GLOBAL CLOUD DATA SECURITY REPORT

THE ESSENTIAL REFERENCE GUIDE: HOW FINANCIAL SERVICES FIRMS ARE PROTECTING THEIR DATA IN THE CLOUD - 2016 EDITION

CipherCloud®
Trust in the Cloud™

TABLE OF CONTENTS

Executive Summary 03

Data Taxonomy 04

Encryption vs. Tokenization 06

- Encryption 06
- Tokenization 06

Encryption Algorithms 08

- Search and Sort of Encrypted Data 08

Protecting Sensitive Data 09

- Protecting Highly Sensitive PII 09
- Protecting Regular PII 11
- Protecting Personal Financial Data 15
- Protecting Business-Sensitive Data 16

Choosing the Right Protection Schemes for your Data 18

Conclusion 19

EXECUTIVE SUMMARY

As cloud adoption continues to expand in enterprises, so does the requirement to protect data in the cloud while supporting business operations. CipherCloud recently conducted an in-depth study across our customer base in the banking and financial services market to investigate how firms in this industry sector protect their data. We looked at which data they protect and how they protect it. The goal is to identify best practices and provide fact-based recommendations to enterprises in the financial services sector.

Our data pool consists of greater than 50 global companies, including banking, investment, wealth management, and financial services. Some of them are fortune 500 companies and all of them are large companies with critical data protection needs.

Our study yielded the following high-level findings:

- **40% of firms with highly sensitive personal-identifiable information use tokenization for protection:** As sensitivity of data goes up, so is the tendency of using tokenization and strong encryption schemes. Tokenization is used progressively less as the criticality of data decreases.
- **64% of firms use searchable encryption to protect sensitive data while supporting business workflows:** Firms trade-off searchability with security strength when the data is used in enterprise workflows to reference or index to other information.
- **Format preserving is key for special-structured data.** Protecting data such as email, URL, phone numbers, where the data has a specific recognizable format requires format-preserving protection, above all other constraints.
- **Protecting sensitive data while supporting business operations is a balancing act.** Depending on the nature of the data – format, sensitivity, and business use cases – you may choose a very different protection scheme. CipherCloud's data protection framework, derived from real-world deployment experiences, guides enterprises through this complex decision making process.

DATA TAXONOMY

Across CiperCloud's install base in the banking and financial services sector, which includes banking, investment, wealth management and finance services, customers are protecting four broad categories of data. They are:

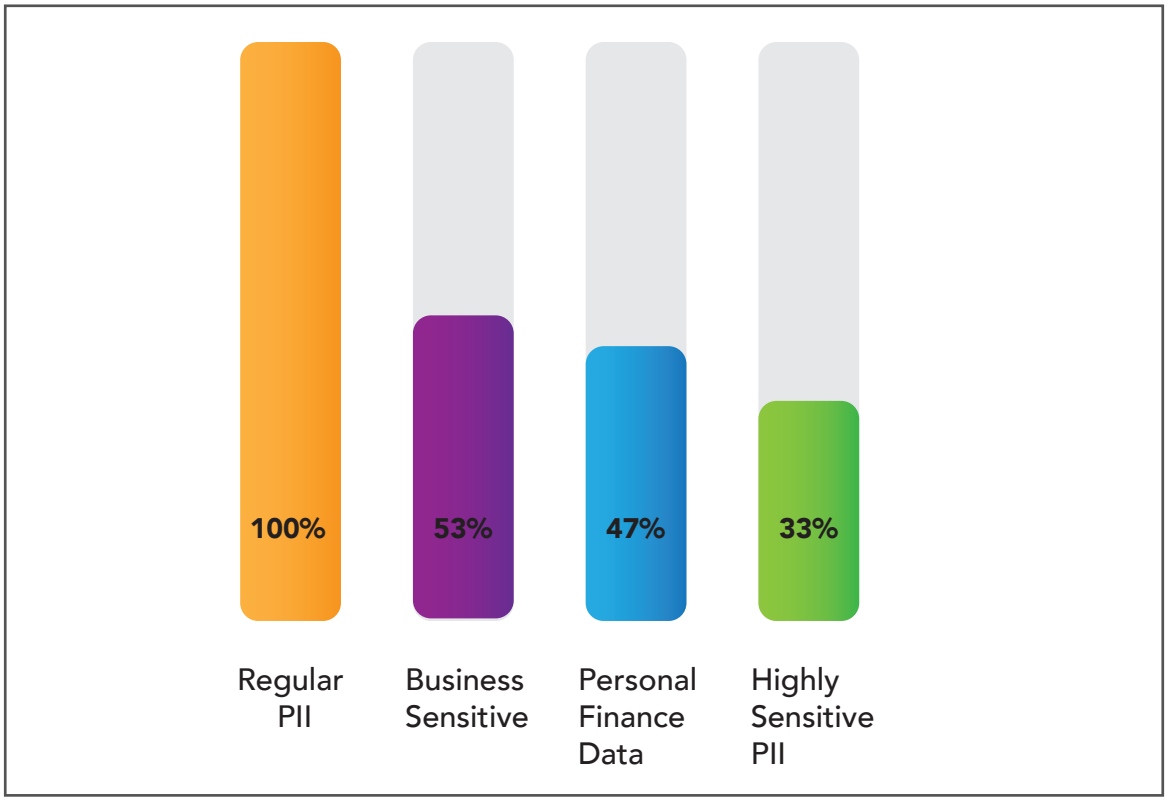
- **Highly sensitive PIIs:** Data in this category are PIIs that are highly sensitive and carry a steep penalty if leaked or compromised. Examples include social security numbers, birthdays, TaxID, banking/trading account numbers, etc.
- **Regular PIIs:** These are personal identifiable information such as names, addresses, phone numbers, emails, physical addresses, regions, etc.
- **Personal finance data:** In this category are individual financial data such as account balance, loan number, financial transaction information, and account statements, etc.
- **Business sensitive data:** In this category are information such as customer comments, activity summary, interaction emails, or business process information.

We asked customers to classify their data using the above taxonomy. It's interesting to note that while in many cases sensitivity of data are universal across different companies, sometimes a piece of "highly sensitive" data at one company can be simply "regular PIIs" at another. A good example is "Name". While many organizations treat Name — first name and last name as regular PIIs, some consider them highly sensitive when they are associated with private banking customers or high net-worth customers.

We use this data taxonomy throughout this report and base much of our analysis on this framework of data sensitivity.

Within the CipherCloud banking and financial services customer base, we see that 100% of the firms use CipherCloud solutions to protect regular PIIs, 53% use CipherCloud to protect business confidential data, 47% personal finance data, and 33% highly sensitive PIIs (see Figure 1).

Figure 1: Distribution of Firms with Protected Data Types



This is fairly consistent with what we see from other data sources. Not every financial services firm collects and maintain highly sensitive PIIs such as social security numbers and birthdays; not everyone deals with personal finance data or puts them in the cloud, but almost every firm has some form of PIIs such as customer contact information – emails, phone numbers, addresses and zip codes.

ENCRYPTION VS. TOKENIZATION

To protect sensitive data in the cloud, Encryption and tokenization are two of the most commonly used protection mechanisms.

ENCRYPTION

Encrption is the process of encoding messages or information in such a way that only authorized parties can read it. In encryption, the original text is referred to as the "clear text" or "plain text" while the encrypted data is "cipher text". The transformation from clear text to cipher text and back is governed by a mathematical algorithm, which typically involves a secret, commonly referred to as the "key". The cryptographic key must be protected for anyone who possesses the key can decrypt the cipher text and reveal the encoded data.

TOKENIZATION

Tokenization is the process of substituting a piece of sensitive data with a random equivalent, referred to as a token. The token has no extrinsic or exploitable meaning or value. In tokenization, there is typically no mathematical relationship between the protected data and the token.

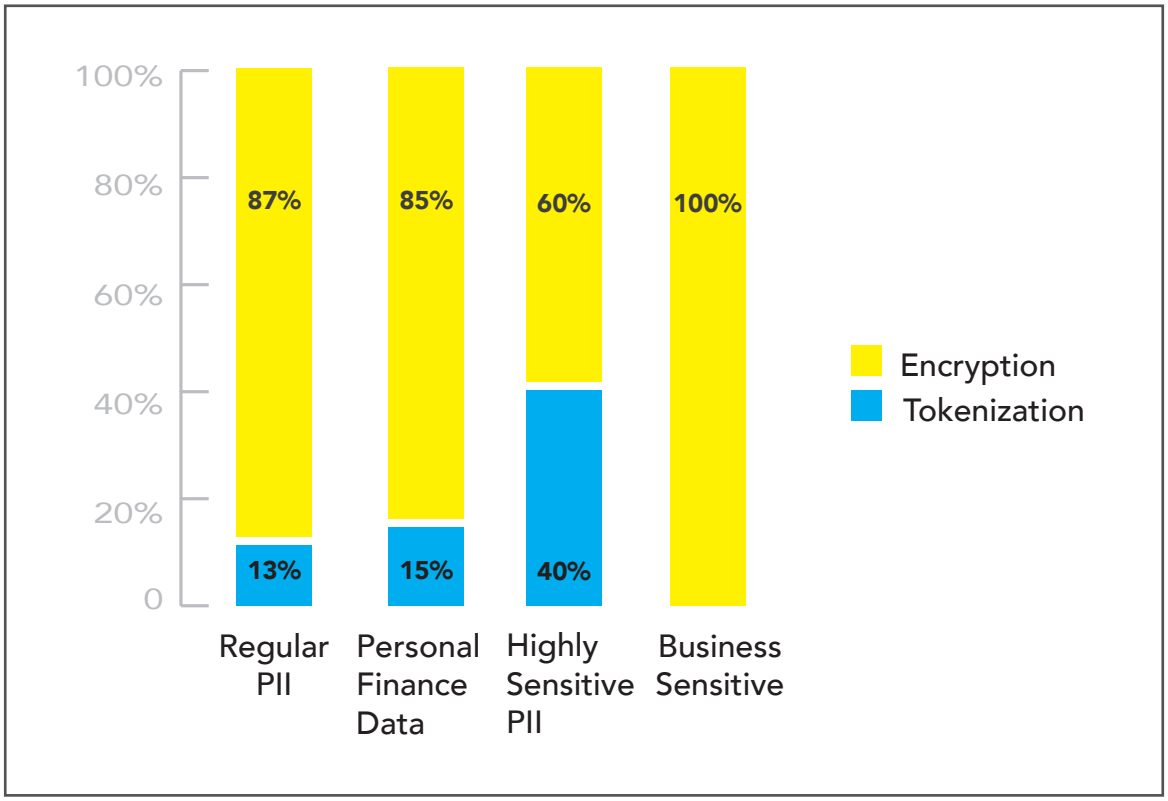
Both encryption and tokenization can be format and length preserving, and therefore both are good candidates for cloud data protection.

A tokenization scheme is commonly stateful. That is, the system must maintain some information – in this case the token lookup database – in order to retrieve the original data. A stateful system has a higher operational overhead as the organization must maintain (and secure) the state, or risk the loss of original data. However, a truly random tokenization scheme is not susceptible to cryptanalysis such as chosen plaintext and chosen cipher text attack, making it appealing for data residency, audit scope reduction, and scenarios with critical security requirements.

Encryption, on the other hand, can be either stateless or stateful. Compared to tokenization, encryption is relatively light in terms of operations overhead and provides adequate security as long as the key is properly secured. Encryption also has the added advantage that there are many standard encryption algorithms such as AES whose cryptographic strength have been peer-reviewed and well understood.

We analyzed the use of encryption vs. tokenization across our banking and financial services customer base. The result is depicted in Figure 2.

Figure 2: Encryption vs. Tokenization Use in Financial Services



As shown, encryption remains the popular choice to protect the different data categories. But 40% of firms that handle highly sensitive PII fields opt to use tokenization to protect these data. This is in contrast to only 13% that use tokenization for regular PII fields and 15% for personal finance data.

It's not surprising to see that encryption is the predominant choice for those seeking to protect business-sensitive data. As this category of data is typically non-critical, few are utilizing heavy-weight tokenization to protect business sensitive data.

CipherCloud provides a wide range of different encryption algorithms to protect cloud data. For the purpose of this discussion, we can think of them as varied implementations of AES algorithms.

SEARCH AND SORT OF ENCRYPTED DATA

Searchable encryption is a way to encrypt data, while maintaining the ability to selectively search it. Searchability is important for a number of use cases, including outsourcing data to a cloud application but needing to perform search on the data in the cloud without having to decrypt all the data.

Searchable encryption must satisfy two properties:

- **Security strength:** The encryption algorithm must withstand a certain level of cryptanalysis attack.
- **Fulfills selective search requirements:** This can be pre-fix search, whole-word search, partial word search, wild card search or search with logical operators.

It's a well-known fact that there is a trade-off between security strength and searchability of the cryptosystem. Not all encryption algorithms support search on encrypted data. In fact, it's rare to have a cryptosystem that supports a wide range of search functions.

Similarly, sorting on encrypted data enables encrypted data items to be placed in an order/list and retrieved from the database without having to decrypt the data first.

A properly-designed cryptosystem that supports search and sort should be zero-knowledge. That is, the system reveals no information to an observer beyond that the data fits in the search criteria or its order in the sorting outcome. If the cryptosystem reveals any information beyond that, it's a weak cryptosystem.

The ability to search and sort encrypted data is especially important given the current political climate where the US intelligence community is pushing for "encryption" backdoors to be installed in popular software/hardware. A system that supports operations on encrypted data may be able to assist the government with legitimate discovery requests without revealing the actual data to the inquiring agency.

In this section, we investigate how banking and financial services firms protect sensitive data. As our data taxonomy suggested, we will be focusing on protection of a) Highly sensitive PIIs, b) Regular PIIs, c) Personal financial data, and d) Business sensitive information.

PROTECTING HIGHLY SENSITIVE PII — SECURITY IS UTTER MOST IMPORTANT

Date-of-birth (DoB), TaxID, Social Security Numbers, and in some cases, national identification numbers, are highly sensitive information. A compromise or leak of such data can lead to disastrous consequences to a company's brand and image.

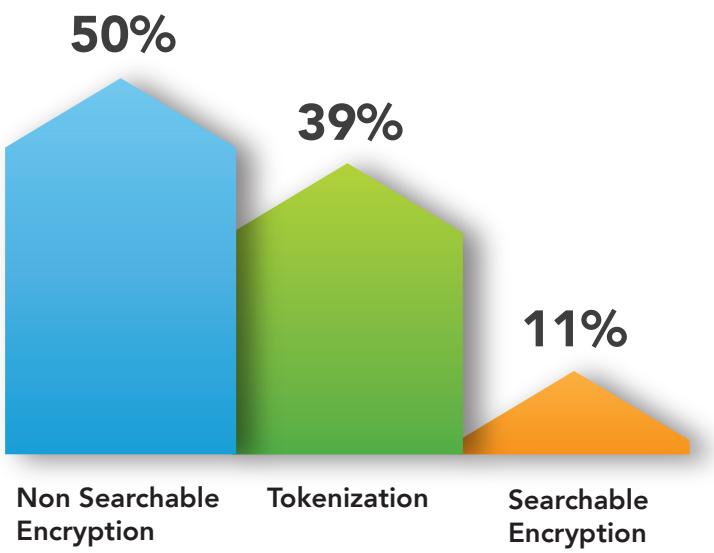
Financial services firms are extremely conscientious of protecting highly sensitive PIIs. Security professionals at these firms told us that the most difficult task is identifying these data items and making sure that they are not used in indexing, searching, and referencing into records.

For example, one of the practices commonly seen in the last decade was to search customer records using social security numbers. This means that those who need to access some mundane aspects of the customer record would have to see the social security number, which goes against the "need to know" principle. Many firms have by and large eliminated such practices, but some pockets still remain.

For the most part, CipherCloud's financial services customers choose the strongest protection available such as strong encryption and tokenization to protect highly sensitive PIIs. Since many business processes no longer use highly sensitive PIIs to index or reference customer records, the requirements for searchability and sortability are not critical.

Figure 3 shows the protection schemes organizations use to protect highly sensitive PIIs. As shown, 50% of the firms protect highly sensitive PIIs with non-searchable encryption, 39% use non-searchable tokenization, while 11% protect their highly sensitive PIIs with searchable encryption.

Figure 3: Protection Mechanisms for Highly Sensitive PIIs



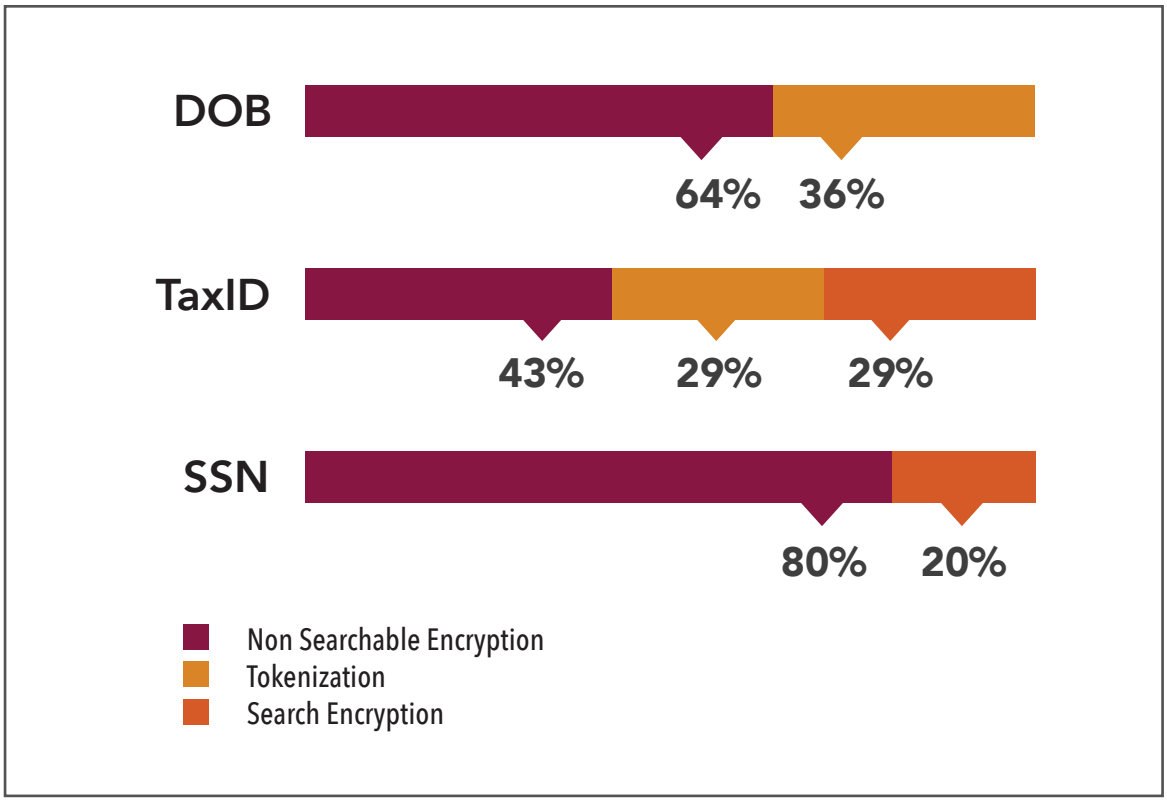
As the data suggests, when it comes to highly sensitive PIIs, banks and financial services firms clearly favor security strength over searchability.

More specifically, we looked at three data items: Date of Birth, Social Security number, and Tax ID. As shown in Figure 4, 64% of the firms that have date-of-birth information protect them with non-searchable encryption, while 36% utilize tokenization. For Social Security Numbers, 80% of the firms use non-searchable encryption and the remainder use searchable encryption. For TaxID, 29% opted for searchable encryption; the remainder is divided between non-searchable encryption and tokenization.

In general, financial services firms seek the strongest protection possible for highly sensitive PIIs. The non-searchable encryption algorithms they choose are typically AES 256-bit with random initialization vectors – one of the most secure secret-key

algorithms there is. Only a small percentage want these protected PIIs to be searchable (e.g., searchable TaxID used in building reports), thereby requiring the use of searchable encryption.

Figure 4: Protection Schemes for Three Specific Data Items



PROTECTING REGULAR PERSONAL IDENTIFIABLE INFORMATION

Beyond the highly sensitive PIIs, a financial services firm may collect other forms of personal data, such as individual names, email addresses, phone numbers, and physical addresses. Regulations and corporate policies mandate that companies protect personal data, regardless if they are stored in the company's own infrastructure or hosted by a third party (e.g. the cloud environment).

In addition, various business processes may utilize these PIIs to index or reference customer records and accounts. As such the protection mechanism must not impede

the operations of these business processes. That means providing search, sort, and alphabetic filtering on top of encryption, as organizations often require those capabilities for reporting, analytics, and enterprise workflow.

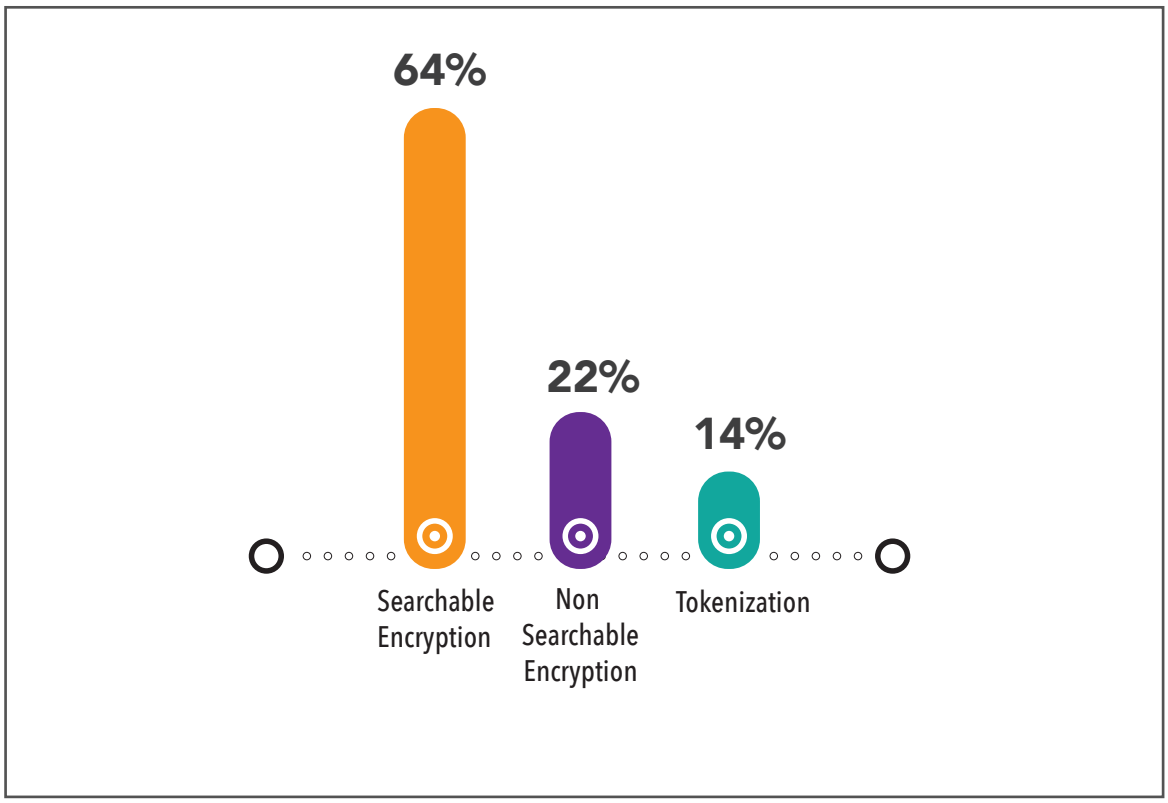
In this section, we looked into a few common types of PIIs and how companies are protecting them in the cloud. The data we are interested includes:

- Customer names (e.g., First name, last name)
- Specially-structured PIIs (e.g., emails, phone numbers, URLs)

Protecting Customer Names — Searchability is a Priority

Names, such as first and last names, are one of the most common personal-identifiable-information that companies collect. Names are also commonly used to search and sort records.

Figure 5: Protection Schemes for Customer Names



Among the financial services customers that collect individual names, 64% use various forms of searchable encryption algorithms to protect the "name" fields with search function. 14% elected to tokenize, and 22% simply deploy encryption protection without searchability (see Figure 5).

Clearly, searchability on individual names, along with encryption, is of high demand for firms seeking to protect customer identities. It is not a surprise that those that use tokenization to protect individual names are primarily European financial institutions and government finance outfits.

Protecting Structured PII Fields — Format-Preserving is Key

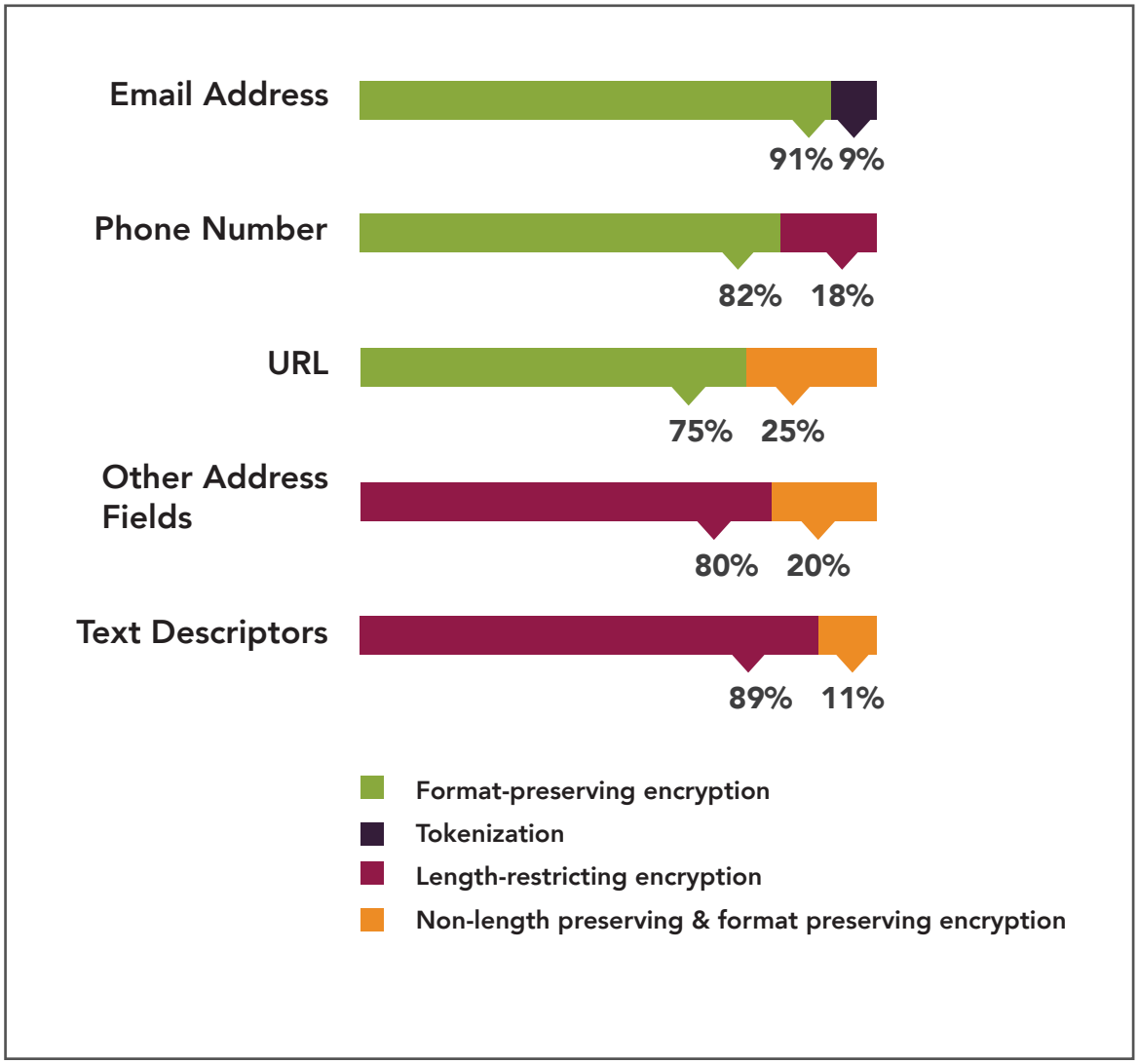
Email addresses, phone numbers, URLs, and dates are examples of structured PIIs with a specific format. Email addresses typically include the symbol "@". Phone numbers in the US are a 10-digit number with certain constraints – 10 zero's do not make a legitimate phone number. Other structured data, such as URLs and zip codes all carry specific formats.

For the cloud application to work with encrypted data, protection of the structured PII fields must be format preserving where the cipher text is structurally equivalent to the clear text. For instance, a clear text of anna_teeples@yahoo.com may get encrypted into 일 ψι=时@間%ゐV.ত৪K5. The cipher text bears no resemblance and has no reversible relationship to the original clear text, except for the fact that both conform to the same format.

CipherCloud provides format-preserving encryption for a variety of structured data types, including Email, URL, Phone numbers, dates, etc. For other structured data, such as mailing street, city, zip code, and text descriptors, length-restricting encryption is often sufficient.

Figure 6 provides a description of the protection schemes used to protect different structured PII fields. As shown, format-preserving encryption is predominately used to secure specially formatted data – over 90% of the firms use format-preserving encryption to protect email addresses, over 80% use the same to protect phone numbers, and 75% for URLs. For some of the more free-formed text fields, such as text fields, additional address information, etc., length-restricting encryption is often the right choice.

Figure 6: Protection Schemes for Structured PII Fields



Note that most of the format-preserving and length-restricting encryption algorithms also support some forms of search.

PROTECTING PERSONAL FINANCIAL DATA — BALANCING BETWEEN BUSINESS FUNCTIONALITY AND SECURITY

Financial services firms often process personal finance data, such as transaction records, loan information, trading information, financial institution data, etc. Protection of these types of data is a slightly different class of problem because the business usage of such data is unique.

Take for example, the information of "Transaction amount" or "Credit Score". This particular information might be used to sort transaction records based on the value or used in search where transactions of certain amounts are included in a report. There are many other examples where business workflows require search, sort, and filtering support along with the protection of personal financial data.

A slightly different example is "Bank account number". This information is often used in search, but rarely needed to support sorting or other operations.

We found that personal financial data are the most commonly used in core business processes and enterprise workflows for our banking and financial services firms. Therefore, the protection mechanism really needs to balance between the need for security and the need to support business operations.

Figure 7: Protection Schemes for Personal Financial Data

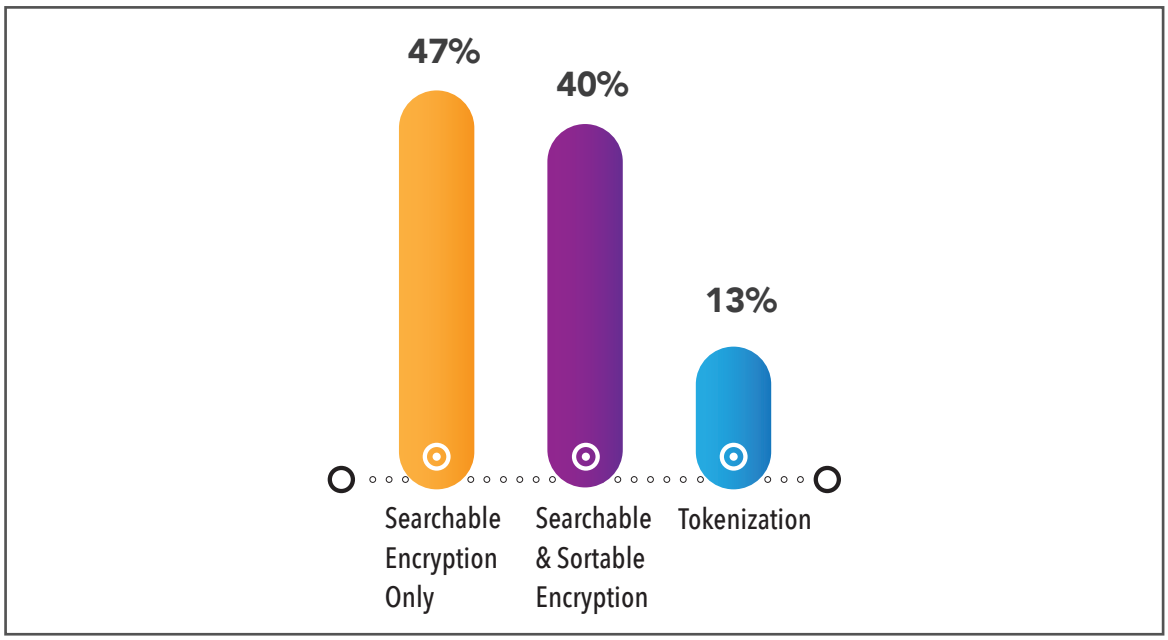


Figure 7 show that 87% of all firms that process and maintain personal financial data in cloud applications support some form of business functions. More specifically, 47% utilize searchable encryption while 40% use encryption mechanisms that support both search and sort.

PROTECTING BUSINESS-SENSITIVE DATA – A BI-MODAL OPERATION

Business sensitive data are internal data specific to the company, such as customer activity records, interaction emails, or business process information. This information can be sensitive, but often is not as security critical as PIIs.

For instance, "Account health rating" is a field that contains the contact health analysis rating for an account – is it an active account? Does the account need nurturing, etc.? This piece of information in itself is useless, but if used together with other context, becomes sensitive.

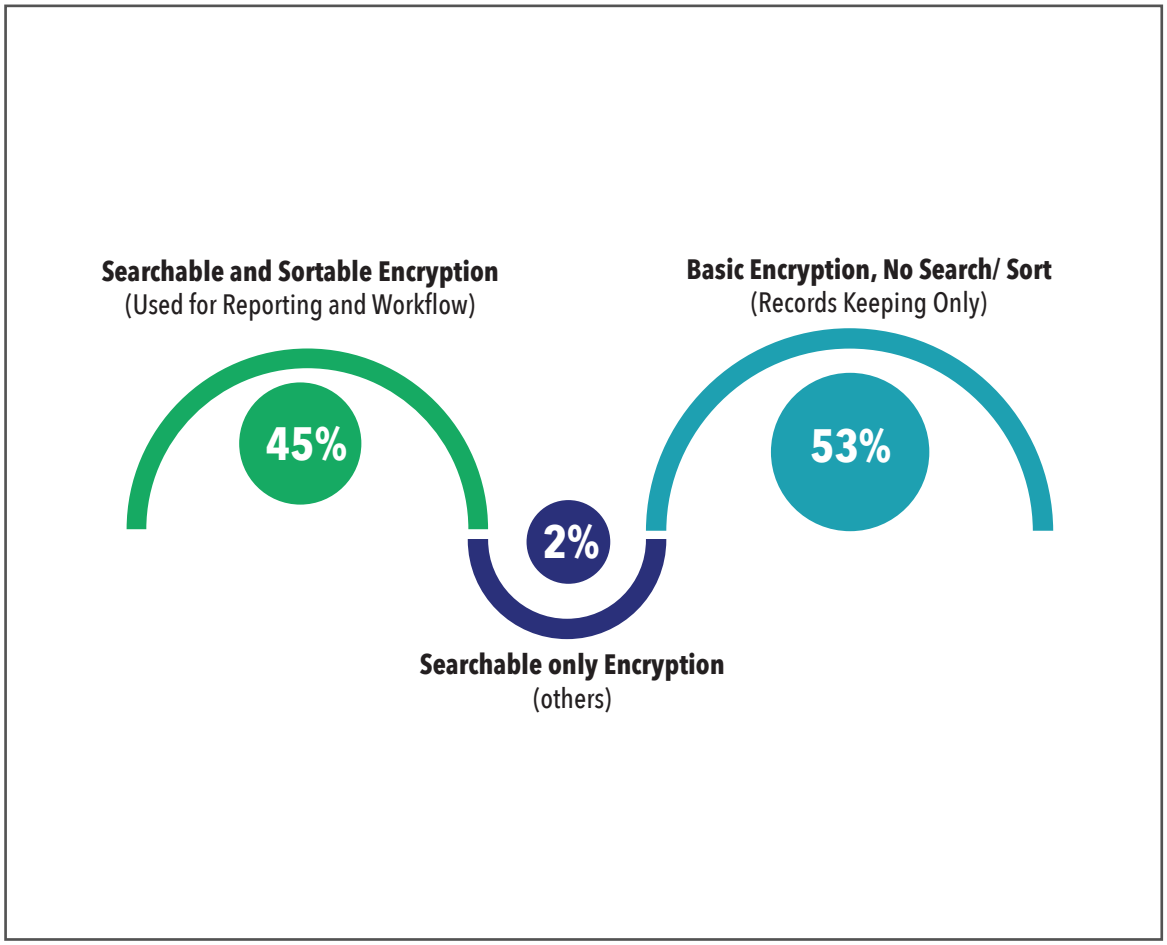
We found that business-sensitive data more or less fall into two different categories: those used primarily for records keeping, and those used for reporting and workflow. The latter requires support for search and sort, and the former really only needs protection – and given it is business-sensitive data, organizations should consider an encryption scheme with good performance as opposed to considering security strength only (see Table 2).

Table 1: Different Types of Business Sensitive Data

Example	Purpose	Requirements
Inbound email, customer call record	Primarily for records keeping	Protection only
Internal assessment	Used to reference into other pieces of data & in reports	Need to support search & sort

Our study shows that CipherCloud's financial services customers fall into a bi-modal distribution exactly as we suggested -- 45% of all firms that handle business-sensitive data use searchable and sortable encryption, while 53% use basic encryption only with no search and sort. There is very little else outside these two modes.

Figure 8: Protection Schemes Used to Secure Business Sensitive Data



CHOOSING THE RIGHT PROTECTION SCHEMES FOR YOUR DATA

To choose the right protection schemes for your data, we devised the following data protection checklist with a list of 5 simple questions. Assuming protection is needed at the data level, ask these questions:

Table 2: Data Protection Qualifying Questions

Questions	Answer Choices
What level of data is it?	Choices: (Highly sensitive PIIs, Regular PIIs, Personal Financial Data, or Business Sensitive Data)
What is it used for primarily?	Choices: (Records Keeping, Indexing/ Referencing, Reporting)
Does the data have residency constraints that the cloud applications cannot meet?	Yes or No
Is the data specifically structured?	Address, Phone Numbers and URLs etc.

Answering these questions will help you quickly narrow down the types of protection schemes you will need for the data. Here are a few of the criteria you can use:

- If the data is business sensitive & it is used primarily for records keeping, you can safely choose an efficient encryption scheme without support for search or sort.
- If the data has residency constraints, you need to consider tokenization.
- If the business requirement is such that the data will be used in indexing or referencing into other records, you will need a searchable encryption algorithm.
- If the data used in comparison, or in reporting, you will need to support sorting of encrypted data.
- If the data is specially-structured PII, you'll need format-preserving and/or length preserving encryption.

CONCLUSION

As cloud computing establishes itself as a compelling enterprise IT delivery and innovation platform, the issue of cloud data protection gets increasingly highlighted.

Today, organizations struggle to extend their on-premises data protection practices into the cloud. The result is inconsistent security practices, data protection gaps, and security blind spots. This may slow down firms' cloud migration journey and impede the business from fully capturing the potential of cloud computing.

Resolving the situation will require a mature data protection framework and concerted efforts on the part of enterprise security professionals and line of business leaders. The two sides must come together to bridge the gap between security and business needs, or risk failure of the cloud experiment.

This study, conducted by CipherCloud, analyzed in detail cloud security deployments across more than 50 global companies, including banking, investment, wealth management, and financial services. We investigated how these firms utilize different data protection schemes to meet regulatory, security, and business requirements. The findings in this report, aggregated over many real-world examples, serve as a foundation for cloud data protection best practices.

The next 18 – 24 months will be critical in capturing cloud competitive advantages to prepare for the next hyper-growth phase. As an enterprise security professional, you must:

Start a data auditing process: You must first determine which data is (or should be) in the cloud. To this end, you need to work with your line of business counterparts to understand the business value as well as criticality of the data.

Craft a policy and data protection framework. Based on the data classification, business requirements, and industry best practices as those detailed in this report, engineer a data protection framework and find appropriate solutions. Avoid products proprietary to specific clouds to prevent interoperability and cloud lock-in issues.

Establish a closed feedback loop with your business group: Business requirements change and so does the cloud ecosystem. A close-loop collaboration between business groups and IT security will allow you to refresh the data protection framework and keep pace with changing business environments. This is a must to ensure your cloud deployments remain meaningful and aligned to your business success.

CipherCloud™
Trust in the Cloud

CipherCloud, the leader in cloud visibility and data protection, delivers cloud adoption while ensuring security, compliance and control. CipherCloud's open platform provides comprehensive cloud application discovery and risk assessment, data protection—searchable strong encryption, tokenization, data loss prevention, key management and malware detection—and extensive user activity and anomaly monitoring services.

CipherCloud is experiencing exceptional growth and success with over 3 million business users across 11 different industries.

The CipherCloud product portfolio protects popular cloud applications out-of-the-box such as Salesforce, Box, Microsoft Office 365, and ServiceNo .

Named SC Magazine's 2013 Best Product of the Year, CipherCloud's technology is FIPS 140-2 validated and the company is backed by premier venture capital firms Transamerica Ventures, Andreessen Horowitz, Delta Partners, and T-Venture, the venture capital arm of Deutsche Telekom. For more information, visit www.ciphercloud.com and follow us on Twitter @ciphercloud.

RP-CC-CloudDPReportQ2-072215v3

Headquarters:
CipherCloud
333 West San Carlos Street
San Jose, CA 95110
www.ciphercloud.com

linkedin.com/company/ciphercloud
@ciphercloud

sales@ciphercloud.com
1-855-5CIPHER (1-855-524-7437)

CipherCloud | © 2016

All trademarks are property of their respective owners.