Tokens and Tokenization

A token is a token is a token

In General, Tokens:

- share some characteristics with the original data elements: character set, length, etc.
- are deterministic: repeatedly generating a token for a given value yields the same token.
- map each data element to a unique value.
- enable searching tokenized data stores by tokenizing query terms and searching for those.

Tokenization is a process by which sensitive data such as PANs (Primary Account Numbers), Personally Identifiable Information (PII) elements, Protected Health Information (PHI), et al. are securely replaced by surrogate values. These surrogate values are called **tokens**. **Tokenization is typically defined as any form of format-preserving data protection.** The terms "format-preserving encryption" ("FPE"), "tokenization", "masking", and others are often used interchangeably to indicate or imply the same method of data protection.

The technology underlying tokenization varies according to use case requirements. Voltage SecureData offers a variety of format-preserving data protection methods, including Format-Preserving Encryption (FPE) Embedded Format-Preserving Encryption (eFPE), Format-Preserving Hash (FPH), and Secure Stateless Tokenization (SST).

Tokenization History

The first tokenization was created by TrustCommerce in 2001 to help merchants protect customer credit card information. Merchants were storing cardholder data on their own servers, which meant that anyone who had access to their servers could potentially view or take advantage of those customer credit card numbers.

TrustCommerce developed a system that replaced primary account numbers (PANs) with a randomized number called a token. This allowed merchants to store and reference tokens, instead of actual PANs, when accepting payments. The merchants submitted payments to TrustCommerce, who converted the tokens back to PANs and processed the payments using the original PANs. This isolated the risk to TrustCommerce, since merchants no longer had any actual PANs stored in their systems.

As security concerns and regulatory requirements grew, such first-generation solutions proved the technology's value, and other vendors offered similar solutions. However, problems soon became clear.

While conceptually simple, any tokenization or detokenization request using this approach must make a server request, adding overhead, complexity, and risk. It also does not scale well. Consider a request to tokenize a value. The server first performs a database lookup to see if it already has a token for that value. If it does, it returns that. If not, it must generate a new random value, then do another database lookup to make sure that value has not already been assigned for a different cleartext. If it has, it must generate another random value, check that one, and so forth. As the number of tokens created grows, the time required for these database lookups increases; worse, the likelihood of such collisions grows exponentially. These implementations must also use multiple token servers—for load-balancing, reliability, and failover—and these must perform real-time database synchronization to ensure reliability and consistency, adding further complexity and overhead. Newer technologies such as Voltage Secure Stateless Tokenization have replaced such approaches.

What Are the Different Types of Tokenization or Format-Preserving Data Protection?

There are two tokenization (format-preserving data protection) categories, **reversible** and **irreversible**. **Reversible** tokenization means a process exists to convert tokens back to cleartext (*pseudonymization* in privacy terminology).

The Payment Card Industry Data Security Standard (PCI DSS) distinguishes between cryptographic and non-cryptographic approaches. This is questionable, as most definitions of cryptography, including the U.S. Export Administration Regulations, consider all such processes to be cryptography. However, some PCI Qualified Security Assessors (QSAs) still make the distinction and prefer non-cryptographic tokens over cryptographic tokens when performing PCI DSS assessments.

- **Cryptographic** tokenization generates tokens.

  NIST-standard FF1-mode AES is an example of cryptographic format-preserving data protection.

  Voltage SecureData offers FF1-mode based Format-Preserving Encryption (FPE) and Embedded Format-Preserving Encryption (eFPE) for cryptographic tokenization.

- **Non-Cryptographic** tokenization is divided into two approaches.

  Dynamic database or *vaulted* tokenization is first-generation technology that randomly generates tokens, mapped to sensitive data elements which are mapped using a lookup table or database. Such implementations suffer from the problems described earlier. There are no industry standards for such approaches, just guidelines such as Tokenization Product Security Guidelines.

  Voltage discontinued database vault-based tokenization due these inherent problems.

  Stateless tokenization securely manipulates randomly generated metadata to build tokens. Such systems can operate disconnected from each other, and scale essentially infinitely, since they require no synchronization beyond copying of the original metadata.

  Voltage SecureData offers Secure Stateless Tokenization (SST), offering optimal performance, security, and scalability.

Irreversible tokenization means there is no practical way to convert tokens back to cleartext (*anonymization* in privacy terminology).

Tokens are created through a one-way function to anonymize data elements for third-party analytics, for using production data in lower environments, etc.

Voltage SecureData offers Format-Preserving Hash (FPH) for irreversible tokenization.

Sensitive Data Type	Sensitive Data	Tokenized Data
Credit Card number	1111-2222-3333-4444	1111-**2287**-**9581**-4444
U.S. Social Security number	**999-88-7654**	740-36-7654
Address	1234 Maple Street	**7321 Uqhap Fbzira**
Phone number	415-555-1234	**819-913-0471**

Tokenization (Format-Preserving Data Protection) vs. Traditional Encryption

Many vendors and service providers sell traditional encryption and tokenization products and services separately. While both provide data protection, encryption services perform standard AES (GCM, CBC, CTR, EME, etc.) encryption, generating outputs that:

- have different characteristics from the original data elements, such as format, length, etc.
- are significantly larger than the original data elements
- may be probabilistically created: repeatedly encrypting a given data item yields different outputs
- are generally used to protect semi-structured or unstructured data, or structured data where there is no requirement to preserve referential integrity

Voltage SecureData offers Identity-Based Symmetric Encryption (IBSE) and asymmetric Identity-Based Encryption (IBE) for select use cases.

A common alternative to field-level tokenization is encrypting data *en masse* at the file/database/file system/storage level. This is appealing because it is transparent to users and applications: data is automatically decrypted when accessed. However, this approach provides much less security than field-level data tokenization, because all it really protects against is physical theft of disk drives. For example, laptops and mobile devices often use whole-disk encryption technologies such as Windows BitLocker, Apple FileVault, or Symantec Endpoint Encryption, because these devices are prone to getting lost or stolen. While there is value in reducing this risk, it does not address today's typical cyber threats.

Voltage SecureData does not offer any transparent database, file system, or storage-level encryption, as these do not offer true data-centric security and only serve to comply with basic data-at-rest protection requirements. To solve this use case, the Voltage solutions portfolio offers file-level encryption with Voltage SmartCipher to protect sensitive personal data in unstructured repositories such as SharePoint, Microsoft Exchange, and many more.

The terms “tokenization,” “masking,” and “obfuscation” are all often used generically to mean any form of format-preserving data protection.

Voltage SecureData offers reversible cryptographic, reversible non-cryptographic, and irreversible tokenization techniques.

Contact us at **CyberRes.com**
Like what you read? Share it.

What Is the Voltage Differentiator?

Voltage offers vast knowledge and expertise in data protection technology, including creating and patenting FF1-mode AES Format-Preserving Encryption. Voltage SecureData is an enterprise, highly scalable, performant solution that offers a variety of tokenization methods free from database-backed tokenization limitations, with extensive integration capabilities. Voltage SecureData offers ***all*** choices of tokenization methods (reversible cryptographic, reversible non-cryptographic, and irreversible cryptographic) to protect any data type. Voltage also provides Stateless Key Management, which removes the burden of key management.

Learn more at
www.microfocus.com/en-us/cyberres/data-privacy-protection/securedata-enterprise

CyberRes
A Micro Focus Line of Business