Protegrity Tokenization

Securing Sensitive Data for PCI, HIPAA and Other Data Security Initiatives

2011 Edition

Who should read it

System architects, security experts, and other IT professionals who are looking to use tokenization to protect massive stores of sensitive data for compliance

What it’s about

- Comparing traditional tokenization with Protegrity Tokenization
- Understanding the benefits of Protegrity Tokenization
- Describing considerations for deploying tokenization in various scenarios
- Reviewing use cases for centralized, distributed, and local deployment
“With escalating IT security costs and continued complexity related to PCI DSS compliance, organizations should take a fresh look at the advantages of tokenization. Protegrity highlights how tokenization can help secure other types of personally identifiable information, such as medical IDs, dates and e-mail addresses that are strictly governed under HIPAA, providing a useful and often more application-friendly alternative to tactics such as encryption. Those who think that tokenization is only about credit account security should explore its advantages more deeply.”

Scott Crawford
Research Director at Enterprise Management Associates
Executive Summary

Tokenization is a process of protecting sensitive data by replacing it with alias values or tokens that are meaningless to someone who gains unauthorized access to the data. Tokenization is one of many technologies used to protect data, such as various techniques for encryption, monitoring/blocking/masking data, truncation, and hashing. Each of these technologies entails tradeoffs in performance, storage, security, and transparency. Tokenization, in particular, has caught recent interest as a way to cost-effectively protect large stores of cardholder and personally identifiable data – protection that is required by the Payment Card Industry Data Security Standard (PCI DSS), HIPAA/HITECH, and other regulations and business mandates.

As these are “early days” of using tokenization, best practices are just emerging. VISA Inc. published its best practices on tokenization for PCI DSS in July 2010. The PCI Security Standards Council will issue related guidance in 2011. While the guidelines are meant to clarify, many relate to traditional tokenization solutions. Those implementations are the prime source of confusion as they require massive computing resources, impose complex operations, deliver poor performance from system latency, and pose practical limitations on the amount of protected data.

Protegrity has invented a new tokenization technology that simplifies and dramatically lowers the cost of deploying this mode of data protection. Tokens are cost effective because they don’t require security controls for protection – they’re not sensitive data, they only represent that data. Yet tokens are transparent to most applications and business processes.

This paper describes Protegrity Tokenization, compares it with traditional tokenization solutions, and illustrates how Protegrity Tokenization addresses their shortcomings. The paper also provides a use case illustrating use of tokenization for simplifying PCI compliance.

Technical Benefits
- Tiny footprint, commodity hardware
- No data replication or collisions, little or no latency
- Reduced key management
- Works in parallel environments with high availability
- Clear segmentation of cardholder data
- Protects any kind of data

Business Benefits
- Lower CAPEX and operational costs
- Meets stringent performance SLAs
- Simplifies and reduces cost of operations
- Supports continuous business operations and continuity
- Simplifies PCI compliance
- Meets multiple compliance mandates and business policies

Key Points
- Massive stores of sensitive data need cost effective protection
- Traditional tokenization is not an optimal solution
- Protegrity Tokenization is cost effective and provides superior performance
Traditional Tokenization

Before describing Protegrity Tokenization, let us first consider fundamental elements of traditional tokenization. Solutions using first generation tokenization typically replace sensitive data in host systems with tokens, and store the actual sensitive data in a database table located inside the token server. The token table usually includes several elements in addition to the encrypted sensitive data and the tokens. The token server must store one unique token value for each sensitive data value that is tokenized. As new sensitive data values are tokenized, the token server size increases dynamically to a very large footprint.

A variation of this dynamic approach is to pre-populate the token server with pre-generated tokens. This can improve performance over the dynamic approach by front loading the tokenization process and thus eliminating it during operational transaction processing.

Several problems result from the implementation approach associated with traditional tokenization primarily due to the large footprint of the token server:

1. The large size of the token server makes it expensive and complex to manage. The large footprint forces a centralized deployment. In a central deployment, redundancy can be implemented to ensure high availability. Multiple (redundant) token servers must be replicated at very high speeds to avoid collisions, a condition where multiple tokens can be assigned to a single sensitive data value.

2. Performance and scalability become a challenge. The distance between the token server and all places where tokenization must occur in the enterprise data flow leads to latency, which hampers performance and scalability.

3. Tokenizing more than a single sensitive data category may not be possible. Adding additional categories of data to be protected with tokenization will amplify the problems listed above, thereby constraining the use of tokenization to a small number of data categories.

An Example of Traditional Tokenization

Consider the challenge of tokenizing all Social Security numbers. There are about 310 million people in the U.S. so a token server for this task will contain up to 310 million records, each consisting of at least hash, a token, and an encrypted Social Security number.

A massive data store like this is challenging for traditional tokenization. Replication quickly overwhelms storage and performance. Optimizing the growing databases for search becomes difficult, which slows performance. And this is just for one type of data
Protegrity Tokenization

Resource requirements become enormous for tokenizing multiple data types such as e-mail addresses, user names and passwords, driver licenses and passport numbers.

The pre-populated variant of traditional tokenization uses pre-generated tokens to improve performance related to tokenization. This approach eliminates the replication requirement and some associated latency issues for distributed deployments because distributed token servers contain all possible tokens. However, when tokenizing Social Security numbers, there is no algorithm that can determine if a pre-generated value is a real Social Security number. To accommodate all possibilities, the pre-populated token solution must create about one billion records for all combinations of nine digits. In the real world, implementing traditional tokenization for multiple data types on this scale is impractical.

Protegrity Tokenization

Protegrity has developed a new tokenization approach that delivers random pre-generated tokens. This patent-pending technology attacks the root cause of many problems associated with traditional tokenization in several ways – including the large footprint.

1. **Keeps token tables small, static and manageable.** Protegrity Tokenization uses pre-generated token blocks in one or more token tables. The token tables contain randomly generated token blocks that are smaller than the target token size. Arriving at the target token size requires multiple lookups into the pre-generated token block tables.

2. **Token block technology provides efficiency for generating tokens.** The Protegrity Tokenization solution can generate any number of tokens for virtually any data type – whether it is a Social Security number with an estimated 300 million unique numbers, an e-mail address with an infinite number of combinations, or a date containing endless possibilities.

3. **Eliminates target of sensitive data.** The Protegrity Tokenization server never grows nor does it store the original sensitive data. Eliminating the encrypted sensitive data from inside the token server erases the target for data thieves.

Token Flexibility

Protegrity Tokenization delivers the flexibility required to protect many different types of data. Protegrity Tokenization can create Numeric, Alpha, Alpha-Numeric, and Date tokens. The token specification can also be used to expose different parts of the token such as the last 4 digits.

“The adoption of tokenization by companies that need to accept, store and transmit credit and debit card data enables those enterprises to greatly reduce the scope of their PCI DSS compliance efforts.”

Avivah Litan
VP Distinguished Analyst, Gartner, Inc.
“Toolkit: Using Tokenization to Limit the Scope of PCI Audits”
July 29, 2010
Protegrity Tokenization also recognizes and preserves delimiters often used in Social Security numbers.

Table 1 shows different types of data that can be tokenized with the Protegrity platform.

<table>
<thead>
<tr>
<th>Type of Data</th>
<th>Input</th>
<th>Token</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSN</td>
<td>075 67 2278</td>
<td>287 38 2567</td>
<td>Numeric</td>
</tr>
<tr>
<td>SSN last 4</td>
<td>075 67 2278</td>
<td>287 38 2278</td>
<td>Numeric, Expose last 4 digits</td>
</tr>
<tr>
<td>SSN delimiters</td>
<td>075-67-2278</td>
<td>287-38-2567</td>
<td>Numeric, delimiters in input</td>
</tr>
<tr>
<td>Credit Card</td>
<td>3872 3789 1620 3675</td>
<td>8278 2789 2990 2789</td>
<td>Numeric</td>
</tr>
<tr>
<td>Medical ID</td>
<td>29M2009ID</td>
<td>497HF390D</td>
<td>Alpha-Numeric</td>
</tr>
<tr>
<td>Date</td>
<td>10/30/1955</td>
<td>12/25/2034</td>
<td>Date</td>
</tr>
<tr>
<td>E-mail Address</td>
<td><a href="mailto:raul.ortega@protegrity.com">raul.ortega@protegrity.com</a></td>
<td><a href="mailto:empo.snaugs@svtiensnni.snk">empo.snaugs@svtiensnni.snk</a></td>
<td>Alpha Numeric, delimiters in input preserved</td>
</tr>
<tr>
<td>Credit Card</td>
<td>3872 3789 1620 3675</td>
<td>8278 2789 2990 3675</td>
<td>Numeric, Last 4 digits exposed</td>
</tr>
</tbody>
</table>

Table 1: Token format flexibility

Comparing Traditional Tokenization & Protegrity Tokenization

Table 2 below compares key elements of traditional tokenization, traditional tokenization using pre-generated tokens, and Protegrity Tokenization using pre-generated tokens with token blocks.

<table>
<thead>
<tr>
<th>Property</th>
<th>Traditional Dynamic</th>
<th>Traditional Pre-Generated</th>
<th>Protegrity Tokenization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Footprint</td>
<td>Large, Expanding</td>
<td>Large, Static</td>
<td>Small, Static</td>
</tr>
<tr>
<td>Replication</td>
<td>Complex replication required</td>
<td>No replication required</td>
<td>No replication required</td>
</tr>
<tr>
<td>Collisions</td>
<td>Prone to collisions</td>
<td>No collisions</td>
<td>No collisions</td>
</tr>
<tr>
<td>Latency / Performance</td>
<td>Will impact performance and scalability</td>
<td>Will impact performance and scalability</td>
<td>Little or no latency Fastest tokenization in the industry</td>
</tr>
<tr>
<td>Tokenizing many data categories</td>
<td>Potentially impossible</td>
<td>Potentially impossible</td>
<td>Can tokenize many data categories with minimal or no impact on footprint or performance</td>
</tr>
</tbody>
</table>

Table 2: Traditional Tokenization compared with Protegrity Tokenization
Performance Metrics of Traditional vs. Protegrity Tokenization

In choosing a data protection technology, two factors are especially important: transparency and performance.

The performance characteristics of traditional tokenization make it a poor candidate for enterprises requiring a high performing data protection solution. Three issues contribute to the poor performance of traditional tokenization:

- The distance between the token server and places in the enterprise data flow needing the tokenization will create some degree of latency that will degrade performance and scalability.
- The large resource footprint gives it poor distribution characteristics that would otherwise allow its location closer to the protected data.
- The optimization demands placed on the token server database continuously rise as the amount of data grows.

By contrast, the tiny footprint and other factors enable Protegrity Tokenization to deliver unmatched performance. Table 3 compares performance between traditional tokenization and Protegrity Tokenization, based on Protegrity lab tests and customer deployments.

<table>
<thead>
<tr>
<th>Traditional Tokenization</th>
<th>Protegrity Tokenization</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 tokens per second (outsourced model)</td>
<td>200,000 tokens per second (single commodity box with 10 connections)</td>
</tr>
<tr>
<td>5000 tokens per second (in house model)</td>
<td>Tokenize 50 million card numbers in 90 minutes</td>
</tr>
<tr>
<td>30 days to tokenize 50 million card numbers</td>
<td>Grows linearly with additional commodity servers and/or additional connections</td>
</tr>
<tr>
<td></td>
<td>9,000,000+ tokenizations per second on a 6 node 5500 Model Teradata Database</td>
</tr>
</tbody>
</table>

Table 3: Contrasting tokenization performance between traditional tokenization and Protegrity Tokenization

An organization may boost system performance simply by adding commodity servers to a token server farm to meet whatever scalability the business processes demand, such as those encountered in batch tokenization. Improvements in scalability can also be achieved by increasing the number of connections that perform tokenization operations concurrently. Both approaches can be combined to achieve the required throughput.

Protegrity also provides for batch tokenization, which allows for bundling multiple data values sent at once to the token server. The token server performs all tokenization and
Protegrity Tokenization de-tokenization. Tokens are returned via the same batch process. Batching reduces the round trips between the calling program and the token server. So if there are 100 values, batching reduces the number of round trips from 100 to one.

**Protegrity Tokenization Opens Up Possibilities**

There are many new types of data protection solutions that become possible with Protegrity Tokenization. Consider the following:

- Protegrity Tokenization allows a large, distributed organization to segment or remove many IT resources from scope of PCI DSS compliance. For example, by implementing tokenization at the point of sale, it would be possible to de-scope all PCs, servers, and kiosks in thousands of retail locations – resulting in a massive simplification of audits by Qualified Security Assessors.

- Create a farm of token servers front-ended by a load balancer that can grow and scale as the business process requires. Servers are commodity hardware. This enables high availability and scale.

- Protegrity Tokenization brings the data protection process close to the protected data. The small footprint tokenization process enables Protegrity to deliver tokenization on a database server without using a token server! Consider what the elimination of latency does to performance.

- Protegrity Tokenization protects most of the HIPAA defined sensitive fields. Blobs can also be tokenized, but encryption may be better.

- The technology enables tokenization of as many sensitive data categories as needed without increasing processing requirements for the token server.

- Business processes have different performance and scalability requirements. One size does not fit all! Protegrity Tokenization allows the creation of purpose built token servers that serve specific business processes and security concerns, such as requirements of the PCI Data Security Standard, HIPAA/HITECH, or Gramm-Leach-Bliley.

**Use Cases Show Deployment Flexibility**

Traditional tokenization, with its large footprint, delivers a single, one-size-fits-all, central distribution option. Protegrity Tokenization delivers the most flexible deployment options in the industry. This flexibility enables customers to deliver tokenization solutions based on the specific needs and data protection requirements of the target scenario.

Consider three different scenarios for tokenization; Credit Card Payment, Personally Identifiable Information, and Health related information. To provide the optimal
tokenization solution for each of these scenarios requires knowledge of the individual security, performance and scalability requirements of each. There is no reason to be forced to deliver a one-size-fits-all solution.

Protegrity Tokenization can deliver tokenization centrally, purpose built, or even without a token server. Different distribution options enable the control of latency and thereby affect performance.

Table 4 describes the data security and the performance requirements of three different types of data for a company; credit card transactions, personally identifiable information in the human resources files of its employees, and medical information about its customers. A pharmacy would be subject to all these requirements; at least one of these will apply to most companies.

<table>
<thead>
<tr>
<th>What needs to be protected?</th>
<th>Security</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Credit Card Payment</td>
<td>PCI DSS</td>
<td>Hundreds of thousands of transactions per day</td>
</tr>
<tr>
<td>Human Resources</td>
<td>State Laws</td>
<td>Thousands of transactions per day</td>
</tr>
<tr>
<td>Healthcare</td>
<td>HIPAA/HITECH</td>
<td>Hundreds of thousands of transactions per day</td>
</tr>
</tbody>
</table>

Table 4: Security and performance requirements of three types of data

Protegrity Tokenization offers deployment choices that will deliver optimal solutions for each type of data to uniquely address the individual security and performance requirements.

The Protegrity Tokenization based solution can deliver an optimal solution for each type of data.

a) **Credit Card Transactions**: The throughput required for processing hundreds of thousands of credit card transactions can be handled by configuring a farm of token servers on commodity hardware. Rendering the transaction database out of scope from a PCI DSS point of view will require the token servers to be isolated on the network. Protegrity Tokenization delivers multiple token servers in each location (Data Center 1 and 2) that do not require an expensive replication or synchronization process. This is critical for reducing complexity and cost and eliminating potential collisions.

b) **Health related data**: Health related information poses a different challenge. It usually has a high volume of processing and the data needs frequent de-tokenization. Since the security requirements don’t require the token servers to be located inside an isolated network segment, the non-token server deployment of tokenization can be used. This deployment of tokenization eliminates the latency and improves performance considerably.
c) **HR Personally Identifiable Information (PII):** The volume of transactions required for protecting personally identifiable information in the HR system is not high and so it can be serviced by a centrally administered token server that is dedicated to these transactions.

Protegrity Tokenization provides other benefits for these scenarios:

- **Smooth initial tokenization:** All tokenization solutions require tokenizing the existing data. Large datasets require using optimized and efficient data protection methods to perform initial tokenization without destabilizing business processes. Protegrity Tokenization offers a high degree of scaling with the choice of deploying a farm of token servers or by eliminating latency.

- **Reduced Liability:** While traditional tokenization stores the encrypted sensitive data and the token, Protegrity Tokenization does not store either. With no data to steal, the security is stronger and the solution and reduces your overall liability.

---

**How a large store chain simplified PCI compliance**

Simplifying PCI compliance was the key goal of a large regional retailer. The Security Project Manager said scope reduction was the path for achievement. The retailer used Protegrity Tokenization to segment, thus removing IT assets from scope.

At the time, the retailer had 18 concurrent projects for PCI compliance. It was imperative that the tokenization solution ease that workload with rapid, painless deployment. The retailer also required tokenization to provide high performance, as its service level agreement for transaction completion was less than one second.

**Deploying the Protegrity Tokenization Solution for PCI Segmentation**

The retailer deployed Protegrity Tokenization as a centralized solution, which uses a dedicated tokenization server on commodity hardware. The token server was deployed outside of the data warehouse to facilitate segmentation. With tokenization, data for a card transaction now follows this path: (a) card data is encrypted at the point of sale; (b) data is transmitted to a centralized host for decryption; (c) data is tokenized on the token server; (d) data enters the data warehouse.

*Continued on page 11*

“We planned on 30 days to tokenize our 50 million card numbers. With Protegrity Tokenization, the whole process took about 90 minutes.”

Security Project Manager
Large Regional Store Chain
The initial tokenization process was expected to take about 30 days for 50 million card numbers. The Protegrity Tokenization process actually required about 90 minutes. Deployment of the solution was non-intrusive as it did not require obtaining third-party modifications to code.

Benefits of the Protegrity Tokenization Deployment

Segmentation with Protegrity Tokenization yielded immediate benefits for the retailer:

- **Faster PCI audits** – The retailer’s PCI audit last year required about seven months. With segmentation, the retailer says the current audit will require half that time.

- **Lower maintenance cost** – “Maintenance is now less expensive because we don’t have to apply all 12 requirements of PCI DSS to every system,” says the Security Project Manager.

- **Better security** – “Everyone agrees the cardholder data is a lot more secure,” says the Security Project Manager. With tokenization, he says the retailer has been able to eliminate several business processes such as generating daily

- **Strong performance** – In addition to the rapid processing rate for initial tokenization, the solution meets the retailer’s sub-second transaction SLA.

Another benefit of Protegrity Tokenization has been no significant changes to the ways this retailer analyzes transactions. As part of the implementation, the retailer elected to leave the first six and last four digits of card numbers in the clear. “This satisfies 98% of our daily requirements in applications, reporting, and answering customer questions,” says the Security Project Manager. Some staffers were concerned about lack of access to the entire card number, so for the two percent of other situations, Protegrity provided the retailer with a tokenizer utility. With this tool, authorized staffers can take a token for access to the full card number, and then re-tokenize after completion of the task. “It’s no longer an issue,” he says.

“We’re happy with the results of Protegrity Tokenization and have a great working relationship with the company,” concludes the Security Project Manager. “This is one of those few cases where a solution ended up the way the sales guys said it would occur – and it’s saved us a lot of time and money.”

“The nice thing about tokenization is it turns the PCI conversation with auditors from gray into black and white. They don’t challenge us like they do with encryption. With Protegrity Tokenization, it’s very clear if you’re in scope of PCI – or not.”

Security Project Manager
Large Regional Store Chain
Conclusion

Protegrity Tokenization provides strong security for cardholder and other sensitive data while dramatically simplifying and lowering the cost of protection. The solution is highly scalable on standard commodity hardware, which lowers CAPEX and operational costs. Technical efficiency of Protegrity Tokenization keeps business processes running smoothly and meets stringent service level agreements. The solution is easier to manage, which simplifies and reduces costs of operations. And the solution is highly secure. With Protegrity Tokenization, your organization can simplify compliance with the PCI Data Security Standard. The solution can protect virtually any type of sensitive data, which enables meeting multiple compliance mandates and business security policies.

We invite you to learn more about Protegrity Tokenization and how this solution can cost effectively improve security and compliance for your organization’s sensitive data. To learn more, please contact Protegrity at 203.326.7200 or request a demonstration of solution capabilities by emailing tokenization@protegrity.com.
Protegrity Data Security Platform

The Protegrity Data Security Platform is a mature delivery vehicle for enterprise data protection solutions. The platform uses a hub and spoke architecture for deployment. Central control occurs with the Enterprise Security Administrator (ESA), which manages policy, keys and reporting. Policy enforcement occurs with Data Protector software deployed at protection points throughout the enterprise.

The foundation for protecting sensitive data is the Protegrity Data Security Policy. Key management is handled inside of the data security policy. This delivers ease of use while providing additional protection for data. Key management is automatic. Reporting delivers a rich set of functions supporting the management of the data protection process.

Data Protectors for Database, File, and Application (API) are deployed to any server in the enterprise requiring interaction with sensitive data. Protection uses the policy deployed from ESA to every protection point. The platform supports the most extensive set of operating systems and databases in the industry. Support for new protection points are added on a continuing basis.

Policy enforcement uses a variety of data protection methods such as strong encryption, Format Controlling Encryption (FCE), monitoring, and tokenization. The umbrella term for these data protection methods is Risk Adjusted Data Protection since use of each method corresponds with the risk profile associated with its data.

The Protegrity Data Security Platform also provides flexibility to successfully solve the challenges of protecting sensitive data throughout the enterprise. Continuous enhancements to the platform are guided by more than a decade of experience in delivering cost effective data protection solutions to large enterprises.

About Protegrity

Headquartered in Stamford, CT, Protegrity provides high performance, infinitely scalable, end-to-end data security solutions that protect sensitive information across the enterprise from the point of acquisition to deletion. The company's award winning software products span a variety of data protection methods, including end-to-end encryption, tokenization, masking and monitoring and are backed by several important data protection technology patents. Currently, more than 200 enterprise customers worldwide rely on Protegrity's comprehensive data security solutions to enable compliance for PCI DSS, HIPAA and other data security requirements while protecting their sensitive data, brand, and business reputation.