opentext™ | Cybersecurity

Solution Flyer

Tokens and Tokenization

Tokenization is de ned as any form of format-preserving data protection. Voltage SecureData Enterprise is a highly scalable, highly performant enterprise class solution that o ers a variety of tokenization methods, including reversible and irreversible cryptographic, and reversible non-cryptographic approaches, to protect any data type in any language.

Attributes of Tokens

- tokenization systems map each data element to a unique value
- tokens may share characteristics with the original data elements, such as character set and length
- tokenization systems are deterministic, such that repeatedly generating a token for a given value yields the same token
- tokenized data can be searched by tokenizing the queried data elements and searching for those

Tokenization is a process by which PANs (Primary Account Numbers), PHI (Protected Health Information), PII (Personally Identi able Information), and other sensitive data elements are securely replaced by surrogate values referred to as tokens. Tokenization is a form of encryption, but the two terms are frequently used in distinct ways. Encryption is commonly meant as the encoding of human-readable data into incomprehensible text that can only be decoded with the right decryption key, while tokenization (or “masking”, or “obfuscation”) is commonly meant as some form of format-preserving data protection; that is, the conversion of sensitive data elements into non-sensitive, replacement values—tokens—that are the same length and format as the original data elements.

Tokenization History

Digital tokenization was rst created by TrustCommerce in 2001 to help a client protect customer credit card information. Merchants were storing PANs on their own servers, which meant that anyone who had access to their servers could potentially view those customer credit card numbers.

TrustCommerce developed a system that replaced the PANs with a randomized number correlated with the PAN through a database. This allowed merchants to store and reference these tokens, instead of the PANs, when accepting payments. When processing the payments, TrustCommerce converted the tokens back to the original PANs. This isolated the risk to TrustCommerce, since the merchants no longer stored PANs in their systems.

As security concerns and regulatory requirements grew, such rst-generation solutions proved the technology’s value, and other vendors o ered similar solutions. However, as discussed in the next section, problems with this approach soon became clear.

What Are the Di erent Types of Tokenization?

There are two types of tokenization: reversible and irreversible.

Reversible tokenization means a process exists to convert the tokens back to their original values. In privacy terminology, data protection via a reversible process is called pseudonymization. Such tokens may be further subdivided into cryptographic and non-cryptographic, although this distinction is arti cial, since tokenization really is a form of encryption.

- Cryptographic tokenization generates tokens using strong cryptography; the cleartext data elements are not stored anywhere—just the cryptographic key.

  NIST-standard FF1-mode AES, pioneered by Voltage Security, now by OpenText in the early 2000s, is an example of cryptographic tokenization.

  Voltage SecureData Enterprise by OpenText’s Format-Preserving Encryption (FPE) and Embedded Format-Preserving Encryption (eFPE) are cryptographic tokenization methods.

- Non-Cryptographic tokenization originally meant that tokens were created by

A common use of traditional encryption for structured data is to protect it at the disk level. This is appealing because it is transparent to users and applications: data is automatically encrypted when written to disk, and automatically decrypted when accessed. However, this approach does not protect the data against modern attack vectors, such as malware, insider threats, and phishing approaches, that take advantage of that automatic decryption “feature.” All it really protects against is access attempts following physical theft of the disk drives. For example, laptops and mobile devices often use whole-disk encryption technologies, such as Windows