CompliancePoint™
A PossibleNOW Company

Data Discovery

Searching for Sensitive Information

PRESENTER

Charles Burke QSA, CISSP, CSSLP – *VP & Chief Solutions Architect*

Matt McClendon – *Director, Security Services*

AUGUST 22, 2011

BIO

- Charles Burke QSA, CISSP, CSSLP
  - VP & Chief Solutions Architect
  - 22 years IT Experience 15 years Security

- Matt McClendon
  - Director Endpoint Security and PIIFinder SaaS

About CompliancePoint

- Information Security & Compliance Company
  - Comprised of two divisions
    - **Information Security Compliance Consulting**
      - Certifications & Assessments to meet mandated requirements (QSA, PA-QSA, HIPAA)
      - SSAE 16 / SOC Reporting Engagements
    - **Information Security Compliance Services**
      - Remediation
      - Technology Integration
      - Staff Augmentation

About You

Quick BIO

Company

Data Discovery experience

Outline

- Background
- Types of Data Discovery / Tools
- Methodology for PII Discovery
- Challenges
- Tools Comparison
- PIIFinder SaaS
- Technical Deep Dive
- Types of Reporting
- Q/A

Background

- eWeek : **10 Biggest Data Breaches of 2011 So Far**
  - **By Fahmida Y. Rashid on 2011-05-25**

1. **Sony's Playstation Network, Qriocity, Sony Online Entertainment**
   - April 26th
   - 101 million user accounts
2. **Epsilon, Alliance Data Systems**
   - April 1st
   - Unknown, Estimated 60 million e-mail addresses
3. **HBGary Federal**
   - Feb 7th
   - 60,000 records
4. **WordPress**
   - April 14th
   - Unknown, 18 million records estimated
5. **University of South Carolina**
   - March 4th
   - 31,000 records

Background

6. **TripAdvisor, Expedia**
   - March 24th
   - Unknown
7. **RSA Security**
   - March 18th
   - Unknown
8. **HuskyDirect.com, University of Connecticut**
   - Jan 11th
   - 18,059 records
9. **Seacoast Radiology**
   - Jan 12th
   - 231,400 records
10. **Ankle and Foot Center of Tampa Bay**
    - Jan 29th
    - 156,000

Background

Organizations are finding it necessary to implement technologies and services to search for sensitive data.

- **Compliance (PCI-Network Segmentation)**
- **Discovery and redaction of sensitive data in non-production environments.**

Outline

- Background
- **Types of Data Discovery / Tools**
- Methodology for PII Discovery
- Challenges
- Tools Comparison
- PIIFinder SaaS
- Technical Deep Dive
- Types of Reporting
- Q/A

Types of Data Discovery / Tools

Forensic Data Discovery

Examining digital media in a forensically sound manner to identify, preserve, recover, analyze and present facts and opinion about the information.

- Often associated with investigations of computer crime and includes practices to create a legal audit trail.
- Target data and systems are typically known prior to activity.
- Typically performed in response to crime, policy violation, etc. - Reactive
- Time sensitive, time consuming, Expensive

Types of Data Discovery / Tools

Forensic Tools (commercial)

Name	Platform	Description
EnCase	Windows	Multi-purpose forensic tool
FTK	Windows	Multi-purpose tool, used to index media
PTK Forensics	LAMP	GUI for The Sleuth Kit
Paraben P2 Commander	Windows	General purpose forensic tool kit
COFEE	Windows	Tools for Windows developed by Microsoft only available to law enforcement
SafeBack	N/A	Digital media acquisition and backup
dtSearch	Windows	Instant search of terabytes of text across desktops, network, internet sites. Includes powerful API

Types of Data Discovery / Tools

Forensic Tools (Open/Free)

Name	Platform	Description
SANS Investigative Forensic Toolkit - SIFT	Windows	Multi-Purpose forensic operating system
Digital Forensics Framework	Windows, Linux, MacOS	DFF is a digital investigation tool and development platform
The Coroner's Toolkit	Unix-like	Suite of programs for Unix analysis
The Sleuth Kit	Unix-like / Windows	Tools for Unix and Windows
Open Computer Forensics Architecture	Linux	Framework for computer forensics Lab environment

Types of Data Discovery / Tools

PII Data Discovery

Examining file systems and databases to identify sensitive data using operating system and application level search functions/utilities to identify specific types of data (personal identifiable data).

- Often associated with risk assessments (Data Leakage) program.
- Target environment may be known but location of data is unknown until search is complete.
- Typically performed proactively to prevent sensitive data exposure / leakage
- Followed by remediation activity to mitigate exposure

Types of Data Discovery / Tools

PII Tools (commercial)

Name	Platform	Description
Identity Finder	Windows	Can identify PII data such as SSN, credit card numbers, bank account numbers, passwords, DOB
dtSearch	Windows	Instant search of terabytes of text across desktops, network, internet sites. Includes powerful API
RegexBuddy	Windows	Popular Regex tool can be used to search file systems for any regular expression match

Types of Data Discovery / Tools

PII Tools (Open/Free)

Name	Platform	Description
Find_SSN	Windows	Searches for matching patters of Social Security Numbers only.
Sensitive Number Finder (SENF)	Windows	Multi-purpose tool, used to index media
Spider	LAMP	GUI for The Sleuth Kit

Outline

- Background
- Types of Data Discovery Tools
- **Methodology for PII Discovery**
- Challenges
- Tools Comparison
- PIIFinder SaaS
- Technical Deep Dive
- Types of Reporting
- Q/A

Methodology for PII Data Discovery

- Information gathering
  - **Following Scan Request, this stage is for gathering information relevant to scanning. The Data Discovery Pre-Engagement worksheet includes a list of questions that provide useful information to engineers before, during, and after scans.**
  - **Collecting this information is necessary before launching any scanning tools to ensure that we have a basic understanding of the environment and avoid any negative impacts.**
  - **This stage will also include the Kick-off meeting/call.**
  - **Discover and document the types of data for searching.**
  - **Determine (based on target specifications) best scanning methods**

- Onboarding & creation of project plan.
  - **The consulting manger will create a detailed project plan for the engagement including onsite and off-site activities.**
  - **Onboard client POC to Project Management Portal for scan request, status, and report delivery**

Methodology for PII Data Discovery

- Scan Preparation
  - Infrastructure provisioning, policy creation based on search criteria
  - User provisioning, Local & Remote access
  - Engineers will run initial scans based on information gathered to test configurations and produce scan list.
  - The scan list(s) will be reviewed as to accurately produce a list of scan targets.
  - Scan list will be created and divided into smaller manageable list if needed.
  - The project plan may require updates based on scan lists created.

- Data Scanning & manual analysis
  - Following the project plan scanning will begin for each environment.
  - Engineers will collect and analyze scan results.
  - Additional scans will be performed if necessary.

Methodology for PII Data Discovery

- Report creation
  - Typically performed offsite, the engineers will create scan reports as a final deliverable to client.
  - The cumulative information from these activities are correlated and analyzed to produce the findings and reporting

Methodology for PII Data Discovery

Pre-Engagement Questions

- Types & Number of Servers (Windows, Unix, etc.)
- Virtual Machines
- Database Server Types (SQL Server, Oracle, DB2, etc.)
- Database Server sizes (estimate of number of records)
- SAN / NAS Storage (estimate of size)
- Remote or Local Scanning
  - **Software distribution method?**
  - **Scan backup or production systems?**
- What types of data will be searched for?
  - **SSN**
  - **Credit Card**
  - **PAN**
  - **Bank Account/Routing Numbers**
  - **Health records**
  - **Passwords**

Outline

- Background
- Types of Data Discovery / Tools
- Methodology for PII Discovery
- **Challenges**
- Tools Comparison
- PIIFinder SaaS
- Technical Deep Dive
- Types of Reporting
- Q/A

Challenges to PII Data Discovery

- Client Request/Requirements
  - Should be captured in scoping (information gathering) exercise
- Delays with client activities
  - infrastructure request
  - User ID provisioning
- System Restarts
- Endless File Types
  - Auto recognition
  - Externalization

Challenges to PII Data Discovery

- Flaws in APIs
  - Zip, Pdf, Image processing
- Scan duration
  - Maintenance windows
- Lots of Data & Results
- False Positives

Outline

- Background
- Types of Data Discovery / Tools
- Methodology for PII Discovery
- Challenges
- **Tools Comparison**
- PIIFinder SaaS
- Technical Deep Dive
- Types of Reporting
- Q/A

Tools Comparison

ISACA Journal Article

Computer Forensics Technologies for Personally Identifiable Information Detection and Audits

http://www.isaca.org/Journal/Past-Issues/2010/Volume-2/Pages/Computer-Forensics-Technologies-for-Personally-Identifiable-Information-Detection-and-Audits1.aspx

Tools Comparison

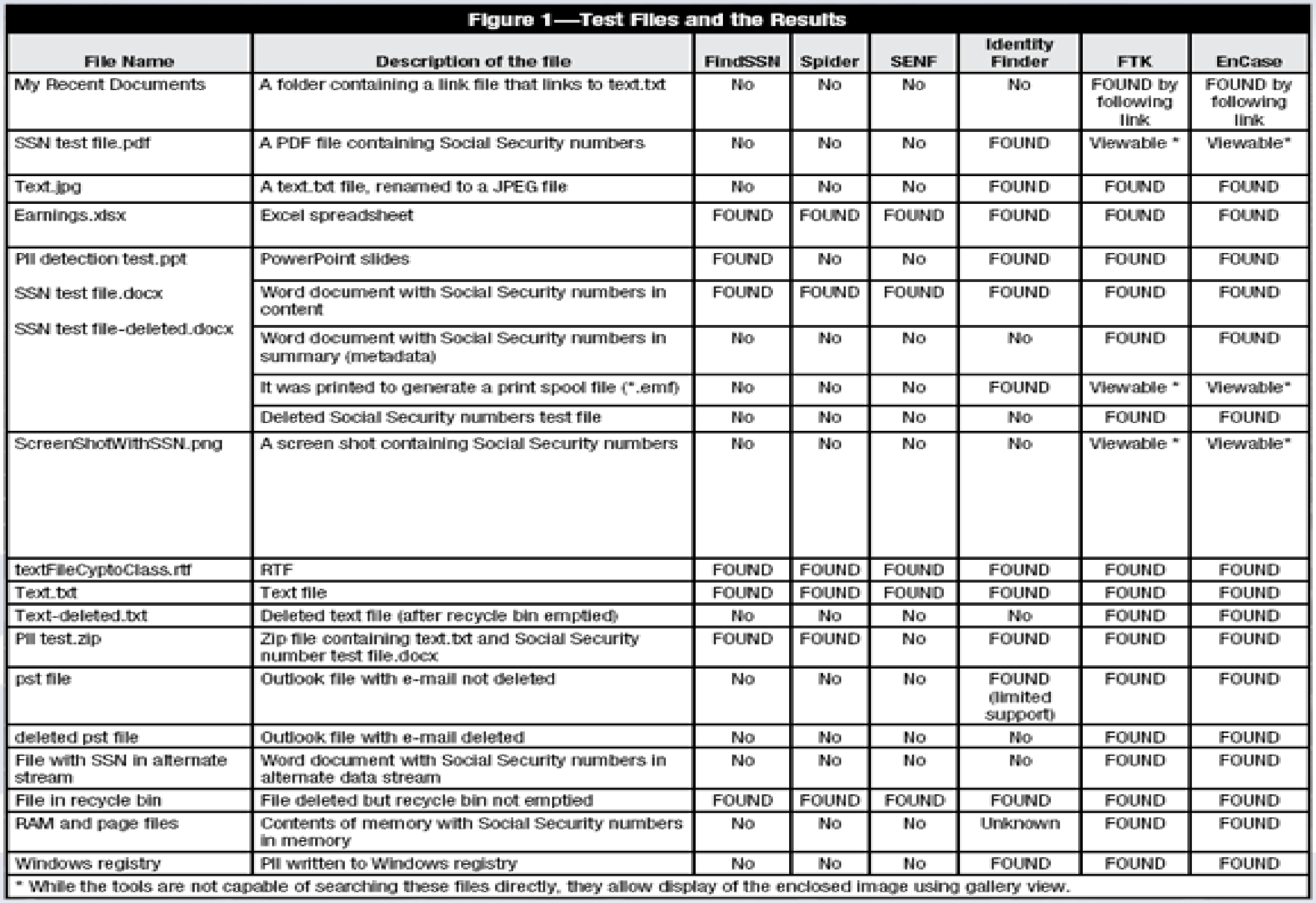


Figure 1—Test Files and the Results

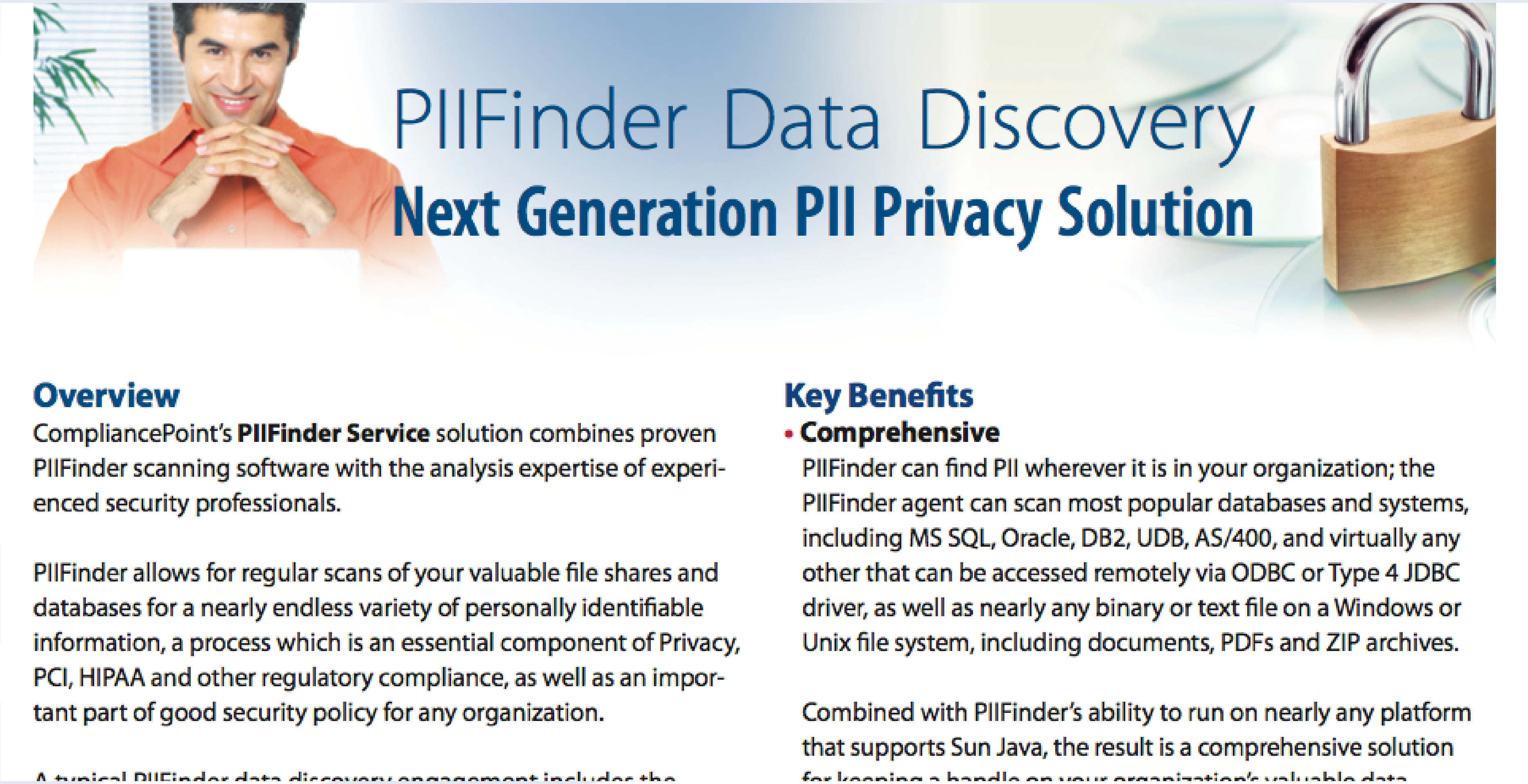
File Name	Description of the file	FindSSN	Spider	SENF	Identity Finder	FTK	EnCase
My Recent Documents	A folder containing a link file that links to text.txt	No	No	No	No	FOUND by following link	FOUND by following link
SSN test file.pdf	A PDF file containing Social Security numbers	No	No	No	FOUND	Viewable *	Viewable*
Text.jpg	A text.txt file, renamed to a JPEG file	No	No	No	FOUND	FOUND	FOUND
Earnings.xlsx	Excel spreadsheet	FOUND	FOUND	FOUND	FOUND	FOUND	FOUND
PII detection test.ppt SSN test file.docx SSN test file-deleted.docx	PowerPoint slides	FOUND	No	No	FOUND	FOUND	FOUND
	Word document with Social Security numbers in content	FOUND	FOUND	FOUND	FOUND	FOUND	FOUND
	Word document with Social Security numbers in summary (metadata)	No	No	No	No	FOUND	FOUND
	It was printed to generate a print spool file (*.emf)	No	No	No	FOUND	Viewable *	Viewable*
	Deleted Social Security numbers test file	No	No	No	No	FOUND	FOUND
ScreenShotWithSSN.png	A screen shot containing Social Security numbers	No	No	No	No	Viewable *	Viewable*
textFileCyptoClass.rtf	RTF	FOUND	FOUND	FOUND	FOUND	FOUND	FOUND
Text.txt	Text file	FOUND	FOUND	FOUND	FOUND	FOUND	FOUND
Text-deleted.txt	Deleted text file (after recycle bin emptied)	No	No	No	No	FOUND	FOUND
PII test.zip	Zip file containing text.txt and Social Security number test file.docx	FOUND	FOUND	No	FOUND	FOUND	FOUND
pst file	Outlook file with e-mail not deleted	No	No	No	FOUND (limited support)	FOUND	FOUND
deleted pst file	Outlook file with e-mail deleted	No	No	No	No	FOUND	FOUND
File with SSN in alternate stream	Word document with Social Security numbers in alternate data stream	No	No	No	No	FOUND	FOUND
File in recycle bin	File deleted but recycle bin not emptied	FOUND	FOUND	FOUND	FOUND	FOUND	FOUND
RAM and page files	Contents of memory with Social Security numbers in memory	No	No	No	Unknown	FOUND	FOUND
Windows registry	PII written to Windows registry	No	No	No	FOUND	FOUND	FOUND

* While the tools are not capable of searching these files directly, they allow display of the enclosed image using gallery view.

Outline

- Background
- Types of Data Discovery / Tools
- Methodology for PII Discovery
- Challenges
- Tools Comparison
- PIIFinder SaaS
- Technical Deep Dive
- Types of Reporting
- Q/A

PIIFinder SaaS



PIIFinder Data Discovery

Next Generation PII Privacy Solution

Overview

CompliancePoint's **PIIFinder Service** solution combines proven PIIFinder scanning software with the analysis expertise of experienced security professionals.

PIIFinder allows for regular scans of your valuable file shares and databases for a nearly endless variety of personally identifiable information, a process which is an essential component of Privacy, PCI, HIPAA and other regulatory compliance, as well as an important part of good security policy for any organization.

A typical PIIFinder data discovery engagement includes the

Key Benefits

- **Comprehensive**

  PIIFinder can find PII wherever it is in your organization; the PIIFinder agent can scan most popular databases and systems, including MS SQL, Oracle, DB2, UDB, AS/400, and virtually any other that can be accessed remotely via ODBC or Type 4 JDBC driver, as well as nearly any binary or text file on a Windows or Unix file system, including documents, PDFs and ZIP archives.

  Combined with PIIFinder's ability to run on nearly any platform that supports Sun Java, the result is a comprehensive solution for keeping a handle on your organization's valuable data.

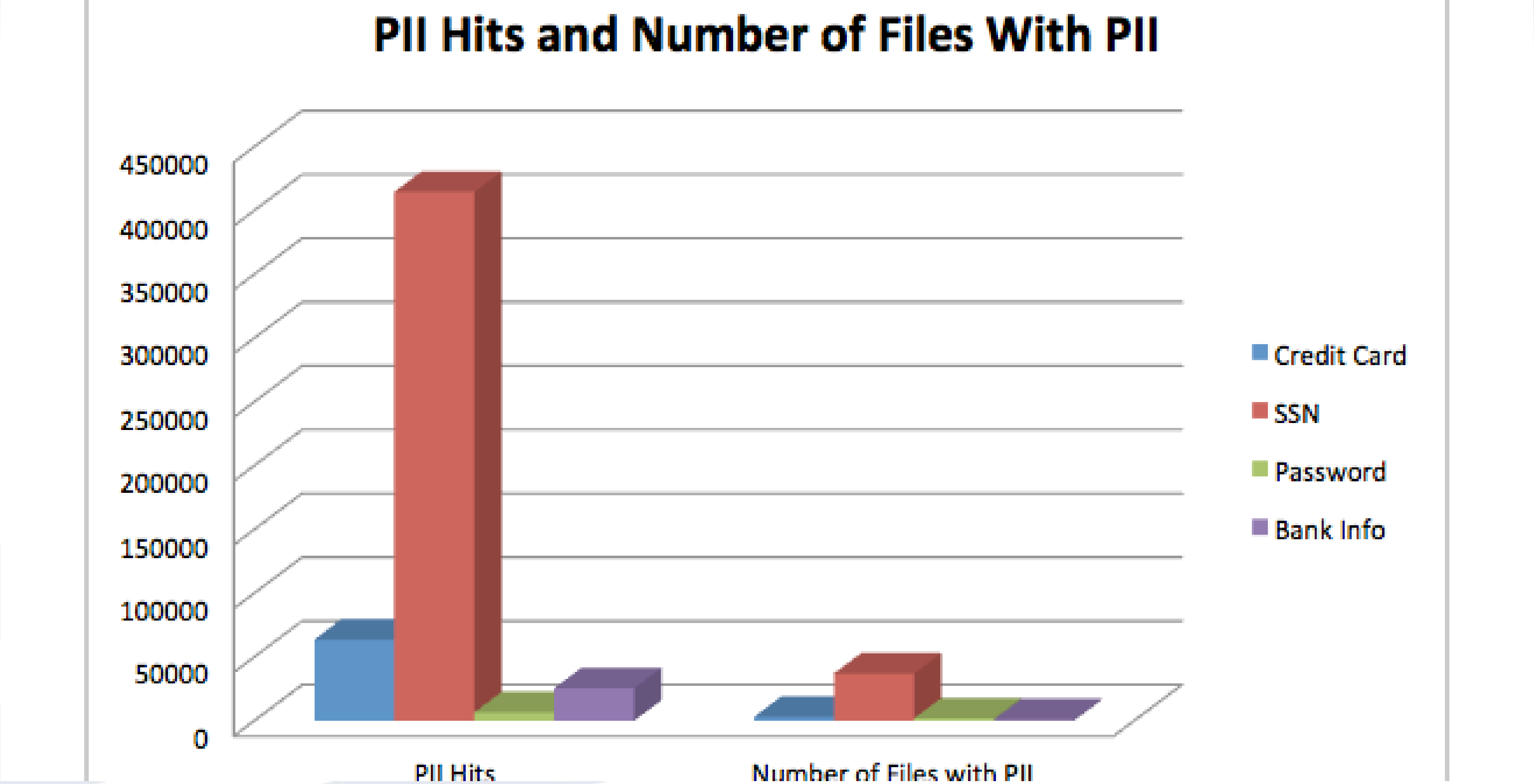
PIIFinder Deep Dive

PIIFinder Deep Dive

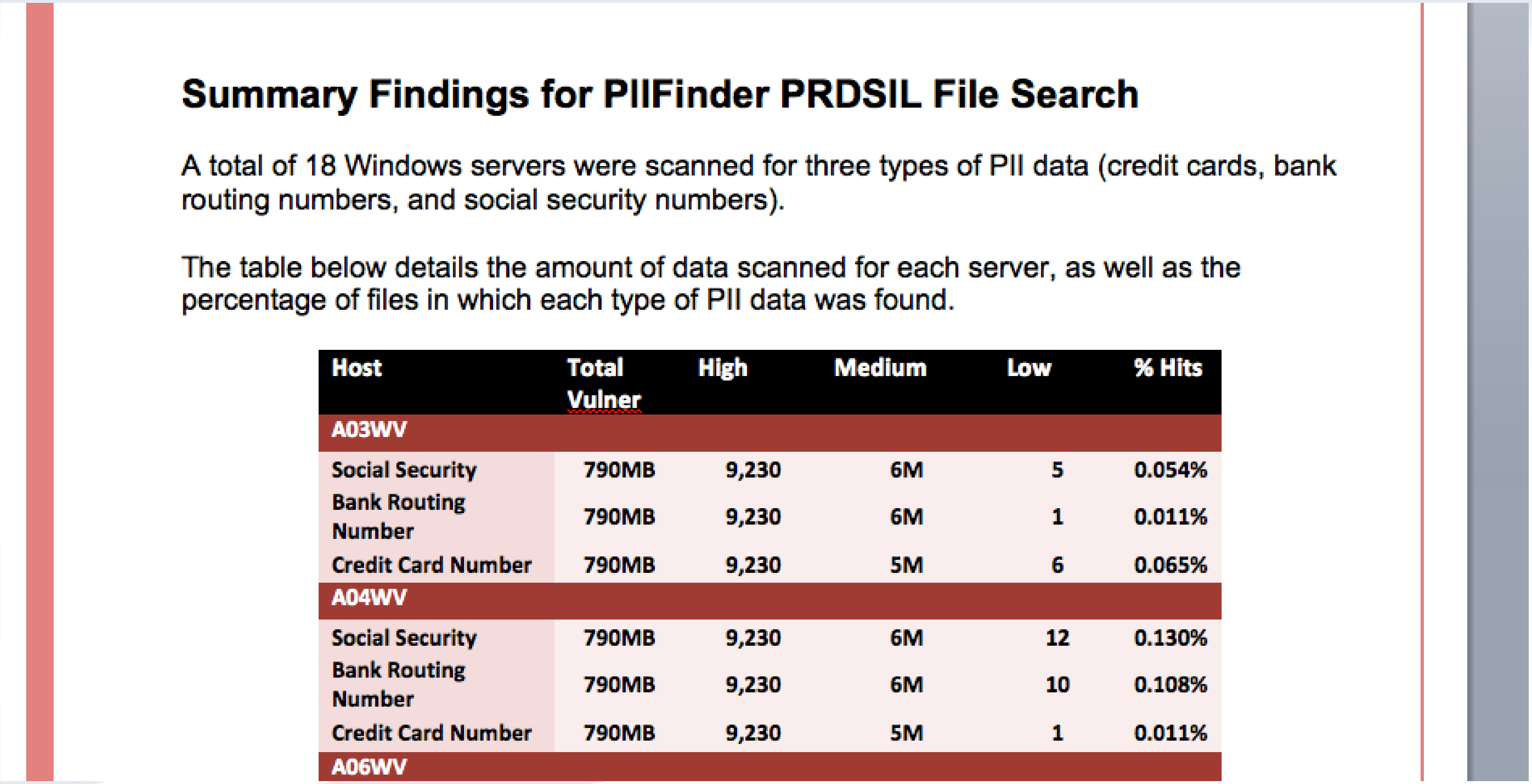
Outline

- Background
- Types of Data Discovery / Tools
- Methodology for PII Discovery
- Challenges
- Tools Comparison
- PIIFinder SaaS
- Technical Deep Dive
- Reporting
- Q/A

Summary Reporting



Summary Reporting



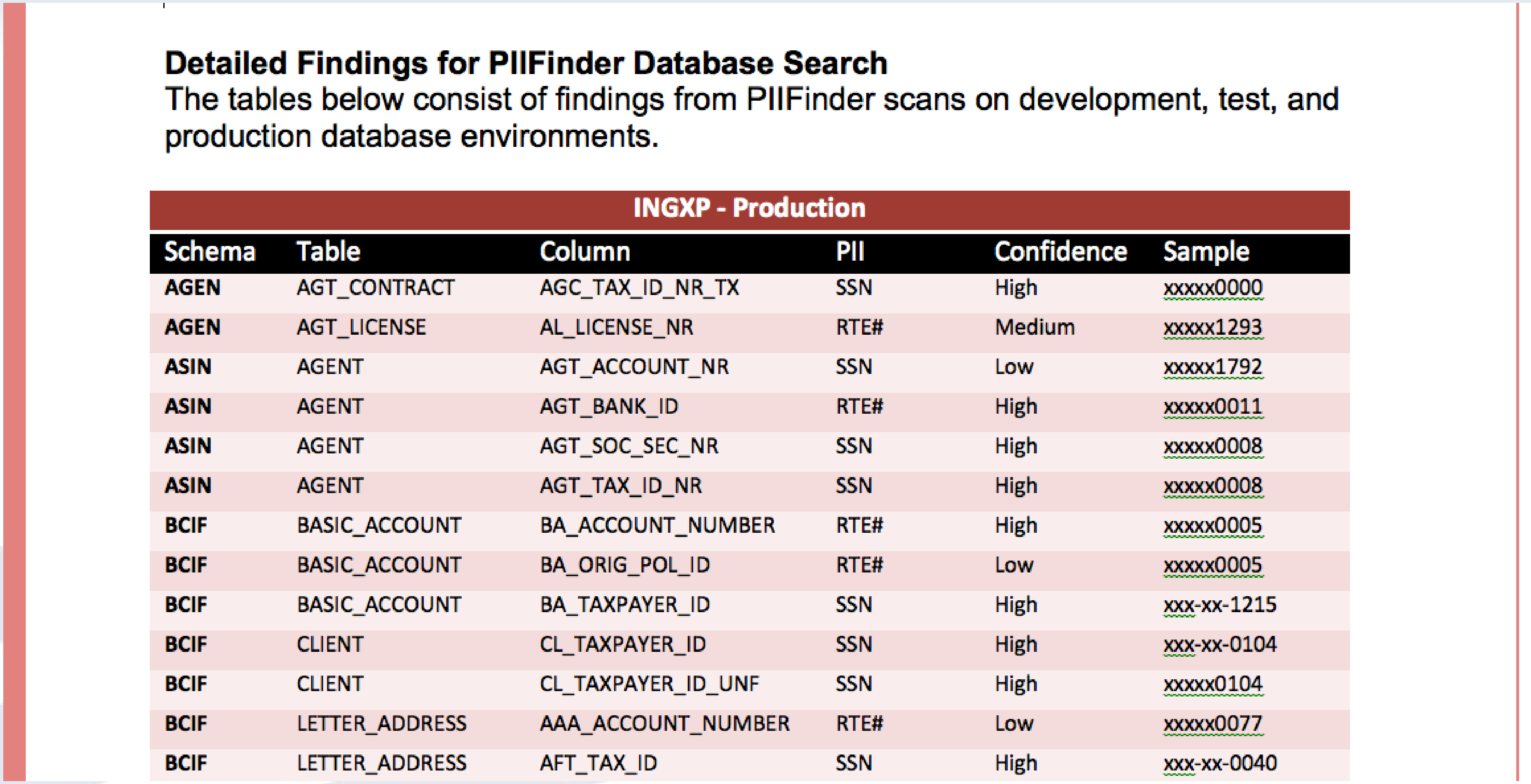
Summary Findings for PIIFinder PRDSIL File Search

A total of 18 Windows servers were scanned for three types of PII data (credit cards, bank routing numbers, and social security numbers).

The table below details the amount of data scanned for each server, as well as the percentage of files in which each type of PII data was found.

Host	Total Vulner	High	Medium	Low	% Hits
A03WV					
Social Security	790MB	9,230	6M	5	0.054%
Bank Routing Number	790MB	9,230	6M	1	0.011%
Credit Card Number	790MB	9,230	5M	6	0.065%
A04WV					
Social Security	790MB	9,230	6M	12	0.130%
Bank Routing Number	790MB	9,230	6M	10	0.108%
Credit Card Number	790MB	9,230	5M	1	0.011%
A06WV					

Detailed Reporting



Detailed Findings for PIIFinder Database Search

The tables below consist of findings from PIIFinder scans on development, test, and production database environments.

INGXP - Production

Schema	Table	Column	PII	Confidence	Sample
AGEN	AGT_CONTRACT	AGC_TAX_ID_NR_TX	SSN	High	xxxxx0000
AGEN	AGT_LICENSE	AL_LICENSE_NR	RTE#	Medium	xxxxx1293
ASIN	AGENT	AGT_ACCOUNT_NR	SSN	Low	xxxxx1792
ASIN	AGENT	AGT_BANK_ID	RTE#	High	xxxxx0011
ASIN	AGENT	AGT_SOC_SEC_NR	SSN	High	xxxxx0008
ASIN	AGENT	AGT_TAX_ID_NR	SSN	High	xxxxx0008
BCIF	BASIC_ACCOUNT	BA_ACCOUNT_NUMBER	RTE#	High	xxxxx0005
BCIF	BASIC_ACCOUNT	BA_ORIG_POL_ID	RTE#	Low	xxxxx0005
BCIF	BASIC_ACCOUNT	BA_TAXPAYER_ID	SSN	High	xxx-xx-1215
BCIF	CLIENT	CL_TAXPAYER_ID	SSN	High	xxx-xx-0104
BCIF	CLIENT	CL_TAXPAYER_ID_UNF	SSN	High	xxxxx0104
BCIF	LETTER_ADDRESS	AAA_ACCOUNT_NUMBER	RTE#	Low	xxxxx0077
BCIF	LETTER_ADDRESS	AFT_TAX_ID	SSN	High	xxx-xx-0040

Reporting

Top environments of unknown PII data

- Production Systems
  - Legacy systems
  - Retired applications
  - Databases
- Network Shares
  - Public folders
  - Admin share
  - Networking share

Reporting

Top environments of unknown PII data

- Q/A
  - Using production data?
  - Databases
- Repurposed resources
  - Server hardware
  - Storage
  - Virtual Machines

Questions?