A closer look at the Lacework Polygraph® Data Platform

How the platform uses data to combat known and unknown threats

LACEWORK®

Introduction

Securing the cloud isn't easy. Rising cyber attacks and poor visibility into cloud infrastructure drive a fear of the unknown. And security teams are in a difficult position. Traditional security solutions weren't built to effectively inventory, understand, or protect cloud environments. Yet organizations must make sense of their cloud activity to make smarter business decisions based on risk, not guesswork. It's time for a new approach.

Frequent, manual audits are costly and will not eliminate compliance gaps. Continuous, automated testing for specific standards will significantly reduce windows of risk. Real-time compliance assessment translates into faster remediation, further narrowing the risk window.

Gartner, Forrester, and IDC are all in agreement that adoption of cloud services will continue to grow at rapid rates through 2020. Minimizing risk with continuous compliance reporting becomes a critical first step towards the larger goal of eliminating that risk altogether with automated remediation and enforcement.

Break the rules

Historically, companies have focused on protecting their systems from known threats by using Indicators of Compromise (IOCs). Unfortunately, this strategy depends on rules — on companies knowing exactly how they will be attacked and manually updating their cybersecurity defenses on those threats.

This approach simply doesn't meet today's reality. New and unknown threats surface every day. By the time companies detect a new attack and learn enough to act on it, the victim has likely been compromised, the data has been breached, and the damage is done.

But what if a company could spot signs of the attack before it happens?

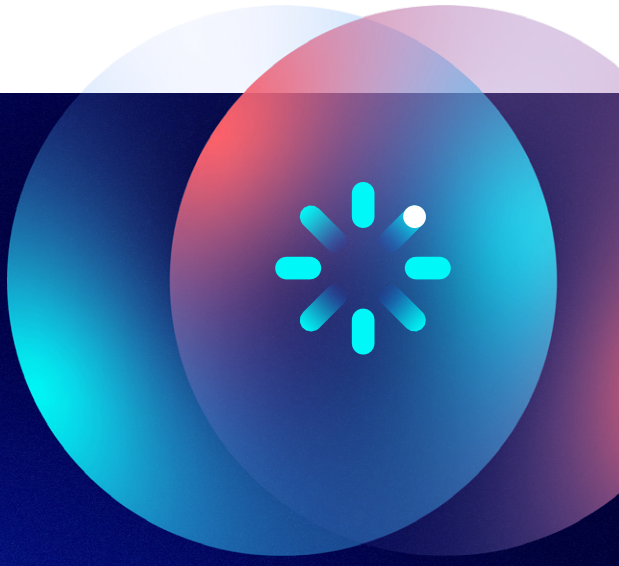
Now you can. This paper will cover:

1. A brief history of anomaly detection
2. Why traditional signature-based security approaches fall short for the cloud
3. How anomaly detection is providing smarter risk-based security for cloud workloads, hosts, containers, and Kubernetes

80% reduction in security breach cost through automation and AI

Breach costs varied from $2.9M with AI and automation vs. $6.71M without.

SOURCE: IBM COST OF A DATA BREACH, 2021



Anomaly detection: A brief history

In 1987, information security researcher Dorothy Denning published ground-breaking research on using anomaly detection for real-time intrusion detection, uncovering security violations in single systems. According to Denning's model, security violations could be detected by observing unusual usage patterns. This assertion has served as the basis for progressive research on anomaly detection ever since.

Traditional security falls short for the cloud

Traditional signature-based security methods face many challenges. These solutions can't detect unknown variants, making it impossible for them to detect attacks that have not previously been seen. They require monitoring for IOCs and building complex rulesets. They require large security teams to interpret and act on the gathered data. But, even then, traditional detection tools are only as good as the strongest security professional's knowledge. They were designed to catch known attacks — not to proactively monitor for signs of trouble.

Rules-based security tools take considerable time and effort to configure, implement, and fine tune. Tired security teams are forced to write new rules or edit existing rules, while faced with an unending queue of non-prioritized (and sometimes false) security alerts. Studies show this leads to burnt out employees in an already talent-shorted field.

But these are challenges encountered even in traditional on-premises environments. Enter cloud computing.

These security methods are no match for the size and scale of the cloud. Dynamic environments, new modes of access, infrastructure that's only as secure as the last DevOps commit. In the cloud, new services are constantly launched, account and resource configurations constantly change, and workloads are scaled up and down without end.

Rules-based security methods are not scalable for the cloud. In the cloud, models based on IoC (file hash and IP address) produce too many false positives — often many thousands per day. This buries critical information, delaying detection of and protection against emerging attacks.

A Cybersecurity Ventures report indicates that there were 3.5 million security jobs left vacated in 2021 and foretells the same amount of vacancies in 2025 — enough jobs to fill 50 stadiums.

SOURCE: THE 2019/2020 OFFICIAL ANNUAL CYBERSECURITY JOBS REPORT, HERJAVEC GROUP

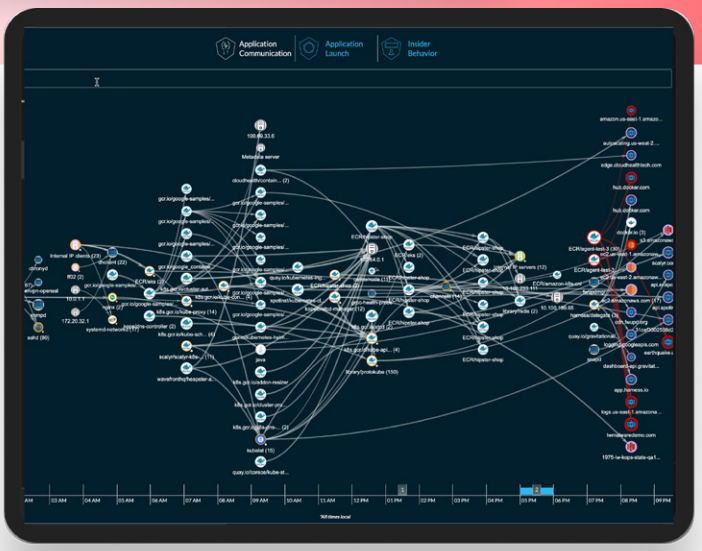
The data that drives our platform

The Lacework Polygraph® Data Platform delivers automated anomaly detection to provide uninterrupted visibility and actionable insights across multicloud environments. The patented Polygraph technology is fed by multiple distinct data sets — including activity data from an extremely lightweight agent and agentless cloud activity log data from your cloud providers. The platform continuously captures hundreds of terabytes of data around processes, applications, APIs, files, users, and networks. Using machine learning combined with behavioral analytics to analyze and cluster unlabeled data sets, this layered approach discovers new behaviors without the need for human intervention.

What's in anomaly detection? A look at the algorithms

The algorithms behind machine learning programs are only as good as the data that companies feed them. Many algorithms have been developed and used to solve these problems. The table below features some machine learning algorithms that are commonly used to address use cases like anomaly detection.

Each of these algorithms is widely used and represents significant engineering advances. However, combining these to create a real picture of large-scale, dynamic cloud environments is a huge engineering task — one that few companies have dared to try.



Algorithm	Function
SimRank	Compares nodes (systems or devices) of a dataset to a network to find similarities. SimRank uses transitive similarity to measure similarity between two nodes based on their neighbors' similarities. It assigns similar ranking and groups information into clusters. For example, if A and B are similar and B and C are similar, then A and C are also considered similar.
Affinity propagation	Used to find how many clusters or groups are in a dataset without needing to know how many clusters of nodes you have. Affinity propagation uses the SimRank similarity matrix to gather similar items then pushes the different items apart.
Term frequency–inverse document frequency (tf-idf)	Used to evaluate how important a word is to a document within a collection of documents, often for information retrieval in document search and spam filtering. Within tf-idf, a word's importance increases as that word appears more frequently within a document, while omitting commonly found words, such as "the," "and," "or," or "etc."
Time series	Used to observe what happened by comparing the current hour to the previous hour.

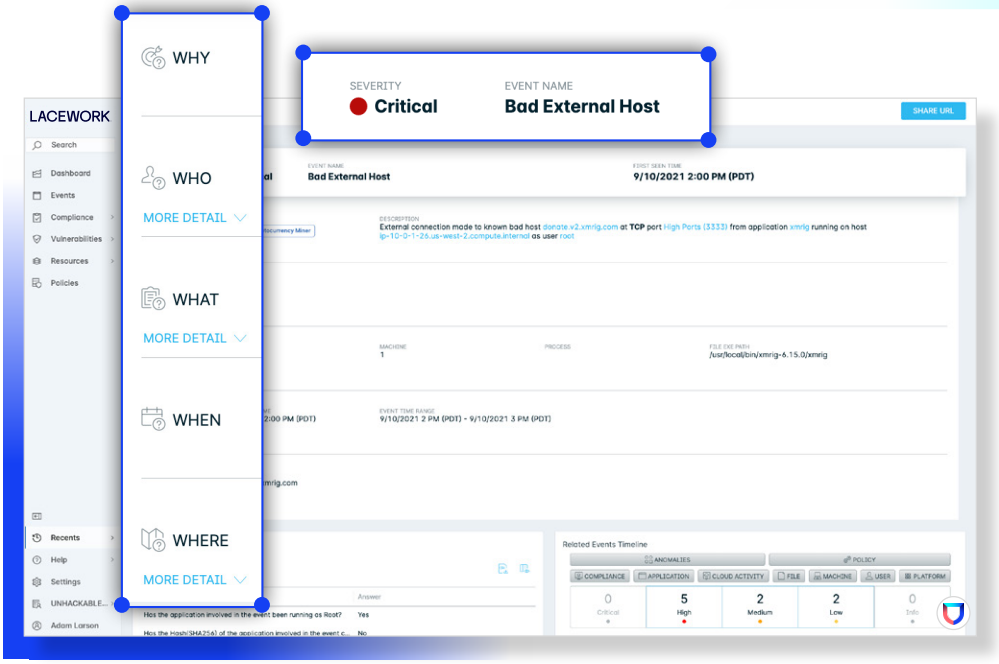
The Lacework Polygraph® Data Platform: Bringing algorithms together

The Lacework Polygraph® Data Platform ingests massive amounts of cloud and workload activity data and analyzes these interactions and behaviors. Using sophisticated algorithms (including those previously outlined and a little bit of secret sauce), the Polygraph technology creates a detailed model of how your company's cloud systems operate. Lacework's platform actually tailors its algorithm to your business, user base, and applications.

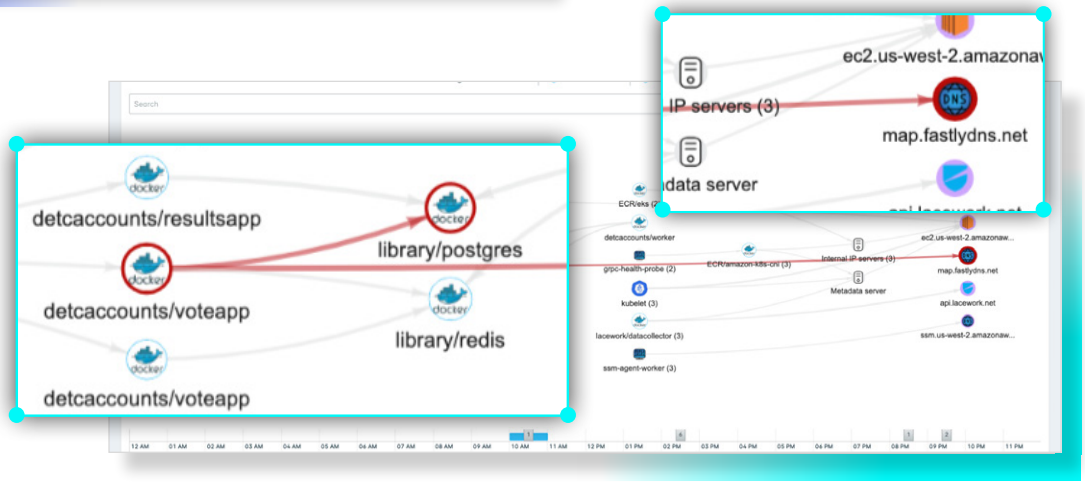
Within three hours, Polygraph creates a baseline for normal cloud activity in your environment by continuously collecting, correlating, and analyzing activity data. Once a baseline is established, Polygraph continues to detect new activities or behaviors every hour. As Polygraph sees new behaviors and changes, those activities are flagged as possible signs of trouble. Polygraph quickly spots trouble in cloud accounts and workloads through behavior-based threat detection and presents this activity within human-browsable graphs.

Alerts are automatically prioritized by criticality to simplify and speed investigations. For any event, Lacework answers five critical questions — who, why, what, when, and where — and provides a visualization to show you exactly what happened, effectively eliminating the manual queries and intensive research typically associated with IOCs and complex rulesets. This important context can help you make quick sense of interactions between resources, services, users, and network activity to detect abnormalities.

Traditional security approaches may require a thousand different models to catch a thousand unique attacks. Our approach automates the manual work required to build those models and understand cloud environments, enabling quick detection of behavior changes, at scale.



Shows critical event details. Get context – the five Ws – to speed incident response and investigation



Shows Polygraph visualization. Quickly identify anomalous communication paths & malicious activity.

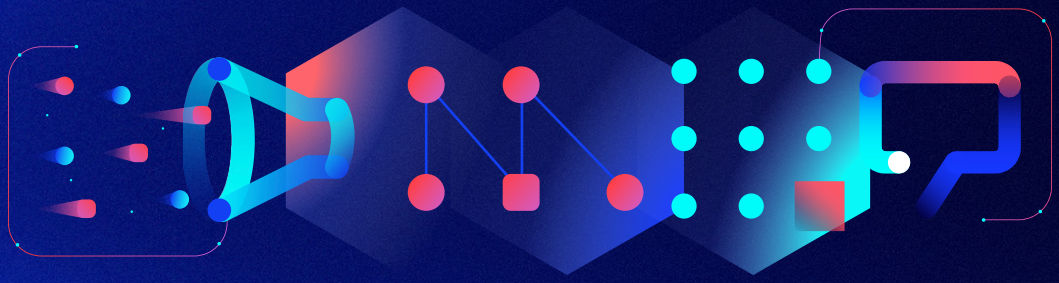
A closer look at Polygraph

In the past, anomaly detection could only identify whether or not something was normal. However, Polygraph can establish the proper context to those deviations by connecting the dots.

Rather than building rules to stop exploits, we use unsupervised machine learning to build a unique baseline for each cloud deployment. Next, we model entities and interactions at a functional level — not just at the network or server level. This step ensures normal cloud changes like the elasticity of machines and containers will not result in false alarms. We only generate alerts for significant changes, not your standard patches or scaling events.

Polygraph uses a variety of behavioral models to answer two simple questions: "Is this normal?" and "Should this be happening?" Within Polygraph, analysis groups enable organizations to streamline data in groups, like applications, processes, and privilege changes, across their cloud environments. This even further simplifies separating normal activity from risky activity, and each group has its own alerts to reduce noise and alert fatigue. This makes it easy to accurately spot changes that accompany an attack.

How it works



Ingest

Polygraph collects data on activity related to:

- Cloud activity
- User and resource behavior
- Application launches

We enrich workload data by mapping network connections to the process that handles them.

Analyze

Polygraph anomaly detection uses data to:

- Create groups for analysis
- Create baseline from activity

Detect

Polygraph anomaly detection detects changes and risks to:

- Identify unusual behavior
- Identify IOC's from threat intelligence feeds

Inform

Polygraph visualizations and alerts provide context to:

- Investigate quickly
- Understand the relevance of what happened
- Integrate with response tools

Polygraph collects and processes real-time activity data in an efficient and scalable way that can result in only a handful of critical results per day. Polygraph builds context between the nodes and edges in order to understand normal relationships and identify anomalous activity. Polygraph then stores a behavior dictionary on each node that describes its features, depending on its type, to determine if a node is new or part of an existing behavior and if this type of behavior is part of the existing baseline. If it is not, it calculates the severity and notifies the security team.

Lacework customer, Hypergiant, particularly appreciates how the Lacework artificial intelligence and machine learning work to surface events. "It's not noise, and we can easily look at the raw data to see what's going on," says Ben Briggs, VP of DevOps and Cybersecurity at Hypergiant. "We're getting more high quality events and the log trash and alert noise has gone away. We get about three alerts a day, and they're always actionable."

Polygraph can do all of this devoid of knowing what the application is or what it is supposed to do. A rules-based system would never be able to do this type of analysis. Security teams can now focus their limited time on the risks that impact their businesses most.

Traditional security approaches rely on rules, built to collect information from datasets such as network activity or processes. However, without understanding the relationships between different dimensions, you are left with a large amount of false positives and missed attacks. Polygraph helps security teams gain a "bigger picture" of their cloud environments, enabling them to analyze and consume data easily, even across multiple clouds. It doesn't just count "things," like the number of logins or traffic; it models behavior and activity.

Polygraph can detect event changes for applications, users, and workloads including the following scenarios:

- Regular application user, redis, launches a new application, curl, and generates a new outbound network connection, pastebin.con
- Your cloud server, which has only ever connected to the S3 service, is suddenly connecting to IAM and trying to create new users from a new geolocation
- There is a new connection to a known bad IP (the Lacework Polygraph® Data Platform checks with about 40 reputation feeds)
- There is an external connection to the mining pool domain(s)
- An attacker on compromised system is downloading tools and creating a reverse shell (from an actual customer's pen-testing exercise)

Other security approaches that rely solely on network logs can only provide a flat environment view, increasing the chance of missed threats and misconfigurations. Our information empowers customers to see and understand cloud changes at scale without requiring manual intervention by security teams.

In other words, you're not going to have to triple your security team to support your workloads. And your existing security team? They will enjoy doing the work that truly matters — not sifting through endless queues of false alerts.

Terms Explained

Node: A connection point. In cloud computing, the term typically means a physical server or another cloud, each of which consists of multiple individual nodes.

Edge: A computing location at the edge of a network, along with the hardware and software at those physical locations.

Cloud computing: The act of running workloads within clouds, while edge computing is the act of running workloads on edge devices.

The team behind the magic

Our team of experienced Data Scientists, ML Engineers, Data Engineers, and Software Engineers uses a combination of data-driven intelligence creation and our own expertise in several key areas, including Machine Learning, Artificial Intelligence, and Expert Systems. We use a combination of both novel and well-known machine learning techniques as well as cloud security concepts to build systems that can learn and evolve over time based on each customer environment.

Get anomaly detection with Lacework

Automation and context are more critical than ever. Lacework can help you automatically detect unknown, unusual activity in your Amazon Web Services (AWS), Azure, Google Cloud, K8s, and hybrid environments and quickly surface unexpected changes that require attention.

About Lacework

Founded in 2015, Lacework is the data-driven security company for the cloud that delivers end-to-end visibility and automated insight into risk across cloud environments, so you can innovate with speed and safety. The Lacework Polygraph® Data Platform ingests data, analyzes behavior, and detects anomalies across an organization's Amazon Web Services (AWS), Google Cloud, Microsoft Azure, and Kubernetes environments. This patented approach significantly reduces noise and turns millions of data points into prioritized, actionable events. Customers all over the globe depend on Lacework to take software services to market faster and more securely, while consolidating overlapping security solutions into a single platform for better visibility and coverage across a multicloud environment.

Get a personal tour.

Request a demo

Dig further into anomaly detection.

Watch a webinar