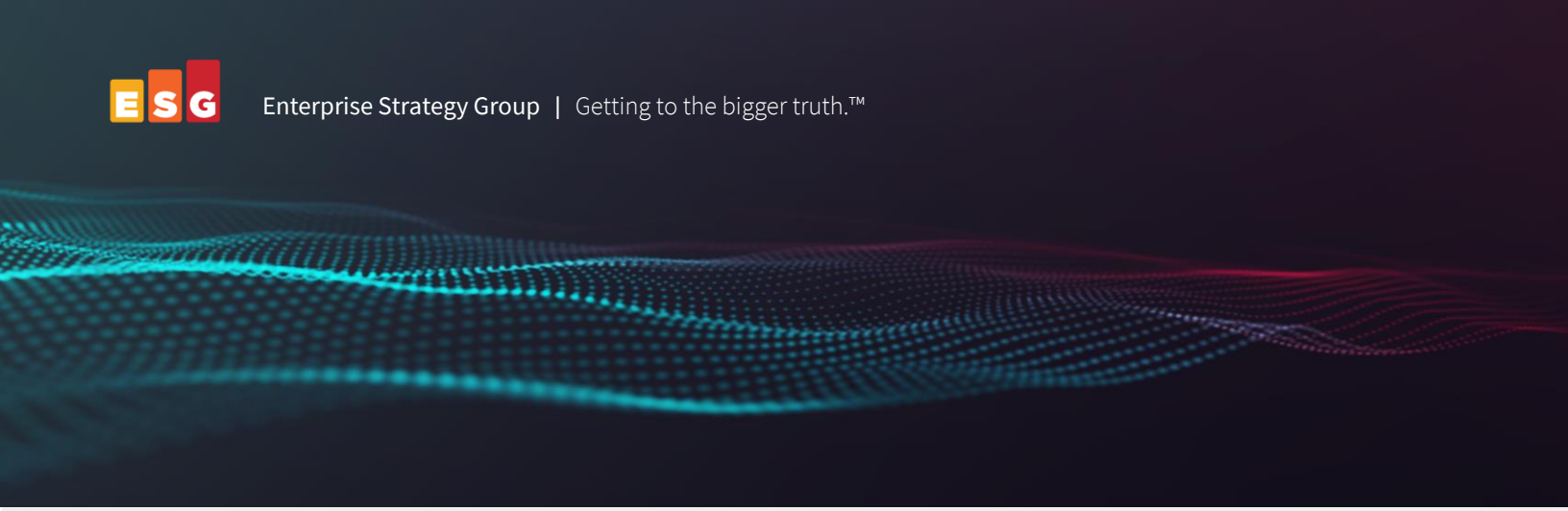


ESG WHITE PAPER

Enterprise Data Loss Prevention, Revisited

Integrated, Cloud-native Data Protection with Palo Alto Networks

By Doug Cahill, Vice President; and Group Director, John Grady, Analyst

October 2020

This ESG White Paper was commissioned by Palo Alto Networks and is distributed under license from ESG.

Contents

Executive Summary 3

The Data Matrix 3

- Cloud-resident Data is Increasingly Sensitive 3
- The Work-from-home Accelerant 4
- Understanding the Bad Actors and Causes of Data Loss 5

Rearchitecting Data Loss Prevention for Modern Workflows 7

- The Discovery and Classification Building Block 7
- Coverage Across Disparate Environments 7
- Protection of Data at Rest and in Motion 8
- Consistent Policy, Distributed Enforcement 8
- The Benefits of a Cloud-native Implementation 8

Palo Alto Networks' Enterprise Approach to Data Loss Prevention 9

- Integrated Control Points and Synchronized Policies 9
- Secure Access Service Edge (SASE) Integrated 9
- Coverage of Data Flows to and from Cloud Stores 10
- Machine Learning-driven Data Classification Engine 10

The Bigger Truth 11

Executive Summary

Digital transformation initiatives, which have resulted in the broad adoption of cloud services to enable more agile business operations, have impacted not only IT processes, but also cybersecurity and corporate compliance programs. Central to this transformation is the need to support and secure the use of a diverse portfolio of cloud applications.

To protect data assets both in motion and at rest, a modern approach to data loss prevention (DLP), designed for today's digital enterprise, is required.

Enterprises with branch offices have been revisiting network infrastructure to support direct-to-cloud workflows, the need for which has been amplified by the increase in remote work. As employees work from distributed locations, access to a diverse range of application data flows through egress points and through the edge direct to cloud stores. Collaboration with colleagues and third parties alike in which data is shared has introduced additional data channels, which are often outside of the purview of existing data protection controls. To protect data assets both in motion and at rest, a modern approach to data loss prevention (DLP), designed for today's digital enterprise, is required. This paper explores the challenges associated with protecting data in the cloud era, the requirements of a contemporary solution, and how Palo Alto Network's Enterprise DLP offering meets these requirements with a cloud-native implementation.

The Data Matrix

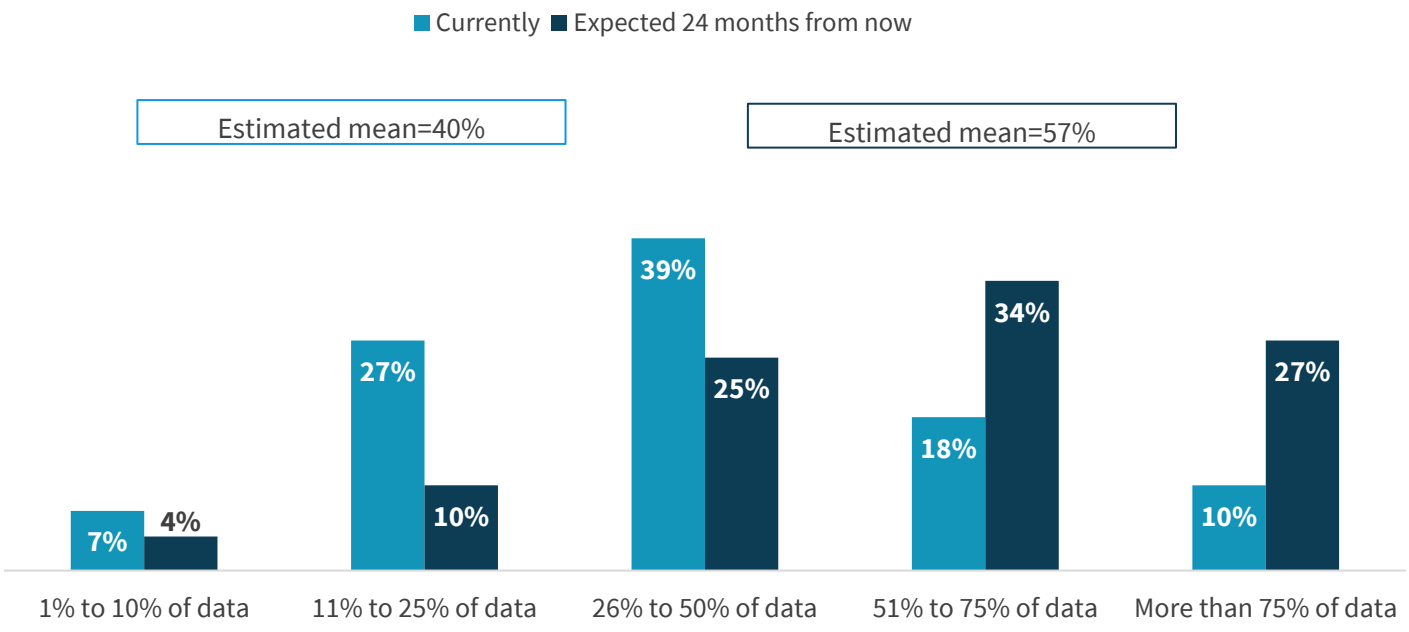
IT and cybersecurity teams are charged with securing an increasingly complex set of situations for how data is created, shared, and stored. Where data resides, how it travels, and who has access characterizes the data matrix.

Cloud-resident Data is Increasingly Sensitive

The complexion of cloud usage has evolved, with cloud applications now serving business-critical functions spanning the front, middle, and back-offices of digitally transformed companies. As a result, the percentage of an organization's data that is currently stored in a public cloud is appreciable and projected to increase. In fact, according to research conducted by ESG, on average, 57% of corporate data will be store in a public cloud in the next 24 months (see Figure 1).[1]

[1] Source: ESG Research Survey, *Cloud-driven Identities*.

Figure 1. Cloud-resident Data



To the best of your knowledge, approximately what percent of your company's data resides in any public cloud (e.g., in a SaaS service or on a PaaS/IaaS platform)? How will this change – if at all – over the next 24 months? (Percent of respondents, N=379)

Source: Enterprise Strategy Group

The diverse nature of an organization's cloud footprint means sensitive data assets are located across a variety of cloud stores spanning software-as-a-service (SaaS) applications and infrastructure-as-a-service (IaaS) platforms. In fact, our research respondents shared that 40% of the data associated with the use of SaaS applications is sensitive and 38% of the data related to their organization's use of IaaS platforms is also sensitive.[2] It is important to note cloud-resident sensitive data often includes that which is subject to industry regulations such as the Payment Card Industry Data Security Standard (PCI DSS) as well as an expanding set of data privacy laws include the European Union's General Data Protection Regulation (GDPR) and the recently enacted California Consumer Privacy Act (CCPA).

The Work-from-home Accelerant

Knowledge workers were already increasingly mobile and working from a variety of locations including insecure WiFi spots, making them susceptible to man-in-the-middle (MITM) attacks. Office closures as a result of the coronavirus pandemic have served as an accelerant for remote work and thus driven the need for direct access to cloud applications and data. As a result, data is in motion more than ever between home offices and cloud stores just as it was and will continue to be from branch offices.

Data is in motion more than ever between home offices and cloud stores just as it was and will continue to be from branch offices.

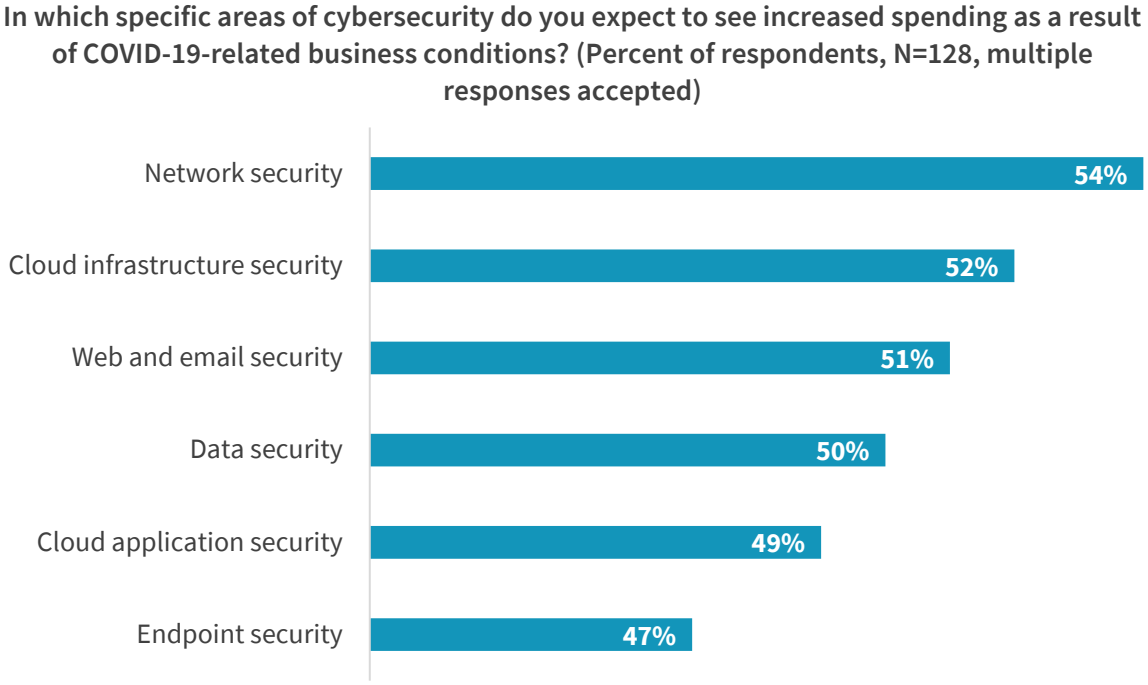
Work-from-home mandates have increased the use of cloud-based collaboration tools—office productivity, document sharing, video conferencing, and instant messaging. The use of online document sharing and collaboration tools alone has increased by 44% as a result of COVID-19. The improper use of collaboration tools in which sensitive data is shared via

[2] Ibid.

these communication channels can leave data exposed, as was the case in the well-publicized Twitter hack in which the password for a highly privileged account was shared in a Slack channel. IT and cybersecurity teams understand the risk, with 40% of organizations citing the security of collaboration platforms as the biggest challenge supporting an increase in remote workers, the top challenge cited. The focus on securing these platforms and the data that flows through them is, unfortunately, well founded. Nearly half of the respondents in ESG's research, 47%, reported an increase of cybersecurity attacks since the initial work-from-home period started.[3]

ESG's research indicates that remote work will be sustained even after the pandemic is over. To support distributed employees, branch offices, and the workers who return to corporate headquarters, investments are required to secure the multiple dimensions of today's perimeter. To do so, network security and data security, along with cloud infrastructure security, are the top areas in which organizations expect to increase cybersecurity spending as a result of COVID-19-related business conditions (see Figure 2).[4]

Figure 2. Top Six Areas of Expected Increased Cybersecurity Spending Due to COVID-19



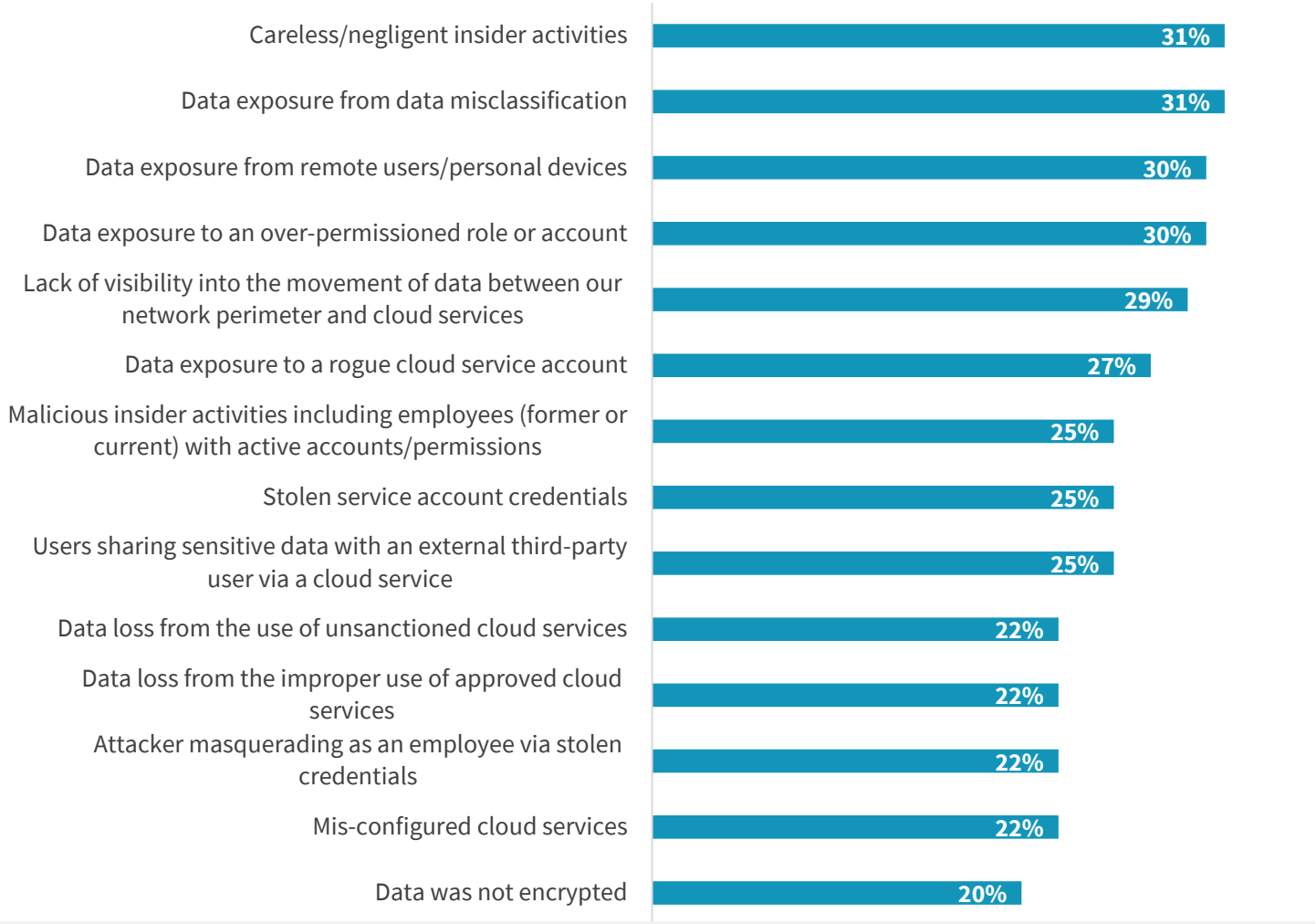
Source: Enterprise Strategy Group

Understanding the Bad Actors and Causes of Data Loss

Among the central aspects of the data matrix are the threat models that depict cybersecurity incidents, which can lead to data loss. That is, it's critical to understand not only who has access to what data assets via what applications and from which locations, but also which bad actors desire access to those same applications and data, why, and by what means. The incident rate of data loss from cloud stores makes understanding the threat landscape that puts such data at risk an imperative, per the 44% of organizations who shared they have, or suspect they have, experienced data loss from a cloud store due to a cybersecurity incident.[5] There are multiple causes of data loss, highlighted by a lack of visibility into data flows and the improper classification of data (see Figure 3).

[3] Source: ESG Master Survey Results, *Technology Impact of COVID-19: IT Decision Maker (ITDM) View*, May 2020.
[4] Ibid.
[5] Source: ESG Research Survey, *Cloud-driven Identities*.

Figure 3. Causes of Data Loss from a Public Cloud



Which of the following do you believe were the biggest contributors to your organization's public cloud-related data loss? (Percent of respondents, N=166, five responses accepted)

Source: Enterprise Strategy Group

The top causes of data loss from cloud stores reveals some themes:

- Negligence is not only related to careless end-user behavior, but also how cloud services are configured, including weak access control lists (ACLs) on cloud-based object stores.
- Credential abuse can take many forms, from the use of stolen credentials by a cyber-adversary to overly permissive user and service accounts.
- There is a clear visibility gap that results in data being exposed in motion as it moves not only through egress points but also directly to cloud stores from remote users. Another indicator of a lack of visibility is the misclassification of data, indicating that sensitive data is not always properly recognized as such, leaving it exposed.

One of the root causes of these issues is identity sprawl, the creation of user accounts, including privileged administrative accounts, specific to SaaS applications. These silos of identities are often not connected to an enterprise directory store

and thus are non-conformant with corporate policies such as roles that denote and enforce data access and usage rights. Bad actors are well aware of this dynamic and often spear phish employees with such credentials that then gives them access to, for example, customer data in a customer relationship management (CRM) SaaS application.

Rearchitecting Data Loss Prevention for Modern Workflows

The first response to securing an ever expanding and evolving use of public clouds is often to try to extend current data loss prevention controls into the cloud. Traditional controls, however, were designed for static, IP-based infrastructure, a fundamentally different environment than dynamic, metadata-based public clouds. These prior generation DLP products also necessitated the provisioning and ongoing maintenance of on-premises management server infrastructure, introducing cost and complexity.

Another approach is leveraging the data security controls provided natively by cloud service providers. A reliance on native controls, however, increases complexity. In fact, just over half of the participants in research conducted by ESG, 51%, cited increased organizational complexity as one of the top issues with using disparate tools that has had the most negative impact on their business.[6]

The broad adoption of public cloud services is necessitating a retooling of all aspects of an organization's cybersecurity program—people, process, and technology. Central to this effort is the shift from a siloed approach in which separate teams employ separate controls for separate environments to a unified posture. As such, modern DLP for today's cloud-centric enterprise must meet a number of core requirements.

The Discovery and Classification Building Block

Data is being created at a blinding pace. At the same time, organizations already have a massive corpus of data which, as noted, is, increasingly, cloud-resident. To focus on that data, which is sensitive, including data that is subject to regulatory compliance requirements and data privacy laws, the classification of both preexisting and new data is a fundamental building block of a data loss prevention initiative. Because misclassification is often the cause of the data loss, accuracy is critical. As such, high fidelity data classification is the underpinning of a data loss prevention policy lexicon.

High fidelity data classification is the underpinning of a data loss prevention policy lexicon.

A key contributor to assure accuracy of classification is context with respect to the associated business application, a user's role, and how data is being accessed, used, and shared. Fortunately, the use of machine learning in newer DLP technologies allow them to consider such contextual attributes when classifying data to operate at scale accurately.

For expedited time to value, a modern DLP product will offer a rich set of predefined data classifiers. Such data classification must also be extensible to enable the creation of custom classes of data tagged by the business to tailor a DLP implementation.

Coverage Across Disparate Environments

The definition of the office is now distributed across multiple locations. That is, the main campus of corporate headquarters, branch offices, and remote work locations, including home offices, now comprise the modern office. As a result, data that flows between these locations, and to and from a varied set of cloud applications and data stores, represents the multiple perimeters of today's hybrid multi-clouds. Because the perimeter is amorphous, a modern data

[6] Source: ESG Research Report, *Transitioning Network Security Controls to the Cloud*, August 2020.

security program must also consider the users, including third parties, their privileges, and the applications and data they use and access as the constant of how we now think of the perimeter. This reality of the data matrix means a core requirement of a modern enterprise data loss prevention (DLP) solution is coverage across locations, users, applications, and types of data stores (i.e., structured and semi-structured).

Protection of Data at Rest and in Motion

Data is transient—it travels through egress points and often lands in a cloud store. That egress point may or may not be a corporate managed and secured control point, given the increase in direct-to-cloud access. As such, a modern enterprise-class DLP solution will secure data in motion and at rest.

Securing data in motion—discovering, classifying, and applying policy—requires the ability to inspect content and enforce policy at line speed scale via highly performant and granularly selective SSL decryption. To protect data at rest in cloud stores, a contemporary DLP solution must support an array of cloud applications via integration with native APIs. API-based DLP also allows for the retrospective classification of data and retrospective application of policy for preexisting data.

Consistent Policy, Distributed Enforcement

The most common issue research respondents cited with regard to securing employee access to corporate applications and resources is ensuring *consistent* data security.[7] Consistency means mapping roles to data classes so that users with certain roles have the same level of access and usage policies by classes of data irrespective of location. But what about enforcement?

Because of the way data flows, policy enforcement needs to be distributed at the control points for data in motion and at rest, which requires enforcement at egress points, at the edge, and in the cloud. As such, a central aspect of moving from a siloed to a unified approach is the ability to federate policies across enforcement points.

Because of the way data flows, policy enforcement needs to be distributed at the control points for data in motion and at rest, which requires enforcement at egress points, at the edge, and in the cloud.

The Benefits of a Cloud-native Implementation

The trend toward software-as-a-service (SaaS) also applies to cybersecurity controls. In fact, nearly half of the organizations who participated in an ESG research study expect that at least 40% of their network security controls will be cloud-delivered within 2 years.[8] A modern DLP implementation will follow suit since it will be based on a cloud-native architecture. There are a few key reasons why cloud-delivered, cloud-native security controls have gained favor:

- Operational efficiency. SaaS applications eliminate the need to manage silos of management servers, including updating the backend management plane. This benefit extends to addressing the complexity associated with maintaining the proxies and databases required with prior generation DLP implementations.
- Scale. A true cloud-native, as-a-service DLP product will be built on a microservices architecture that allows it to scale as needed, a critical architectural consideration given the need for a modern enterprise DLP solution to inspect content and enforce policy at line speed.

[7] Ibid.
[8] Ibid.

- **Expedited time to value.** Delivered as a service means new updates, including new functional capabilities, are quickly available to customers, providing expedited time to value. This core value tenet of SaaS is in sharp contrast to the traditional approach of first testing a new release for compatibility and, under change control guidelines, planning the deployment of major and minor software releases.

- **Economics.** Modern enterprise DLP solutions also provide a cloud-native financial model, not just by enabling a shift from CapEx to OpEx, but also with metered, consumption-based pricing so subscribers to a DLP service pay as they go.

Palo Alto Networks' Enterprise Approach to Data Loss Prevention

Palo Alto Networks recently announced a data loss prevention offering, which meets the aforementioned architectural and functional requirements of a modern, enterprise-grade DLP solution. The product also effectively leverages the multiple form factors of the company's next-generation firewall and secure access service edge (SASE) for distributed enforcement of policy vis-à-vis the integrations cited below.

Integrated Control Points and Synchronized Policies

Central to the Palo Alto Networks DLP solution is a single DLP engine that delivers policies to control points, enabling a unified approach via policy consistency. Policies are automatically synchronized to distributed control points for enforcement at egress points, the edge, and in the cloud, as follows:

- **Egress:** The product leverages Palo Alto Networks' Strata family of next-generation firewalls, inclusive of physical and virtual form factors as the control points for policy enforcement at physical perimeters.

- **Cloud:** There is also integration with Prisma SaaS and Prisma Cloud for enforcement with cloud applications.

- **Edge:** Palo Alto Networks' secure access service edge (SASE) offering, Prisma Access, is also a control point for the enforcement of policy at the edge.

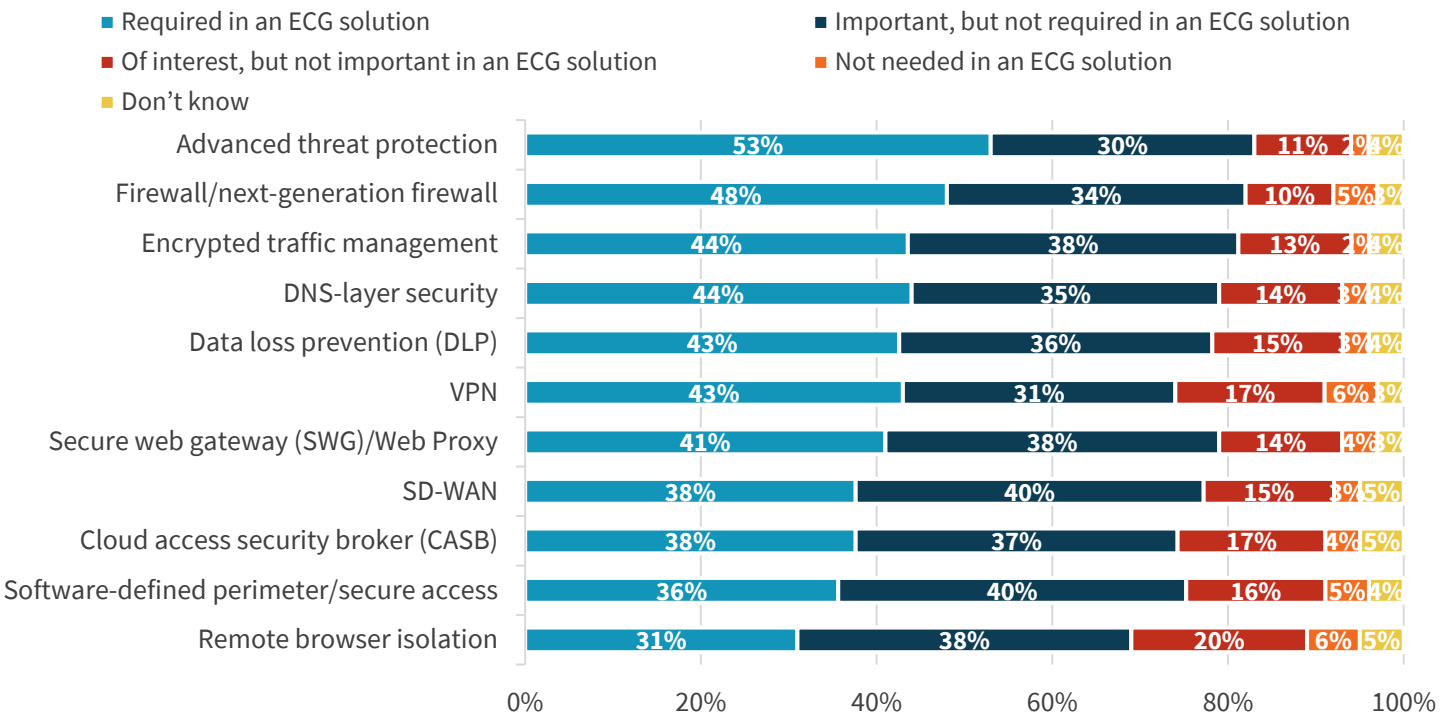
Secure Access Service Edge (SASE) Integrated

Integration is a theme with Palo Alto Networks' approach to enterprise DLP. That theme extends to Prisma Access, the company's multifunction SASE gateway, which integrates a series of controls into a set of cloud-delivered services. These services include zero-trust network access (ZTNA), firewall-as-a-service (FWaaS), SD-WAN, DNS Security, threat prevention, secure web gateway (SWG), cloud access security broker, and now, data loss prevention.

As a cloud-delivered platform comprising a rich set of services, SASE solutions will be employed by organizations in stages as a means to provide secure access to cloud resources from a variety of locations. As such, ESG suggests that customers evaluate a SASE offering in the context of an elastic cloud gateway (ECG) reference architecture. Recent research conducted by ESG on elastic cloud gateways highlights the importance of an integrated DLP service, with 79% of respondents citing DLP as either a required or important capability of an ECG-based SASE solution (see Figure 4).[9]

[9] Ibid.

Figure 4. ECG/SASE Requirements



Please rate the level of importance of each of the following security tools or functions as they pertain to a consolidated elastic cloud gateway solution. (Percent of respondents, N=376)

Source: Enterprise Strategy Group

On a related note, research respondents view encrypted traffic management as essential, a capability required for an integrated DLP engine to inspect content and one Palo Alto Networks provides in Prisma Access.

Coverage of Data Flows to and from Cloud Stores

Via the integration with the range of control points noted previously, the Palo Alto Networks Enterprise DLP solution provides the requisite of an enterprise DLP product by securing data in motion as it flows from remote users, branch offices, and those on the corporate network to other users and cloud-based data stores. Once at rest, the product also protects cloud-resident data via API-level integration with a range of prominent cloud brands, including unsanctioned and sanctioned cloud apps as well as public object stores. This list of cloud applications includes Dropbox, Facebook, YouTube, Office 365, Salesforce.com, G Suite, as well as object stores offered by AWS, Azure, and Google Cloud.

Machine Learning-driven Data Classification Engine

Palo Alto Networks Enterprise DLP solutions come with hundreds of predefined data patterns to categorize data based on standard classifications. These predefined classifications include those standard types of sensitive data such as personally identifiable information (PII) and protected health information (PHI), as well as others for the following purposes:

- Industry regulations: Payment Card Industry Data Security Standard (PCI DSS), Health Insurance Portability and Accounting Act (HIPAA), and Sarbanes-Oxley (SOX).
- Data privacy laws: General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA).

These out-of-the box classes can be augmented by the ability to create custom data classes, including integration with third-party data classification products.

As noted previously, misclassification of data too often leads to data loss. As such, high fidelity is critical, a requirement Palo Alto Network's DLP offering aims to meet via the use of machine learning algorithms trained in massive data sets so data can be quickly and accurately identified and classified.

The Bigger Truth

The nature of cloud usage has crossed a tipping point, serving fundamentally critical business functions with the data associated with that consumption intrinsically essential. At the same time, the expansion of the perimeter, inclusive of branch offices and a remote workforce who now rely on cloud applications, has challenged data security initiatives. As a result, cybersecurity teams face a strategic imperative—the need to protect distributed, in-flight data assets via a unified approach.

To meet this remit, organizations require an integrated approach that enables consistent policy management and distributed enforcement. Palo Alto Networks' Enterprise Data Loss Prevention (DLP) offering represents a modern implementation that leverages a cloud-native implementation to protect data assets by enforcing data access and usage policy where data moves and lives, at the perimeter, at the edge, and in the cloud.

Enterprise Strategy Group is an IT analyst, research, validation, and strategy firm that provides market intelligence and actionable insight to the global IT community.

www.esg-global.com contact@esg-global.com 508.482.0188