Managing PHI in the Cloud
Best Practices

Executive Whitepaper

Recent advances in both Cloud services and Data Loss Prevention (DLP) technology have substantially improved the ability of healthcare organizations to manage Personally Identifiable Health Information (PHI) in the Cloud while maintaining compliance with HIPAA and other regulatory requirements. This paper reviews what to look for in selecting and utilizing these advances. It is addressed to Healthcare Industry executives responsible for assuring the proper control of their organization’s regulated data.

Code Green Networks has been providing superior Data Loss Prevention solutions to the Healthcare Industry since 2004.

©2015 Code Green Networks.
Executive Summary

Benefits available from Cloud computing have increased significantly over the last year. Not only do these include the initial motivations of cost reductions and ease of scaling through the sharing of technology. Now the Cloud’s ability to significantly improve the sharing of information is understood and is rapidly being implemented. Examples in healthcare range from collaboration in medical research to providing for the patient across multiple locations and time zones.

However, in order to realize these benefits, healthcare organizations must meet the regulatory standards necessary to insure protection of sensitive data. In particular, this includes protecting Personal Health Information (PHI) from malicious or inadvertent exposure. With this in mind, IT and security executives are concerned that they may lose track of their data if they implement a cloud-based architecture. To address these issues, several important product advances are now available that extend proven Data Loss Prevention (DLP) technology to the protection of data across the enterprise, including the Cloud.

This white paper discusses how best to protect PHI data in the Cloud. First it reviews the business and technology benefits of using the technology and then summarizes some of the obstacles that healthcare executives must consider before implementing a cloud-based architecture. Next this paper focuses on the role DLP can play in protecting healthcare data in the Cloud. It reviews the technology and provides guidance on how best to implement this solution in a healthcare organization.

Why Move to the Cloud?

Cloud Computing Defined

In 2011, The National Institute of Standards and Technology (NIST) published this definition: “Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.”

Sharing information

Initially, the cost savings and scalability benefits from sharing a pool of computing resources were sought by many organizations having limited IT resource. Now, however, with appropriate security, the cloud is seen by the Healthcare Industry as a means to enable more efficient and productive sharing of information between multiple care providers, their associates and with the patients as well. The Cloud provides a means to facilitate collaboration, data exchange, and improved workflow from research to the delivery of care itself.
Following are representative examples:

**Collaboration** - Sharing current studies or intellectual property among faculty, staff and researchers through the Cloud, can dramatically improve the flow of information and accelerate progress.

**Coordination** - Many parties communicate and exchange information while caring for patients or billing and accepting payment for care. The Cloud can facilitate productivity in these processes by providing immediate access to required documents.

**Clinical Information** - The Cloud can help coordinate care between teams in a variety of locations including ambulatory settings where this is often difficult. Files may be speedily handed off from one medical professional to another.

**Centralize Information** - Clinical journals, peer-reviewed articles in development and grant projects may all be centralized in the cloud in one place and tracked in real-time.

**Patient Communication** - Healthcare organizations can deliver personalized content directly to patients through the Cloud. Patients can consume, interact with and assemble follow up questions for care teams—all on cloud.

**Research** - University hospitals and many other organizations conduct research across their internal departments as well as with external researchers, vendors and partners. The Cloud can help collaboration across multiple departments as well with outside contracted research partners. Using the cloud multiple research locations can upload study results or lab work, organizing topics into folders, adding comments, assigning tasks, and searching quickly for previous data. This keeps everyone in sync on status of projects and required resources.

**Teaching** - The Cloud can be utilized to make continuing medical education more efficient, collaborative, and rewarding. Learning materials, videos, publications, and case discussions may be accessed using tablets, mobile phones, or laptops.

**Obstacles Preventing Cloud Migration**

**Visibility Concerns**

Concerns over the loss of visibility of data in the cloud have remained a barrier to many organizations seeking these benefits. Security officers and top executives are asking “how can I protect sensitive patient information or our own proprietary data if I don’t know where exactly that information is?”
**Regulation Confusion**

The improper release of regulated patient medical records can result in fines and penalties from the Health and Human Services regulatory agencies as well as a drop in confidence by patients and the general public. Maintaining compliance with these regulatory acts is a vital concern for every organization handling electronic health records.

**Encryption is not Enough**

HIPAA regulations require, as just one example, that, whether data is encrypted or not, the organization must know where patient data is stored or sent. In other words, encryption, alone, is not enough to meet the necessary standards or to provide the visibility required to govern this type of sensitive data.

**Business Associates**

In September of 2013 the HIPAA Omnibus ruling took effect which, among other guidance, provided significant clarity in spelling out what is required of the Business Associates of a health care organization. In a nutshell, they are required to provide the same protection to personal health information as a covered entity.

**PHI Discovery: Things to Consider**

**Be Wary of Incomplete Solutions**

Many products currently exist in the market that can help scan for PHI in outbound email and web transactions, on internal file stores, and exiting to external devices. But these solutions are often incomplete and may be unsatisfactory in their ability to accurately detect PHI wherever it may be, and, particularly, in the Cloud.

**Methods of Detection are not all Equal.**

Traditional methods used to find sensitive information are prone to false positives and false negatives, typically have high staffing costs, and often result in drawn-out implementations that don't lead to accurate actionable detection. The methods most commonly employed are usually combinations of pattern matching (e.g., looking for 9 digits in the usual SSN xxx-xx-xxxx format) or dictionary matching (e.g., looking for specific words such as diagnosis names). Unfortunately, these methods will present a tradeoff to users forced to choose between: a) wading through a sea of false positives, or, b) allowing more false negatives in order to avoid getting overwhelmed with manual inspections.

History shows that organizations will commonly opt to not spend an inordinate amount of time manually weeding out false positives.
and, instead, simply accept failing to identify significant instances of regulated information.

Many compliance solution vendors have simply chosen to err on the side of accepting false negatives in order to reduce the burden on the user to have to check false positives. Their logic being that many organizations handling PHI will not be aware of, won’t see, and, therefore won’t complain, about, false negatives – until such time as a loss of data becomes public.

Despite these problems, each of these methods may be applicable in limited use cases. However, significantly stronger methods are now available to more accurately detect sensitive data.

**Implementing Cloud Services**

**Compliance Signatures**

Make sure that your cloud services provider supports HIPAA and HITECH compliance and will sign your HIPAA Business Associate Agreements (BAA). The cloud provider should provide encryption of data in transit and at rest, offer a full audit trail on user access to content, offer multiple levels of controls to set access to files, have state-of-the-art practices for identity management, and offer reporting and analytic tools to get up-to-the-minute information on what’s happening with your content.

**Eliminate Unsecure File Sharing Solutions**

Replace the inherent risks of corporate email, FTP, USB drives and other vulnerable consumer-grade document sharing technologies with secure, connected cloud access.

**Enable Mobility**

Provide doctors, researchers and administrators with quick, secure access to critical information on any device—in an exam room, an office or even at home. Enable collaboration seamlessly on any device while files contain sensitive data stay secure. Make sure Cloud services integrate with your mobile device management solutions so data is available, while maintaining the control and security that your healthcare organization requires.

**Data Loss Prevention**

Make sure that your cloud provider supports an appropriate DLP solution that will allow uniform application of information policies across the entire enterprise network, including the Cloud.
Advances in Data Loss Prevention Technology

Data Loss Prevention, DLP, technology has been employed over the past ten years in successfully reducing risks from loss of sensitive data. In particular, considerable experience has been gained in DLP applications managing regulated health care data.

Not all DLP Offerings in the Market are Equal

Because of its unique advantages and powerful capabilities, “Content Aware DLP” is often referred to as “Enterprise DLP”. Gartner, Inc. has provided this definition in its IT Glossary:

“Content-aware data loss prevention (DLP) tools enable the dynamic application of policy based on the content and context at the time of an operation. These tools are used to address the risk of inadvertent or accidental leaks, or exposure of sensitive enterprise information outside authorized channels, using monitoring, filtering, blocking and remediation features.”

In addition, more recent DLP technology, now available, is capable of extending this coverage to the cloud as well as the enterprise’s own centers and networks.

Implementing Cloud Services

Employ a DLP Solution Covering the Entire Network

By selecting a DLP solution that provides coverage across the enterprise including cloud storage, the organization’s ongoing management of regulated or other sensitive information is greatly simplified. Policies will be enforced with consistency and from single administrative control in all locations.

Employ a DLP solution Addressing Important Healthcare Needs

Some examples:
- Policy Templates specifically designed for HIPPA and HITECH compliance
- Unique protocol handlers to identify HL7 v2, HL7 v3, or e-PHI transmitted over X12
- Specialized connectors to all major Electronic Health Record Systems
- Built in healthcare code sets (e.g. HCPCS, ICD-9, ICD-10, LOINC, and NDC)
**Employ a DLP Solution that Accurately Detects Sensitive Data**

Database Record Matching, for example is a method of using the actual sensitive data in inspecting other sources such as an email, a file share, the cloud, a web posting; anywhere that information would be problematic if found there.

Medical Record or Medical Insurance policy “numbers”, for instance, don’t follow the same format across institutions, leading to an array of differing possible formats. Some of these are numeric. Some are alphanumeric. And, they can differ in length and other formatting aspects. Thus, pre-built pattern matching solutions can require significant tailoring to produce reliable results with such data. In contrast, DBRM, by its nature, easily provides an extremely accurate means to detect an actual ID in all inspected alphanumeric forms.

**Operating in the Cloud**

**Discovery and Audit**

Discovery Data Loss Prevention is a key tool to locate, identify and secure sensitive data throughout the network. An appropriate DLP, using advanced technology will locate and identify sensitive data at rest on endpoints and servers across the network and in the cloud providing visibility and audit reporting of potentially unsecured information. Once discovered the decision to Alert, Move, or Remove the data can be made.

For efficiency it may sometimes be appropriate to scan entire file systems when there are uncertainties regarding content. Or, the file systems may be so complex that it is desirable to scan them prior to the uploading transmissions which will look at each record at a time as it is sent. Before release to the cloud, sensitive information may be denied passage or automatically encrypted. Or, other proscribed remediation may be applied.

**Apply Remediation Selectively at Each Step**

It may or may not be most effective to encrypt everything sent to the cloud. An appropriate DLP will allows, at every stage in the process, the appropriate remediation to be automatically applied according to the policies established by the enterprise for that particular information and where it is being stored or transmitted.
Cloud Brings Sharing; DLP Provides Protection

Data Loss Prevention has proven to be an invaluable resource in protecting regulated PHI even as advances in technology have migrated information from secure data centers to distributed file servers to the desktop and to mobile computing devices. Now, this technology may be applied to help manage the sharing of data in the cloud as well.

There are many resources to assist organizations in sorting out the options available for data protection. But, the most important criteria for a healthcare organization is to evaluate solutions that will apply consistent and uniform policy enforcement to patient data across the entire enterprise, no matter where it is stored, including cloud storage. A good way to evaluate the best course of action is to ask any vendor for a proof of concept on site.

Code Green Networks delivers solutions that help enterprises protect and manage regulated and other sensitive digital information across their data network, whether local, remote, mobile or in the cloud. The company’s solutions have been tested and proven through daily use by hundreds of deployments in large and small organizations across the United States and around the globe. By working closely with customers since 2004, Code Green Networks has applied innovative technology to produce Data Loss Prevention solutions with the most advanced capabilities available to locate, identify and manage regulated data.

Contact Code Green Networks
385 Moffett Park Drive, Suite 105, Sunnyvale, CA 94089
Phone: +1 (408) 716-4200
Email: info@codegreennetworks.com
www.codegreennetworks.com

TrueDLP™, Database Record Matching™, DBRM™ are trademarks of Code Green Networks.