USE CASE

ANOMALI

Navigating the Cybersecurity Risks of Dark Data

Dark data refers to the digital information generated and stored by businesses but not actively used or analyzed. It can encompass a variety of formats, including log files, old versions of documents, unused database entries, emails, and other data artifacts. There are several reasons why dark data is significant:

Potential Value:
Dark data can contain insights that could be beneficial if analyzed. For instance, historical transaction data might reveal buying trends over time.

Storage Costs:
Maintaining data, especially in large volumes, incurs costs. Organizations might be spending resources on storing data that doesn't offer them any tangible benefits.

Security Risks:
Dark data can pose cybersecurity threats. If businesses aren't aware of the data they have, they can't protect it adequately. It might contain sensitive or private information that, if exposed, could lead to breaches and legal repercussions.

Compliance Concerns:
Regulations like the General Data Protection Regulation (GDPR) mandate that organizations know what personal data they hold and ensure its protection. Dark data can make compliance challenging.

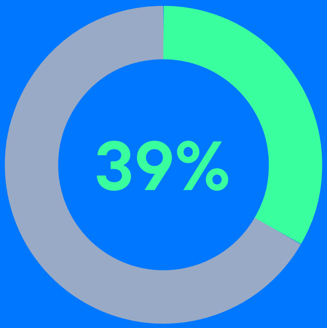
In the realm of cybersecurity, understanding and managing dark data is crucial to both harness its potential value and mitigate associated risks.

The number of tools in a Security Operations Center (SOC) can vary widely depending on the size of the organization, its industry, compliance requirements, and threat landscape. However, it's not uncommon for an average SOC to use anywhere from 20 to 50 different tools. These may include:

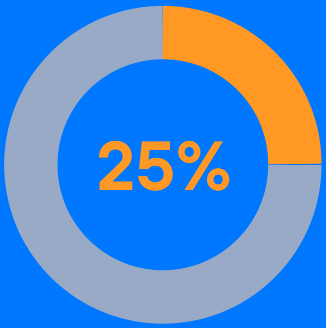
- Security Information and Event Management (SIEM): For centralized logging and analysis.
- Endpoint Detection and Response (EDR): For managing endpoint security.
- Network Detection and Response (NDR) or Intrusion Detection Systems (IDS)/Intrusion Prevention Systems (IPS): To monitor network traffic for suspicious activity.
- Firewalls: For controlling incoming and outgoing network traffic based on predetermined security rules.
- Vulnerability Assessment Tools: For identifying weaknesses in the infrastructure.
- Threat Intelligence Platforms: Like Anomali, for aggregating and correlating threat data.
- Identity and Access Management (IAM): To manage authentication and authorization.
- Data Loss Prevention (DLP): To monitor and control data transfers.
- Incident Response Tools: For organizing and automating responses to security incidents.
- Forensic Tools: For post-incident analysis.
- Ticketing Systems: For tracking incidents and responses.

Having a multitude of tools can sometimes lead to "tool sprawl," which complicates management and can even create security gaps if not carefully managed. And given the learning curve for each of these tools, it's nearly impossible for any SOC analyst to become fully proficient on any one of them.

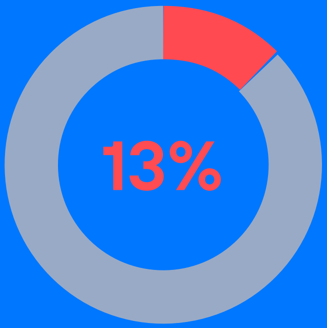
How and why does data go dark?



Too much data coupled
with insufficient analysis



of security systems can only
access structured data sets



of available data is
understandable to security tools

We cannot overstate the significance of the risk associated with dark data in the realm of cybersecurity. As data continues to grow exponentially, so does the exposure to the enterprises where it resides.

There are also significant cost implications; the average data breach currently averages $3.5 million per incident - and may not be detected for several months. This not only impacts your bottom line but it also sullies your hard-earned reputation.

The proper management of dark data isn't merely a best practice but a necessity. Understanding the value of your data assets and managing them securely can reduce the likelihood of a costly breach. Securely storing and managing dark data is a multi-faceted challenge. Complexities can range from technological limitations to compliance issues.

While dark data presents challenges, it also offers opportunities for enhanced cybersecurity measures when managed correctly.

Dark data is essentially unstructured Big Data with untapped potential; it's a potential goldmine waiting to be discovered but one that comes with inherent risks. Dark data is very much 'uncharted territory.' This can include undetected fraud, unnoticed command-and-control servers, and invisible viruses, which often go unnoticed but could be malicious activities hiding in plain sight. These elements are not merely statistical outliers but represent real and significant cybersecurity threats that can't be ignored.

You should begin by understanding the anatomy of threats:

- Break down what each type of hidden risk means.
- Undetected fraud could be financial anomalies or unauthorized access.
- Unnoticed Command-and-Control Servers could signify that your network is part of a botnet.
- Invisible viruses may be lying dormant, ready to be activated.

These are not hypothetical issues. They are clear and present dangers that pose a significant risk to data security. These threats are time bombs waiting for the opportune moment to strike, exacerbating the need for proactive measures. Assess and manage your dark data now to effectively mitigate these risks.

Obstacles to Recovering Dark Data
(% Affected)

39%

Volume of dark data

34%

Lack of necessary skill sets

20%

Difficulty in coordinating across departments

26%

No one is dedicated to finding data the org possesses

23%

Difficulty in coordinating with data generating third parties

21%

Lack of interest from organization leaders

19%

Lack of creativity

As a relevant and timely example, SEC Form 8K rules require public companies to disclose 'material' data breaches in 4 days. Are you able to search all your data in four days?

In order to react to this new mandate without the risk of fines, we need a generational leap in security analytics to truly unlock its potential. This can include the following:

- Large Language Models (LLM). These sophisticated tools can greatly assist security analysts by automating mundane tasks and offering actionable insights. The goal here is not to replace analysts but to make them more effective. LLMs can also help in the 'Getting Data In' problem—streamlining the process of ingesting dark data from diverse sources into a unified platform for analysis. This is critical because data collection is often the first stumbling block in security analytics.
- Next would be Threat Detection and Prediction, driven by log analysis. Hidden within server and network logs may be signs of a cyber attack or internal threat. Machine learning algorithms can analyze this dark data to predict vulnerabilities or recognize suspicious patterns.
- Forensics and Incident Response driven by historical data correlation. Old emails or archived documents can sometimes provide invaluable insights during a cybersecurity incident investigation, helping analysts piece together how an attack happened or who might be responsible.
- Compliance and Reporting in its many varied forms. For industries with strict regulatory requirements like finance or healthcare, dark data can be used to prove compliance or to reconstruct events in case of an audit.
- Insider Threats and User Behavior Analysis. Idle user activity logs or database access logs can help in identifying unusual behavior that may indicate an insider threat.
- Supply Chain Security enabled by communication analysis. Emails and other forms of communication with suppliers or third-party vendors can be analyzed to identify potential threats or weak links in the supply chain.
- Risk Assessment and Management of Asset Utilization Data. Information about how hardware and software assets are used can help organizations better understand their cybersecurity risk posture. Unused or underutilized assets may present unnecessary risks and can be eliminated or better secured.
- Data Loss Prevention through analysis of file metadata. Metadata for files stored on company servers can reveal sensitive data stored in unprotected locations, helping plug data leaks.
- Zero-Day Exploit Prediction and code repository analysis. In the case of software companies, analyzing version history and bug reports in code repositories can sometimes reveal patterns that predict where a zero-day exploit might occur.

Conclusion

Dark data serves as the invisible framework that permeates the vast expanse of cyberspace. While it may appear to be inconsequential or dormant, this category of data performs a critical function: it connects disparate data points, such as network logs, unstructured files, and unused databases, to form a more complete picture of the cybersecurity landscape.

This interlinking is not merely an academic exercise. It serves to illuminate hidden patterns and correlations that can be crucial for understanding the full scope and scale of cyber threats. For instance, an isolated piece of dark data may seem harmless, but when connected to other data points, it could reveal the existence of an advanced persistent threat or an ongoing data exfiltration.

In essence, dark data is far more than just dormant information gathering dust in your storage systems. It is an untapped resource that, when analyzed and contextualized, can provide invaluable insights. It holds the key to unlocking a more nuanced understanding of cyber threats, from identifying vulnerabilities to predicting future attacks. Far from being a liability, it has the potential to become one of your most powerful assets in the fight against cybercrime.

ANOMALI