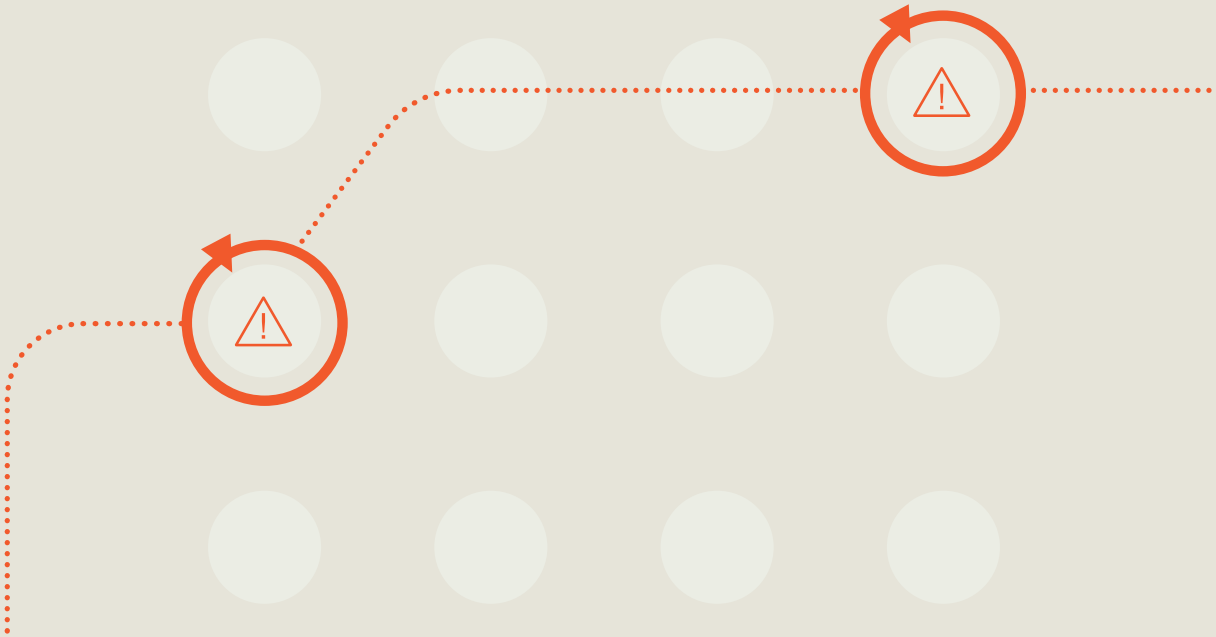
BITSIGHT

REPORT

A Data-Driven Approach to Asset Discovery and Risk Measurement



Introduction

One of the principal reasons the world's highest-profile organizations turn to Bitsight to understand, measure, and mitigate risk is that our solutions are powered by data—a lot of it. Data is the foundation that enables us to deliver everything from the highest-quality security ratings to exposure risk insights across your expanding attack surface. Insights that correlate with real-world security outcomes.

But the volume of data that we collect is only part of the equation. The quality of that data, and more importantly, the insights that we derive from it, is what sets us apart. We recognize that trust in that process – our methods to transform billions of security observations into actionable recommendations – is earned through transparency.

The purpose of this guide is to help you understand more about:

- How we think about risk data broadly
- Our unique methods of collecting and enriching data
- How our data informs the services we deliver to customers

We'll accomplish this by exploring the two essential concepts:

Data Collection: How can we collect as much accurate and meaningful data about entities, assets, and risks as possible?

Mapping and Attribution: How can we use different types of data for different purposes to understand more about the internet as a whole and the risks faced by individual organizations?

We use a wide range of tools and methodologies to perform these critical functions, including substantial ongoing investments in two proprietary technologies that help us accelerate and scale our data collection and categorization efforts:

- ***Bitsight Groma*:** A proprietary internet scanning engine that operates at a global scale
- ***Bitsight Graph of Internet Assets*:** An AI-driven asset and entity discovery and relationship mapping service

While these innovations do not replace the human expertise that is at the core of our research, product development, and service delivery, they act as a force multiplier that allows us to achieve unprecedented speed and scalability while continuously improving the quality of the data powering our offerings.

Data Collection

Continuously collecting a broad and detailed view of internet-connected assets is essential to our ability to understand our customers' risk posture. Our approach includes a diverse set of activities that come together to create a rich and expansive view of the world's ecosystem of interconnected entities, their individual digital footprints, and their risk posture.

Types of Data We Collect

We collect data from a diverse set of sources, including but not limited to:

- Publicly available information
- Observable Internet communications
- External scanning of internet-connected systems
- Asset and vulnerability details and context

Our Collection Methods

Across these areas, our data collection methods fall broadly into two categories:

Passive collection techniques, where we listen for signals that provide additional context about asset ownership, relationships among entities, and risk posture

Active collection techniques, where we proactively query specific data repositories and scan internet-connected assets to deepen our understanding of risk posture

In addition to independent data collection activities, we increasingly conduct collaborative collection with our customers, including syncing their data from multiple cloud infrastructure providers. In these instances, we integrate with customer environments with their permission to automatically and continuously collect their data, further enhancing data completeness and accuracy.

Passive Collection Techniques

Our passive data collection efforts help us gain insights into assets, relationships, and observable indicators of organizational risk.

Examples include:

- Using sinkholes, malware emulators, honeypots, and similar techniques to discover ransomware precursors, worms, botnets, greyware, adware, malware distribution, malicious internet scanning, and vulnerability exploits.
- Assessing the version levels of endpoint browsers, operating systems, and desktop software.
- Listening to network advertisements, such as those performed by the BGP routing protocol, to determine IP address ownership.
- Analyzing WHOIS records, certificate transparency logs, DNS queries from endpoints, and related information to determine the affiliations of hostnames and subdomains.
- Monitoring the behavior of endpoint devices, such as movements between locations, to develop baselines of normal workforce computing behavior.
- Observing the speed and effectiveness of organizations' hardware and software lifecycle management activities.
- Monitoring for publicly exposed account credentials.

Active Collection with *Bitsight Groma*

Our active data collection efforts target areas where we can proactively gather information about asset ownership, relationships between organizations, configurations such as open ports, TLS settings, software versions, and specific vulnerabilities that individual organizations have using a variety of publicly accessible sources. This is accomplished primarily through continuous, internet-scale scanning.

Our internet scanning approach has evolved over time from the curation of third-party vulnerability and configuration information to a hybrid approach that combines third-party signals with extensive first-party data from our proprietary internet scanning engine, ***Bitsight Groma***.

Bitsight Groma **scans the entire internet on a continuous basis from over 130 points of visibility globally to:**

- Give us a comprehensive view of the global ecosystem of internet-connected assets and entities
- Help our customers understand their external risk posture in detail
- Provide critical input into the measurements that support security ratings, external attack surface management, and other products

Using Sinkholing to Discover and Understand Malware

A couple of Bitsight's most powerful passive data collection methods include sinkholing and the operation of malware emulators. Sinkholing is a technique that intercepts command and control communications from malware and botnets and redirects it to a sinkhole instead of its intended destination. Both approaches allow researchers to analyze the communication patterns to learn more about the malware and track the source IP addresses of the infected machines.

Bitsight operates one of the world's largest sinkhole infrastructures, enabling unmatched visibility into botnets and malware and the risks they pose to organizations globally. In addition to adding value directly for customers, the Bitsight sinkhole infrastructure is used regularly in our collaboration with law enforcement agencies to identify victims of threat actor groups, gather essential evidence, and dismantle botnet operations.

The insights we gain through *Bitsight Groma* include risk factors such as:

- Presence of known software vulnerabilities
- Susceptibility to widely used web application exploit techniques
- Authentication weaknesses
- Misconfiguration of SSL/TLS
- Exposure of sensitive industrial control system (ICS) and operational technology (OT) assets
- Firewall misconfiguration
- Email security protocols

Performing our own internet scanning with *Bitsight Groma* allows us to:

- Innovate more rapidly through greater control over the scanning process
- Accelerate mean-time-to-detection for new vulnerabilities and asset updates
- Respond faster to changes in customer environments through on-demand rescans
- Accommodate specialized requests from our security researchers

These benefits are already visible in our product offerings. For example, ***Bitsight Groma*** contributed to our ability to adjust the lifetime of the Patching Cadence considered by our security ratings algorithm from 300 days to 90 days in our 2024 ratings algorithm update. This was possible due to the fact that higher volume, higher quality scan data can achieve similar levels of correlation to bad outcomes with a shorter finding lifetime.

Security Observables: A Critical Output of Data Collection

Collectively, our passive and active data collection produce a rich set of security observables that give us the data points we need to understand and measure risk. Many of the data points above are examples of security observables.

Whenever possible, we seek out security observables that target specific areas from multiple directions, using a combination of passive and active collection techniques, to give us a more precise view of organizational risk.

The following examples illustrate this.

RISK AREA	PASSIVELY CAPTURED SIGNALS	ACTIVELY CAPTURED SIGNALS
Vulnerability Exposure	Evidence of active vulnerability exploitation Insight into software versions and patching practices	Identified presence of known vulnerabilities
Credential Compromise Risk	Observation of credentials in credential dump databases	Observation of credentials in active spam campaigns

Mapping and Attribution

The data we collect is only actionable if it can be viewed and analyzed in the proper context. While mapping and attribution are tightly integrated with our data collection process, our approach to this critical activity is useful to understand in more detail. The goal of our mapping and attribution efforts is to use data to demonstrate ownership, authority, responsibility, or use of the digital assets we discover. This allows Bitsight to perform extensive monitoring of customers' critical assets and those of companies they do business with.

Bitsight Graph of Internet Assets (GIA)

The focal point of our mapping and attribution efforts is the ***Bitsight Graph of Internet Assets*** (GIA). GIA gives us a continuously updated view of internet-connected assets and organizational attribution mappings, drawing from sources such as:

- Domain and IP WHOIS details
- Organizational details included in SSL/TLS certificates
- Other corporate identifiers that can be discovered from websites.

It combines graph technology, machine learning, and artificial intelligence to:

- Discover new assets
- Map relationships between assets and entities
- Listen for changes and keep mapping details up to date

GIA takes the insights, logic, and mapping strategies developed by our researchers and allows them to be applied:

1. At a global scale

2. With extremely high frequency

GIA's AI models were trained using years of human-curated mapping data that now spans over 1 million entities and millions more underlying assets and points of evidence. It uses this base of knowledge to create high-confidence mappings of assets at 10x the speed of our previous automation tools. This creates a significant force-multiplying effect, allowing us to continuously improve the quality and timeliness of the data powering our services.

1 million

entities mapped

10x

faster creating high-confidence mappings

In addition to accelerating infrastructure discovery by orders of magnitude, GIA gives our team a much more flexible platform for data exploration. GIA's graph technology enables rich visualizations that can begin with a domain and expand to find linkages across the entire internet that can reach thousands of levels. All linkages are given a confidence level, and human feedback loops further train GIA's model and continuously improve its accuracy.

For example, GIA provides a feed of probable assets of specific entities to Bitsight's Technical Research team. This accelerates the time it takes to build views of customers' externally visible attack surface. Our researchers also have the ability to mark any assets on the threshold of our confidence level as yes or no. GIA uses this feedback to further train its AI model, improving the accuracy and confidence levels of future data refreshes. In the future, this workflow will be extended directly to customers of Bitsight's External Attack Surface Management (EASM) offering under a self-service model.

Human Intelligence

While technology innovations such as ***Bitsight Groma*** and ***Bitsight Graph of Internet Assets*** are critical to scaling and optimizing our data collection, mapping, and attribution efforts, human experts continue to play a highly strategic role across several key areas.

Bitsight's data efforts are directed by a team of **over 150 technical researchers, security researchers, and data scientists**, who collaborate with an extended team of engineers and product managers to:

- Perform foundational research about the cybersecurity threat landscape and possible methods of measuring risk
- Explore ways to apply foundational research to deliver new products and other value to Bitsight customers
- Supervise the mapping and attribution process by hand-curating data and using it to train our machine learning and AI models
- Map relationships between entities to help customers understand third- and fourth-party risks
- Curate vulnerability and device fingerprint information to bring focus and context to our data collection
- Enable the multiplying effect of automated processes

This team is organized into three critical functions:

Technical Research

Our Technical Research team is composed of individuals with extensive knowledge about the global domain name service (DNS) infrastructure and IP addressing. Drawing from data collected from Bitsight Graph of Internet Assets and other sources, this team manually curates the assignment of IP addresses and domain names to organizations. This provides an essential foundation for the automated mapping and attribution activities performed by GIA.

Security Research

Another major focus of Bitsight's human intelligence activity is Security Research. This group is staffed by security and big data experts who focus on critical areas such as:

- Researching techniques to remotely detect and measure security configurations and properties that provide insight into organizational security risk, exposure, and performance
- Devising and testing new techniques for associating discovered infrastructure with specific entities
- Developing techniques to remotely detect the presence of vulnerabilities at Internet-scale using ***Bitsight Groma*** to assess exposure and evaluate patching and remediation practices
- Reverse engineering malware and intercepting command and control communication to support the collection of compromised system data

Data Science

Finally, our Data Science team holistically examines the data we collect and develops strategies for turning it into meaningful insights about risk. The best example of these efforts is the Bitsight Security Rating, which is informed by carefully considered application and weighting of the data we collect to arrive at ratings measurements that correlate with real-world security outcomes. This team is also responsible for working with, improving, and tuning GIA's AI models to continuously improve the speed and quality of our data attribution efforts.

Using Data to Understand Risk and Guide Security Investments

In the era of AI, high-quality data is everything. Bitsight spent the last decade amassing a large, high-fidelity collection of data about internet-connected assets and the entities associated with them.

This data is:

- Global in scope
- Continuously updated
- Inclusive of historical views
- Human-curated for quality
- Scaling at internet speed with the help of technology

Now, modern AI technologies allow us to use this data in even more compelling ways to unlock value for customers. ***Bitsight Groma*** and ***Bitsight GIA*** are the first examples of this, and they are already enhancing the products our customers rely on every day in significant ways.

We will continue to focus our innovation efforts onthe intersection of data and AI, with the goal of helping organizations like yours better understand risk and base security decisions and investments on data instead of guesswork.

Ready to take the first step? Learn more about how Bitsight can help you:

Reduce risk and improve cybersecurity execution →

Understand and reduce exposure to third-party risks →

Build specialized risk research and analysis capabilities →

Bitsight is a cyber risk management leader transforming how companies manage exposure, performance, and risk for themselves and their third parties. Companies rely on Bitsight to prioritize their cybersecurity investments, build greater trust within their ecosystem, and reduce their chances of financial loss. Built on over a decade of technological innovation, its integrated solutions deliver value across enterprise security performance, digital supply chains, cyber insurance, and data analysis.

BOSTON (HQ) RALEIGH NEW YORK LISBON SINGAPORE

BITSIGHT