



Achieving PCI Compliance on the Mainframe

A White Paper by Xbridge Systems

Author: Michael Kibort

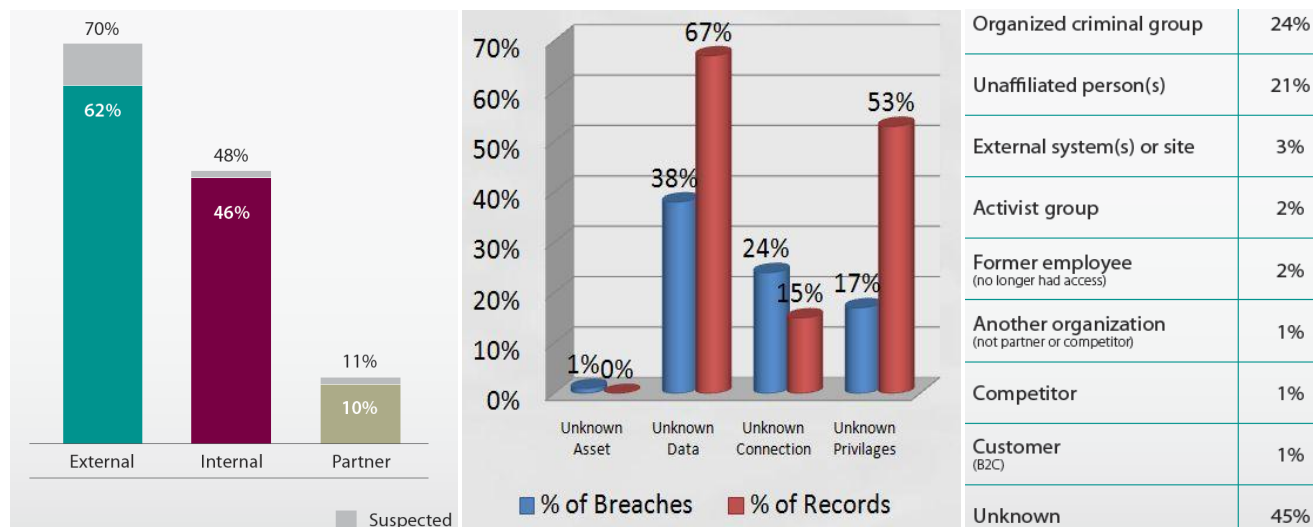
Publish Date: April 2011

| Contents | Page |
|--|-------------|
| Executive Overview | 2 |
| Introduction to PCI Data Security Standard | 3 |
| PCI DSS V2.0 Requires Mainframe Compliance | 3 |
| Mainframes Have Flown Below the PCI Compliance Radar | 4 |
| Access Controls Are Only Part of the Solution | 5 |
| The Mainframe Challenge – A Technical Discussion | 5 |
| <i>Where is the Data?</i> | 5 |
| <i>Mainframe Data is Not Like Windows™ and Open Systems Data</i> | 5 |
| <i>Mainframe Data Unrecognizable by Manual Inspection</i> | 6 |
| <i>Mainframe Data: Unstructured within Structured Data Sets</i> | 7 |
| <i>Mainframe Data: Unstructured within Various Unstructured File Types</i> | 7 |
| <i>Mainframe Scale and Complexity Requires an Automated Data Discovery Solution</i> | 7 |
| <i>An Automated Data Discovery Process for Meeting Compliance Must Not Impact Mainframe Operations</i> | 7 |
| Introducing DataSniff Mainframe Data Discovery Software | 8 |
| How DataSniff Works | 8 |
| <i>The Basics</i> | 8 |
| <i>Data Types that DataSniff Supports</i> | 8 |
| <i>Mainframe Subsystem Architecture Diagram</i> | 9 |
| <i>PC Server Software Architecture Diagram</i> | 9 |
| <i>Management of Network Traffic between the Mainframe and PC Server</i> | 9 |
| <i>One Day Installation and Setup</i> | 10 |
| <i>DLP Home Page</i> | 10 |
| <i>Scan Analysis Status and Results</i> | 10 |
| <i>Scan Scheduling</i> | 11 |
| <i>Final Scan Results</i> | 11 |
| <i>Patent Pending Scan Management Functions</i> | 12 |
| <i>Other Notable Capabilities and Functions</i> | 12 |
| Remediation | 13 |
| DataSniff Enables Mainframe Data Remediation Processes | 13 |
| Summary | 13 |
| About Xbridge Systems | 14 |
| Copyright and Disclaimer | 14 |
| References | 14 |

Executive Overview

In the world of IT security, identity theft remains the top cyber crime with over \$56 billion lost to identity theft in 2009.¹ With continued growth in credit card use and electronic data transactions, the number of data breaches and personal information lost to cyber crime continues to increase at an alarming rate.

The most recent study by Verizon Business shows a steady increase in percentage of data breaches perpetrated by company insiders, with substantial growth in this activity by organized crime.²



This study also highlights the disturbing reality that almost 70 percent of records lost to cyber crime were records that were unknown to the enterprise (either by content or storage location). However, the most concerning fact from this report was that almost half of insiders and external entities responsible for data breaches did not act alone. This clearly shows a growing trend that organized crime is now more likely to gain access to confidential data from *within* the enterprise.

The mainframe is no exception to this trend. In fact, the mainframe is the biggest offender when considering the unknown amount, content, and location of data within the enterprise. With over 70% of business critical, transactional, and customer data being stored in the mainframe³, it is an obvious high-value target for organized crime and insider attack. The sheer volume of data that is processed and stored, as well as unique storage methodologies used for over 40 years, present daunting challenges to any corporation looking to locate and protect *all* sensitive data stored in mainframe environments. Mainframes are typically massive and complex environments where manual discovery is impossible and automated solutions have not existed.

The Payment Card Industry Data Security Standard (PCI DSS) was created by major credit card companies to address protection of all transmitted and stored cardholder data. Mainframes have largely been exempt from processes required for compliance to PCI DSS due to the challenges mentioned earlier. In an effort to allow mainframe data to pass a compliance audit even if data is not truly protected, “compensating controls” have been implemented on the mainframe as a temporary stop-gap measure. The enterprise and auditors alike have widely ignored mainframe data during the PCI compliance process because there has been no viable method or tool available by which to verify presence and/or protection of *all* cardholder data within mainframe environments.

Here lies the challenge: Unless a *complete* Cardholder Data Environment (CDE) has been defined, an enterprise cannot properly implement a comprehensive data protection strategy, be it encryption, tokenization, or access controls. Inability to search for, and identify the location of cardholder data within the mainframe prevents successful identification and implementation of an enterprise-wide CDE, undermining the effectiveness of *any* protection strategy, and making it impossible to provide adequate protection of *all* stored cardholder data from compromise.

It could be said that the only way to search through large volumes of mainframe data and ultimately deploy mainframe remediation technologies is to automate the process of discovering and mapping credit card data locations. A comprehensive PCI compliance initiative is ONLY possible with deployment of an automated mainframe data discovery tool, due to the scope, complexity, high level of CPU utilization, and business critical nature of applications running within mainframes. Equally important to an automated solution, is one that DOES NOT impact mission-critical applications.

Xbridge Systems' DataSniff software is the world's *first* and *only* automated mainframe data discovery solution. When dealing with the challenge of creating a complete and detailed map of a CDE within the mainframe, DataSniff provides Qualified Security Assessors (QSA's), Auditors, and the enterprise with the capability to meet all requirements of this critical first step in PCI compliance, and assures that *all* cardholder data within the enterprise is identified for protection.

Introduction to PCI Data Security Standard

The Payment Card Industry Security Standards Council (PCI SSC) was launched on September 7, 2006 to manage on-going evolution of PCI DSS with focus on improving payment account security throughout all transaction processes. The PCI SSC is an independent body that was created by major payment card brands (Visa, MasterCard, American Express, Discover and JCB). It is important to note, payment brands and acquirers are responsible for enforcing compliance, not the PCI council. The PCI DSS is a set of requirements for security infrastructure, policies, and practices intended to improve security of cardholder and account data throughout all industries.

The PCI SSC guides efforts of Chief Information Security Officers, Compliance Officers, and others who need to protect cardholder information on behalf of payment card issuers, merchants, banks, processors, and service providers.

PCI applies to ALL organizations or merchants, regardless of size or number of transactions, that accepts, transmits or stores any cardholder data.

Companies that must comply with PCI DSS have worked aggressively to answer questionnaires, conduct pre-assessments, hire QSA's for compliance audits, evaluate results of the audit and incorporate all changes necessary to comply with PCI DSS. These tasks are not trivial for the largest retailers, insurance companies and financial institutions. According to a 2009 Visa report, these companies each spent an average of over \$2 million on compliance efforts in 2009, and that amount continues to rise.

PCI DSS is not a one-time event, and compliance is required in perpetuity for companies that continue to process credit card transactions and store credit card data. As a result, companies are required to have yearly audits by qualified PCI auditors called QSA's who follow stringent guidelines outlined in the most recent PCI DSS. If customers do not pass their audit, they may be subject to fines by credit card companies, and may lose the ability to take credit card orders from their customers.

PCI DSS V2.0 Requires Mainframe Compliance

PCI DSS Version 2.0 (effective January 1, 2011) requires ALL stored cardholder data to be identified and protected. Version 2.0 expands the scope of assessment and emphasizes the need for companies to know their data *before* they begin the PCI compliance process. The requirement explicitly states:

The First step in a PCI DSS assessment is to accurately determine the scope of the review. At least annually and prior to an annual assessment, the assessed entity should confirm the accuracy of their PCI DSS scope by identifying all locations and flows of cardholder data and ensuring they are included in the PCI DSS Scope. To confirm the accuracy and appropriateness of PCI DSS scope, perform the following:

- The Assessed entity identifies and documents the existence of all cardholder data in their environment, to verify that no cardholder data exists outside of the currently defined CDE.
- Once all locations of cardholder data are identified and documented, the entity uses the results to verify that PCI DSS scope is appropriate (for example, the results may be a diagram or an inventory of cardholder data locations).
- The entity considers any cardholder data found to be in scope of the PCI DSS assessment and part of the CDE unless such data is deleted or migrated/consolidated into a currently defined CDE.
- The entity retains documentation that shows how PCI DSS scope was confirmed and the results, for assessor review and/or for reference during the next annual PCI SCC scope confirmation activity.

**** PCI DSS Version 2.0: Scope of Assessment Definition Requirements, Page 10
(October 2010)

Most Fortune 1000 corporations process credit card numbers using mainframe systems. In fact, mainframe systems operate at the core of many businesses and handle the bulk of credit card data transactions and storage. The PCI Standard makes no distinction between data types or platforms and mainframes have not been excluded, so **mainframe data must be identified and protected** to comply with PCI DSS Version 2.0. Although PCI has been around since late 2006, mainframes have not participated in the process of PCI compliance for a number of significant reasons. Companies were forced to use specific language within the standard to keep mainframe data outside of the scope of PCI, knowing that data within the mainframe was not sufficiently protected.

This paper provides insight into mainframe compliance:

- I. Mainframes have flown below the PCI compliance radar
- II. Inherent aspects of the mainframe that make compliance a challenge
- III. Technology needed to bring the data into scope
- IV. Remediation of cardholder data to achieve compliance

Mainframes Have Flown Below the PCI Compliance Radar

Compensating Controls

Without availability of an automated data discovery tool, companies with mainframes have been forced to use stop-gaps to position their mainframe systems outside the scope of PCI DSS. The common stop-gap has typically been “compensating controls.” The problem with this approach is that it puts the cart before the horse. How can you determine which data sets need compensating controls until you know what data resides within them?

In the PCI compliance process, mainframes typically receive “special” treatment. Most often, compensating controls are put in place as a method to allow mainframe data to pass an audit even if that data is not truly protected. Compensating controls have never been considered a permanent solution for protecting mainframe data. Specifically, PCI DSS has always stated the following:

Only those companies that have performed a risk analysis and have legitimate technical or documented business constraints can consider the use of compensating controls to achieve PCI compliance. Compensating controls may be considered when an entity cannot meet a requirement explicitly as stated, due to legitimate technical or documented business constraints, but has sufficiently mitigated the risk associated with the requirement through implementation of other controls.

Compensating controls must satisfy the following criteria:

1. Meet the intent and rigor of the original stated PCI DSS requirement;
2. Provide a similar level of defense as the original PCI DSS requirement;
3. Be "above and beyond" other PCI DSS requirements (not simply in compliance with other PCI DSS requirements); and
4. Be commensurate with the additional risk imposed by not adhering to the PCI DSS requirement.
5. The assessor is required to thoroughly evaluate compensating controls during each annual PCI assessment.

**** PCI DSS and PA-DSS Glossary of Terms, Abbreviations, and Acronyms Version 2.0 (October 2010)

As technology evolves to discover PCI data on the mainframe, that data must be included within the enterprise CDE to meet the intent and rigor of the original PCI DSS requirement.

Access Controls Are Only Part of the Solution

Access control technologies have been readily available to mainframe shops with mature technologies like RACF, Top Secret, and ACF2. Access controls are only part of the overall PCI DSS solution. For a full definition of what access control means to a complete compliance initiative, please review requirement 7 of PCI DSS V2.0. The section is titled: *“Implement Strong Access Control Measures.”*

Access controls have been used as a method to “sufficiently mitigate” risk associated with unprotected data. The intent of PCI DSS is to protect the DATA, with access controls being deployed ONLY AFTER a map of the data is documented in the CDE.

Again, access control technologies within RACF, Top Secret, and ACF2 are only part of an overall PCI DSS solution. To meet the intent of PCI DSS, the overall solution must be a combination of data remediation (encryption, tokenization, or deletion) paired with appropriate access controls. Additional PCI DSS requirements are not addressed in this document, but can be viewed at the PCI Security Council website www.pcisecuritystandards.org.

With respect to this White Paper, the following are considered paramount in reaching adequate protection for cardholder data:

- Scope of Assessment for Compliance with PCI DSS Requirements to understand and manage the people, processes and technology that store, process or transmit cardholder data or sensitive authentication data
- Discover, define and create an inventory of all locations of cardholder data – create a CDE.
- Encrypt, tokenize, or delete all cardholder data
- Create and manage access controls relating to all cardholder data

Compensating controls, as well as access control technologies, may have been implemented and ultimately accepted as a stop-gap for PCI compliance because mainframe customers lacked tools required to support mainframe audit processes. A fundamental problem with achieving compliance on the mainframe has been the ability (or lack thereof) to create a comprehensive CDE that includes mainframe data.

The Mainframe Challenge – A Technical Discussion

Where is the Data?

The sheer volume of data that is processed and stored by mainframes, as well as unique storage methodologies used for over 40 years, present a daunting challenge to any corporation that must meet compliance requirements like PCI. The issue: An enterprise cannot effectively implement an access control initiative or an encryption strategy unless a complete CDE has been defined.

So, why is it so hard to discover credit card data in the mainframe?

Mainframe Data is Not Like Windows™ and Open Systems Data

There has been no tool available for searching arbitrary text or business-critical information within all mainframe files, unlike Windows and open systems. Mainframes have unique storage methodologies that present challenges to existing search and data discovery technologies. Data “owned” by subsystems (for example: IBM IMS and DB2 Data Base Management Systems) are not accessible by other applications. Many mainframe databases (e.g. IMS, IDM) and proprietary applications have no established standards for identifying structure of the data they store, unlike what is common for open systems relational databases. Subsequently, searching these data stores require sophisticated mechanisms that infer structure from common practices, most of which are never employed on non-mainframe systems, such as encoding data in a Packed Decimal format.

Mainframes do not utilize directory structures typically seen in distributed and open-computing environments. Instead, there are “pointers” to data locations. These pointers are stored within the mainframe by various methods, and point to data locations that are assigned to a wide range of programs operating simultaneously

Mainframe Data: Unstructured within Structured Data sets

```

Menu  Utilities  Compilers  Help
-----
BROWSE  XBR.PCIDEMO.ORDERSN.TXT          Line 00000000 Col 017 096
Command ==>                               Scroll ==> PAGE
***** Top of Data *****
eborah 101 Main St      San Jose      CA
        7779 Flatbush   New York     NY use card on file.
ark     455 Santa Cruz     Los Angeles  CA Use personal ccn:5466 130048039417
ly      5675 Akers         Kingston     NY cash only
        2387 Sunset      Los Angeles  CA
ttany   6345 San Carlos    San Jose     CA Problem customer handle with care
        3277 Lincoln     Los Angeles  CA use fathers CCN 5466-1300-4803-9417
hn      4355 State         Madison      WI
es      4567 Lombardi    Green Bay    WI
in      8754 Ocean        Carmel       CA
***** Bottom of Data *****

F1=Help  F2=Split  F3=Exit  F5=Rfind  F7=Up    F8=Down  F9=Swap
F10=Left F11=Right F12=Cancel

```

Figure 4: This view shows data being stored in an unstructured format within a structured data set environment. This is most common within unstructured “note” fields within large-scale customer databases, where help-desk and other support employees can and often do write customer notes and comments in the notes field of a customer file, mistakenly including sensitive information. Even with a fully automated software application managing all sensitive customer data under concrete policy management, the most stringent security policies cannot prevent accidental input of sensitive customer data within a database notes field.

Mainframe Data: Unstructured within Various Unstructured File Types

Data can be stored or resident within unstructured data types, such as flat-files. These files must be identified, accessed, opened, and analyzed to discover any potentially sensitive data that may be stored within those files.

```

Menu  Utilities  Compilers  Help
-----
BROWSE  XBR.DLPDEMO.BIGNOTES.TXT          Line 00000000 Col 001 080
Command ==>                               Scroll ==> PAGE
***** Top of Data *****
temp file: Created when data entry system crashed.
Still taking calls. Will enter data in when system comes back up.
DELETE WHEN FINISHED.

Wednesday May 19, 1979
9:18 am
Blue Moon Parts called. (567) 555-2304.
emergency failure needs to reorder part immediately
customer Id: 994574
user personal credit card number
Part number: 2730 CCH: 5466-1300-4803-9417
12:17 emergency order line.
Acme Products Wile E. Coyote, Genius 914-555-7734
Part number: 2830 CCH: 5466 13004803 -9417
13:05 System back on line orders entered into system.
temp file: Created when data entry system crashed.
F1=Help  F2=Split  F3=Exit  F5=Rfind  F7=Up    F8=Down  F9=Swap
F10=Left F11=Right F12=Cancel

```

Figure 5: This view shows sensitive data on the mainframe in unstructured form. During a “system down” scenario, a customer support application is down. Access to the comment field normally associated with the support application is not available. To document the support call process, the call center employee may create an ad hoc flat file to record information that customers give them over the phone. When the system regains operational status, the created data is not deleted and stays stored within the system without any structured security controls to protect it.

Mainframe Scale and Complexity Requires an Automated Data Discovery Solution

Due to the aforementioned challenges outlined in this document, it could be argued that the only feasible way to perform access, discovery, analysis, and then preparation of all mainframe data for exemption or remediation is to automate the entire process. This can ONLY be accomplished by deploying an automated mainframe data discovery tool.

An Automated Data Discovery Process for Meeting Compliance Must Not Impact Mainframe Operations

Mainframes are responsible for the bulk of mission-critical, operational, and financial applications. As a result, there are not a lot of spare CPU cycles available to support security and compliance initiatives. One of the challenges for mainframe data discovery is how to find cardholder data without adversely impacting business.

Mainframes are the most powerful computing platform available. Unlike distributed/open computing environments, mainframe processors typically run at a very high level of CPU utilization. Due to cost and availability of CPU overhead, the enterprise cannot support additional CPU utilization for compliance processes. Therefore, an automated data discovery solution requires several key functional parameters in order to provide

data access, discovery, and mapping operations without adversely impacting mainframe computing environments or the networks to which they are attached:

- Flexible scheduling must be available for the data discovery process to operate only during periods where both mainframe and network have available bandwidth. An automatic “pause and resume” function must be available in order to stop and start data scan/analysis mid-stream, giving priority to all other mainframe and network load requirements defined by business operations of the enterprise.
- Limits must be set for scanning operations to assure all scanning processes are defined and controlled
- Throttling by media type – This is the management of maximum load seen on each mainframe subsystem via multiple execution threads. An example of media type is DB2 DBMS. Many DB2 DBMS have a maximum number of threads for all applications. Production applications and ad-hoc queries require these threads. As a result, execution threads are a limited and critical resource requiring assured availability. Thus, applications requiring threads must be prevented from grabbing all threads available. Without this program limitation, business critical applications could be adversely effected.
- Error handling capabilities – A data scanning processes could encounter errors by attempting to read data currently in-use, or other factors that could affect accessing or reading data sets and/or records. Again, an automatic “pause and resume” function must be available in order to stop and start data scan/analysis processes mid-stream. An ability to re-scan specific locations with errors is necessary to assure a complete data discovery process. This must be supported by automatic, high-volume techniques.

Introducing DataSniff Mainframe Data Discovery Software

DataSniff software from Xbridge Systems is the world’s *first* and *only* mainframe data discovery solution. DataSniff allows mainframe customers to identify credit card data within their mainframe data sets. DataSniff is now available to the largest mainframe shops just as the PCI Security Council has emphasized the need for “knowing where your data resides before the process of PCI assessment begins.”

Verizon Business, author of the most comprehensive “2010 Data Breach Investigations Report” has a team of experts that support compliance initiatives like PCI, Hitech Act, HIPAA, etc... Their expertise in data protection is well respected in the PCI market and they confirm that PCI protection is about the data, and the need to locate and protect it at the source.

“We tell all of our customers that they need to know where their data resides and to protect sensitive data at the source. We also tell them that access controls are only part of an overall data protection strategy.”

How DataSniff Works

The Basics

Mainframe data sets to be analyzed are identified and mapped into meta-data templates through automated and semi-automated processes that utilize system catalogs, COBOL copy books, control blocks, and source blocks that are already used within the mainframe as definitions for meta-data locations owned by various applications. This is normally done during installation. Once defined, users may then create scans of data set groups to search for sensitive information that may reside within any part of the group being scanned. Scan results are then displayed in a simple, easy to view, browser-based user interface dashboard. Results are shown as multi-data set scans, and can be expanded to show individual data set scan results, as well as individual record results. All results can be assigned a disposition and tagged in preparation for exemption, remediation, or deletion.

Data Types that DataSniff Supports

IBM IMS (HDAM, HIDAM, HISAM, SHISAM, SHSAM, and INDEX), IBM DB2, VSAM, QSAM, IDMS, BDAM, PDS, ORACLE, SQL, ODBC, Migrated Data (Virtual and Real Tape).

Mainframe Subsystem Architecture Diagram

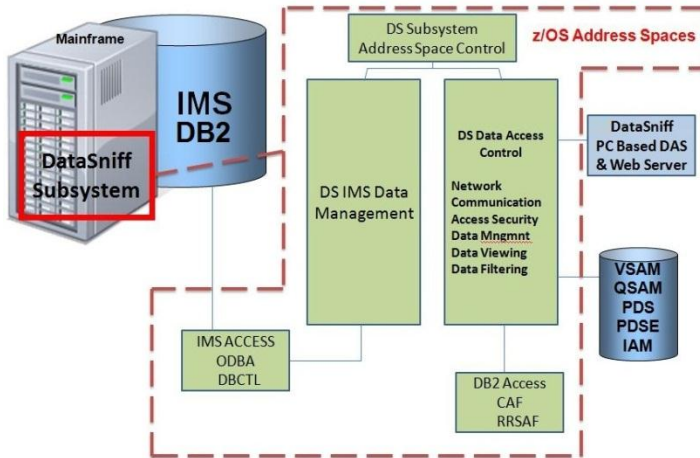


Figure 6: DataSniff was designed to maximize capability of all components in the system, thus maximizing performance of data access and discovery processes without any noticeable performance degradation at the mainframe CPU. DataSniff accomplishes this by off-loading data access, analytics, and mapping processes to a PC server. This architecture was based upon feedback from customer installations of a previous generation of Xbridge Systems' mainframe data access, query, and reporting software known as "Host Data Connect". Proven technology developed for "Host Data Connect" is the core technology upon which the mainframe portion of DataSniff software has been developed to operate as a fully authorized z/OS subsystem.

PC Server Software Architecture Diagram

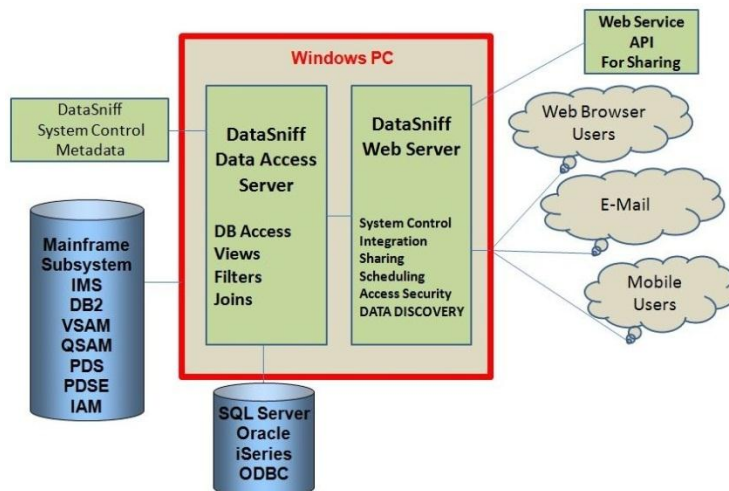


Figure 7: The PC serves to provide all data analysis, scan management, scan scheduling, and reporting functions, and is designed to expand in capability as customer requirements' expand. Support of functional expansion without creating additional burden on the mainframe is again possible through architectural separation of DataSniff analytics from the DataSniff z/OS subsystem software residing on the mainframe.

Management of Network Traffic between the Mainframe and PC Server

Functionality is designed within DataSniff to minimize burden on network traffic. DataSniff accomplishes this by providing flexible and precise control over many operational details of data scanning and analysis processes.

Key aspects of how DataSniff controls scanning processes are:

- Increasing scan performance by providing the capability to analyze multiple data sets/tables simultaneously
- Scheduling of scanning activity only during hours where both mainframe and network have available bandwidth
- Manipulation of scan definitions
- Defining scan parameters and limitations
- Utilizing user-supplied scan filters
- Implementation of a "pause and resume" function for all scan operations, halting one or more analysis request associated with scan definitions by storing the point of interruption for each analysis request, and resuming the servicing of analysis requests at the stored position for each request

One Day Installation and Setup

DataSniff software is loaded onto the mainframe and operates as a fully-authorized subsystem within z/OS. The DataSniff PC server is connected to the mainframe, communicating with the subsystem via TCP/IP. Mainframes typically have several LPARs. The locations of each will be user-defined as Host-Sources (IP address, subsystem port, name, description, and set of credentials). Locations of DB2, IMS, and other databases are also user-defined as Host Sources (IP address, subsystem port, DB2 subsystem name, and DB2 Schema name). Locations of data to be analyzed are identified through automated and semi-automated processes. By utilizing system catalogs, COBOL copy books, control blocks, and source blocks already being used by the mainframe to define meta-data locations owned by various applications, DataSniff software can be fully functional within hours.

Once host sources have been identified, the user can then partition analysis of all data on each Host Source into one or more scan definitions. By creating multiple scan definitions for a single Host Source, organizations can prioritize which data sets to analyze. It also provides organization of scan/analysis results.

DLP Home Page

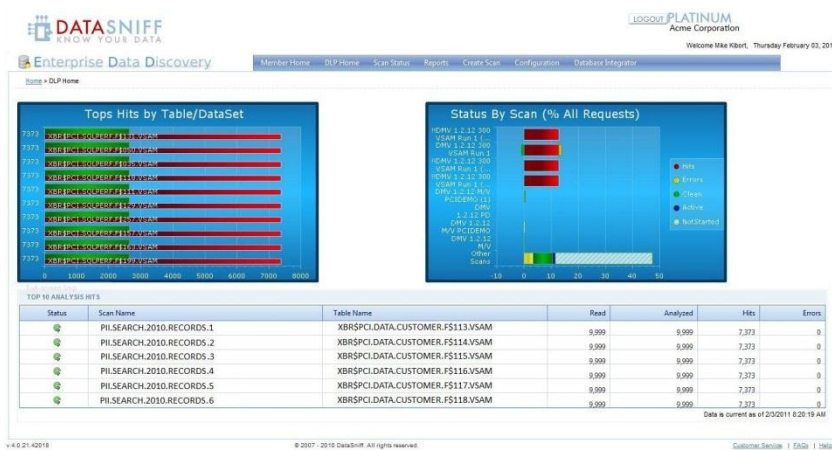


Figure 8: This is the DataSniff Data Loss Prevention (DLP) home screen. The left chart provides an overview of data sets or tables that are highest risk, highlighting those that contain the largest number of “hits”. By using this chart, users can compare relative risk with respect to the overall number of records, and then identify and prioritize data sets for remediation. Once necessity of remediation is determined, the user can disposition the results. Upon disposition, the data set/table will be removed from

“top hits” and replaced with a data set with the next highest number of “hits”. This enables focus on remediation of datasets posing the highest risk.

The chart on the right provides a view into scan/analysis progress, allowing the user to see scans with the most hits (to enable more efficient remediation approaches), scans with the most errors (facilitating removal of transient errors more effectively), and scans with the largest number of data sets/tables awaiting analysis.

Scan Analysis Status and Results

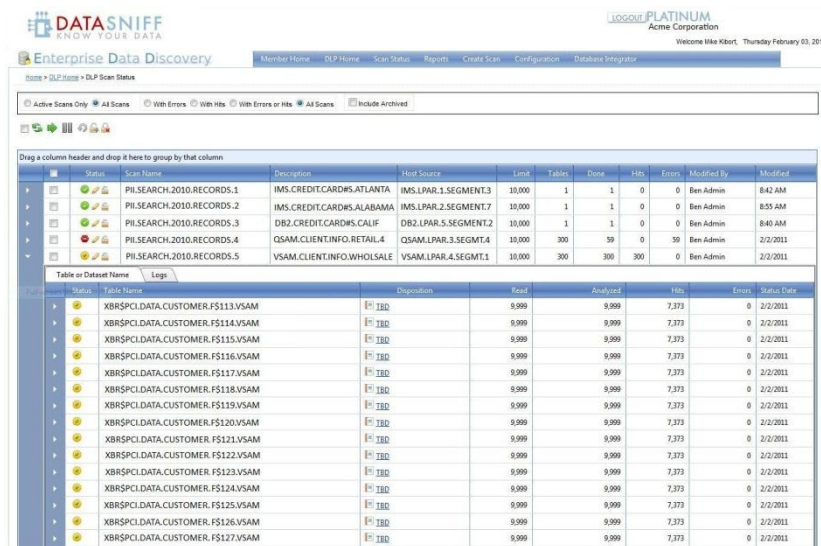


Figure 9: In this view, several scan definitions have been defined (name, description, host source, and limits) and summary results are visible at the top level of drill down, including number of data sets/tables associated with each scan definition, number of data sets analyzed, and number that have reached a hit threshold. It also includes data sets/tables that encountered errors during analysis processes.

Users can drill down to see what specific data set or tables are associated with each scan. This level of hierarchy shows data set/table name, disposition, progress and

scan results relating to that data set or table, including number of records read and analyzed, number of hits found during analysis, and number of errors encountered. Users can also access any log entry associated with scans, including documentation of any changes to scan definitions, errors encountered by analysis of any associated data set/table, and a result summary of any data set/table.

Users can further drill down into specific analysis results for each data set or table. This level of hierarchy includes logs that define progress of the analysis performed and a results summary, a listing of the actual hit including record location, column, position within the column, and a masked value of what was found, and a list of errors (if any) encountered.

Scan Scheduler

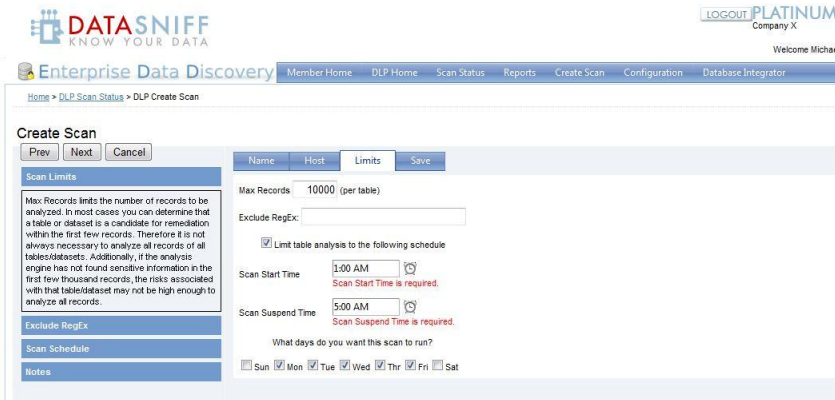


Figure 10: The scan limit function allows the user to limit the number of records to be analyzed. In most cases, a determination can be made that a table or dataset requires remediation within the first few records. Therefore, it is not always necessary to analyze all records of all tables/datasets. If analysis has not discovered sensitive information in the first few thousand records, risks associated with that table/dataset may not warrant analysis of all records.

Users can also exclude any tables/data sets from analysis by entering a regular expression in the “Exclude RegEx” field. When tables/datasets are analyzed, those names that match the RegEx term are automatically given a disposition (as specified by configuration). This effectively causes those tables/datasets to never be analyzed.

Scan scheduling is typically done to minimize the potential impact of performing analysis on a production system or a system that has restricted availability. If selected, the scan analysis will start only after the specified time and end before the specified time, and only on the days of the week that are checked.

Final Scan Results

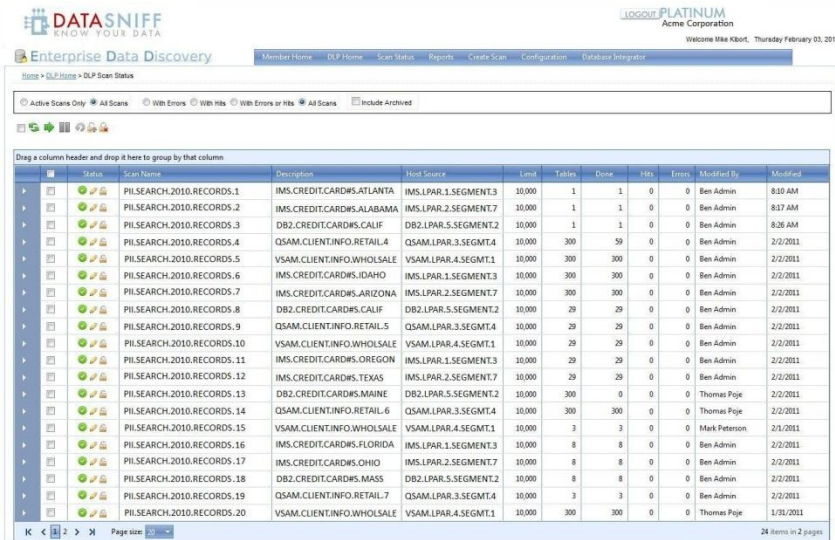


Figure 11: Upon completion of disposition and remediation processes, a final scan of data sets provides confirmation that all sensitive data within scope of PCI DSS has been remediated and/or risk-assessed. Note that all data set scans are displayed as confirmed with green scan status icons.

Patent Pending Scan Management Functions

By employing the following patent pending scan management capabilities within its architecture, DataSniff has virtually no impact on mainframe performance and minimizes total overhead on network traffic. The following points relate to the flexibility of DataSniff to throttle and control the process of running multiple scans. These are all important considerations in building predicate-based search technology and actively managing large amounts of data in a mainframe environment:

- **Implementation of a pause and resume function for all scan operations** – termed “redrive” operation
 - The term “redrive” refers to the capability to pause one or more analysis requests associated with scan definitions at the point of interruption, then pick up the scan process at the exact point from where it was paused. DataSniff does this by storing the point of interruption for each analysis request, and resuming the servicing of analysis requests at the stored position for each request. This is of paramount importance for the effective implementation of a scan scheduling system operating within the mainframe environment, performing large-scale data discovery activities without adversely affecting CPU performance and/or network traffic.
- **Scan Schedules** – Specific blocks of time may be allocated for scanning purposes. No scans will be performed outside of these allocated blocks of time. This assures that scanning activity occurs only during periods when both mainframe and network have available bandwidth.
- **Default Analysis Limit** – Maximum number of records DataSniff will read and analyze per data set. Normally, hits will be seen well before this, if at all. Reducing this number reduces packet traffic.
- **Hit Limit Per Request** – Minimum number of hits per data set after which DataSniff will stop scanning. If this minimum is reached, the data set usually is in serious need of mitigation.
- **Active Requests Per Scan** – Maximum number of data sets allowed for concurrent scanning. Decreasing this can decrease rate of packet traffic and mainframe load.
- **Total Active Requests Per Host** – Maximum number of data sets allowed per mainframe for concurrent scanning. Decreasing this can decrease rate of packet traffic and mainframe load.
- **Inter-Read Delay** – Number of seconds delay between each data set record read. Increasing this can decrease rate of packet traffic and mainframe peak load.
- **Hits Per Request Threshold** – Minimum number of hits per data set, below which will not be reported.
- **Mainframe throttle on concurrent scans per media type** – DataSniff mainframe subsystem will not schedule more concurrent scans than the configurable throttle limit.
 - Examples of Media Types: Primary Disk, Tape (Virtual & Real), Migrated Data, IBM IMS, IBM DB2, and VSAM.
- **Use of switches rather than router** – Clients may utilize switches between the DataSniff PC server and mainframes. This removes the traffic from the overall shared enterprise network.

Other Notable Capabilities and Functions

- **Manipulation of scan definitions** to assure focused scanning approach
- **Utilizing user-supplied scan filters** assure data sets selected for scanning are scanned with a focused approach

Remediation

Once an enterprise has successfully identified the location of all stored cardholder data, creating a comprehensive CDE.... What is the next step?

Section 3 of PCI DSS V2.0 defines the requirements to protect stored cardholder data once it has been located:

Protection methods such as encryption, truncation, masking, and hashing are critical components of cardholder data protection. If an intruder circumvents other security controls and gains access to encrypted data, without the proper cryptographic keys, the data is unreadable and unusable to that person. Other effective methods of protecting stored data should be considered as potential risk mitigation opportunities. For example, methods for minimizing risk include not storing cardholder data unless absolutely necessary, truncating cardholder data if full PAN is not needed, and not sending unprotected PANs using end-user messaging technologies, such as e-mail and instant messaging.

*****PCI DSS Requirements and Security Assessment
Procedures, Version 2.0*

DataSniff Enables Mainframe Data Remediation Processes

Upon completing the process of identifying the location of cardholder data in the mainframe, the DataSniff Mainframe Data Discovery tool generates a detailed map of all identified potential cardholder data.

In many cases, there will be no need to analyze every record within every data set before moving forward with remediation. The reason: During analysis of the first portion of the entire data set/table (defined by “Analysis Limit” and/or “Hit Limit” functions mentioned earlier), the determination was made that a specific column within a data set/table included enough sensitive records to flag all data within that column for remediation.

The map of data sets, tables, and/or columns targeted for remediation can be presented within the DataSniff user interface, generated as a SQL Server database for import to various encryption or tokenization software systems, or accessed via API. These methods may also be used to provide cardholder data location information to other Data Loss Prevention (DLP) monitoring software systems. As remediation progresses, periodic rescans can be done with DataSniff and scan dispositions can be updated. DataSniff DLP reports can be run to track/audit remediation progress.

Summary

Mainframe computers have historically been the backbone of corporate computing and will be with us for many years to come. The IT world is coming full circle as mainframe systems, once again, are increasing in utilization. The inherent strengths of their processing capabilities, proven security, integral virtualization, energy efficiency, and a reduction in cost per MIPs, make them a viable computing platform for today and tomorrow.

Many corporations will continue to keep the majority of their critical data in the mainframe, presuming that it is protected from external threats. However, internal threats continue to increase in frequency and sophistication, while mainframe data is made increasingly available to distributed systems and cloud computing environments. Without the ability to identify and locate *all* sensitive data stored in the mainframe, protection of that data and compliance with PCI DSS is all but impossible.

DataSniff software from Xbridge Systems is the world's *first* and *only* automated mainframe data discovery solution. When dealing with the challenge of creating a comprehensive map of a mainframe CDE, DataSniff provides QSA's, Auditors, and the enterprise with the capability to meet all the requirements of this critical first step in PCI compliance, and assures that *all* cardholder data within the enterprise is targeted for protection.

About Xbridge Systems

Co-founded in 1994 by Dr. Gene Amdahl and Raymond A. Williams, Jr., Xbridge provides the IT market with world-class Data Loss Protection, Data Discovery, Data Access, Data Query, and Data Delivery solutions, helping companies locate, monitor, and protect their sensitive information. In 2010, the company focused on data discovery and DLP, building upon its proven mainframe data access technology. The result of this effort is the recent release of DataSniff Mainframe Data Discovery Software. From open systems to the mainframe, Xbridge enables the foundation for meeting comprehensive data security and compliance initiatives.

Copyright and Disclaimer

This document is copyright © 2011 Xbridge Systems, Inc. No part of this publication may be reproduced by any method whatsoever without the prior consent of Xbridge Systems.

Due to the nature of this material, numerous hardware and software products have been mentioned by name. In the majority, if not all, of the cases, these product names are claimed as trademarks by the companies that manufacture the products. It is not Xbridge Systems' intent to claim these names or trademarks as our own. Likewise, company logos, graphics or screen shots have been reproduced with the consent of the owner and are subject to that owner's copyright.

Whilst every care has been taken in the preparation of this document to ensure that the information is correct, the publishers cannot accept responsibility for any errors or omissions.

References

1. Javelin Strategy and Research- Identity Theft Report, 2009
2. Verizon Business – Data Breach Investigations Report, 2010
3. IBM / SHARE Mainframe Executive Study, 2007