Privacy-Preserving Analytics and Secure Multiparty Computation

Organizations are increasingly concerned about data security in several scenarios, including collecting and retaining sensitive personal information; processing personal information in external environments, such as the cloud; and information sharing. Commonly implemented solutions do not provide strong protection from data theft and privacy disclosures.

Privacy and risk management professionals are particularly concerned about the privacy and security of data used in analytics and shared externally. Compliance to privacy regulations such as the US State of California Consumer Privacy Act (CCPA), the EU General Data Protection Regulation (GDPR) and other emerging regulations around the world require techniques for secure processing of sensitive data. New approaches to privacy preserving computing that are transparent to business processes can open new opportunities and help find the right balance between privacy, security and compliance (**figure 1**).

Encrypting data at rest is not sufficient to avoid data breaches. Data-at-rest encryption creates a “crypto boundary,” outside of which data are accessible in plaintext. Because plaintext data are normally needed for processing, this boundary often exists below the point at which a compromise is possible. Data-at-rest encryption also does not support scenarios in which data has to be shared with other organizations. For data to be useful, they usually must be accessible as plaintext within applications, and this significantly reduces encryption’s protection capability. A drawback of typical data masking techniques is that they do not broadly support the protection of transactional or behavioral data. These limitations of data-at-rest encryption and data masking are driving an increased focus on finding new techniques for data protection—particularly advanced approaches that can protect data in contexts where traditional encryption and data masking approaches cannot.

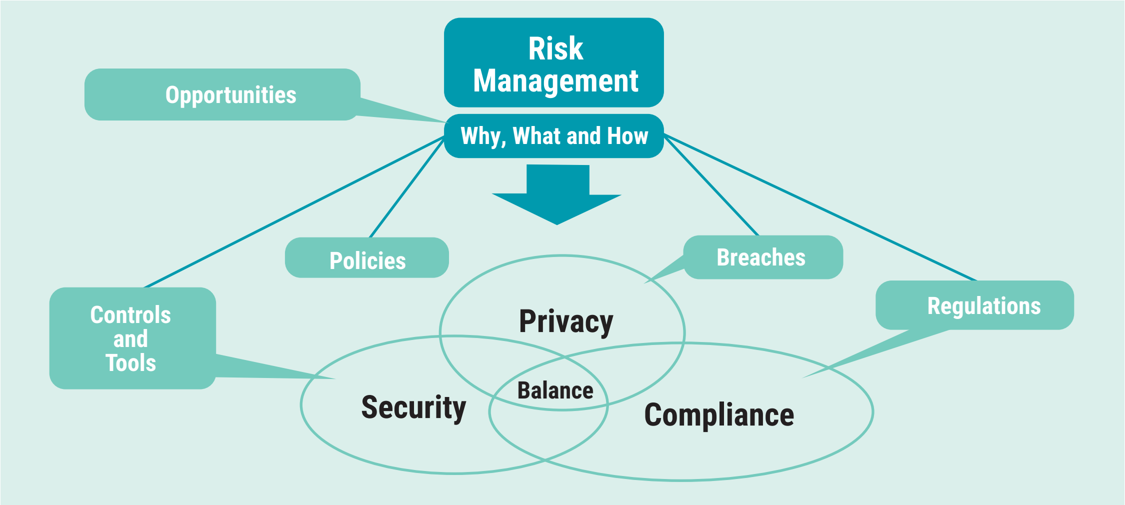
Sharing Sensitive Information Securely

Different industries are taking advantage of secure data sharing techniques. New privacy-preserving computing approaches are needed to meet legal requirements and provide privacy for data sharing.

The Benefits of Secure Data Sharing in Healthcare
Consider an example from the healthcare domain.

Ulf Mattsson, MSE
Is chief security strategist at Protegrity and contributed to the development of the Payment Card Industry Data Security Standard (PCI DSS), American National Standards Institute (ANSI) ANSI X9 and Cloud Security Alliance (CSA). He also developed products and services when working at IBM, Protegrity and other technology companies in the areas of robotics, enterprise resource planning, data encryption and tokenization, data discovery, cloud application security brokers, web application firewalls, managed security services, and security operation centers. Mattsson has worked with data protection projects in several different countries, including compliance solutions for EU cross-border data protection laws. He is a regular speaker at international security conferences and has written more than 100 articles for the *Institute of Electrical and Electronics Engineers (IEEE) Xplore, IBM Journals, ISACA® Journal*, and the *Information Systems Security Association (ISSA) Journal*. He is an inventor who holds more than 70 issued US patents. He can be reached at ulf@ulfmattsson.com.

Figure 1—Balance Between Privacy, Security and Compliance



The fairly recent ability to fully map the human genome has opened endless possibilities for advances in healthcare. Data from DNA analysis can test for genetic abnormalities, empower disease-risk analysis, and help discover family history and the presence of an Alzheimer's allele. These studies require very large DNA sample sizes to detect accurate patterns; however, sharing personal DNA data is a particularly problematic domain. Many citizens hesitate to share such personal information with third-party providers, uncertain of if, how and to whom the information might be shared downstream. Moreover, legal limitations designed to protect privacy restrict providers from sharing this data as well. Homomorphic encryption (HE) techniques enable citizens to share their genome data and retain key privacy concerns without the traditional all-or-nothing trust threshold with third-party providers.

Benefits of Secure Data Sharing for Financial Organizations

A typical financial institution may see only up to 25 percent of its customers' activity. Secure collaboration across institutions, business lines and borders helps to speed processes, cut false positives, lower operational costs and catch more criminals by having a more complete views of all activities. Gaining these insights requires navigating a minefield of private client information and sharing confidential financial data between independent financial institutions.

Financial institutions can benefit from three forms of data sharing:[1]

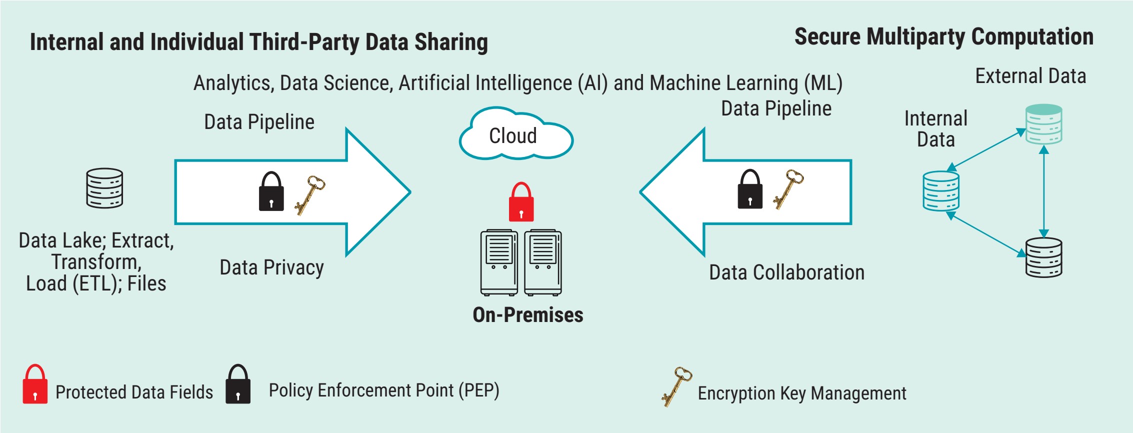
1. Inbound data from third parties
2. Owned outbound data with third parties
3. Collaborative data that can be similar forms of data inbound and outbound

Inbound data sharing allows institutions to enrich their decision-making systems[2] with additional information, leading to higher-quality outputs and more accurate operations. For example, trading firms can use third-party services such as Thomson Reuters MarketPsych Indices[3] to inform their buy/sell decisions with social media data, hypothetically leading to a more accurate understanding of market sentiment. Outbound data sharing, on the other hand, enables institutions to draw on capabilities (and offer customer benefits) that they may not own internally. For example, Wealthsimple, a robo-adviser, allows its clients' portfolio information to be pulled into Mint.com through a secure connection,[4] so that customers can see their investment balances alongside their day-to-day spending and build a comprehensive understanding of their finances.[5]

Privacy-Enhanced Computation

Figure 2 illustrates a data flow that brings together different privacy-preserving techniques that can provide security for data in use and data sharing.

Figure 2—Security for Data in Use and Data Sharing


Internal and Individual Third-Party Data Sharing
Secure Multiparty Computation
Analytics, Data Science, Artificial Intelligence (AI) and Machine Learning (ML)
External Data
Data Pipeline
Cloud
Data Pipeline
Internal
Data
Data Lake; Extract,
Transform,
Load (ETL); Files
Data Privacy
On-Premises
Data Collaboration
Protected Data Fields
Policy Enforcement Point (PEP)
Encryption Key Management

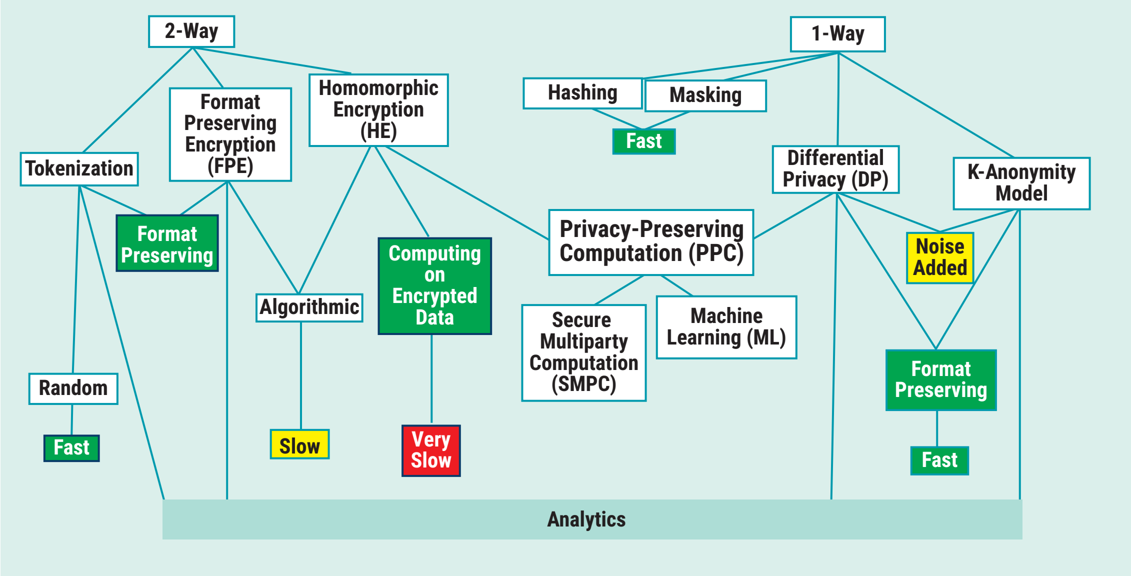
Positioning Different Data Protection Techniques

De-identification techniques and formal privacy measurement models are defined in the International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) standard ISO/IEC 20889:2018 *Privacy Enhancing data de-identification terminology and classification of techniques*.[6] Some of the techniques include two-way reversable methods and non-reversable one-way methods.[7, 8] **Figure 3** illustrates the positioning of different data protection techniques.

Characteristics of different data protection techniques include:

- Algorithmic techniques use encryption keys and encryption algorithms.
- Differential privacy and k-anonymity models add noise that may impact the correctness of statistical operations.
- Homomorphic encryption offers computing operations on encrypted data that can provide privacy during flow and processing between computers, suitable for the training of machine

Figure 3—Positioning Different Data Protection Techniques


2-Way
1-Way
Tokenization
Format Preserving Encryption (FPE)
Homomorphic Encryption (HE)
Hashing
Masking
Fast
Differential Privacy (DP)
K-Anonymity Model
Format Preserving
Privacy-Preserving Computation (PPC)
Noise Added
Computing on Encrypted Data
Algorithmic
Secure Multiparty Computation (SMPC)
Machine Learning (ML)
Format Preserving
Random
Fast
Slow
Very Slow
Fast
Analytics

learning models and secure multiparty computation (SMPC).

- Format preserving techniques also preserve the length of data fields.
- Analytical applications may require fast search on encrypted data values, sometimes also fuzzy search.

Techniques for preserving privacy can be divided into three categories, each with its own benefits and constraints: field-level data transformations, software-based secure computation algorithms, and architectures that use cryptographic data transformations and hardware-based security mechanisms.

However, privacy-preserving computation comes at a cost. Current versions of these technologies are often computationally costly, rely on specialized computer hardware, and are difficult to program and configure directly.[9]

Secure Multiparty Computation

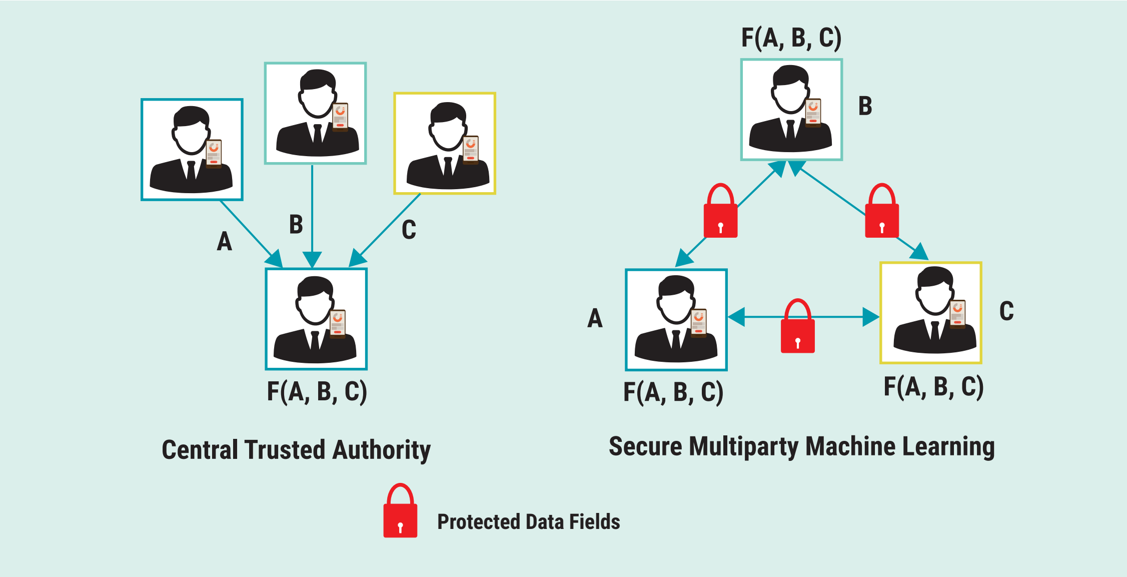
In SMPC, computations can be performed on data contributed by multiple parties without any individual party being able to see more than the portion of the data they contributed. This enables secure computation to be performed without the need for a trusted third party. **Figure 4** illustrates that participants collaborate on the computation knowing only the results of that computation and not the specific data others contributed, without the need of a central entity.

Example of Multiparty Computation: Retail

A large aggregator of payment card transaction data wanted to open a new revenue stream by using its data with its business partners in retail and banking. The aggregator helped their partners achieve a better ad conversion rate, improve customer satisfaction and provide more timely offerings.

By using secure multiparty computation, the aggregator could respect user privacy and specific regulations and enable the retailer to gain insights while protecting the organization's Internet protocol (IP). An analyst at each organization's office first used the software to link the data without exchanging any of the underlying data and using the protected data to train the machine learning and statistical models. The aggregator split the data set into secret shares and trained the model without needing to put together the pieces. The information that was communicated between peers was always encrypted. As a result, the retailer was able to get a better picture of its customers' buying habits.

Figure 4—Participants Collaborate on Computation



Example of Multiparty Computation: Average Salary

"Allie's salary is US$100K. In additive secret sharing, US$100K is split into three randomly generated pieces (or secret shares): US$20K, US$30K, and US$50K for example.[10] Allie keeps one of these secret shares (US$50K) for herself and distributes one secret share to each Brian (US$30K) and Caroline (US$20K). Brian and Caroline also secret share their salaries while following the same process (**figure 5**). Each participant locally sums their secret shares to calculate a partial result; in this example, each partial result is one third of the necessary information to calculate the final answer. The partial results are then recombined, summing the complete set of secret shares previously distributed. Allie, Brian and Caroline's average salary is US$200K."[11]

Standards in Privacy-Preserving Computation Techniques

ISO/IEC 29101:2013 *Information technology—Security techniques—Privacy architecture framework*, is "one of the oldest standards efforts that handles secure computing."[12] It presents architectural views for information systems that process personal data and shows how privacy-enhancing technologies, such as secure computing, pseudonymization and query restrictions, could be deployed to protect personally identifiable information (PII).

ISO/IEC 19592-1:2016 *Information technology—Security techniques—Secret sharing—Part 1: General*, focuses on "the general model of secret sharing and the related terminology."[13] It introduces properties that secret sharing schemes could have (e.g., the homomorphic property that is a key aspect for several SMPC systems).

ISO/IEC 19592-2:2017 *Information technology—Security techniques—Secret sharing—Part 2: Fundamental mechanisms*, "introduces specific schemes."[14] All schemes are systematically described using the terms and properties from part one.

Homomorphic Encryption

Homomorphic encryption (HE) plays a role in a family of privacy-preserving computation techniques (PPCT) that address and eliminate the classic compromise of sharing data while retaining privacy. HE expands the role of encryption by extending its scope from data at rest and data in transit to data in use (i.e., data being processed, viewed, updated). HE can better enable enterprises to leverage the services of third-party providers (typically but not restricted to the cloud) by reducing or eliminating privacy concerns. HE provides the ability to compute on data while the data are encrypted. This has enabled "industry and government to provide capabilities for outsourced computation securely."[15]

> "HE PROVIDES THE ABILITY TO COMPUTE ON DATA WHILE THE DATA ARE ENCRYPTED."

HE Applications

HE enables private queries to a search engine—the user submits an encrypted query and the search engine computes an encrypted answer without exposing the query in the clear text. "It also enables searching on encrypted data—a user stores encrypted files on a remote file server and can later have the

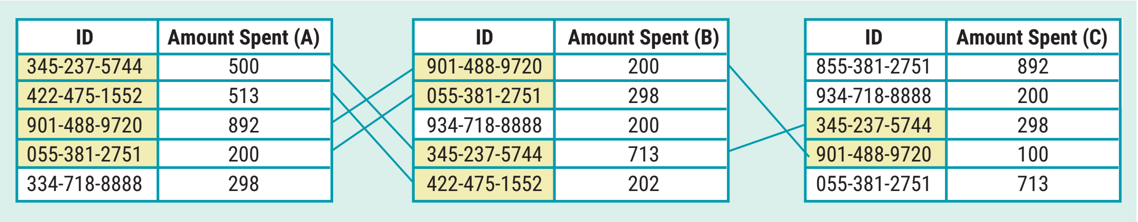
Figure 5—Example of Multiparty Computation: Average Salary

	Allie	Brian	Caroline
A=$100	50	30	20
B=$200	-80	100	180
C=$300	0	350	-50

Allie	Brian	Caroline	Sum	Average
50	30	20		
-80	100	180		
0	350	-50		
-30	480	150	$600	$200

Source: Adapted from Inpher, "What Is Secure Multiparty Computation?" *https://www.inpher.io/technology/what-is-secure-multiparty-computation*

Figure 6—Private Set Intersection



ID	Amount Spent (A)
345-237-5744	500
422-475-1552	513
901-488-9720	892
055-381-2751	200
334-718-8888	298

ID	Amount Spent (B)
901-488-9720	200
055-381-2751	298
934-718-8888	200
345-237-5744	713
422-475-1552	202

ID	Amount Spent (C)
855-381-2751	892
934-718-8888	200
345-237-5744	298
901-488-9720	100
055-381-2751	713

Source: Adapted from "A Privacy-Preserving Way to Find the Intersection of Two Datasets," OpenMined, 29 April 2020, *blog.openmined.org/private-set-intersection*

server retrieve only files that (when decrypted) satisfy some Boolean constraint, even though the server cannot decrypt the files on its own."[16]

Private Set Intersection

Private set intersection (PSI) is a powerful cryptographic technique that enables two parties to compute the intersection of their data without exposing their raw data to the other party. PSI identifies common elements between data sets held by different parties (**figure 6**). This replaces simplistic approaches such as one-way hashing functions that are susceptible to dictionary attacks. Applications for PSI include identifying the overlap with potential data partners (i.e., Is there a large enough client base in common to be worthwhile to work together?) as well as aligning data sets with data partners in preparation for using MPC to train a machine learning model.[17]

Differential Privacy

Differential privacy is a form of field-level data masking designed such that data can be used for querying aggregate statistics while limiting the exposure of individuals' specific information. This approach supports data-sharing scenarios and has the capability to process data in untrusted environments (**figure 7**).

Differential privacy can be implemented in six different types of transformation algorithms that are suitable for different use cases (**figure 8**). They provide mathematical definitions of how the algorithms hide the presence or absence of any individual's data in a data set.

Example of Differential Privacy: Banking

A bank wanted to broaden access to its data lake. Stakeholders found that "current approaches to de-identify data such as masking, tokenization, and aggregation can leave data unprotected."[18] Current approaches to de-identifying data did not fulfill the compliance requirements and business needs, which led to several bank projects being stopped. The issue with these techniques is that they do not sufficiently protect the data without overly degrading data quality.

This approach enables the creation of privacy-protected data sets that retain their analytical value for data science and business applications. The solution automatically enforces the compliance policies before the data are consumed by data science and business teams from the data lake. The analytical quality of the data is preserved for machine learning purposes by using artificial

Figure 7—Differential Privacy Data Flow

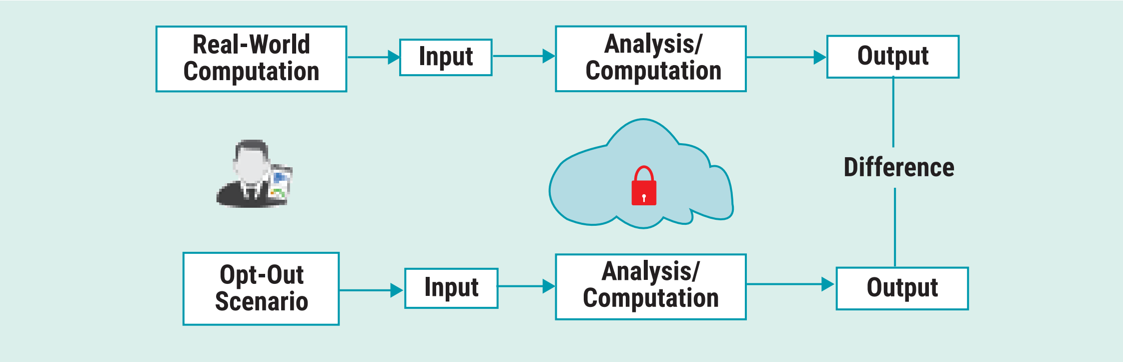
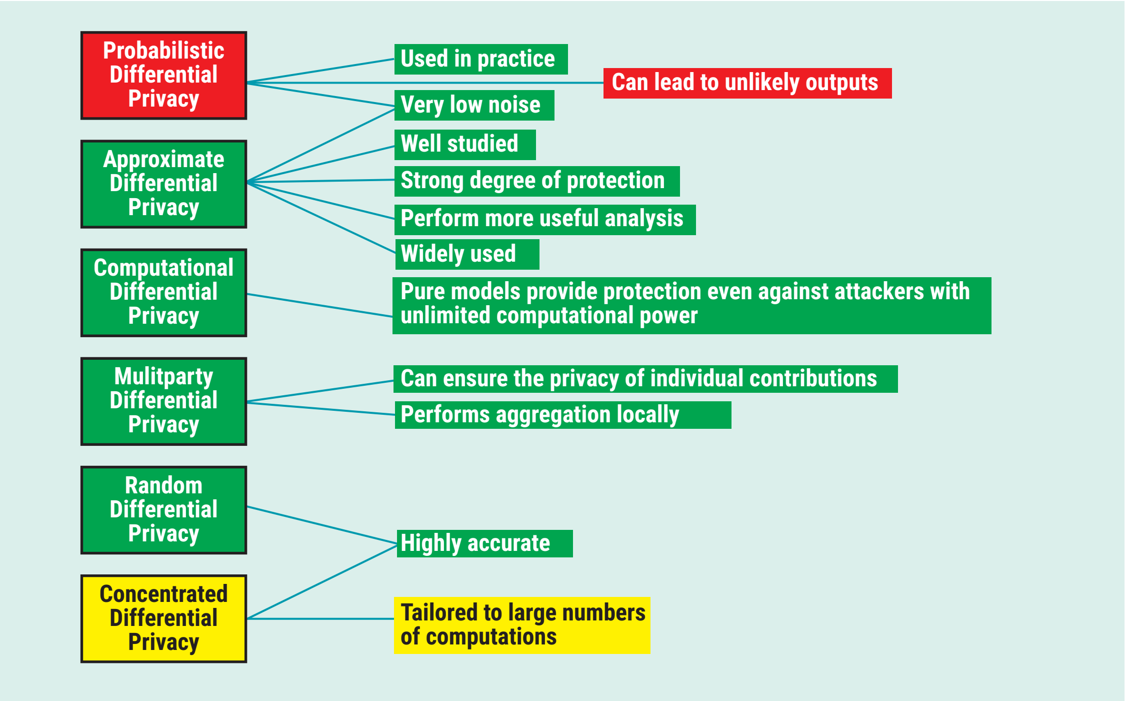


Figure 8—Differential Privacy Models


Probabilistic Differential Privacy
Approximate Differential Privacy
Computational Differential Privacy
Mulitparty Differential Privacy
Random Differential Privacy
Concentrated Differential Privacy
Used in practice
Can lead to unlikely outputs
Very low noise
Well studied
Strong degree of protection
Perform more useful analysis
Widely used
Pure models provide protection even against attackers with unlimited computational power
Can ensure the privacy of individual contributions
Performs aggregation locally
Highly accurate
Tailored to large numbers of computations

intelligence (AI) and leveraging privacy models such as differential privacy and k-anonymity.

Improved data access for teams increases the enterprise's bottom line without adding excessive infrastructure costs while reducing the risk of consumer information exposure.

K-Anonymity Model

The k-anonymity model ensures that groups smaller than k individuals cannot be identified. Queries will return at least k number of records. K-anonymity is a formal privacy measurement model that ensures that for each identifier there is a corresponding equivalence class containing at least k records.

L-diversity is an enhancement to K-anonymity for data sets with poor attribute variability. It is designed to protect against deterministic inference attempts by ensuring that each equivalence class has at least L well-represented values for each sensitive attribute. This variant of K-anonymity is subject to attacks, which have led to the development of T-closeness. T-closeness is an enhancement to L-diversity for data sets with attributes that are unevenly distributed, belong to a small range of values or are categorical.[19]

Privacy-Preserving Search on Data

Sensitive data that are encrypted on the local premises before outsourcing them to the cloud hinder searching on the encrypted data, which is of critical importance for many use cases. Searchable encryption techniques need to provide a balance between performance, privacy and functionality.

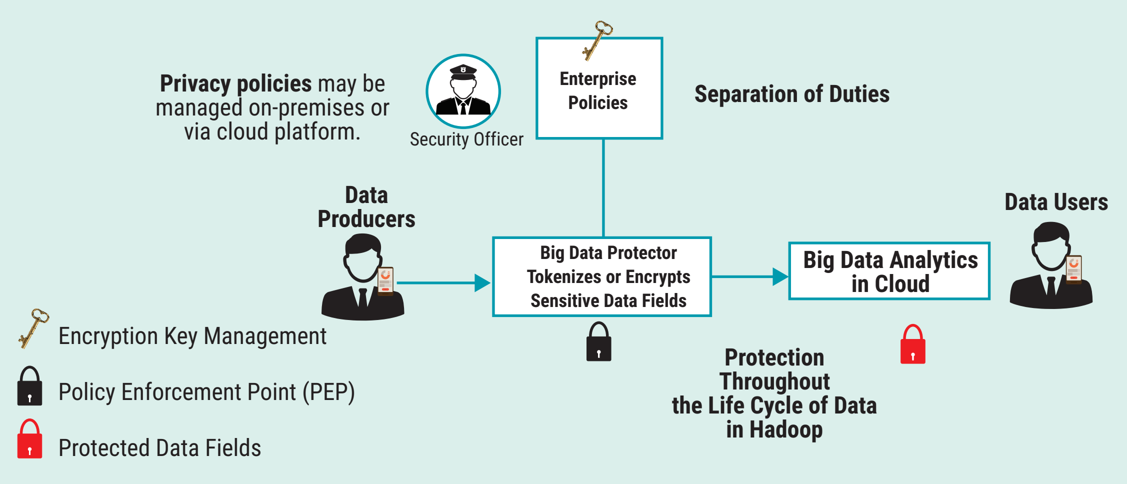
Outsourcing Data That Are Confidential or Regulated

A medical center that owns patients' health records cannot outsource its data to a cloud that is vulnerable to attacks due to legal regulations. A law enforcement agency that keeps sensitive criminal records should hesitate to use cloud storage. One way to overcome this confidentiality problem is to encrypt data on the local premises before outsourcing them to the cloud. However, although this approach preserves data confidentiality, it hinders data processing.[20] It is important to enable searching, which is of paramount importance, for outsourced data.

Data Protection for Cloud

Another example is of a data warehouse that analyzes encrypted data using built-in machine learning (ML) capabilities. Dremel technology[21] is a scalable, interactive *ad hoc* query system for analyzing read-only nested data.

Figure 9—Data Protection for the Cloud


Privacy policies may be managed on-premises or via cloud platform.
Security Officer
Enterprise Policies
Separation of Duties
Data Producers
Big Data Protector Tokenizes or Encrypts Sensitive Data Fields
Big Data Analytics in Cloud
Data Users
Encryption Key Management
Policy Enforcement Point (PEP)
Protected Data Fields
Protection Throughout the Life Cycle of Data in Hadoop

Data privacy is provided by tokenization or encryption of sensitive data fields throughout the life cycle of data in Hadoop. Privacy policies are managed on-premise or in the cloud (**figure 9**).

Approaches to Searching Encrypted Data in the Cloud

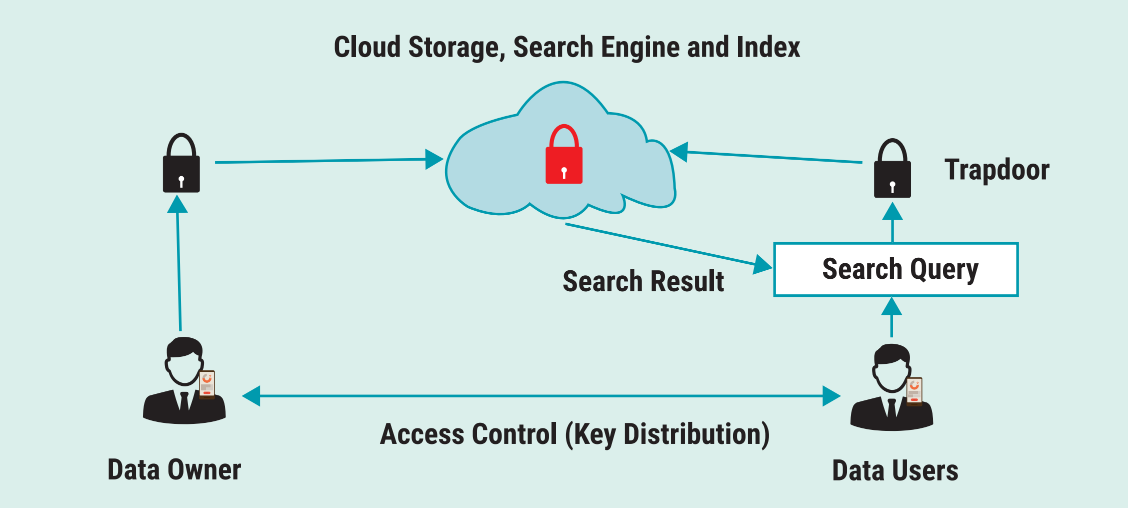
Searchable encryption techniques were studied in early 2000.[22] Since then, much research has been done to understand different types of searchable encryption. Although the studied systems are different in their searching approaches, security level and performance, they share certain architectural similarities. There are several survey studies on different searchable encryption systems.[23, 24]

Utilize an Index Structure

"Searchable encryption systems commonly utilize an index structure to keep track of occurrences of keywords in documents." The process of initializing this index takes keys from a collection of documents as inputs. Then it extracts keywords from the documents and inserts them into the index structure.[25]

Figure 10 illustrates a build-index process that is used by the data owner to generate a secure and searchable structure that enables search over the encrypted data. An index structure is generally implemented in the form of a hash table, metadata (markup) or an inverted index where each unique

Figure 10—Search of Encrypted Data in the Cloud


Cloud Storage, Search Engine and Index
Trapdoor
Search Result
Search Query
Access Control (Key Distribution)
Data Owner
Data Users

Source: Adapted from Pham, H.; J. Woodworth; M. A. Salehi; "Survey on Secure Search Over Encrypted Data on the Cloud," *Concurrency and Computation,* vol. 31, iss. 17, 10 September 2019, *https://onlinelibrary.wiley.com/doi/10.1002/cpe.5284* keyword is mapped to the document identifiers in which it appears.

Expansion to the Keyword-Based Search

One expansion to the keyword-based searchable encryption is to allow users to perform regular expression searches on encrypted data. A preliminary approach proposes to create all possible variations of a given regular expression.[26] For instance, for ab[a – z] query, it generates all 26 possible search queries that are aba, abb, . . . , abz. This approach only works for simple regular expressions and is not scalable for those with high degrees of variability, e.g., a*b*.

Fuzzy Keyword Search

Fuzzy keyword search can allow searchable encryption systems to accept minor typographical errors, but it may not exactly cover the semantic perspective.

> *Fuzzy keyword search greatly enhances system usability by returning the matching files when users' searching inputs exactly match the predefined keywords or the closest possible matching files based on keyword similarity semantics, when exact match fails.*[27]

Different Security Levels in Search Techniques

In semisecure searchable encryption systems, the index structure could be partially encrypted, and some information about the documents or the keywords can be leaked from the index structure. Fully secure searchable encryption systems do not trust any part of the system, except the client's device. Also, the auxiliary index is properly secured and does not expose any plain text data to the server. Keywords in the index structure can be hashed.

> *Somewhat secure searchable encryption systems in this category often deploy a trusted server (also known as a private cloud or a gateway) in between the third-party server (e.g., public cloud) and the client device.*[28]

Cryptographically Protected Database Search

Protected database search systems "cryptographically isolate the roles of reading from, writing to, and administering the database.[29] This separation limits unnecessary administrator access and protects data in the case of system breaches.

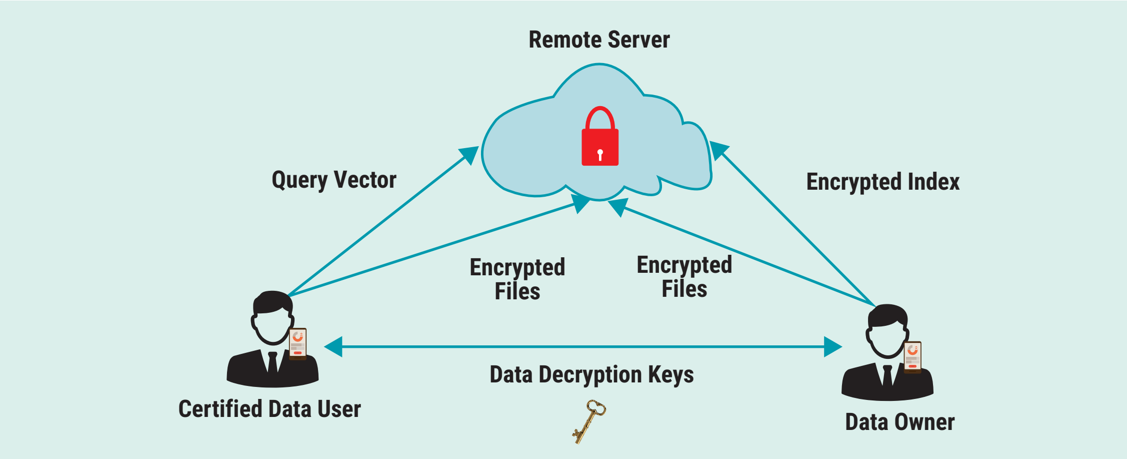
> *Design of such systems is a balancing act between security, functionality, performance and usability. This challenge is made more difficult by ongoing database specialization, as some users will want the functionality of [Structured Query Language] SQL, [not only Structured Query Language] NoSQL, or NewSQL databases. This database evolution will continue, and the protected search community should be able to quickly provide functionality consistent with newly invented databases.*[30]

Fuzzy Search Over Encrypted Data

To meet both ends of security and searchability, search-supported encryption is proposed. However, many previous schemes suffer severe vulnerability when typos and semantic diversity exist in query requests. To overcome such flaws, higher error tolerance is always expected for search-supported encryption design, sometimes defined as "fuzzy search." This approach introduces a new mechanism to map a natural language expression into a word-vector space. Compared with previous approaches, this "approach can work well for both accuracy and efficiency and will not hurt the fundamental security."[31] **Figure 11** illustrates that searchable encryption approaches can be divided into three steps:

1. **Represent**—Keywords are extracted from outsourced files or received queries and transferred into word-vectors, a combination of which builds the final representation of files or queries.
2. **Encrypt and index**—Files and queries are both encrypted to enhance security. They are suggested to be encrypted in heterogeneous ways. The encryption algorithm and key are provided usually by data owners. With some data structure, encrypted files are organized and stored for indexing.
3. **Search**—Users send queries and data holders perform some search algorithms on the query and stored encrypted data. Search consists of the calculation of relevance score and ranking by the score. The data user usually only asks for the top-k most relevant files with the query instead of all relevant files.

Figure 11—Search Over Encrypted Data



Source: Adapted from Cao, J.; J. Zhu; L. Lin; Z. Xue; R. Ma; H. Guan; "A Novel Fuzzy Search Approach Over Encrypted Data With Improved Accuracy and Efficiency," 2019, *https://arxiv.org/abs/1904.12111*

Bloom Data Search Filters

In 1970, the Bloom filter technique was introduced for applications where the amount of source data would require an impractically large amount of memory if conventional error-free hashing techniques were applied. It is "a space-efficient probabilistic data structure"[32] that is used to test whether an element is a member of a set.

Popular databases use Bloom filters to perform Bloom searches of partitions for certain queries, for example, when joining a data dimension table with a large fact table. **Figure 12** illustrates that false positive matches are possible but false negatives are not. Elements can be added to the set but not removed. The more items added, the larger the probability of false positives.

Hybrid Cloud Considerations

Organizations may be familiar with on-premises encryption and key management systems, so they often prefer consistence to leverage the same tool and skills across multiple clouds. Organizations often adopt a "best of breed" cloud approach. Some customers simply do not trust their vendors. A common concern is vendor lock-in, an inability to migrate to another cloud service provider.[33]

Figure 12—False Positive Matches With Bloom Filters

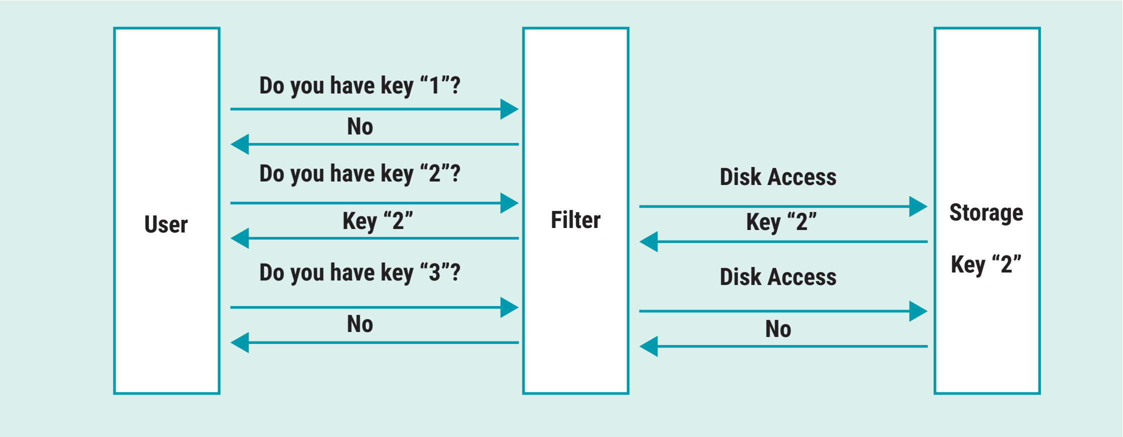
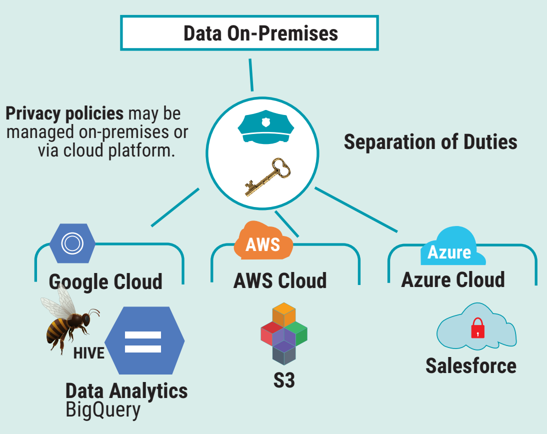


Figure 13 illustrates the central management of privacy across clouds and on-premises.

Figure 13—Central Management of Privacy Across Clouds and On-Premises



Conclusion

Organizations are increasingly concerned about data privacy; however, new techniques make it possible to securely share data and protect the privacy of individuals. These techniques can allow searches on encrypted data in data lakes and the cloud without compromising data privacy and while still preserving the data's analytical quality.

Commonly implemented solutions do not provide strong protection from data theft and privacy disclosures. Encrypting data at rest is not sufficient to avoid data breaches. Different industries have already started taking advantage of new privacy-preserving techniques. New privacy-preserving computing approaches are needed to help pursue new opportunities and find the right balance between privacy, security and compliance.

HE efforts remain diverse and fragmented, and a lack of standardization inhibits consistency to create scale and simplify and standardize APIs and SDKs. HE technology must be abstracted and simplified by incorporating it into familiar developer languages, frameworks and platforms.

> “NEW PRIVACY-PRESERVING COMPUTING APPROACHES ARE NEEDED TO HELP PURSUE NEW OPPORTUNITIES AND FIND THE RIGHT BALANCE BETWEEN PRIVACY, SECURITY AND COMPLIANCE.”

Endnotes

[1] The World Economic Forum, *The Next Generation of Data-Sharing in Financial Services: Using Privacy Enhancing Techniques to Unlock New Value*, Switzerland, September 2019, *www3.weforum.org/docs/WEF_Next_Gen_Data_Sharing_Financial_Services.pdf*

[2] *Ibid.*

[3] Thomson Reuters, “Thomson Reuters Expands Sentiment Data to Track Top 100 Cryptocurrencies,” 13 June 2018, *https://www.thomsonreuters.com/en/press-releases/2018/june/thomson-reuters-expands-sentiment-data-to-track-top-100-cryptocurrencies.html*

[4] Wealthsimple, “Wealthsimple Announces Partnership With Mint,” Cision, 3 May 2016, *https://www.newswire.ca/news-releases/wealthsimple-announces-partnership-with-mint-577957751.html*

[5] *Op cit* World Economic Forum

[6] International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC), *ISO/IEC 20889 Privacy Enhancing data de-identification terminology and classification of techniques*, Switzerland, 2018, *https://www.iso.org/standard/69373.html*

[7] Mattsson, U.; “Data Security: On Premise or in the Cloud,” *ISSA Journal*, December 2019, *https://www.issa.org/journal/december-2019/*

[8] Mattsson, U.; “Practical Data Security and Privacy for GDPR and CCPA,” *ISACA® Journal*, vol. 3, 2020, *https://www.isaca.org/archives*

[9] Big Data UN Global Working Group, *UN Handbook on Privacy-Preserving Computation Techniques*, *http://publications.officialstatistics.org/handbooks/privacy-preserving-techniques-handbook/UN%20Handbook%20for%20Privacy-Preserving%20Techniques.pdf*

10 Inpher, "What Is Secure Multiparty Computation?" *https://www.inpher.io/technology/what-is-secure-multiparty-computation*

11 *Ibid.*

12 International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC), ISO/IEC 29101 *Information Technology—Security Techniques—Privacy Architecture Framework*, Switzerland, 2013, *https://www.iso.org/standard/45124.html*

13 International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC), ISO/IEC 19592-1 *Information Technology—Security Techniques—Secret Sharing—Part 1: General*, Switzerland, 2016, *https://www.iso.org/standard/65422.html*

14 International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC), ISO/IEC 19592-2 *Information Technology—Security Techniques—Secret Sharing—Part 2: Fundamental Mechanisms*, Switzerland, 2017, *https://www.iso.org/standard/65425.html*

15 Homomorphic Encryption Standardization, "Homomorphic Encryption," *https://homomorphicencryption.org/*

16 Gentry, C.; "A Fully Homomorphic Encryption Scheme," Stanford University, California, USA, September 2009, *https://crypto.stanford.edu/craig/craig-thesis.pdf*

17 "A Privacy-Preserving Way to Find the Intersection of Two Datasets," OpenMined, 29 April 2020, *blog.openmined.org/private-set-intersection*

18 "Privacy-Protected Cloud Migration," 12 September 2019, *https://betakit.com/cryptonumerics-launches-tool-to-help-companies-identify-dataset-privacy-risks/*

19 Li, N.; T. Li; S. Venkatasubramanian; "t-Closeness: Privacy Beyond k-Anonymity and l-Diversity," IEEE 23rd International Conference on Data Engineering, 2007, *https://ieeexplore.ieee.org/document/4221659*

20 Poh, G. S.; J. Chin; W. Yau; K. R. Choo; M. S. Mohamad; "Searchable Symmetric Encryption: Designs and Challenges," *ACM Computing Surveys*, vol. 50, no. 3, 2017, *https://dl.acm.org/doi/10.1145/3064005*

21 Melnik, S.; A. Gubarev; J. J. Long; G. Romer; S. Shivakumar; M. Tolton; T. Vassilakis; "Dremel: Interactive Analysis of Web-Scale Datasets," Proc. of the 36th International Conference on Very Large Data Bases (VLDB), vol. 3, no. 1–2, September 2010

22 Song, D. X.; D. Wagner; A. Perrig; "Practical Techniques for Searches on Encrypted Data," Proceedings of IEEE Symposium on Security and Privacy, 2000

23 Bosch, C.; P. Hartel; W. Jonker; A. Peter; "A Survey of Provably Secure Searchable Encryption," *ACM Computing Surveys*, vol. 47, no. 2, 2014, p. 18:1–18:51

24 *Op cit* Poh

25 Pham, H.; J. Woodworth; M. A. Salehi; "Survey on Secure Search Over Encrypted Data on the Cloud," *Concurrency and Computation*, vol. 31, iss. 17, 10 September 2019, *https://onlinelibrary.wiley.com/doi/10.1002/cpe.5284*

26 *Op cit* Song *et al*.

27 Li, J.; Q. Wang; C. Wang; N. Cao; K. Ren; W. Lou; "Fuzzy Keyword Search Over Encrypted Data in Cloud Computing," 2010 Proceedings IEEE INFOCOM, 14-19 March 2010, San Diego, California, USA, *https://ieeexplore.ieee.org/document/5462196*

28 *Op cit* Bosch

29 Bloom, B. H.; "Space/Time Trade-Offs in Hash Coding With Allowable Errors," *Communications of the ACM,* vol. 13, no. 7, 1970, p. 422–426

30 Fuller, B.; M. Varia; A. Yerukhimovich; E. Shen; A. Hamlin; V. Gadepally; R. Shay; J. D. Mitchell; R. K. Cunningham; "SoK: Cryptographically Protected Database Search," 2017, *https://arxiv.org/abs/1703.02014*

31 Cao, J.; J. Zhu; L. Lin; Z. Xue; R. Ma; H. Guan; "A Novel Fuzzy Search Approach Over Encrypted Data With Improved Accuracy and Efficiency," 2019, *https://arxiv.org/abs/1904.12111*

32 *Op cit* Bloom

33 Securosis, *Multi-Cloud Key Management*, USA, 22 March 2019, *https://securosis.com/research/multi-cloud-key-management-2019*