution would be deployed in the cloud, but be able to analyze the telemetry from both cloud and on-premise sources, modern and legacy tools, systems and applications for a wide range of security use cases.

Now, it is worth clarifying that a cloud native toolset has plenty of advantages over something built for the on-premise world and then “lifted and shifted” to public IaaS. Auto-scaling, elasticity and deployment automation of the cloud native tools cannot be beaten legacy software deployed in the cloud.

Chronicle: A Different Approach to Security Analytics

Chronicle is a cloud service, built as a specialized layer on top of core Google infrastructure, designed so that enterprises can privately retain, analyze and search the massive amounts of security and network telemetry they generate today. Chronicle normalizes, indexes, correlates, and analyzes the data -- against itself and against third party and curated threat signals -- to provide instant analysis and context
regarding any risky activity. We can drill down from this description into some of the platform’s key functions:



A) Data Ingestion

Chronicle can ingest a variety of telemetry types, through a variety of methods:

- The most common is the Chronicle Forwarder, a lightweight software component, deployed in the customer's network, that supports syslog, packet capture, and existing log management / SIEM tools. The Forwarder can be installed on Windows platforms and also as a container on Linux platforms.

- Chronicle also offers an ingestion API that enables logs to be sent directly to the Chronicle platform, eliminating the need for additional hardware or software in customer environments. MSSPs and technology partners can leverage the Chronicle ingestion APIs to forward raw logs as well as structured logs that adhere to the Chronicle normalized format, directly to the Chronicle data pipeline. The Chronicle Ingestion API is a RESTful API with a JSON payload and API keys are used to authenticate calls.

- Additionally, Chronicle can also pull telemetry from other cloud services, including data not originating in the cloud. For example, some EDR solutions push endpoint logs to an Amazon S3 bucket, and Chronicle can be configured to ingest logs directly from that location. In contrast, Carbon Black's EDR uses an event forwarder to ingest telemetry directly to Chronicle . Simply put, there are many ways for customers to upload their telemetry.

- Chronicle also integrates with 3rd party cloud APIs to facilitate ingestion of logs. This includes sources like Office 365 and Azure AD.

Regardless of the ingestion path, a key architectural goal during ingestion is high throughput and this is partly achieved by first writing ingested logs in the format received to disk and processing (e.g., normalizing, indexing) them thereafter. This approach also has the advantage of raw log access for correction of any parsing or other errors at the processing stage.

A critical advantage of a cloud-native security analytics is that logs can be collected and then parsed (or re-parsed once improved parsers become available) in the cloud. Those who operated traditional SIEM tools have long lamented about the collector updates and changes. With Chronicle, the raw logs are collected and retained and then can be "magically" parsed, normalized and enriched as needed.

Finally, the Chronicle security analytics platform supports more than logs - and it won't cost you any extra. EDR data, network traffic captures (such as those from Zeek and other network metadata tools) can be collected and retained at no extra cost beyond the initial investment. Moreover, such endpoint and network telemetry (combined with logs) creates an effective triad of security visibility for an organization.

B) Data Analysis

The analytical capabilities of Chronicle are delivered to security professionals as a simple, browser-based application. Many of these capabilities are also accessible programmatically via read APIs and can be triggered from other security tools. At its core, the purpose of Chronicle is to give analysts a way, when they see a potential threat, to determine what it is, what it's doing, whether it matters, and how best to respond. Next, let's understand Chronicle's analytical capabilities using a real life example:

What is the threat? Chronicle combines the scale of its core Google infrastructure backend with unique data enrichment to surface all IoC matches automatically and continuously.

For example, if a threat feed just informed Chronicle about a new APT network domain, the Chronicle Enterprise Insights dashboard will instantly surface every hostname that accessed that domain going back a full year, regardless of the data volume.

Similarly, Chronicle continuously enriches the incoming event stream by correlating IPs to hostnames so that analysts have full information. It does so instantly; and without the need to write complex queries - or, in fact, without any user action whatsoever.

With other security analytics solutions, data is rarely if ever kept in a hot state for a full year due to storage and license costs. Even in cases where customers are willing to bear higher storage costs for fast access, retroactive IoC matches are manually initiated by an analyst and take an inordinate amount of time due to infrastructure (compute) costs.

Furthermore, these searches may simply yield IP address matches which in turn have to be manually correlated with DHCP data to uncover the hostname that leased that particular IP address at that time. Hence if the log data is stored for a year in raw format with no associated context data, it will become useless much sooner than a year. In short, even a common and seemingly simple security operations task of determining hostnames that have reached out to a malicious domain consumes significant time, infrastructure, and expertise (proprietary search language proficiency).

What is it doing? Following the above example, let's drill down into the IoC match for the US DHS (Dept. of Homeland Security) threat feed which is one of several threat intelligence sources embedded into the Chronicle platform. Once an asset match is found for a known malicious domain (in this case Todd Fields' PC), analysts generally want to understand what was happening on that asset around the time of access to look for corroborating signals that confirm whether the outreach led to compromise.

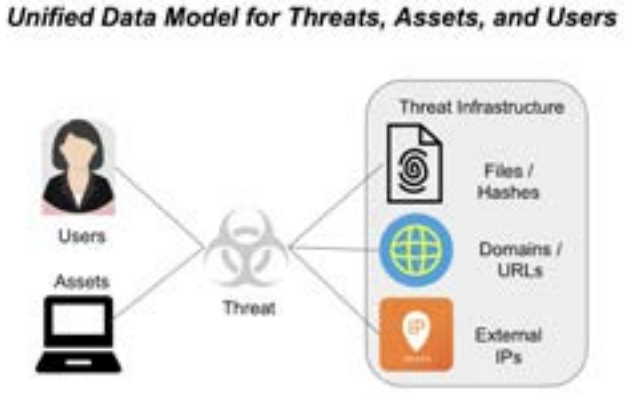
Common questions an analyst would want to answer include:

- What other domains did this asset access around the same time? Were any of those other domains malicious, rare or entirely new to my enterprise?
- What was the pattern or frequency of access to the domain in question as well as other rare domains. For example, was the outreach representative of beaconing behavior?
- Were any alerts fired around the same time for the given asset from Firewalls, CASB, EDR or other security solutions?
- Do endpoint events reveal activity that suggests the malicious domain access was followed by a successful compromise?

Modern risks lie at the intersection of targeted assets, users and threat campaign infrastructure (domains, URLs hosted on specific external IP addresses and malicious files deployed from there).

In Chronicle , curated views for internal assets and users as well as threat campaign infrastructure (domains, URLs, file hashes, external IPs) operate on a Unified Data Model which enables analysts to investigate threats quickly, intuitively, and without having to rely on a proprietary query language.

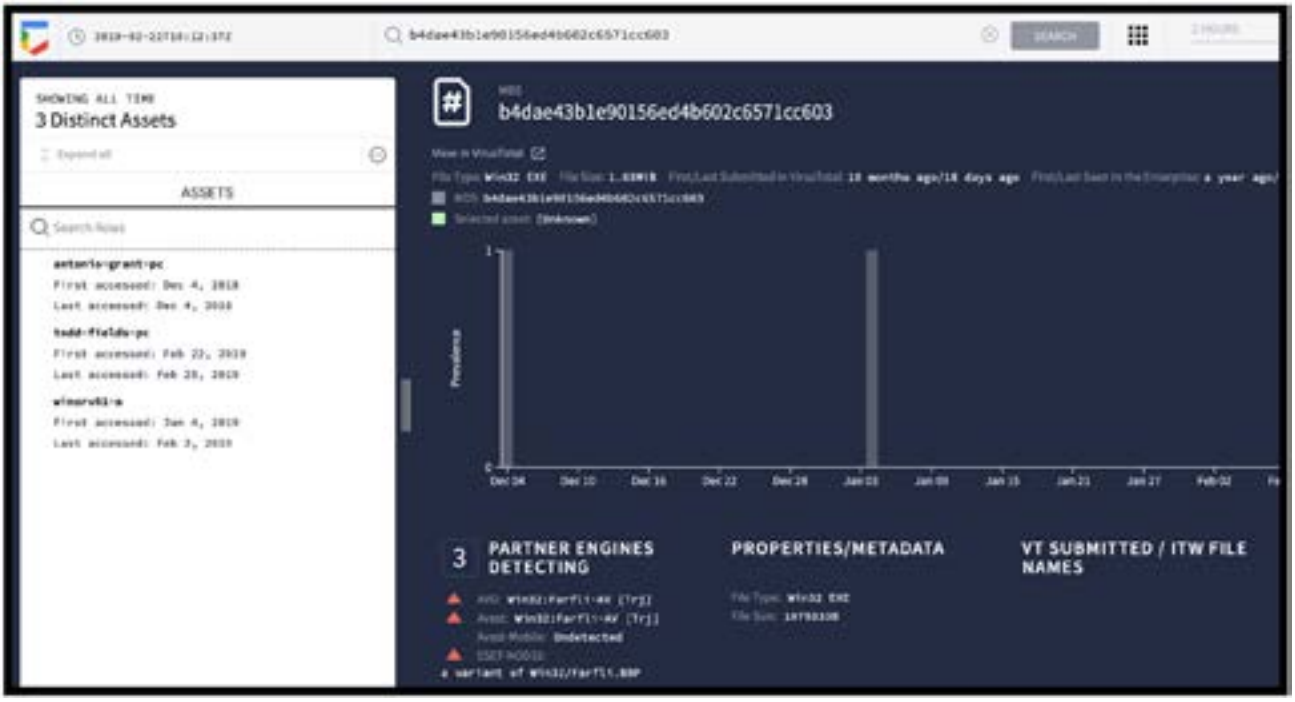
In this example, we drill down into the Asset View for the endpoint that accessed the APT domain and can quickly see a beaconing pattern to rare (low prevalence) domains graphically as well as a suspicious process in the timeline panel (ghost.exe) that is launched immediately after. Subsequently, a file (notepab.exe) is written to disk and a new registry run key is created to gain persistence across reboots. Chronicle uses stateful URLs so at any time an analyst can capture the exact set of filtering conditions and pass it on to the next analyst for continued triage across shifts or escalation tiers. Similarly, the timeline of events can be easily filtered on data model dimensions and attributes and the relevant set of events can be easily exported.



With traditional security analytics and logging platforms, painting this summary picture across network and endpoint events for a single asset can take days of complex queries that have to be created and run in sequence after which results have to be combined. In fact, it is very common that it will require access to multiple systems such as one for SIEM and another for EDR. Chronicle provides the triad of visibility - logs, network data, endpoint data - and enriches and correlates such multi-source data into a coherent story.

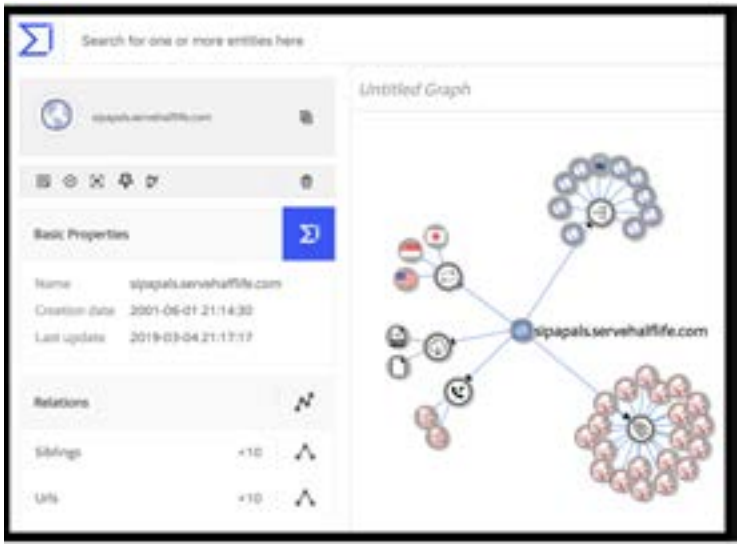
Does it matter? Security telemetry, and logs alone, rarely provides the full picture needed to hunt, investigate or detect threats. Context is critical to giving analysts the ability to prioritize real threats and dismiss false positives. In this example, we have an endpoint (todd-fields-pc) that did reach out to a known malicious domain and is likely compromised by a malicious file (ghose.exe) that was downloaded.

Each curated view provides relevant context and insights to aid the investigation or hunt. The Asset View shows insights about domains accessed by an asset that are rarely seen in the enterprise (low prevalence); domains that are new to the enterprise; alerts from other security tools; known vulnerabilities for the given asset and more. Selecting the malicious file in the timeline panel surfaces the hash and known verdicts from Avast's 400 million consumer AV endpoint agents, and ESET's AV results -- as well as embedded VirusTotal metadata. By clicking into the hash itself we pivot out from the Asset View and into a Hash View which tells us that two other assets also have a file with the same fingerprint, indicating a potentially wider compromise.



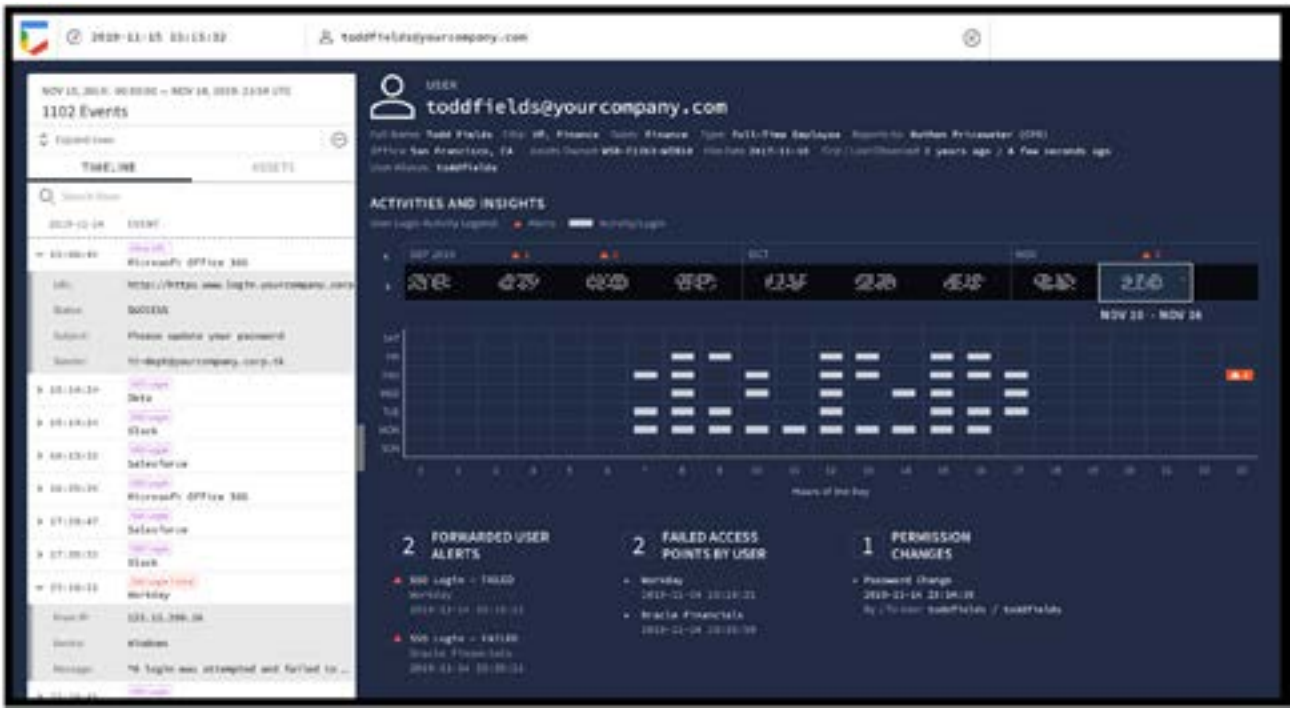
To evade detection, threat actors constantly change the domains, URLs, files and to a lesser extent - the IP addresses used in their campaigns. VirusTotal is the largest repository of malware samples with the unique ability to explore artifacts based on their relationships.

This is why VirusTotal is integrated into Chronicle not just to provide summary context in the Domain, URL, External IP and Hash Views but also to enable analysts to pivot into VirusTotal (contextually) and uncover related IoCs that may be part of the same campaign. For example, an analyst investigating the APT domain in this scenario within Chronicle's Domain View, can click into "see details in VirusTotal" in the VT Insights panel and start exploring related artifacts. Sure enough, there are several other sub-domains, URLs, and files that are likely tied to this campaign. The analyst can now start looking for the existence of variants of the initial threat that has been detection.



Analysts can also pivot to the Chronicle User View which provides context on who the user is (from directory sources like Active Directory) and whether their behavior is anomalous. User View operates on distinct data sources and events such as Windows (login events), Office 365, and Azure AD. In this example, we see that Todd Fields is a VP of Finance and that there are a few user centric security alerts associated with his account. The graph tells us that while Todd normally only logs into his account between 7am and 5pm, after the compromise of his laptop there has been anomalous account activity past 11pm.

The combination of his role, recent account related alerts from 3rd party SaaS applications (which may also come from a CASB), and anomalous access by time of day suggest Todd's account may have been compromised. The likelihood is only higher because we know Todd's endpoint has been compromised by an APT that successfully downloaded a malicious payload to his laptop



To summarize, in a matter of seconds, the analyst has seen a threat, understood its behavior, gained context and judged the severity, and generated a list of all affected machines that need remediation. None of this required a single query to be written, and all can be performed with a single console.

How to respond? Large organizations with mature SOCs have had documented playbooks for investigation and remediation of threats for years. SOAR tools (Security Orchestration, Automation and Response) or other orchestration technologies that automate these playbooks are now growing in adoption. Across the board, the sheer difficulty of hiring trained security professionals has increased reliance on outsourced or managed security services. Chronicle's Read APIs expose some of the analytical capabilities described in the scenario above and enable enterprises and MSSPs alike to integrate Chronicle findings into their security playbooks for automation and into other technologies such as ticketing systems or dashboarding tools.

The Chronicle Read or Search API uses the OAuth 2.0 protocol for authentication and authorization. As a simple example, the ListAssets method returns assets that have accessed a specified artifact (a domain for example) within a specified time period, including the first and last time those assets accessed the artifact. An example of the sample request and response for the ListAssets method follows:

Sample Request

```
https://backstory.googleapis.com/v1/artifact/listassets?start_time=2019-10-15T00:00:00Z&
end_time=2019-10-17T00:00:00Z&artifact.domain_name=www.google.com&page_size=1
```

Sample Response

```
{assets: [{asset:                {hostname: "rick"},
           firstSeenArtifactInfo: {artifactIndicator: {domainName: "www.google.com"},
                                   seenTime:          "2018-09-14T20:10:27.157476Z"},
           lastSeenArtifactInfo:  {artifactIndicator: {domainName: "www.google.com"},
                                   seenTime:          "2019-10-24T22:04:04.327829Z"}},
          {asset:                {hostname: "morty"},
           firstSeenArtifactInfo: {artifactIndicator: {domainName: "www.google.com"},
                                   seenTime:          "2019-06-17T21:22:44.812738Z"},
           lastSeenArtifactInfo:  {artifactIndicator: {domainName: "www.google.com"},
                                   seenTime:          "2019-10-24T20:40:54.846676Z"}}]}
```

C) Security & Compliance:

As a specialized, private layer built over core Google infrastructure, Chronicle inherits compute and storage capabilities as well the security design and capabilities of that infrastructure

(Chronicle's "Core Infrastructure"). The underlying design of our Core Infrastructure is described in more detail in a Google whitepaper.

The Chronicle service has also directly received specific certifications and attestations:

- SOC 2 Type 2 and SOC 3
- ISO 27001
- HIPAA BAA

Summary of Chronicle capabilities and benefits

Feature	Description	Benefit
Continuous Telemetry Enrichment	• Automated IP to host correlation	• Faster time to investigate • Greater analyst productivity • More useful answers to common security questions
YARA-L Threat Detection	• Use YARA-L language to define and run rules to detect modern threats	• Easy rule writing • Chronicle rules deliver detection value immediately • Reveal insights beyond threat intel matching
Context and Insights (Threat / IoC, Vulnerability, Asset, User)	• Embedded threat intelligence sources (Proofpoint, DHS AIS, OSInt, Avast, ESET) • Customer provided threat intelligence sources • Vulnerability context • User context • Derived Insights	• Faster time to investigate • Greater analyst productivity • Detection of threats without writing rules
Read APIs	• High performance APIs that expose Chronicle functionality to downstream enterprise and MSSP SOC playbook stages and tools (ticketing, orchestration, dashboarding)	• Automation of SOC playbooks • Integration with MSSP portals • Faster time to remediation
Ingest APIs and Unified Data Model	• High throughput APIs that enable sending data directly to the Chronicle data pipeline without the need for a Forwarder.	• Faster time to value • Zero deployment footprint • Faster telemetry source integration
Raw Log Scan	• Access to all fields in supported data sources. • Search support for data types not yet in scope in curated views	• Faster onboarding of all security telemetry • Enabled for threat hunting use cases
Security / Compliance	• Full stack adherence to GCP common controls inheritance • SOC 2 and SOC 3 • ISO 27001 • HIPAA BAA	• Documented, stringent controls to protect your data at every layer • Key attestations and certifications in place