FROST & SULLIVAN

Automation and Comprehensiveness Form the Pillars of Fourth Generation Data Leak Prevention

An Executive Brief Sponsored by GhangorCloud

Michael Suby, VP of Research

INTRODUCTION

In today's digital age, data is an asset. Digital winners actively collect volumes of data and transform their undulating data oceans into competitive advantage. Yet the data leak prevention (DLP) technologies that enterprises have relied on to ensure their data assets and sensitive consumer and business partner data entrusted to them does not "walk out the door" with cyber thieves and malicious insiders may be significantly antiquated and inadequate. Older generational DLP technologies simply lack the automation, comprehensiveness, and integrity that today's data-dependent enterprises require.

Saddled with past generation DLP, enterprises are in heightened peril of data breaches and at a competitive disadvantage. Their DLP investments have not kept pace with the sophistication of cyber threats and, according to a survey conducted on Big Data Analytics (BDA), concern over inadequate data protection and privacy is a significant impediment to enterprise adoption and use of BDA solutions.[1] From this same survey, 88% also indicated that their BDA environments now or in the future will involve sensitive customer or company information. Outside the realm of BDA, common use cases raise similar data protection and privacy red flags, such as the movement of sensitive data into unsanctioned or unmonitored cloud services (e.g., file sharing) or as email attachments.

Saddled with past generation DLP, enterprises are in heightened peril of data breaches and at a competitive disadvantage.

Enterprises should assess their current DLP solutions. Assuming what is in place is the best available option, it is still missing a significant and favorable advancement in DLP—the Fourth Generation DLP. In this paper we start with a review of the evolution in DLP from First to Fourth Generation. We then add context on why taking action in assessing DLP is of growing importance, and we end with an introduction of GhangorCloud, a provider of a true Fourth Generation DLP.

GENERATIONAL EVOLUTION IN DLP

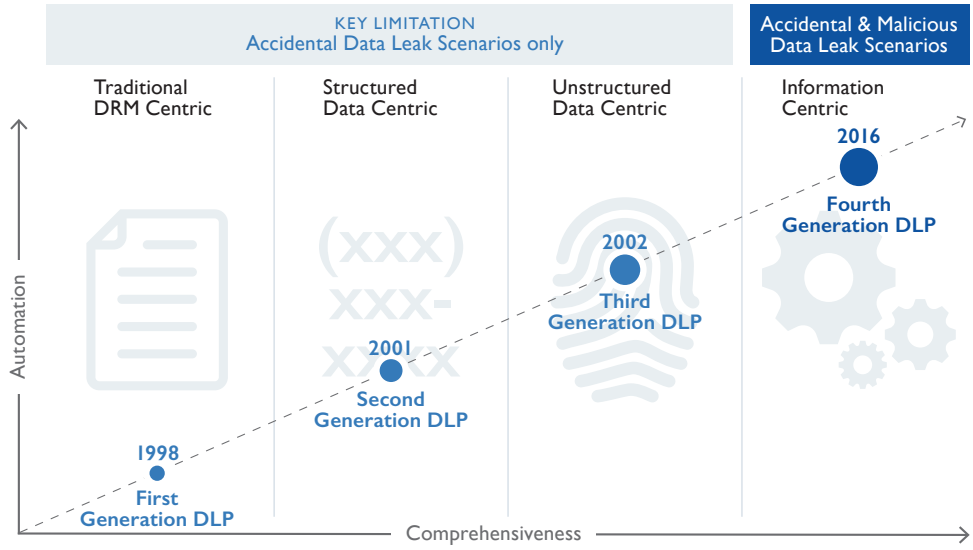
DLP is a nearly two-decades-old security category. While DLP has evolved, its evolution from the first through the Third Generation produced only incremental advancements. Significant limitations in comprehensiveness (i.e., data leak scenarios addressed) and automation remain.

As illustrated in Figure 1, the first through Third Generation DLPs have had a central focus on preventing accidental data leaks. These are the inadvertent scenarios, such as accidental selection of unintended email recipients or selecting the wrong file as an email attachment. In each, sensitive data is shared with recipients that are not authorized to receive the sensitive data. Also, in the circumstance of the email recipient being external to the enterprise (i.e., different email domain), the sensitive data moved from an enterprise-controlled environment to an uncontrolled environment.

1 In a 2017 Frost & Sullivan survey, 49% of the surveyed BDA decision makers categorized their concern about data quality, privacy, security, or audit compliance in BDA solutions as either 6 or 7 on a 7-point "No Impact" to "High Impact" scale. Conversely and reflecting widespread concern, a scant 5% selected 1, 2, or 3 in this 7-point scale.

...manual tendencies of pre-Fourth Generation DLP leave bare openings for malicious insiders and cyber thieves who have stolen credentials of legitimate employees to succeed in their data exfiltration exploits.

Figure 1: Evolution of DLP Solutions over Four Generations



Source: Frost & Sullivan

More precarious is the malicious insider scenario: an employee authorized to access sensitive data purposefully leaks it to the outside through any number of means, including sending the data to an external email account; uploading to an unsanctioned, cloud-based, file-sharing service; and copying to mobile and unsanctioned storage media (e.g., a USB thumb drive or CD). Adding further seriousness to this scenario is the case where a malicious insider may not actually be an insider, but a cybercriminal masquerading as a legitimate employee. Examples of this are Business Email Compromise (BEC) schemes. In one variety of BEC, the cybercriminal hijacks the email account of a high-ranking official within the enterprise to trick employees with authorized access into replying to the email with the requested sensitive data (e.g., W-2 information).[2] A non-routine request, the targeted employee nevertheless complies with the request out of a sense of duty to a superior. Because the email account is under the control of the cybercriminal, the sensitive data received is quickly exfiltrated.

Depending on the sophistication of earlier DLP generations and the enterprise's effectiveness in applying the technology, mitigating portions of malicious insider exploit is feasible. Even so, manual tendencies of pre-Fourth Generation DLP leave bare openings for malicious insiders and cyber thieves who have stolen credentials of legitimate employees to succeed in their data exfiltration exploits. The reasons for these openings include:

- **Manual processes are the antithesis of scalability –** Without scalability in an age of growing volumes of sensitive content, more sensitive content falls through the cracks and does not benefit from the protection capabilities of DLP. Consequently, pockets and pools of sensitive content are undefended against exfiltration exploits of malicious insiders and cyber thieves.

2 Based on its investigations, the FBI has noted a 1,300% increase in identified BEC exposed dollar losses since January 2015, *Business E-Mail Compromise:The 3.1 Billion Dollar Scam* (June 14, 2016).

- **Manual processes are subject to human error –** Whether the errors are systematic or the less detectable non-systematic (i.e., random), both contribute to sensitive data not being protected as it should.

- **Manual processes are dependent on employees acting in the best interests of the enterprise –** For example, "moles" within the enterprise could intentionally omit tagging sensitive data when they should. Also, in incidents where administrator controls and activity monitoring are weak, moles and cyber thieves can clandestinely modify user access permissions, and security policies and enforcement, to be less effective. All of these intentional malicious actions increase the potential of data breaches and difficulty in detecting breaches.

With past generation DLP (i.e., First Generation to Third Generation DLP solutions), manual tendencies span all DLP functions: (1) identifying and classifying sensitive data, (2) generating security policies, and (3) enforcing policy outcomes and responding to incidents.

- **Identifying and classifying sensitive data –** As DLP evolved from the First to the Third Generation, identification and classification of data expanded from files only (a Document Rights Management (DRM) approach), to structured data within and outside files (e.g., relational database fields and in email messages), and then unstructured data. Missing in each generation was full automation in identifying and classifying sensitive content. In the First Generation DLP, authoring linkages are used to identify and tag documents containing sensitive content—an approach that is not scalable and not automatically adaptive as data sensitivity changes. The Second Generation DLP relies extensively on regular expression and pattern matching algorithms to identify, tag, and classify structured data. While a logical approach, matching algorithms necessitates intervention (manual tuning) to inject context and differentiate between sensitive and non-sensitive data that correspond to the same regular expression or pattern. Also, matching algorithms are not viable for unstructured data. The Third Generation DLP came into the picture to address unstructured data. Yet, this generation reverted to fingerprinting the container (e.g., document) that holds the unstructured data rather than the data itself. Although a positive evolution, the Third Generation DLP did not fully overcome the technological challenges present in the First and Second Generations. As such, unreliable manual intervention is required for enterprises to extend data leak and exfiltration prevention to more of their sensitive content.

- **Generating security policies –** Unidimensional security policies are inadequate in balancing the needs of enterprise data-hungry operations against the risks and consequences of data leaks. Instead, policies should be multidimensional based on who (who the policy applies to), what (type of sensitive content), where (where the sensitive content is going), and how (the communication channel). Also, as each dimension's trustworthiness varies, the risk of data leak is determined by the real-time combination of all four. Policy generation, therefore, should accommodate this multi-dimensional risk fluctuation. However, when this type of multi-dimensional dynamism is not supported in older generation DLP, the enterprises are confronted with two unappealing options with both leading to the same outcome: manual intervention. One option is to set static policies that err on the side of caution with overly restrictive policies and then overlay exceptions to circumstantially lessen restrictions. The alternative option is to establish policies that are excessively lenient but incrementally overlay exceptions to reduce risk—risk that likely became known after an audit uncovered policy weaknesses or a data breach was attributable to weak DLP policies. With either option, as exceptions grow in number and tenure, they add to policy management complexity, time, and error.

- **Enforcing policy outcomes and responding to incidents –** Cross-system integration is common in the cybersecurity discipline. In DLP, cross-system integration varies in enforcing policy outcomes (e.g., allow, block, quarantine, and monitor) and gathering incident forensics and then responding. The implications on manual effort and human error also vary. For example, if the DLP technology is factory-equipped to seamlessly integrate with essential collaboration systems (e.g., proxies and security information & event management (SIEM) platforms), that lessens the cross-system integration the enterprise IT security team needs to create, deploy, and then monitor and maintain.

CIRCUMSTANCES COMPELLING ENTERPRISES TO RE-EVALUATE THEIR PAST DLP INVESTMENTS

The generational evolution in DLP is in response to circumstances that heighten risks to sensitive data. In this section, we zero in on the following key circumstances:

- Data leak and exfiltration attack scenarios are evolving;
- Data volumes are growing in size and dispersion; and
- Data protection regulations have teeth.

DATA LEAK AND EXFILTRATION ATTACK SCENARIOS ARE EVOLVING

Cybercriminals are never complacent. They are always evolving their tactics and techniques to compromise enterprise systems, take over user and administrator accounts, and exfiltrate data. Advanced Persistent Threats (APTs) are representative of this evolution. In APTs, the attacker's objective is to exfiltrate valuable information. To accomplish this, the attacker will use several techniques to succeed, such as social engineering to establish a foothold, obfuscation to evade detection during lateral movement, and account takeover to access sensitive data. Moreover, the attacker will modify any and all of its techniques to circumvent changes in the target's defensive shields.

The rapid increase in business email compromise schemes, as previously noted, is another example of this evolution. How malware delivery is morphing to evade detection is yet another example and one of particular concern as malware is a prominent tactic used in data breaches.[3] Two examples of this are firmware malware and weaponized user files. Uncovered in 2015 by McAfee, hackers are injecting malware into firmware.[4] Coming in below the software layers that anti-virus programs scan, malware firmware evades standard detection. In weaponized user files, malicious code is embedded into the structure of popular user file types (e.g., Microsoft Office documents and PDF files).[5] Also evading common detection approaches, the embedded malware moves among unsuspecting users as files are shared and opened.

3 Fifty-one percent of data breaches in the Verizon 2017 Data Breach Investigations Report included malware as a tactic.

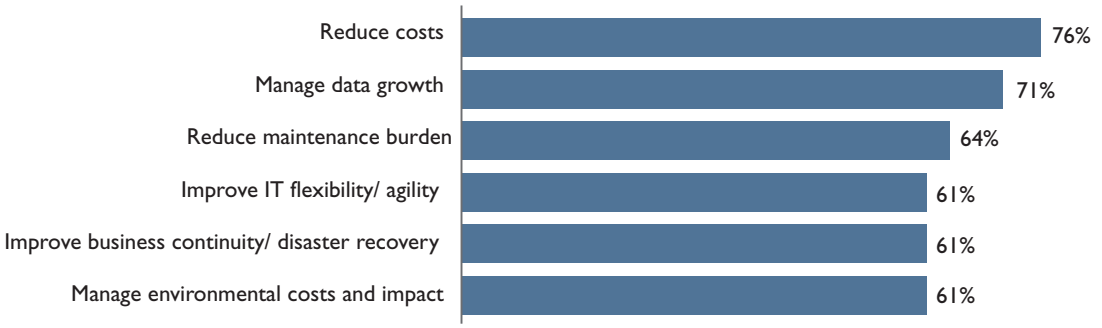
4 McAfee Labs Threats Report, June 2017.

5 A recent example of weaponized files is described in Votiro Labs exposed a new hacking campaign targeting Vietnamese organisations using weaponized Word documents (August 17, 2017).

DATA VOLUMES ARE GROWING IN SIZE AND DISPERSION

There is no doubt that data volumes are growing rapidly. Nowhere is this more evident than how enterprises are choosing to respond to data volume growth. According to Frost & Sullivan's latest Cloud User survey, 71% cited "managing data growth" as either an important or very important tactical driver in their cloud decisions (see Figure 2). This driver was second in priority to "reducing costs" (76%). Also telling of growing data volume is business network traffic. According to Cisco, global business IP traffic will grow 21% annually from 17,804 petabyte (PB) per month in 2016 to 44,452 PB per month in 2021.

Figure 2: Top-rated Tactical Drivers for Cloud



Percent of respondents citing driver as "important" or "very important" to cloud decision

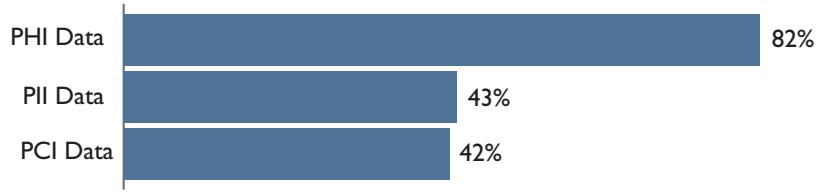
The adoption of cloud services also highlights the growing dispersion of data and the risks that follow. From the same Frost & Sullivan Cloud User survey, "unauthorized access to my data or applications" was the most often cited concern for not placing workloads in the cloud (65% citing as either important or very important). Despite these concerns, cloud adoption and use has risen rapidly, as seen in the cloud services of Amazon Web Services (AWS) and Microsoft; each exceed $15 billion on an annualized run rate and are growing at double-digit rates year-over-year.

Cloud apps (i.e., Software as a Service) further highlight the risks of data dispersion. According to cloud app analysis conducted by Symantec and reported in its 2H 2016 Shadow Data Report, 25% of all files stored in the cloud and 27% of all emails in the cloud are broadly shared, and 3% and 8% of these files and emails, respectively, contain compliance-related data. Even more disconcerting are these findings on sensitive data exposure with file-sharing cloud apps (see Figure 3):

- 82% of all Personal Health Information (PHI) in file-sharing cloud apps is exposed;
- 43% of all Personally Identifiable Information (PII) in file-sharing cloud apps is exposed; and
- 42% of all Payment Card Information (PCI) in file-sharing cloud apps is exposed.

Add into the data dispersion equation personal smartphones used in business and the extensive storage capacity and always-on connectivity of corporate-issued laptops and tablets, and the risk of data leakage mounts. All of these circumstances point to critical attributes needed in DLP solutions: agility and adaptability for rapid deployment, and scalability in an evolving and expanding IT footprint.

Figure 3: Exposed Data in File-sharing Cloud Applications



As learned by the industry the hard way, data breach discovery is the unfortunate consequence of a data breach that has already occurred. An equally sobering aspect of data breaches, particularly when they are perpetrated by a malicious insider or an external attacker that is masquerading as an insider, is breach discovery frequently does not transpire for months or even years. Verizon, in its 2017 Data Breach Investigations Report, determined that 15% of all of its investigated breaches (excluding miscellaneous errors) were attributed to insider or privilege misuse. In the vast majority of these data breaches, the time to discover took months (42% of insider and privilege misuse breaches) or years (39%). With this lengthy time to discover, prevention should be a higher priority.

DATA PROTECTION REGULATIONS HAVE TEETH

Data protection and privacy regulations have historically been a catalyst in enterprise use of DLP. The looming European Union (EU) General Data Protection Regulation (GDPR) adds further impetus for enterprises to evaluate their technologies and practices in data protection and privacy, and take fortifying actions as needed.

EU members developed and ratified GDPR for several reasons, including (1) rising concerns about escalating volumes of personal data collected, stored, and processed in digital form; (2) varied attention to data protection among data collectors and processors; and (3) the need to harmonize data privacy regulations. Reflecting these reasons, GDPR was created to provide individuals across the EU with stronger rights and controls over their personal data and bolster protections of this data from cyber threats and malicious insiders. The enforcement of this regulation takes effect on May 25, 2018.

Salient "call-to-attention" aspects of this dense, 88-page regulation are:

- **Both data controllers and data processors are subject to this regulation.**
- **The definition of protected personal data is extensive.** Personal data is any information relating to an identified or identifiable natural person that originates in the EU (by EU residents and EU visitors) or is of EU residents. Location of data controllers and processors and processing locations have no bearing on whether an enterprise must comply.
- **The penalties for non-compliance can be significant.** They can reach the greater of €20 million or 4% of global annual turnover of the preceding fiscal year.
- **Data breach notification timeframe is short.** Data controllers are required to notify the GDPR's supervisory authority within 72 hours after becoming aware of a personal data breach.
- **Non-compliance penalties are not the only monetary costs with GDPR.** As part of a post-breach investigation, the regulator will assess whether the enterprise utilized "state-of-the-art technologies" that could have prevented the breach. If determined that the enterprise did not, the enterprise risks having to make material and costly changes to its data use and handling infrastructure and processes as defined and mandated by the regulator.

With GDPR not yet the law of the land in EU but approaching quickly, its full impact on enterprises is not yet known with absolute certainty. What is certain is that GDPR raises the expectations on enterprises that are subject to this regulation and for those that do not heed these expectations, the costs can be significant. Projecting on the potential shadow effect of GDPR, if GDPR becomes a template for data protection and privacy regulation in other parts of the world, more enterprises will be impacted and need to take action. And similar to breach prevention being preferable to after-the-fact breach detection, building compliance-supporting technologies and processing within an enterprise's handling of sensitive data is preferable to penalties and retrofitting, or overlaying additional data protection technologies and modifying operations.

DLP FOR THE FUTURE

These outlined circumstances compelling enterprises to re-evaluate their past DLP investments are not reversing. More likely, these trends will strengthen over time. As such, businesses must not only re-evaluate the past, but define DLP capabilities and attributes needed now and into the future. Following is our list of those capabilities and attributes:

- **End-to-end, seamlessly integrated components –** Data identification and classification, policy generation, and policy enforcement should be a tightly integrated system rather than pieces of a loose-fitting puzzle. Independent components lead to operational inefficiencies and are speed bumps to maximum effectiveness—the type of laboring technology sprawl enterprise IT security organizations are encountering with greater regularity. Even so, DLP cannot be an isolated system. DLP is dependent on other components of the enterprises' broader IT and security environment in its functioning (e.g., user directories) and provides valuable and unique telemetry to other components (e.g., SIEM and incident detection and response). Therefore, DLP should also "snap together."

- **Automation in and across all three components: Data Classification, Policy Generation, and Control Enforcement –** Automation is essential in matching the speed of data creation and data movement, reducing error-prone and time-consuming manual processes—the very processes that open the door for malicious insiders—and building a culture that supports controlled and auditable use of sensitive data.

- **Full scenario protection –** The sensitive data types and data leak and exfiltration scenarios have expanded and will likely continue to expand. The evolution from First to Third Generation DLPs is proof of past expansion. Fourth Generation DLP incorporates the data type range of the previous generations with expanded focus on the data leak and exfiltration risk of malicious insiders and cyber criminals masquerading as legitimate insiders.

- **Transparency –** Optimal security is effective without interfering with legitimate business operations and end-user activities. Negatively, security's hindrances can limit its deployment and use, and can hamper end-user productivity to the extent that end users pursue risky workarounds. DLP, as a process-intensive operation that produces its best prevention when functioning in real time during data movement, should not produce noticeable latency or operational friction. Also, in the incidents when DLP policy enforcement suspends data movement, the reasons are unambiguous and decisive. In other words, false positives should be rare occurrences.

- **Administratively lightweight –** Security administrators have a daunting responsibility. Threats continue on the pathway of greater sophistication and subtlety, data-dependent operations are more

dynamic, regulations are growing and intensifying, and the range of technologies to manage is expanding. DLP should not add effort and complexity to administrators. As enterprises move from older generations of DLP to Fourth Generation, the outcome should be better data leak and exfiltration prevention with lower administrative effort. Integrated comprehensiveness and automation underpin administrator ease, effectiveness, and productivity.

- **Pervasively deployable –** Data movement contributes to DLP blind spots unless the DLP technology is pervasively deployed at the strategic junctures of data movement and captures the full spectrum of data-use actors: end users and their affiliations, on-net and off-net devices, and applications. In a digital world increasingly consisting of humans, robots, self-service cloud services, and automated but dynamically adaptable systems (i.e., machine learning), the DLP technology must be inline wherever there is movement of sensitive data.

INTRODUCING GHANGORCLOUD: A FOURTH GENERATION DLP

GhangorCloud's Information Security Enforcer (ISE) is an innovative approach to DLP and one that epitomizes a Fourth Generation DLP as well as the DLP capabilities listed in the previous section. Four architectural elements form ISE's DLP prowess: inline processing, end-to-end automation, accidental and malicious insider scenarios protection, and centralized management.

ISE operates in line with data flows. Deployed next to enterprise data repositories and applications, and as an agent on end-user devices (i.e., PCs, laptops, tablets, and smartphones), ISE interrogates network data flows and the data flows within end-user devices. From this strategic vantage point, ISE conducts real-time identification and classification of sensitive data and enforces DLP policies.

ISE's automated identification and classification consists of semantic analysis for identifying unstructured data, and a combination of regulatory expression, pattern, and keyword matching algorithms for structured and semi-structured data. With its real-time, touchless data identification and classification, ISE virtually eliminates the need for manual data classification pre-tagging. Of equal importance, all data flowing through ISE is evaluated for sensitivity, even "virgin" data.

Leveraging ISE's inline data flow position in enterprise networks and on end-user devices, DLP policies are enforced based on four dimensions gathered and synthesized in real time:

1. **Actor –** Who the policy applies to, an individual or a device, and the actor's pre-defined organizational role (e.g., integrated from enterprise user directories).
2. **Auto-data Classification –** What the content is, as determined through the real-time Auto-identification and Auto-classification process without any human intervention or pre-tagging.
3. **Source & Destination Categorization –** Where the content is being sent and/or from where it originates.
4. **Communication Channel –** How data is transmitted (e.g., communication protocol).

DLP policies, correspondingly, are automatically generated from these four dimensions. ISE comes equipped with default policies based on these dimensions and enterprises can further tailor to suit their unique needs.

ISE is built on a unique paradigm that combines and addresses the traditional data leak scenarios and the emerging sophisticated data exfiltration attacks (e.g., via Advanced Persistent Threats). Also due to its inline, real-time data

flow processing, ISE protects enterprises from malicious insider data leak and exfiltration risk. Since ISE virtually eliminates the need for manual data tagging, an exploitable operation for malicious insiders ceases to exist (i.e., malicious insiders intentionally not tagging data or documents that should be tagged, or tagging them as less sensitive). Furthermore, since policy is independently enforced on hard-to-conceal actions of actors, the means for malicious insiders and cyber thieves that have stolen end-user credentials to exfiltrate data is narrowed.

ISE's Centralized Command Control Collaboration and Intelligence (C4I) module is a full-spectrum CyberSecurity Operations Center (CyberSOC). It provides administrators with a complete set of tools to configure, monitor, and control all functional aspects of ISE. C4I also provides administrators and security analysts with real-time incident forensics and response capabilities.

STRATECAST: THE LAST WORD

Limits on scenario protection in past generation DLPs and their manual requirements have restrained DLP deployments to specific areas of the business and to enterprises that could justify the personnel needed to be intimately involved in DLP operations. This situation is no longer tolerable. Malicious insiders and cyber thieves will find sensitive content and holes in DLP protection, regardless of company size. Moreover, regulations and post-breach damage have no bearings on company size. At the same time, large and small enterprises do not have the luxury to double up on their DLP expenditures, even as their sensitive data volumes grow and the location of that data becomes more dispersed.

Fourth Generation DLP, as offered by GhangorCloud, offers a fresh and distinctive approach. By automating and conducting data identification, classification, policy generation, and policy enforcement in real time as data moves through the enterprise network and within end-user devices, manual operations are significantly lessened, gaps in data leak scenarios narrowed, and more pools of sensitive data are prevented from leaking. For enterprises that have First, Second, or Third Generation DLP solutions in place and enterprises that can no longer be without DLP, now is the time to take action in reviewing the merits of Fourth Generation DLP.

Michael Suby
VP of Research
Stratecast | Frost & Sullivan
msuby@stratecast.com

NEXT STEPS

Schedule a meeting with our global team to experience our thought leadership and to integrate your ideas, opportunities and challenges into the discussion.

Interested in learning more about the topics covered in this white paper? Call us at 877.GoFrost and reference the paper you're interested in. We'll have an analyst get in touch with you.

Visit our **Transformational Health** web page.

Attend one of our **Growth Innovation & Leadership (GIL)** events to unearth hidden growth opportunities.

SILICON VALLEY
3211 Scott Blvd
Santa Clara, CA 95054
Tel 650.475.4500
Fax 650.475.1571

TORONTO
2001 Sheppard Ave. E, Suite 504
Toronto, ON M2J 4Z8 Canada
Tel: +1.416.490.1511
Fax: +1.416.490.1533

BUENOS AIRES
1061 Luis Maria Campos, 9th Floor
Buenos Aires, Argentina
(C1426BOI)
Tel +54 (0) 11 4777-1550
Fax +54 (0) 11 4777- 0071

877.GoFrost
myfrost@frost.com
www.frost.com

Frost & Sullivan, the Growth Partnership Company, works in collaboration with clients to leverage visionary innovation that addresses the global challenges and related growth opportunities that will make or break today's market participants. For more than 50 years, we have been developing growth strategies for the Global 1000, emerging businesses, the public sector and the investment community. Is your organization prepared for the next profound wave of industry convergence, disruptive technologies, increasing competitive intensity, Mega Trends, breakthrough best practices, changing customer dynamics and emerging economies?