How to Identify Datasets Containing PCI, PII or Other Sensitive Information

David Wade
CIO/EVP
Primerica, Inc.
david.wade@primerica.com

February 5, 2013
Session 12943
Primerica: Who We Are

- Life Insurance and Investments
- Middle-income families
- Main Street, not Wall Street
- Entrepreneurial business model
- 4.3 million lives insured
- 2 million investment clients
- Publicly traded “PRI”
## Infrastructure Architecture

<table>
<thead>
<tr>
<th>Applications</th>
<th>Field Registration &amp; Licensing</th>
<th>Field Compensation</th>
<th>Field Technology</th>
<th>Life Administration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image Technology</td>
<td>PeopleSoft / Cognos TM1</td>
<td>HR Administrative Systems</td>
<td>IBM TSO/E v1.12</td>
<td>National Benefit Life</td>
</tr>
<tr>
<td>Securities Administration</td>
<td></td>
<td></td>
<td>MS .NET Framework v4.0</td>
<td></td>
</tr>
<tr>
<td>National Benefit Life</td>
<td></td>
<td></td>
<td>Application Infrastructure Support</td>
<td></td>
</tr>
</tbody>
</table>

### Presentation Layer
- **Microsoft Internet Explorer, Firefox, Safari, Chrome**
- IBM Websphere Application Server v6.1 & v7.0
- IBM Transaction Server v4.2 (CICS) Transaction Gateway v7.2
- IBM TSO/E v1.12
- MS .NET Framework v4.0
- Tomcat v6

### Database Management
- IBM DB2 v10 for z/OS
- Oracle 11g
- IBM DB2 for LUW v9.7
- Microsoft SQL Server for Integrated 3rd Party Application
- VSAM – QSAM
- NAS
- MySQL v5.5

### Information Security
- IBM Security Server v1.12 (RACF)
- Native UNIX & CA eTrust v8
- Novell v6.5 eDirectory v8.7
- MS Active Directory 2003

### Telecom
- Diverse, AT&T Multi-link 155/100 Mbps Internet Access
- AT&T OC48 SONET Metropolitan Area Network
- Single & Multi-Mode Fiber & CAT 5e Wiring
- Siemens 9751 PBXs
- CISCO 65xx & 72xx Equipment

### Operating Systems Platforms
- PR/SM IBM Mainframe Z/OS v1.12
- Power VM IBM Unix-AIX v5.3/6.1
- VMWare Vsphere 4 Microsoft Windows 2003/2008
- SUSE V10 SP3 Edirectory v8.8 and Novell v6.5 Edirectory v8.7
- Red Hat v6 Update 2

### Disaster Recovery
- Mature Recovery Plans
- End-to-End Application Testing
- Network Connectivity
- IBM Out of Region Recovery Center
- Business Recovery Center - Hembree
DataSniff

What’s in your mainframe?

Finding PCI, PII and other sensitive information in your legacy mainframe
Mainframe Legacy

- 30+ years of legacy data
  - Production data classified manually
  - End user and Test data not classified

- Structured Data
  - VSAM
  - DB2 (includes user-owned tables)

- Unstructured Data (Disk and Tape)
  - QSAM / Sequential, GDG
  - PDS, PDSE
  - Packed Decimal
  - DFHSM ML1 / ML2
Data Loss Prevention Initiative

Effective DLP Programs Utilize a Multitude of People, Process and Technology

Data Governance

- Policies and Standards
- Identification
- Risk Assessment
- Classification
- Architecture
- Quality

Source: Ernst & Young, LLP, “Data Loss Prevention, Keeping your sensitive data out of the public domain”
# Data Loss Prevention Initiative

## Data Control

### Key Focus Areas for DLP Controls

<table>
<thead>
<tr>
<th>Structured data</th>
<th>Unstructured data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data in Motion</strong></td>
<td><strong>Data in Use</strong></td>
</tr>
<tr>
<td>Perimeter security</td>
<td>Privileged user monitoring</td>
</tr>
<tr>
<td>Network monitoring</td>
<td>Access/Usage monitoring</td>
</tr>
<tr>
<td>Internet access control</td>
<td>Data anonymisation</td>
</tr>
<tr>
<td>Data collection and exchange</td>
<td>Use of test data</td>
</tr>
<tr>
<td>Messaging (Email, IM)</td>
<td>Data redaction</td>
</tr>
<tr>
<td>Remote access</td>
<td>Export/Save control</td>
</tr>
</tbody>
</table>

Source: Ernst & Young, LLP, “Data Loss Prevention, Keeping your sensitive data out of the public domain”
### Data Loss Prevention Initiative

**Supporting Information Security Processes**

<table>
<thead>
<tr>
<th>Identity/Access Management</th>
<th>Security information/event management</th>
<th>Configuration management</th>
<th>Vulnerability management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digital rights management</td>
<td>Incident response</td>
<td>Physical security</td>
<td>Training and awareness</td>
</tr>
<tr>
<td>Asset management</td>
<td>Data privacy/document protection</td>
<td>Employee screening and vetting</td>
<td>Third-party management and assurance</td>
</tr>
<tr>
<td>Business continuity</td>
<td>Disaster recovery</td>
<td>Regulatory compliance management</td>
<td>Change management/SDLC</td>
</tr>
</tbody>
</table>

Source: Ernst & Young, LLP, "Data Loss Prevention, Keeping your sensitive data out of the public domain"
If You Can’t Measure It…

...You Can’t Manage It!
The Task at Hand

Classify all mainframe data files

- Who owns them?
- What do they contain?
- Where is the confidential, PII, PCI or other sensitive information?
DataSniff by XBridge Systems

- Data Discovery Tool
  - Runs on the mainframe
  - Automated, native scanning of mainframe data
  - Schedule by Media Type

- Analytical Engine
  - Appliance or Windows-based

- Browser-based User Interface
  - Schedule scans
  - Manage reporting
DataSniff Evaluation

- Identify and secure
  - Payment Card Numbers (PCN)
  - Personally Identifiable Information (PII) especially Social Security/Tax ID numbers, home address

- Support all Mainframe data structures
  - DB2 databases
  - VSAM KSDS & ESDS
  - QSAM, GDG files on both disk and tape
  - PDS & PDSE
  - DFHSM ML1 & ML2 on tape
  - Packed decimal data
Features Evaluated

- Locate SSN/PCN format data in DB2 tables, QSAM/VSAM, & PDS(E)s
- Selection of data using regular expressions
- User control of mainframe resources
- HSM re-migration
- User-written regular expressions
- Results reporting
- Impact on mainframe performance
Management Console
Management Console
Successes

- PCN/SSN data located with minimum of “false positives”
- User-defined selection filters
- Scanned all required data structures: DB2, VSAM, Sequential / QSAM, PDS/PDSE, HSM migrated data sets
- Scanned packed decimal structures
- Controlled use of scarce mainframe resources (tape, disk)
- HSM migrated files automatically recalled and re-migrated to original migration level (ML1, ML2)
- Scanned large numbers of data stores with no performance impact
Issues

Resolved during evaluation

✓ Dynamic allocation of HSM migration control dataset caused contention with HSM CDS reorgs.

✓ Excessive runtimes for PDS/PDSE scans. Fix has been developed.
Performance

- Quick data analysis; select group of datastores

<table>
<thead>
<tr>
<th></th>
<th>Test LPAR</th>
<th>Prod LPAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Files Scanned</td>
<td>34,960</td>
<td>225,195</td>
</tr>
<tr>
<td>By File Type:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PDS/E</td>
<td>669</td>
<td>1,804</td>
</tr>
<tr>
<td>QSAM</td>
<td>5,049</td>
<td>176,852</td>
</tr>
<tr>
<td>VSAM</td>
<td>28,447</td>
<td>37,112</td>
</tr>
<tr>
<td>DB2 Tables</td>
<td>795</td>
<td>9,428</td>
</tr>
<tr>
<td>By Media Type:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DASD (non-migrated)</td>
<td>32,169</td>
<td>67,812</td>
</tr>
<tr>
<td>Tape</td>
<td>115</td>
<td>41,905</td>
</tr>
<tr>
<td>Migrated (ML1/ML2)</td>
<td>2,676</td>
<td>115,478</td>
</tr>
</tbody>
</table>

- Minimal mainframe resource usage
- User-controlled mainframe resources:
  - Tape
  - HSM recalls
Custom Regular Expressions

- Identify SSNs with specific state codes
- Addresses and zip+4’s in the form “xxxxxx-xxxx”.
- Example of SSN regular expression

```plaintext
(?:[0-6]\d{2}|7[0-6]\d|77[0-2])(?: {0,2}(?(-)(?:-\ {0,2})))(?:\d{2})(?: {0,2}(?(-)(?:-\ {0,2})))(?:\d{4})

((253|263|260)[0-9]{6})|((253|260|263)-[0-9]{2}-[0-9]{4})
```
Conclusion

✓ Data identification and classification
✓ Mainframe data at rest
✓ Easy to use
✓ Quick and efficient
✓ Little or no impact to the operational environment

DataSniff is a valuable part of Primerica’s comprehensive DLP program.