UNSTRUCTURED DATA AND PCI DSS COMPLIANCE

Unstructured data—data contained in files, rather than in a database—remains a largely unsolved problem for organizations that are subject to PCI DSS requirements.

PCI standards call for protection of cardholder data at rest and in transit, but in many organizations, data is only protected while it resides in (or is transmitted between) database environments. When cardholder data is extracted from a database and stored in files, it often drops out of view, while remaining a significant risk to the organization's compliance and data protection goals.

PROTECTING CARDHOLDER DATA

Since its introduction in 2004, the Payment Card Industry Data Security Standard (PCI DSS) has provided a common framework for banks, merchants, payment processors, and other organizations that handle credit card data.

PCI DSS requirements are enforced around the globe, and have been incorporated into data protection laws in some US states. Companies that fail to comply face fines, increased transaction costs, and other consequences.

While the majority of PCI DSS requirements address issues related to system design and user activity, requirements #3 and #4 deal directly with cardholder data itself.

The provisions of requirement #3 (***"Protect stored cardholder data"***) call for organizations to place limits on data retention, implement measures that render account numbers unreadable when stored or displayed on screens, and follow appropriate key management procedures when using encryption.

Requirement #4 (***"Encrypt transmission of cardholder data across open, public networks"***) prescribes encryption protocols to be used when transmitting cardholder data, and prohibits organizations from sending unprotected account data via email or other messaging technology.

These requirements are intended to prevent inappropriate use of credit card numbers by making it impossible for unauthorized users to see the information, even if they gain access to a device or transmission that contains credit card data.

PCI DSS explicitly states that its data protection requirements apply equally to credit card data in databases and in files. However, most organizations have focused their compliance efforts on database environments, while the most serious risk may lie elsewhere.

WHEN CARDHOLDER DATA IS EXTRACTED FROM A DATABASE AND STORED IN FILES, IT OFTEN DROPS OUT OF VIEW, WHILE REMAINING A SIGNIFICANT RISK TO THE ORGANIZATION'S COMPLIANCE AND DATA PROTECTION GOALS.

STRUCTURED DATA IS ONE THING...

PCI requirements do not typically pose a significant challenge where structured data is involved. Most organizations already have solutions in place to protect cardholder data within their database environments, and are able to demonstrate their compliance to auditors and other stakeholders.

PCI compliance is relatively easy to maintain within a database for several reasons:

- The database schema determines where credit card numbers and other forms of sensitive data will reside, and defines how data can be moved or manipulated within the environment. As long as the database is configured and administered correctly, it is difficult for data to exist in a state that violates organizational policy.
- Data encryption and masking fall within the capabilities of most commonly-used enterprise database administration tools.
- Processes for transmitting data between databases are typically automated, minimizing the possibility that human error could compromise data protection.
- Database reporting tools provide instant visibility into the amount and type of data being stored, as well as the types of protection applied.

In fact, for most organizations, compliance with PCI requirements #3 and #4 is simply another use case for the database administration tools already in place.

...BUT UNSTRUCTURED DATA IS ANOTHER

Protecting sensitive data outside the database environment, on the other hand, can be much more challenging.

Data in files—whether stored on employee devices or on file servers—is harder to control, because most organizations lack visibility into file contents and lack good options for remediating sensitive data within files. These challenges leave many companies with unprotected credit card data in their environments, and with unmanaged risks related to data governance and PCI compliance.

Unstructured data also dramatically increases the scope of compliance activities and audits. Many organizations attempt to simplify compliance by minimizing the number of systems that fall within the scope of PCI requirements. However, when users extract credit card numbers from databases and save the data on laptops, desktops, and file servers, those systems are pulled into scope for PCI compliance. This increases the burden on IT and security resources, and can also increase the risk of an audit failure.

30 COMPUTERS, 70 MILLION UNPROTECTED CREDIT CARD NUMBERS

A recent experience by a PKWARE customer demonstrates how large the risk of credit card numbers in unstructured data can be.

A global company had acquired a smaller company, and was beginning to integrate the acquired company's systems into its IT infrastructure. Prior to a PCI audit, security administrators discovered credit card numbers in unprotected files on several user devices.

To determine the severity of the problem, administrators conducted a trial implementation of PKWARE's automated data redaction solution on 30 of the acquired company's laptops and desktops.

On those 30 user computers, PKWARE discovered 4,100 unprotected files containing more than 70 million credit card numbers in all.

Based on this result, the company installed PKWARE's solution on all user laptops and desktops, and remediated millions of files in time to achieve 100% compliance on its PCI audit.

THE PROBLEM WITH FILES

Unstructured data presents a complex security and compliance challenge precisely because it lacks structure—it grows and spreads in unpredictable ways, and can be put into a wide variety of formats.

Uncontrolled Sprawl

Organizations can control the amount and location of structured data that exists in their environments. They define the processes—manual and automated—by which information is added to a database, and define specific tables and columns that will contain credit card numbers. This level of control, however, does not extend to unstructured data.

Employees at banks, payment processors, merchants, and other organizations are constantly extracting credit card information from databases and saving it in spreadsheets, text files, PDFs, and other formats. Once data is pulled from a database, it might be copied dozens or even hundreds of times, and saved in file shares, cloud drives, removable media, and other locations.

In large organizations, where thousands of users work with credit card data on a daily basis, the result is a massive volume of unstructured data—typically accounting for 80% of the organization's total data volume—extending across multiple platforms, over which the company has little visibility or control.

Limited Options

Endpoint data loss prevention (DLP) products—designed to monitor file activity on user devices and servers—might appear to be a viable solution for unstructured data, but they tend to produce unsatisfactory results, Many tools are able to scan files and detect the presence of credit card data, but they lack the capabilities necessary for effective remediation.

One typical endpoint DLP workflow involves detecting credit card data within a file, copying the file to a quarantined location, and then deleting the file from the user computer or file server that originally contained it. Users who look for the deleted file will find a "stub file" in its place, with information on how (or whether) the original file can be accessed. This process is frustrating and time-consuming for end users, and can disrupt critical business processes.

Some endpoint DLP products offer encryption as a remediation option, but this approach creates issues of its own. Encryption requires careful administration, and can break workflows and disrupt user activities when improperly implemented or managed.

Encryption also fails to address the issue of PCI scope. Because encryption is a reversible process, encrypted files (and the systems where they reside) cannot be taken out of scope for PCI compliance purposes.

These challenges discourage many organizations from implementing endpoint DLP technology, or from allowing their DLP systems to take action on the sensitive data they discover. This approach keeps existing workflows intact, but does not address the problem of unprotected credit card data in files.

THE ANSWER: DATA REDACTION

Data redaction—the process of permanently removing or obscuring sensitive data—keeps cardholder data safe without breaking internal workflows, and allows organizations to keep their unstructured data out of PCI scope.

With an automated data redaction solution, organizations can remove the middle six digits from credit card numbers as soon as they are extracted from a database and saved in a file (PCI DSS refers to this approach as "truncation"). Removal of these digits renders the card data unusable, and since the process cannot be reversed, files with redacted credit card data no longer fall under the scope of PCI requirements.

Redaction leaves other file contents unchanged, and eliminates the need to quarantine or encrypt files, keeping the remaining data accessible to authorized users. In use cases where employees may require access to unredacted credit card numbers, files can be copied to a quarantine location prior to redaction.

Organizations can use automated redaction technology to eliminate sensitive data from existing files, and to scan and remediate new data in real time as files are created and modified.

Redaction also eliminates obstacles to external data exchange. Since redacted files do not contain usable credit card data, they can be shared via email or other means without the need for potentially complex encryption, and without violating PCI requirement #4.

A SOLUTION FOR THE LONG TERM

As security breaches become more common and data volumes continue to expand, unstructured data is likely to attract more scrutiny from PCI auditors and other industry stakeholders.

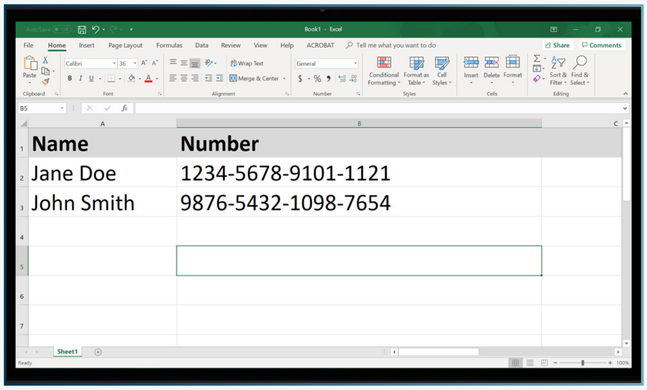
Unlike "compensating controls" that—at best—provide a temporary solution for security gaps, data redaction provides a permanent solution, allowing organizations to gain the same level of control over credit card numbers in unstructured data that they exert over structured data today.

AUTOMATED REDACTION FOR UNSTRUCTURED DATA

PKWARE's automated data redaction solution removes credit card numbers and other sensitive data from files, leaving other file contents unchanged.

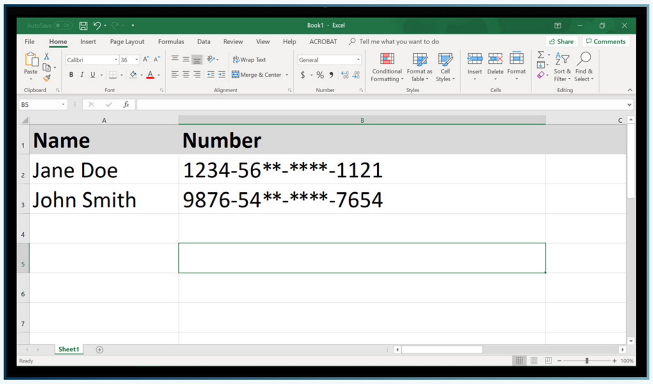
Redaction takes files out of PCI scope, and ensures that cardholder data will not be exposed in the event of a computer theft or other security event.

File with unredacted card data



Name	Number
Jane Doe	1234-5678-9101-1121
John Smith	9876-5432-1098-7654

File after automated redaction by PKWARE



Name	Number
Jane Doe	1234-56**-****-1121
John Smith	9876-54**-****-7654

PKWARE®

www.PKWARE.com

PKWARE solutions help organizations eliminate security gaps, manage sensitive data, and meet their data compliance goals.

CORPORATE HEADQUARTERS

201 E. Pittsburgh Ave.
Suite 400
Milwaukee, WI 53204

+ 1 866 583 1795

EMEA HEADQUARTERS

79 College Road
Suite 221
Harrow HA1 1BD

+ 44 (0) 203 367 2249