How Innovative Enterprises Win With Secure Machine Learning

亦有中文简体译本
www.isaca.org/currentissue

New solutions help innovative enterprises succeed in an ever-changing, increasingly competitive digital economy. These solutions can empower enterprises to activate and extract value from sensitive data and engender trust by preserving the privacy of customers and employees. This enables enterprises to use private data—including their application in advanced analytics, machine learning (ML) and artificial intelligence (AI)—effectively without worrying about putting customers, employees or intellectual property at risk.

Enterprises can gain an edge over less innovative competitors by understanding how using ML on protected data is beneficial, how recent advances in quantum computing can significantly impact opportunities and threats to new and historical data, and how current and future technologies and a longer-term road map with future technologies can optimize and protect ML code. All of this propels enterprises forward by giving them a competitive advantage over less innovative competitors.

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 on data protection projects in several different countries, including working on 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.

Secure AI-Extracting Value From Protected Data

Secure AI solutions create opportunities to harness the sensitive data that are proven to be most effective in activating advanced analytics and ML. With the confidence that sensitive data are protected, enterprises can quickly extract value, apply insights in real time and predict outcomes that accelerate growth. Sensitive data should be secured wherever they are and whatever they are—in the cloud or on-premises, at rest or in use—so they can be leveraged across the enterprise by frontline employees, analytics teams and anyone who needs the information to make business decisions. Data know no boundaries nor should data protection. Whether encrypting, tokenizing or applying privacy methods, the solution should secure the data behind the many operational systems that drive the day-to-day functions of the enterprise and the analytical systems behind decision-making, personalized customer experiences and AI modeling.

Case Study: Reducing Fintech Risk With Data Protection Tools

Anonymization is one way to minimize the risk of identification.

Anonymization is a nonreversible method of data protection that can advance data-intensive business applications, such as analytics, by using differential privacy or k-anonymity.

In the example in **figure 1**, a bank requiring credit card approval for a transaction by a customer reduced the privacy risk from 26 percent to 8 percent and provided 98 percent accuracy compared to the initial ML model.

This approach can be used for analysis, insight, dashboarding, reporting, predictions, forecasts, simulation and optimization with values to be expected in savings and revenue adds.

Another data protection technique in ML models is pseudonymization, which is a reversable approach that can be based on encryption or tokenization. Encryption uses mathematical algorithms and cryptographic keys to change data into binary ciphertext, and tokenization substitutes cleartext data with a deterministic random string of characters. **Figure 2** illustrates the positioning of some characteristics of different protection approaches.

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.[1]

Differential privacy is a rigorous mathematical definition that emerged from lengthy work applying algorithmic ideas to the study of privacy. In the simplest setting, this algorithm analyzes a data set and computes statistics about it (such as the data set's mean, variance, median and mode) and it is differentially private—looking at the output, one cannot tell whether any individual's data were included in the original data set. In other words, the guarantee of a differentially private algorithm is that its behavior hardly changes when a single individual joins or leaves the data set; anything the algorithm might output on a database containing an individual's information is almost as likely to have come from a database without that individual's information. This guarantee holds for any individual and any data set;

Figure 1—Anonymization Process to Reduce the Risk of Identification

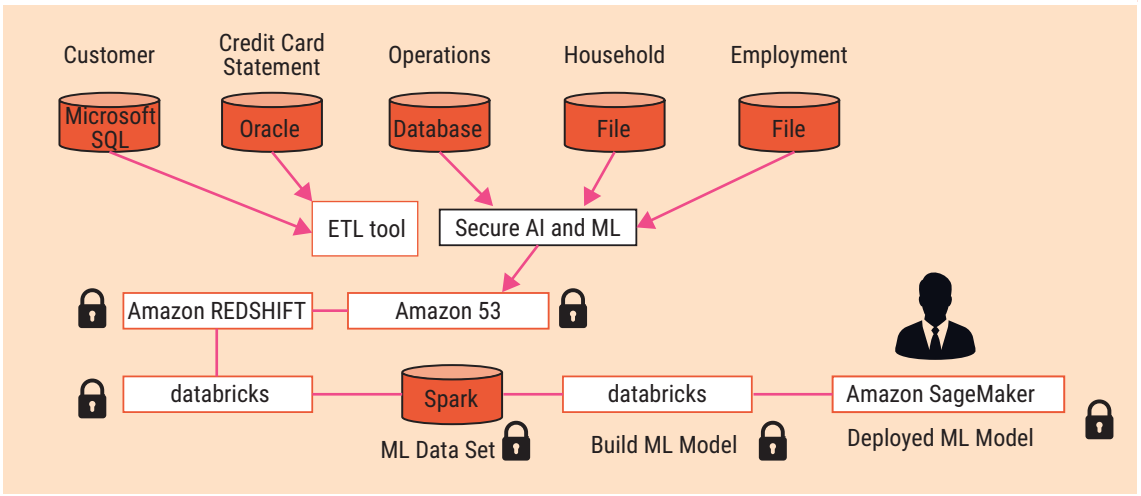
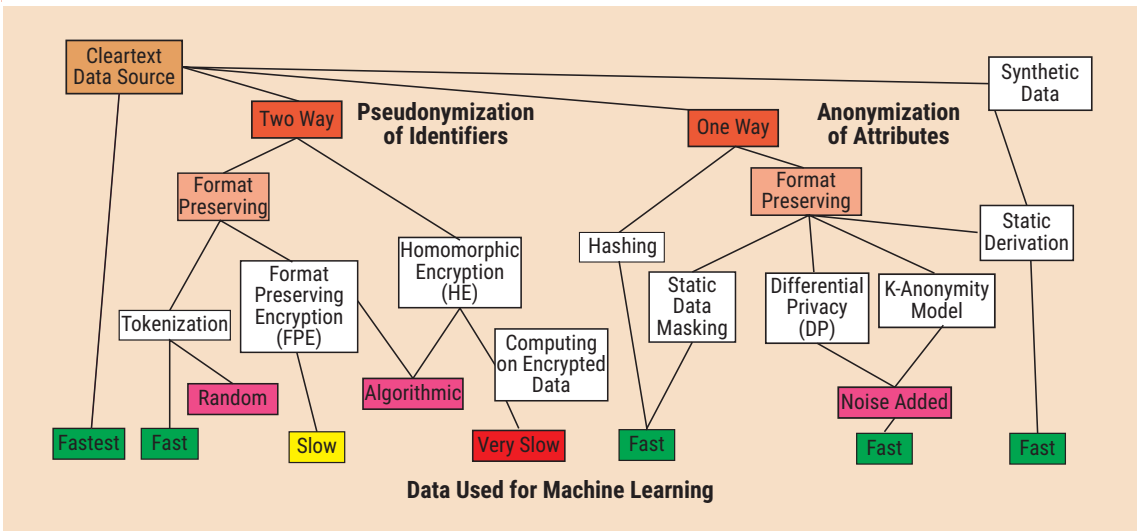


Figure 2—Positioning Characteristics of Different Data Protection Techniques



therefore, regardless of how eccentric any single individual's details are and the details of anyone else in the database, the guarantee of differential privacy still holds. This means that individual-level information about participants in the database is not leaked. This approach supports data sharing scenarios and can be applied to processing data in untrusted environments.[2]

Synthetic Data

Synthetic data are used for a nonreversible approach to generating microdata artificially to represent a predefined statistical data model.[3] By definition, a synthetic data set does not contain any data collected from or about existing data principals, but the data look realistic for the intended purposes. If the synthetic data fit the original data too closely, they can reveal information about the genuine data principals, such as personal data. There are various ways to create synthetic data. Theoretically, data can be generated randomly based on a number of selected statistical properties. Key characteristics of such a model are the distributions of each attribute (overall and in subpopulations) and the internal relationships among the attributes. In practice, the generation of synthetic data can involve multiple or continuous transformations on real data sets using randomization techniques and sampling. Typically, synthetic data are used for testing tools and applications, for developing queries, in some applications and as surrogates for real data. The data curator should reproduce queries performed on synthetic data on actual data to ensure that inferences drawn on the synthetic data are correct when drawn on real data. The privacy guarantees of synthetic data can be evaluated using the differential privacy model.[4]

> "A SYNTHETIC DATA SET DOES NOT CONTAIN ANY DATA COLLECTED FROM OR ABOUT EXISTING DATA PRINCIPALS, BUT THE DATA LOOK REALISTIC FOR THE INTENDED PURPOSE."

Anonymization in Healthcare

K-anonymity can be used to generalize data. 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. For k-anonymity to be achieved, there needs to be at least "k" individuals in the data set who share the set of attributes that might be identifying for each individual.[5] **Figure 3** shows an example of k-anonymity.

K-anonymity might be described as a "hiding in the crowd" guarantee: If each individual is part of a larger group, then any of the records in this group could correspond to a single person.[6] **Figure 4** shows the data anonymized. This is achieved by generalizing some quasi-identifier attributes and redacting some others.

Synthetic Data in Fintech

When historical data are not available or when the available data are not sufficient because of lack of quality or diversity, organizations rely on synthetic data to build models. A random sample of any distribution can be generated. The utility of synthetic data varies depending on the analyst's degree of knowledge about a specific data environment. Fitting real data to a known distribution by generating synthetic data can be done to generate synthetic data. There are also various tools such as the CA Technologies Datamaker and the Informatica Test Data Management Tool that can be used to generate data. Static derivation of real data to synthetic data can provide data that are not regulated but highly useful for sharing with third parties (**figure 5**).[7]

Quantum Computing: The Pros and Cons

There are pros and cons to using quantum computers. Quantum computers can improve

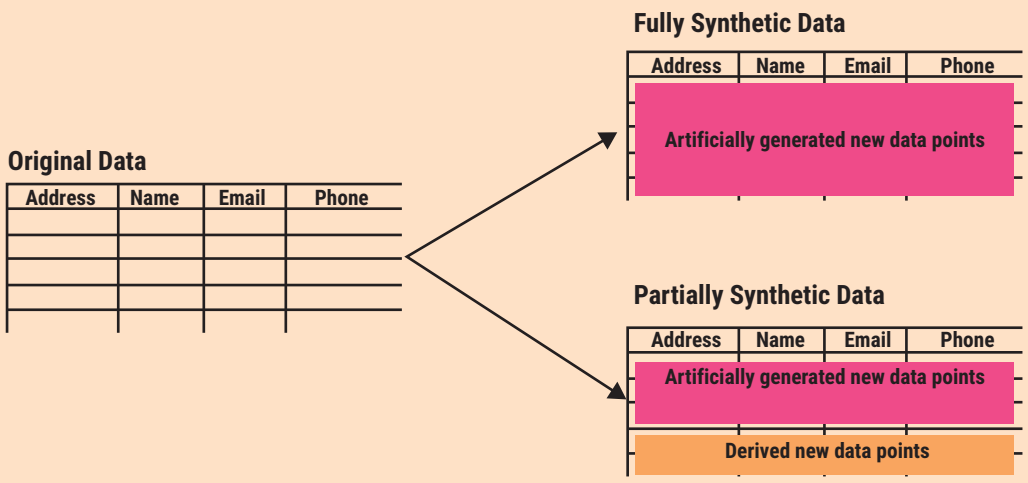
Figure 3—Sample Data Set

Identifier	Quasi Identifiers			Confidential Attributes	
Civic Number	Gender	Age	Post Code	Wage	Affiliation
123-55-1321	M	22	94123	22	Socialist
321-33-4321	F	26	94321	33	Conservative
876-89-6543	M	24	94654	44	Conservative
345-56-6789	M	40	90222	55	Socialist
876-34-4322	F	38	90654	43	Conservative
837-45-1256	F	42	90876	32	Socialist

Figure 4—K-Anonymity Applied to a Data Set

Identifier	Quasi Identifiers			Confidential Attributes	
Civic Number	Gender	Age	Post Code	Wage	Affiliation
*	M	40	94***	22	Socialist
*	M	38	94***	33	Conservative
*	M	42	94***	44	Conservative
*	F	22	90***	55	Socialist
*	F	26	90***	43	Conservative
*	F	24	90***	32	Socialist

Figure 5—Static Derivation of Real Data to Synthetic Data



Enjoying this article?

- Read *Privacy by Design and Default. www.isaca.org/Privacy-by-Design*
- Learn more about, discuss and collaborate on privacy in ISACA's Online Forums. *https://engage.isaca.org/onlineforums*

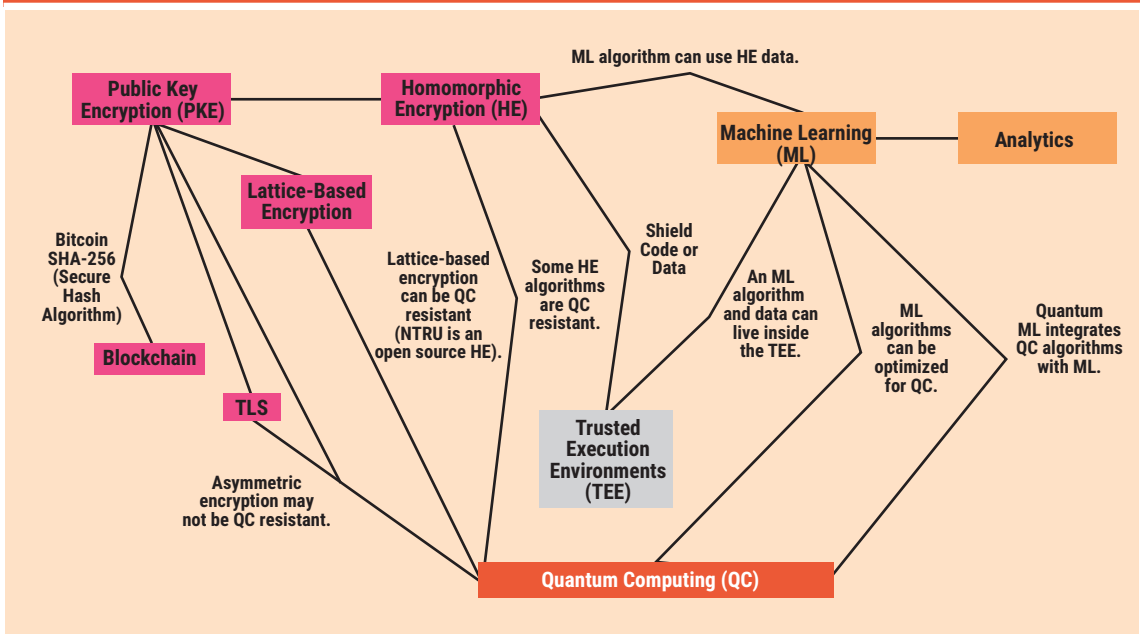
performance in computing, some quantum ML algorithms can be optimized for quantum computers, and quantum computers can break algorithms and patterns in encrypted data, in particular public key cryptography. **Figure 6** illustrates how some of these techniques are related. In a scenario where an organization is using analytics in ML models, the organization likely wants to protect some sensitive data ML models that it runs in the cloud. This can be done with a trusted executing environment (TEE).

> "THE CLOUD CAN BE USED WITH AI TO REINVENT HOW ORGANIZATIONS MAKE DECISIONS."

Homomorphic Encryption and Quantum Computing

New homomorphic encryption (HE) algorithms can be secure from quantum computer-based attacks, and ML algorithms can be optimized for quantum computers.[8] HE, which allows computations on encrypted data, and ML are growing in popularity. Both HE and quantum computing can be applied to provide privacy and security for sensitive data and confidential ML models in vulnerable environments, such as different cloud models. The EU General Data Protection Regulation (GDPR), the US State of California Consumer Privacy Act (CCPA) and other regulations already enacted—with more sure to come—only stress the need for enterprises to protect their data. Data must move in a protected form through an enterprise's many hybrid cloud databases and applications. The cloud can be used with AI to reinvent how organizations make decisions. The types of data that are most critical in driving innovation—with advanced analytics, ML and AI—are those deemed most sensitive and, therefore, they must be safeguarded. ML models and data in them can be shielded in TEEs. A TEE is similar to a hardware security module (HSM) but faster and typically not evaluated against US National Institute of Standards and Technology (NIST) Federal Information Processing Standard (FIPS) 140 requirements. This is complementing protection provided by encryption of the nonlinear ML models and data for outsource environments, such as different cloud models and IoT devices. Linear ML models can be protected by HE with reasonable performance degradation. Nonlinear ML models can be extremely slow when protected with homomorphic encryption.

Figure 6—Scenario With Data Protection for Analytics



Opportunities With Quantum Computing

HE and ML allow enterprises to use sensitive data to fuel advanced analytics, ML and AI, even as those initiatives migrate to cloud environments.[9] This leads to opportunities as well as threats with current and future computers.

Quantum Computers Can Break Blockchain and Public-Key Cryptography

One of the biggest challenges surrounding digital technology is securing systems and data. For decades, computer scientists have worked to develop increasingly sophisticated algorithms designed to encrypt data and protect them through frameworks such as public-key cryptography (PKE), which is also known as asymmetric cryptography.[10] These frameworks function relatively well, and billions of transactions and interactions use these algorithms every day.

As quantum computers advance and creep into the mainstream, they introduce a level of computing power that raises the stakes.[11] Although there are many potential benefits, a major disadvantage is the ability to crack today's PKE, including widely used Rivest-Shamir-Adleman (RSA) and Diffie-Hellman frameworks. This impacts everything from routers and virtual private networks (VPN) to the ability to verify digital signatures.

In 2016, the US National Security Agency (NSA) issued an alert and recommended that organizations begin looking at ways to switch to more advanced cryptography. A year later, NIST began soliciting new and more advanced algorithms that could withstand cracking by quantum computers and become standard.[12]

Quantum computers lack the processing power to succeed in a brute-force assault on classical cryptography algorithms. However, in a few years, once these machines hit a threshold of approximately 10 million physical qubits, they will possess this power. The risk is palpable for enterprises, universities and governments. If quantum computers crack PKE algorithms, more than just devices would be affected; an enterprise's historical data could be exposed. Consequently, mathematicians and computer scientists are developing new and far more advanced cryptographic algorithms that use both classical and lattice-based frameworks.[13] The former relies on noncompact code; the latter uses mathematical formulas or proofs to ensure the integrity of the algorithm. In fact, lattice-based algorithms are part of a broader move toward formal (verified) software.

> "QUANTUM SUPREMACY, WHICH DESCRIBES THE POINT IN TIME WHEN QUANTUM COMPUTERS EXPLICITLY OUTPERFORM CLASSIC ONES, IS ON THE VERGE OF REALIZATION."

Threats With Current Computers

Outside of quantum computers, current Intel Xeon computers are also a threat to RSA encryption. The security of RSA relies on the practical difficulty of factoring the product of two large prime numbers (the factoring problem). Breaking RSA encryption is known as the RSA problem. Factoring is demonstrated by an RSA key that has 240 decimal digits and a size of 795 bits.[14]

For the short term, enterprises can keep safe from improvements in Intel processors and similar processors by moving to at least 2048-bit RSA, Diffie-Hellman or DSA keys. Although Transport Layer Security (TLS) negotiation times are slower with a larger key, it is likely only noticeable on busy sites. Most sites that are busy enough for the slowdown to affect it can likely afford to buy or rent the hardware needed to help support it. For short-term use, Curve25519 can be used; it is very fast and is unaffiliated with NIST. It is the approximate equivalent of 128-bit Advanced Encryption Standard (AES) encryption.

Blockchain Security

Although it may not be as inherent as some believe, resilience is one of the main motivations for enterprises to use blockchain technology.[15]

> IN MANY CASES, ENTERPRISES WILL NEED TO UPDATE CERTIFICATE MANAGEMENT FRAMEWORKS, DEVICES AND SOFTWARE TO SUPPORT NEW ALGORITHMS.

Blockchain relies on Internet connectivity and public key infrastructure (PKI) based on symmetric encryption such as RSA algorithms. The blockchain framework relies on the security of the cryptographic processes underlying it. Without trusted hash functions and public key signatures, there can be no blockchains. Quantum computers threaten several of the cryptographic primitives used in blockchains. Scalable quantum computers, which are necessary to attack the mathematical problems behind the cryptographic primitives, are not yet available; however, small-scale quantum computers have already been built by several enterprises and governments. Some are even accessible on the Internet and can be used to test quantum algorithms. Quantum supremacy, which describes the point in time when quantum computers explicitly outperform classic ones, is on the verge of realization, if not already attained. Therefore, it is important to understand the threat posed to blockchains and to outline possible solutions.

Quantum-Resilient Algorithms

Within the next couple of years, NIST is expected to finalize new standards for quantum-resilient algorithms.

For now, enterprises can prepare for this next phase of cryptography by staying updated on the NIST initiative and keeping an eye on breaking news in the field. It is not too early to begin assessing systems and devices and considering when and where quantum-resilient algorithms make sense. In many cases, enterprises will need to update certificate management frameworks, devices and software to support new algorithms. It is also a good idea to upgrade older systems to 256-bit keys to maximize data protection.

Fortunately, symmetric-key cryptography (which relies on private keys) is not as susceptible to being cracked by quantum computing and is not considered at risk for now. However, it is impossible to rely on symmetric key cryptography to handle many of the interactions and transactions that take place in today's computing environment. Once quantum-safe algorithms appear, it would be wise to migrate to them as soon as possible.

Many PKE algorithms rely on extremely large numbers that are the product of two prime numbers. Other encryption algorithms base their security on the difficulty of solving certain discrete logarithm problems. With sufficiently big enough key sizes, there is no known way to crack the encryption they provide. The factoring of the large numbers and the computing of a discrete logarithm defeat the cryptographic assurances for a given key size and force users to ratchet up the number of bits of entropy they use.

Post-quantum cryptography research focuses on six different approaches:[16]

1. **Lattice-based cryptography**—This approach includes cryptographic systems such as learning with errors (LWE), ring learning LWE, the ring LWE key exchange, the ring LWE signature, the older NTRU or Goldreich-Goldwasser-Halevi (GGH) encryption schemes, and the newer NTRU signature and Bimodal Lattice Signature Scheme (BLISS) signatures.
2. **Multivariate cryptography**—This includes cryptographic systems such as the Rainbow Unbalanced Oil and Vinegar (UOV) scheme, which is based on the difficulty of solving systems of multivariate equations. Various attempts to build secure multivariate equation encryption schemes have failed.
3. **Hash-based cryptography**—This includes cryptographic systems such as Lamport signatures and the Merkle signature scheme and the newer eXtended Merkle Signature Scheme (XMSS) and SPHINCS schemes. Hash-based digital signatures were invented in the late 1970s and have been studied ever since as an interesting alternative to number-theoretic digital signatures such as RSA and Digital Signature Algorithm (DSA).
4. **Code-based cryptography**—This includes cryptographic systems that rely on error-correcting codes, such as the McEliece and Niederreiter encryption algorithms and the related Courtois, Finiasz and Sendrier signature scheme. The original McEliece signature using random Goppa codes has withstood scrutiny for more than 30 years.

5. **Supersingular elliptic curve isogeny cryptography**—This cryptographic system relies on the properties of supersingular elliptic curves and supersingular isogeny graphs to create a Diffie-Hellman replacement with forward secrecy.
6. **Symmetric key quantum resistance**—If a sufficiently large key size is used, the symmetric key cryptographic systems such as Advanced Encryption Standard (AES) and SNOW 3G are already resistant to attacks by quantum computers.

Beyond Breakable Encryption

The US State Department and several other US government agencies mandated the move from 128-bit AES to 256-bit AES and the cessation of certain secure hashes associated with 256-bit AES.[17] **Figure 7** is an example of a cryptography road map in preparation for quantum computing.

The Road to Randomness

Quantum computers and other strong computers can break algorithms and patterns in encrypted data. Alternatively, random numbers can be used to secure sensitive data, as they are not based on an algorithm or pattern.

> "THE BLUE TOKENS REPRESENT TEMPORARY RESULTS AND THE FINAL TOKEN VALUES ARE GREEN."

Random numbers should be validated by the NIST statistical test suite for random numbers. NIST Special Publication (SP) 800-22 offers 15 statistical tests that assess the presence of a pattern that, if detected, indicates that the sequence is nonrandom.[18] The focus of the test is the proportion of zeros and ones for the entire sequence. The purpose of the test is to determine whether the number of ones and zeros in a sequence are approximately the same as would be expected for a truly random sequence. The test suite includes tests for frequency and approximate entropy.

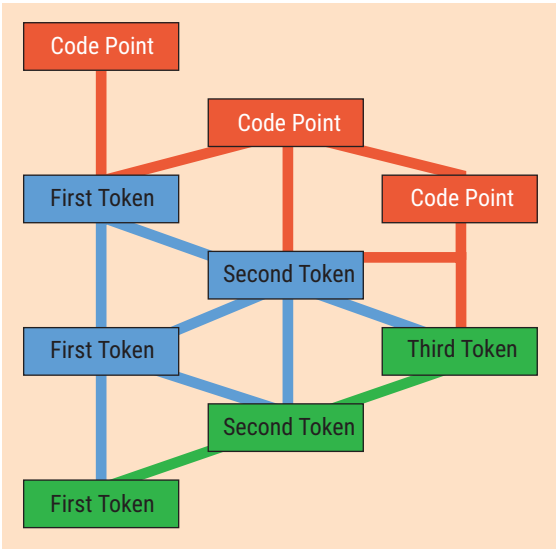
Figure 7—Example of a Cryptography Road Map

Time frame	Area	Comment
Short	Upgrade to AES, preferably AES-256 with strong random seed	Immediate-medium step
Short	Use SHA-512 for hashing	Immediate step
Short	Short use stateful hash-based signatures for signing	Immediate review
Short	Use hybrid cryptography to protect against both weaknesses in RSA/ECC and potential weaknesses in post-quantum algorithms	Immediate steps
Medium	Lattice-based algorithms	Tools study and integration plan
Medium	HE	Tools and partner integration
Medium	Operation on encrypted data	Integration of protocols
Medium	Secure multiparty computing (SMPC)	Integration of protocols
Medium	2022 NIST to complete review safe algorithms	Tools integration
Medium	2022 NIST standards to be released	Tools integration
Long	2024 industry standards based on NIST algorithms from NIST Standards	Tools integration
Long	Analytics and ML	ML algorithms optimized for quantum processors
Long	Full industry adoption 2019+	Tools integration

Building a Token Fabric of Protected Data Elements

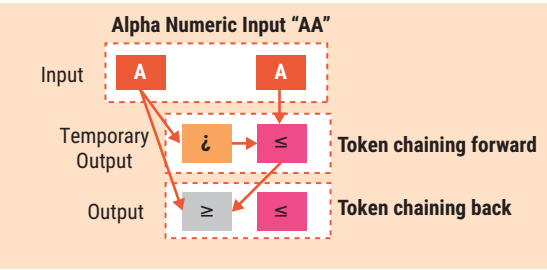
Protecting short data is always a challenge, since the entropy or variation of possible values is small. A fabric can be used to increase the entropy when protecting short data. A fabric of intermediate data tokens that replaces the clear text data can be created. These are randomized values where each layer gradually increases the entropy of each final token that will replace the original input data. **Figure 8** is an example of a token fabric of intermediate tokens. The blue tokens represent temporary results and the final token values are green.

Figure 8—Example of a Fabric of Intermediate Tokens



A tokenization function based on randomized lookup tables can be used. The chaining of tokens can add entropy via additional input data to the tokenization process in each step. **Figure 9** is an example of short data with a two-character input-string ("AA") that will generate the middle layer tokens that are temporary results and the final tokens on the bottom layer. Each token is based on unique initialization values and is chained forward and backward with the others to increase high entropy and randomness. The security of the specific design and implementation should be validated by leading experts and universities.

Figure 9—Example of Chaining to Add Entropy to the Tokenization Process



Conclusion

Innovative enterprises can stay competitive by implementing solutions that help extract value from sensitive data. New techniques such as TEE and tokenization fabrics make it possible for enterprises to securely use private information—including its application in advanced analytics, ML and AI—to be successful without worrying about putting customers, employees or intellectual property at risk.

Commonly implemented solutions do not provide strong protection from quantum computers. Proper planning for and understanding of available technologies such as tokenization fabrics and enhancement options offered by evolving technologies of quantum computers can provide realistic approaches to data protection that give enterprises a competitive advantage over less innovative competitors.

> "COMMONLY IMPLEMENTED SOLUTIONS DO NOT PROVIDE STRONG PROTECTION FROM QUANTUM COMPUTERS."

Endnotes

1 Zhao, J.; T. Jung; Y. Wang; X. Li; "Achieving Differential Privacy of Data Disclosure in the Smart Grid," IEEE Conference on Computer Communications, Toronto, Canada, April 2014

2 Nayak, C.; "New Privacy-Protected Facebook Data for Independent Research on Social Media's Impact on Democracy," Facebook Research, 13 February 2020, *https://research.fb.com/blog/2020/02/new-privacy-protected-facebook-data-for-independent-research-on-social-medias-impact-on-democracy/*

3 Reiter, J. P.; "Using CART to Generate Partially Synthetic, Public Use Microdata," *Journal of Official Statistics*, vol. 21, iss. 3, January 2005

4 Watson, A.; "Using Generative, Differentially-Private Models to Build Privacy-Enhancing, Synthetic Datasets From Real Data," *Medium*, 2 March 2020, *https://medium.com/gretel-ai/using-generative-differentially-private-models-to-build-privacy-enhancing-synthetic-datasets-c0633856184*

5 Privitar, "K – Anonymity: An Introduction," 7 April 2017, *https://www.privitar.com/blog/k-anonymity-an-introduction/*

6 *Ibid.*

7 Sarkar, T.; "Synthetic Data Generation—A Must-Have Skill for New Data Scientists," *Towards Data Science*, 19 December 2018, *https://towardsdatascience.com/synthetic-data-generation-a-must-have-skill-for-new-data-scientists-915896c0c1ae*

8 Mattsson, U.; "New Technologies for Data Protection That Arm Innovative Businesses to Win," ISACA® San Francisco Chapter (California USA), USA, 21 April 2021, *https://engage.isaca.org/sanfranciscochapter/events/eventdescription?CalendarEventKey=7fc789f0-0538-4887-a6e0-8668bcdc68c1&CommunityKey=f510bd50-4fdc-46b1-a329-d6ce8a64bae7&Home=%2Fcommunities%2Fcommunity-home%2Frecent-community-events*

9 Mattsson, U.; "Homomorphic Encryption Will Take on the Challenge of AI," RSA Conference, 25 February 2021, *https://www.rsaconference.com/Library/blog/Homomorphic%20Encryption%20Will%20Take%20on%20the%20Challenge%20of%20AI*

10 The National Academies Press, *Decrypting the Encryption Debate: A Framework for Decision Makers*, USA, 2018

11 IBM, "Quantum Computing: Tomorrow's Computing Today," *https://www.ibm.com/quantum-computing/?p1=Search&p4=43700050386405608&p5=b&gclsrc=aw.ds&gclid=EAIaIQobChMIuLe0jcOR8AIVhrLICh1lCQZ0EAAYASAAEgI2avD_BwE*

12 National Institute of Standards and Technology (NIST), "NIST Kicks Off Effort to Defend Encrypted Data From Quantum Computer Threat," USA, 28 April 2016, *https://www.nist.gov/news-events/news/2016/04/nist-kicks-effort-defend-encrypted-data-quantum-computer-threat*

13 The National Academies Press, *Quantum Computing: Progress and Prospects*, USA, 2019

14 Goodin, D.; "New Crypto-Cracking Record Reached, With Less Help Than Usual From Moore's Law," *Ars Technica*, 12 March 2019, *https://arstechnica.com/information-technology/2019/12/new-crypto-cracking-record-reached-with-less-help-than-usual-from-moores-law/*

15 Deloitte, "Security Controls for Blockchain Applications," *https://www2.deloitte.com/ch/en/pages/risk/articles/security-controls-for-blockchain-applications.html*

16 Raffaelli, F.; R. Denman; R. Collins; J. C. Faugere; G. De Martino; C. Shaw; J. Kennard; R. Sibson; L. Perret; C. Erven; "Combining a Quantum Random Number Generator and Quantum-Resistant Algorithms Into the GnuGPG Open-Source Software," *Advanced Optical Technologies*, vol. 9, iss. 5, 2020

17 *Op cit* Goodin

18 Rukhin, A.; J. Soto; J. Nechvatal; M. Smid; E. Barker; S. Leigh; M. Levenson; M. Vangel; D. Banks; A. Heckert; J. Dray; S. Vo; *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*, National Institute of Standards and Technology (NIST) Special Publication (SP) 800-22, USA, 2010, *https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-22r1a.pdf*