TAG CYBER

SELF-PROTECTING DATA AS A MEANS FOR BUSINESS RESILIENCY

KATIE TEITLER, TAG CYBER

TAG CYBER

SELF-PROTECTING DATA AS A MEANS FOR BUSINESS RESILIENCY

KATIE TEITLER

Enterprise resiliency, a cornerstone of sustainability, has a new partner: Self-protecting data. With cloud usage at ubiquitous levels, and cybercriminals leveraging vulnerable infrastructures to target valuable data, organizations need greater control over how their data is accessed and used. Traditional security and privacy technologies — especially those built for on-premises networks — do not go far enough to prevent tampering and ensure end-to-end data file confidentiality, availability, and integrity.

INTRODUCTION

Data has been called the "crown jewels" of organizations, meaning, data is the foundation upon which organizations plan, produce, profit, and prosper. As such, many data protection mechanisms have been developed over the years, from kludgy DLP to fussy encryption to zero trust-based access controls. All these protection capabilities (and more) have a place in the data lifecycle, but none of them (as standalone controls) fulfills end-to-end data protection and control, from creation to network traversal and storage and, finally, through secure data disposal or destruction. Furthermore, any visibility into how and when data is used throughout the numerous stages relies on the technologies implemented to protect it, not the data itself. When those technologies are circumvented, or they don't function as intended, the data itself remains largely vulnerable.

Encryption is the best mechanism by which organizations can shroud data from unauthorized and malicious individuals, whether that data is stored on a network, traversing a network, or traveling between networks. Encryption allows data

owners to obfuscate data, rendering it unreadable to unauthorized individuals and systems. However, encryption is neither resilient to reverse engineering (in fact, many legitimate security technologies must decrypt data in order to protect it) nor straightforward to manage. Attackers generally do not attempt to defeat encryption – instead, they steal the passwords, other authentication credentials, and basically exploit the complexities of key management or authentication management or privilege management systems.

What's more, data is the lifeblood of a business. How valuable is your business if no one can read your data, analyze it, update it, sort it, process it, share it, etc.? These functions are part and parcel of working life. Thus, the very existence of data implies the expectation of access and handling. Though data can be encrypted during many stages of the data lifecycle, it cannot be used (processed, updated, etc.) in an encrypted form. Simply leaving it unencrypted creates a vulnerability. As a result, even when organizations have rigorous encryption practices, data ends up going through oscillating stages of encryption-decryption-encryption-decryption. Therefore, other layers of protection and governance must be applied over the top.

Furthermore, the increased dependency on traversing disparate supply chain networks and the abundant use of cloud networks necessitates another level of data protection. Though the major cloud providers are very good at protecting their infrastructure, the Shared Responsibility Model means that organizations must assure the security and privacy of any data and data files placed into cloud environments as well as the access controls and configurations that allow authorized users to access the environments, including data in files.

In 2020, cloud misconfigurations were cited as the attack vector that caused the exposure of 33.4 billion records. It should come as no surprise, then, that utilities to directly protect data and files must be a business priority in 2021 and beyond. In this report, we explore how companies can create and utilize data that can defend itself to ensure that it's tamperproof and resilient to the cleverest of cyber attackers but accessible and usable by intended users.

WHY TRADITIONAL DATA PROTECTION AND ACCESS CONTROLS AREN'T ENOUGH

The prevalence of cloud usage and the vulnerabilities associated with it, as described in the introduction, are evidence that new mechanisms for protecting and controlling cloud data files are needed. A few other prime use cases also bubble to the top:

- ***Data migration:*** Many companies are reducing the amount of on-premises infrastructure they manage and thus must develop a strategy for secure migration of data files to the cloud. The biggest risk is transmission to and retrieval from the cloud. Protecting data in transit can be accomplished through encrypted connections (HTTPS, TLS, FTPS, etc.) and encryption of the data, itself. Yet, as data transits the TCP/IP comms layer, node2node2node, "securing" multiple, sometimes disparate channels is complicated. Companies should also implement robust network security and endpoint controls to ensure those vectors cannot be attacked in the process. While this layered approach is assumed "trusted," it fails to include any data self-reporting, meaning, if a savvy attacker is able to exploit any stage in the migration process, an operations team may not know that an attack is happening at the data layer until the exploit has already succeeded.

- ***Data creation in cloud (including software development):*** More and more, data is created in cloud environments. Data in use — which includes writing, testing, and deploying code — cannot be encrypted. Most businesses therefore use access controls to try to protect software and data files. But when those controls are compromised or entrusted to others, the data remains highly vulnerable to tampering if they are not self-controlling and do not have self-awareness or self-reporting mechanism.

- ***More sensitive data in the cloud:*** As with data and data files created in cloud environments, the addition of more highly sensitive data requires organizations to have end-to-end processes for protecting and controlling data files, whether they are in use, at rest, or in transit. No stage can be ignored and policy enforcement should be autonomous.
- ***Backup, recovery, and resiliency***: Ransomware is a top-line business threat. As demonstrated with the Colonial Pipeline attack, as well and many others, ransomware can bring a business and its customer eco-system to its knees if they are not adequately prepared. While we at TAG Cyber advocate for proactive cyber security, it would be foolhardy to ignore the fact that some cyber attacks will be successful. Therefore, we believe that a part of a proactive cyber security program is building and maintaining a robust backup and recovery process. Doing so ensures resiliency in the event of a cyber attack. It is not enough, though, to simply backup data for recovery purposes. While a novice may go after data files on users' devices or corporate/cloud servers, savvy cybercriminals understand that it is more devastating to tamper with all instances of data files, including backups. If backups are not properly protected — and standalone encryption may not be an option given its useless in a ransomware scenario — organizations would be wise to develop alternatives which utilize data that can defend itself.
- ***Privacy and compliance***: The GDPR and the CCPA ushered in a wave of privacy regulations worldwide. In their wake, and because of the persistent frequency of data breaches coupled with a burgeoning "big brother," citizens are demanding greater privacy protection from not only governing bodies but also the organizations within which they work. In turn, businesses have begun to realize the competitive advantages of provable data privacy. *Demonstrating* data privacy when asked or required is a necessity, and manual, traditional efforts not only do not scale but cannot be trusted in today's digital world.

THE IMPACT OF DATA THAT CANNOT DEFEND ITSELF IN THIS GLOBAL, DATA-CENTRIC FRONTIER

When security pros think of data management, it's generally in the form of data protection, a "shift left" mindset. Resiliency is a less-considered aspect of security programs, though arguably a more important one. Far too often, organizations do not realize the criticality of resiliency until a breach occurs and (either or both) data and systems that house said data are unavailable or unreliable.

"Resiliency is truly the cornerstone of operational sustainability and should be considered a top priority of security teams."

Resiliency is truly the cornerstone of operational sustainability and should be considered a top priority of security teams. When a breach occurs, corporate brand attrition leads to loss of shareholder value, and the regulations intended to preserve enterprise security don't go far enough in their efforts nor do they combat the core problem: data resiliency.

While certain data protections, such as those mentioned above, are a good start, they are wrappers, of sorts, that hover around the data. In contrast, data that protects itself by means of embedded controls does not rely on third-party mechanisms, either for protection or for reporting on the status of the data, and therefore supplies a hardened security and privacy layer. It is this hardening that allows companies to mitigate data compromise and remain resilient when another system vulnerability — such as a cloud misconfiguration or stolen credentials — is exploited.

DETECTING AND PREVENTING DATA MISUSE AND ABUSE

As is consistent with zero trust principles, data protection and privacy controls must be independent of the environment or network in order to prevent data misuse and abuse. Modern business requirements simply won't allow for network-based controls alone, as they have proven exploitable. Today, companies need policy and control throughout the OSI stack, however, very few (effective) security technologies exist at layer 2, the data layer.

Nonetheless, the data layer has taken on new importance, especially in the last year as businesses shifted to work-from-home and now to hybrid work operating environments. Employees, customers, and systems all need fortified data access, but in such a way that doesn't hinder productivity. Yet, preventing unauthorized access to and use of data is a top-line business requirement for any company that wants to stay out of the breach headlines. But protecting the networks on which the data reside, transport mechanisms, or even the access and permissions to the data isn't enough; far too many breaches have occurred when these controls were in place.

Self-protecting data that is decisioning and controlling at layer 2 conform to the principles of zero trust and preserve privacy and enforce security, wherever the data is — at rest, in transit, or in use. What's more, an intelligent approach to data protection and privacy incorporates self-awareness and self-protection and allows businesses to immediately identify — or prevent entirely — when code is changed without authorization, when malicious code is inserted, and prevent any sort of tampering, thereby assuring the integrity of the data. This fulfills the need for data, and more importantly, business resilience.

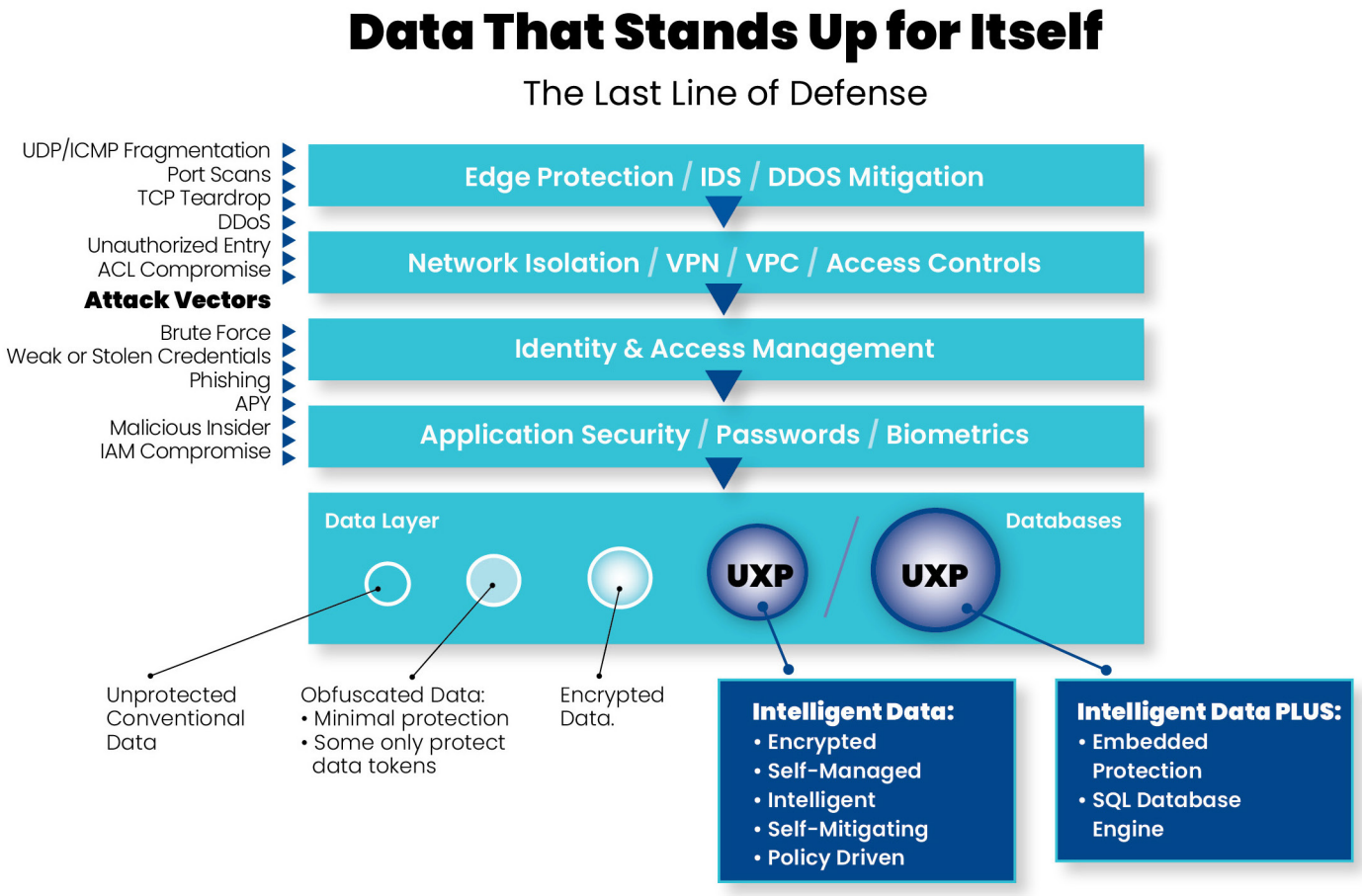

FIGURE 1. Smart Data vs. The New Standard for "Intelligent Data"

HOW DO YOU MAKE DATA INTELLIGENT?

What, exactly, should be the new standard for "intelligent" data? There are many solutions today that embed "intelligence" into data-files. But that definition of intelligence is essentially information relating to rules, encryption keys and in some cases, authentication credentials. This standard for "intelligence" built into data-file does not translate into self-protecting data and does not empower the data to decision and control enforcement of security and privacy policies? The new standard for intelligent data uses an embedded decisioning and control engine which identifies, authenticates, and governs what a user can do with the data file. Decisioning and control functions are exclusively the purview of the embedded engine, while execution of the functions is carried out by the application. This intelligence capability assures the integrity of data and affords data resiliency even when other parts of an organization's digital systems have been compromised.

The embedded "intelligence" instructs and controls the application to carry out both proactive and reactive processes. This intelligence capability assures the integrity of data and affords data resiliency even when other parts of an organization's digital systems have been compromised.

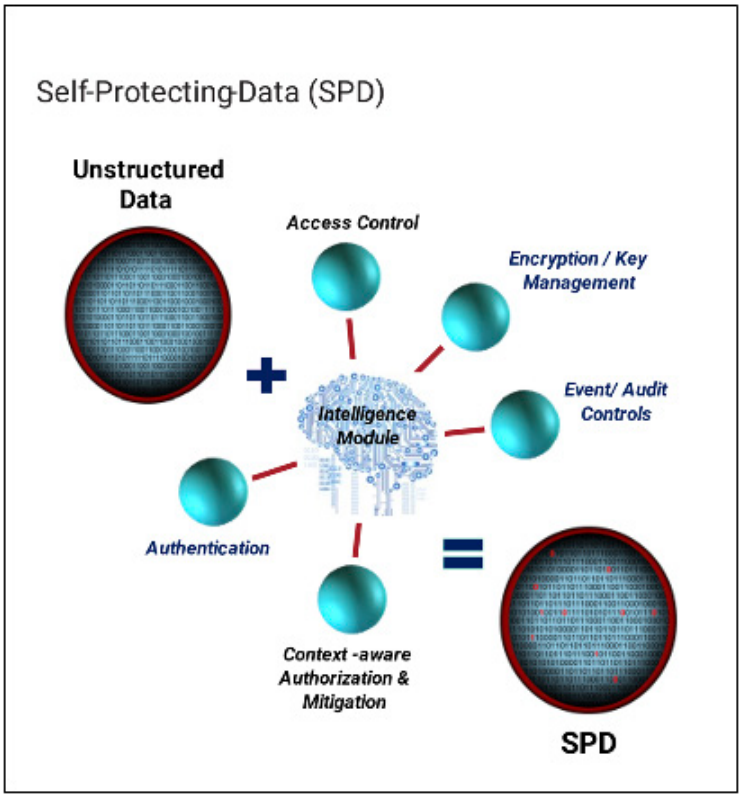

FIGURE 2. Data Self-Protection Intelligence Module

Some examples of these proactive and reactive processes are given below.

- Self-protecting "digital identities" are utilized to control selective access to sensitive information within the data-files
- Track who is attempting access, to what information, when, from where, how
- Enforce stepped-up authentication credentials
- Inform and/or alert the data-file owner of key events
- Self-shred

WHAT IS SELF-PROTECTING DATA?

In the last few years, terms like "self-healing" and "self-protecting" have cropped up in cyber security marketing materials. It's easy to dismiss these terms as hype, hyperbole, or outright exaggeration. However, when combined with credible and demonstrable technological capabilities, they become powerful mechanisms for protection and privacy.

So, what is "self-protecting data," practically speaking?

- ***Data that can defend itself:*** As opposed to solutions that put security information "in" data or protections around data, self-protecting data has intelligent software code written directly into it that beacons back to data owners, giving them insight into how the data or data file is being used, accessed, and by whom. This embedded defense includes important protection techniques like encryption, but unlike traditional encryption where keys must be managed, self-protecting data stores the (cloaked) keys inside the data file so they cannot be intercepted by an attacker.
- ***Data that can authenticate itself and its users:*** The embedded programmable code supports zero trust and allows only authorized users access to data and data files. Further, authentication protocols are independent of the underlying infrastructure of the networking environment, mitigating the possibility of an attacker hijacking or hiding in easily spoofed network protocols.
- ***Data that can enforce policy and take mitigative action in real-time:*** Because the code is part of the data/data file, not a protective layer around it, it provides an intelligent decisioning and controlling service at the data-layer to alleviate suspicious attempts to access information in the data-file, in real time, wherever the data is, and at whatever stage of the data lifecycle it's in. Policy enforcement and remediation at the data layer result in governance policies that follow data, wherever it goes. Examples of actions self-protecting data can take include:
    - Alerting the data owner of access or edit attempts
    - Providing only partial access to non-sensitive data files
    - Denying access
    - Requiring step-up authentication for access
    - Shredding the data (in extreme cases)
- ***Data that can track events, end-to-end, and assure its providence:*** From who the data owner is to how data files are being shared and stored, self-protecting data keeps records all throughout the data lifecycle and ensures data integrity.

EVALUATING DATA PRIVACY AND PROTECTION PLATFORMS

The total number of data records compromised in 2020 grew by 141% over the previous year. Correspondingly, the number of cloud breaches also continues to rise, due often to misconfiguration, inappropriate data access, and misuse (i.e., mis-delivery). As businesses shift more of their data to cloud environments — or create it there — the need to implement strong controls around data and data files is essential. Encryption, alone, can be difficult to manage and expensive, not to mention that circumventing encryption doesn't require special talents or techniques.

Today, a thorough data protection and privacy program incorporates not only a multi-layered approach -- including data-layer policies, transport -layer controls, and strict data access governance and reporting – but also a more granular approach. For example, how does one control selective access to important information within a file without at the same time compromising the PII in that file? Further, a data security strategy must incorporate elements of real-time visibility and reporting so that

data owners can adjust policies and controls when necessary.

When evaluating business applications for provable data privacy, the following questions will help determine which type of tool your organization needs:

1. *Risks* – How is your company currently assessing data risk?
    a. *How are you measuring your data risk?*
    b. *How much data do you have?*
        - *Where is it located?*
        - *How is it protected?*
        - *Who has access to it?*
    c. *How long would it take your organization to identify a data breach and quantify the scope?*
    d. *How are you creating and maintaining a data audit trail?*
    e. *What are the impacts to your business if certain types of data are lost, stolen, irreparably modified or unavailable?*
2. *Assets* – What tools and techniques for data security and privacy do you maintain?
    a. *How many tools/techniques do you need to use?*
    b. *Which data access technologies and processes are implemented?*
    c. *What data file access policies do you maintain?*
        - *How easy/hard are they to maintain?*
        - *How frequently do they need to be updated?*
        - *Do you have one place for central management of data file access -policies or do you need disparate systems for different data types and locations?*
    d. *What systems do you have for backup and recovery?*
3. *Compliance* – Which regulatory requirements are your organization subject to?
    a. *How are you meeting compliance requirements?*
    b. *How are you auditing compliance requirements?*
    c. *Are there additional industry standards (e.g., ISO 27701) that are required or desired?*

CONCLUSION

Given the amount of data in use at enterprises today, the requirements for such use, and the mandates imposed by regulatory bodies as well as employees, partners, and customers, enterprises must look for ways beyond encryption and access controls to protect data and ensure its resiliency. Self-protecting data is a new approach to data protection and resiliency that will enable organizations to safeguard the privacy of its stakeholders, its investments, and cloud deployments. By implementing a self-protecting data program, enterprises and SMBs alike will be better prepared for and able to recover from technological, security, and environmental disruptions, irrespective of the infrastructure in use.

ABOUT TAG CYBER

TAG Cyber is a trusted cyber security research analyst firm, providing unbiased industry insights and recommendations to security solution providers and Fortune 100 enterprises. Founded in 2016 by Dr. Edward Amoroso, former SVP/CSO of AT&T, the company bucks the trend of pay-for-play research by offering in-depth research, market analysis, consulting, and personalized content based on hundreds of engagements with clients and non-clients alike—all from a former practitioner perspective.