Identification of Known Files on Computer Systems

AAFS 2005

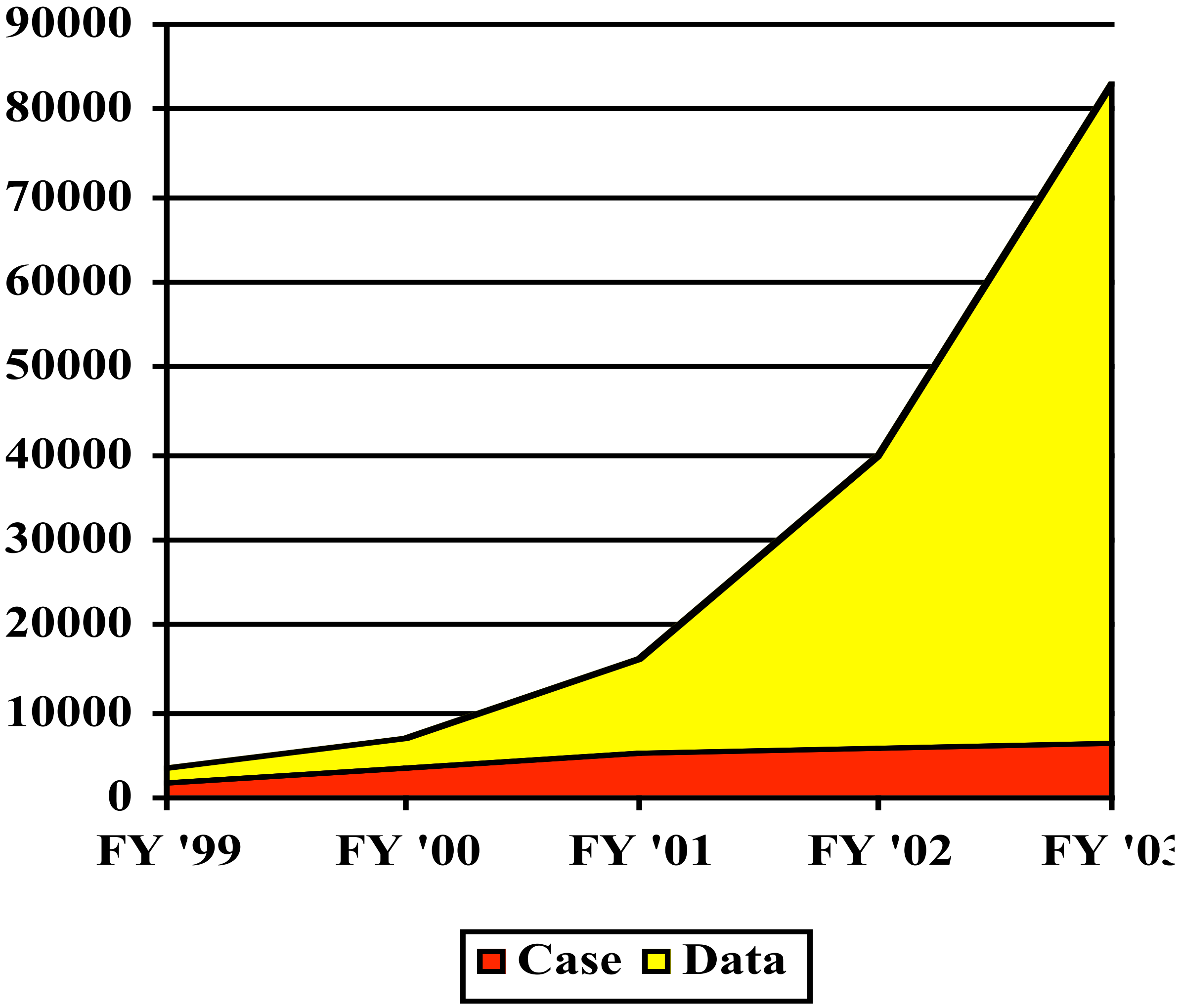
Douglas White
Michael Ogata

NIST United States Department of Commerce
National Institute of Standards and Technology

Disclaimer

Trade names and company products are mentioned in the text or identified. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products are necessarily the best available for the purpose.

Problem: Data Inflation



FBI's Cyber Caseload and Dataset Size Growth

Source: FBI CART, Oct 2003

NIST Digital Forensics Goals

- Provide standard reference data that investigators and tool makers can use
- Assist in reducing manual processes in case loads, reducing case processing time
- Identify known files, allowing investigator to focus on user-generated data

Known File Identification

Digital fingerprint, or “hash”

- Cryptographic function: MD5, SHA-1
- Like human fingerprint, can’t rebuild original from this information
- Extremely hard to circumvent
  - Be aware of collision research

Related History

- CRC concept dates from 1960’s
- MD5 algorithm published in 1991
- Tripwire open source tool 1992
- Hash command “md5sum” available
- FIPS 180-1 (SHA-1) published in 1995
- Hash command “sha1sum” available
- Known File Filter project 1998
- FIPS 180-2 (SHA-512) published in 2002
- Hash command “sha2sum” available

Hash Examples

Filename	Bytes	SHA-1
NT4\ALPHA\notepad.exe	68368	F1F284D5D757039DEC1C44A05AC148B9D204E467
NT4\I386\notepad.exe	45328	3C4E15A29014358C61548A981A4AC8573167BE37
NT4\MIPS\notepad.exe	66832	33309956E4DBBA665E86962308FE5E1378998E69
NT4\PPC\notepad.exe	68880	47BB7AF0E4DD565ED75DEB492D8C17B1BFD3FB23
NT31WS\I386\notepad.exe	57252	2E0849CF327709FC46B705EEAB5E57380F5B1F67
NT31SRV\I386\notepad.exe	57252	2E0849CF327709FC46B705EEAB5E57380F5B1F67
contract.txt		0BD71F653A5B83E61D66DB6D29B9B46655D77F42

Hash Application

Which was the original?

```
contract1.txt
John Doe owes Rachel Roe $15.00
```

```
contract2.txt
John Doe owes Rachel Roe $1500.
```

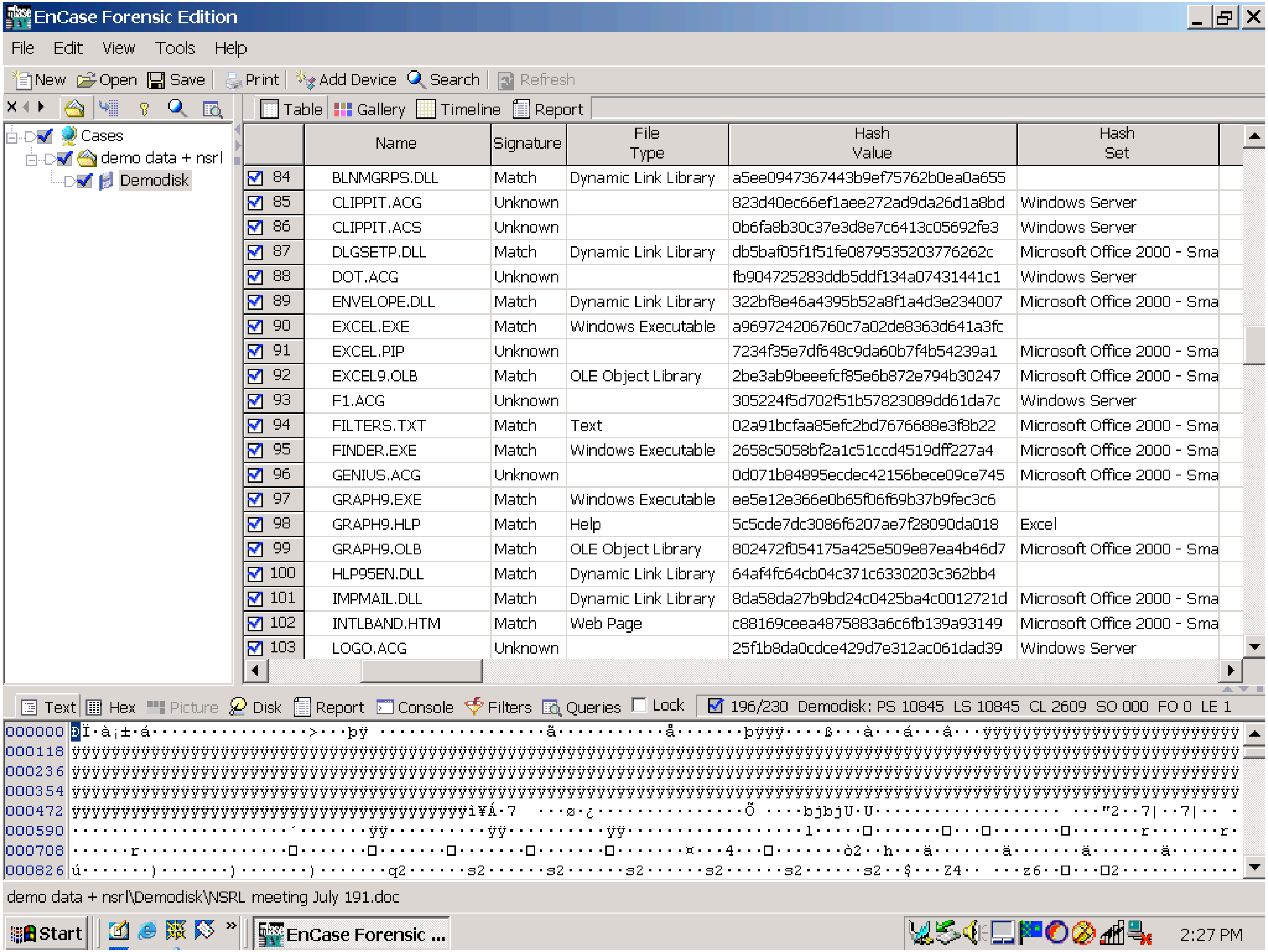
Hash Application

```
sha1sum contract*
0BD71F653A5B83E61D66DB6D29B9B46655D77F42    contract1.txt
B10A4DEDC819737E7D62363ADE0A2F035A2CC20F    contract2.txt


0BD71F653A5B83E61D66DB6D29B9B46655D77F42    contract.txt
```

Hashset Sources

- NIST NSRL
- NDIC HashKeeper
- Maresware
- Tripwire FSDB
- Known Goods website
- Vendors, e.g. Sun Solaris Fingerprints
- CFTT, iLook, CFID email lists
- Professional connections

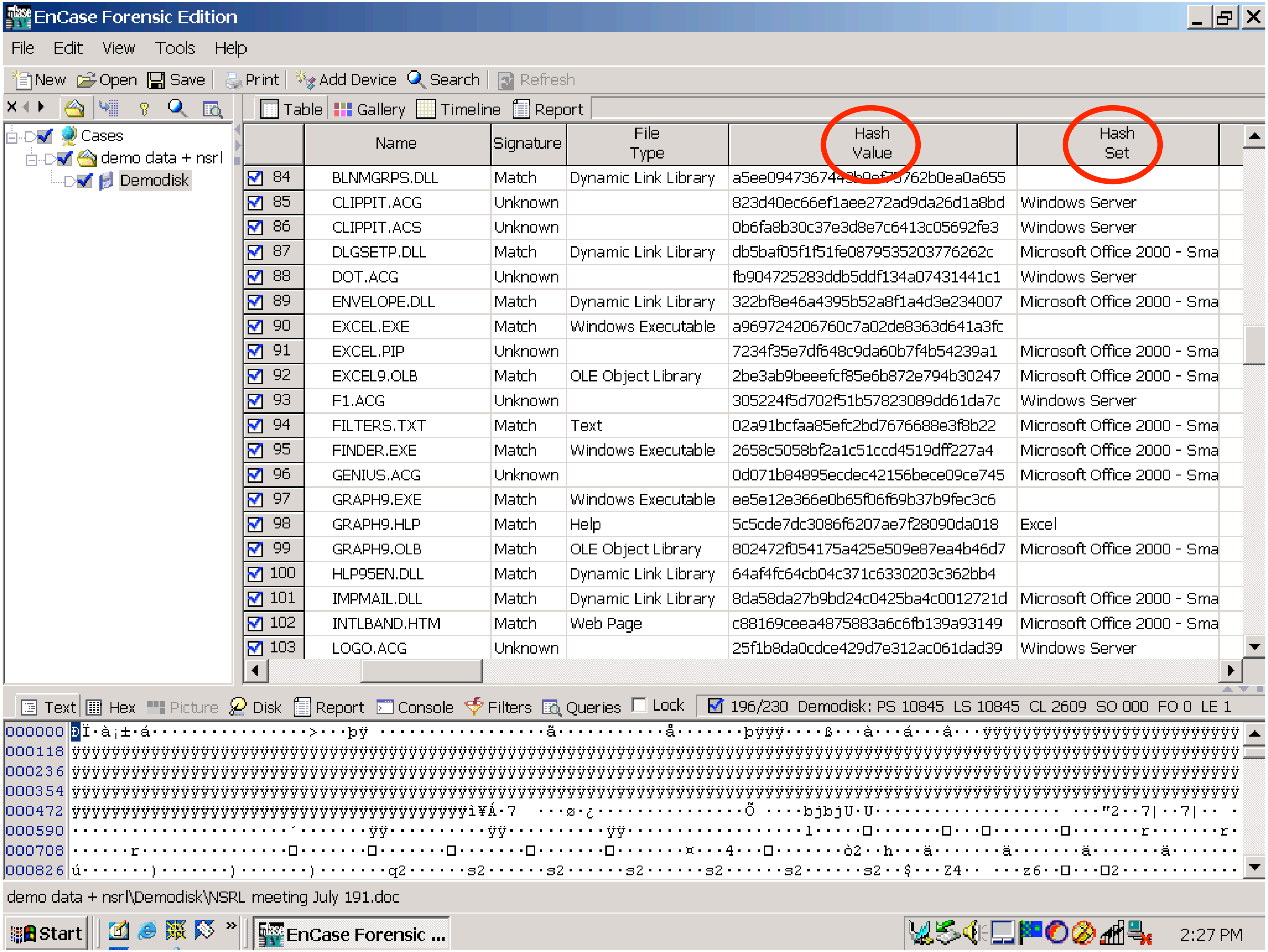


	Name	Signature	File Type	Hash Value	Hash Set
☑ 84	BLNMGRPS.DLL	Match	Dynamic Link Library	a5ee0947367443b9ef75762b0ea0a655	
☑ 85	CLIPPIT.ACG	Unknown		823d40ec66ef1aee272ad9da26d1a8bd	Windows Server
☑ 86	CLIPPIT.ACS	Unknown		0b6fa8b30c37e3d8e7c6413c05692fe3	Windows Server
☑ 87	DLGSETP.DLL	Match	Dynamic Link Library	db5baf05f1f51fe0879535203776262c	Microsoft Office 2000 - Sma
☑ 88	DOT.ACG	Unknown		fb904725283ddb5ddf134a07431441c1	Windows Server
☑ 89	ENVELOPE.DLL	Match	Dynamic Link Library	322bf8e46a4395b52a8f1a4d3e234007	Microsoft Office 2000 - Sma
☑ 90	EXCEL.EXE	Match	Windows Executable	a969724206760c7a02de8363d641a3fc	
☑ 91	EXCEL.PIP	Unknown		7234f35e7df648c9da60b7f4b54239a1	Microsoft Office 2000 - Sma
☑ 92	EXCEL9.OLB	Match	OLE Object Library	2be3ab9beeefcf85e6b872e794b30247	Microsoft Office 2000 - Sma
☑ 93	F1.ACG	Unknown		305224f5d702f51b57823089dd61da7c	Windows Server
☑ 94	FILTERS.TXT	Match	Text	02a91bcfaa85efc2bd7676688e3f8b22	Microsoft Office 2000 - Sma
☑ 95	FINDER.EXE	Match	Windows Executable	2658c5058bf2a1c51ccd4519dff227a4	Microsoft Office 2000 - Sma
☑ 96	GENIUS.ACG	Unknown		0d071b84895ecdec42156bece09ce745	Microsoft Office 2000 - Sma
☑ 97	GRAPH9.EXE	Match	Windows Executable	ee5e12e366e0b65f06f69b37b9fec3c6	
☑ 98	GRAPH9.HLP	Match	Help	5c5cde7dc3086f6207ae7f28090da018	Excel
☑ 99	GRAPH9.OLB	Match	OLE Object Library	802472f054175a425e509e87ea4b46d7	Microsoft Office 2000 - Sma
☑ 100	HLP95EN.DLL	Match	Dynamic Link Library	64af4fc64cb04c371c6330203c362bb4	
☑ 101	IMPMAIL.DLL	Match	Dynamic Link Library	8da58da27b9bd24c0425ba4c0012721d	Microsoft Office 2000 - Sma
☑ 102	INTLBAND.HTM	Match	Web Page	c88169ceea4875883a6c6fb139a93149	Microsoft Office 2000 - Sma
☑ 103	LOGO.ACG	Unknown		25f1b8da0cdce429d7e312ac061dad39	Windows Server

☑ 196/230 Demodisk: PS 10845 LS 10845 CL 2609 SO 000 FO 0 LE 1

```
000000 ÐÏ·à¡±·á·····················>···þÿ ···················ã···········å········þÿÿÿ····ß···à···á···â···ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
000118 ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
000236 ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
000354 ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
000472 ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿì¥Á·7  ···ø·¿··············Õ ····bjbjU·U····················  ···"2··7|··7|··  ·
000590 ·························´········ÿÿ·············ÿÿ·············ÿÿ····················l·····□·········□···□·········□·········r·········r·
000708 ······r··················□·········□·········□·········□·········□·········¤···4···□·········ò2··h···ä·········ä·········ä·········ä·······
000826 ú·······)·······)·······)·······q2······s2······s2······s2······s2······s2······s2··$···Z4··  ···z6··□···□2·············
```

demo data + nsrl\Demodisk\NSRL meeting July 191.doc

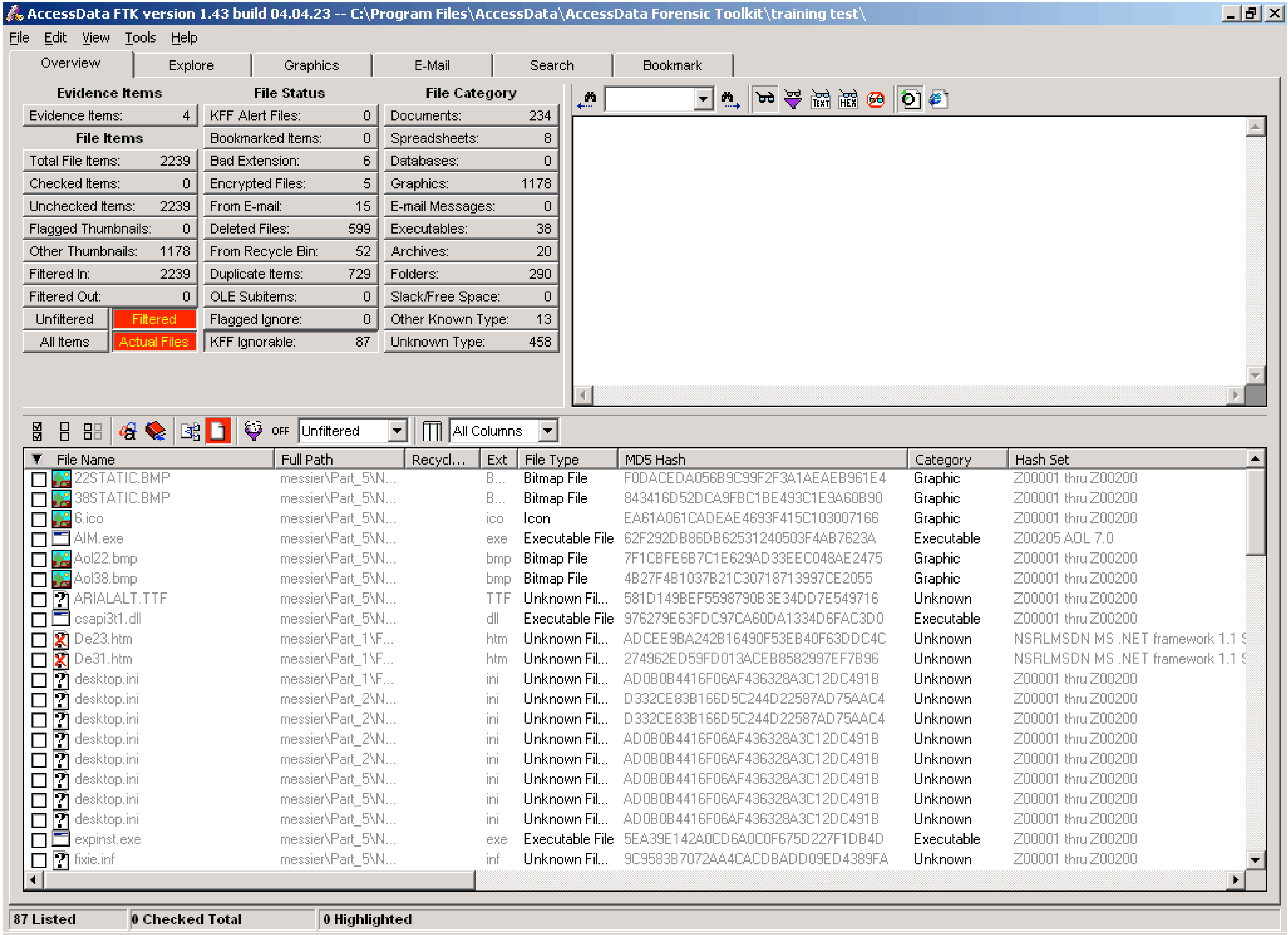

EnCase Forensic Edition
File Edit View Tools Help
New Open Save Print Add Device Search Refresh
Cases
demo data + nsrl
Demodisk
Table Gallery Timeline Report
Text Hex Picture Disk Report Console Filters Queries Lock
Start
EnCase Forensic ...
2:27 PM

	Name	Signature	File Type	Hash Value	Hash Set
☑ 84	BLNMGRPS.DLL	Match	Dynamic Link Library	a5ee0947367443b0ef75762b0ea0a655	
☑ 85	CLIPPIT.ACG	Unknown		823d40ec66ef1aee272ad9da26d1a8bd	Windows Server
☑ 86	CLIPPIT.ACS	Unknown		0b6fa8b30c37e3d8e7c6413c05692fe3	Windows Server
☑ 87	DLGSETP.DLL	Match	Dynamic Link Library	db5baf05f1f51fe0879535203776262c	Microsoft Office 2000 - Sma
☑ 88	DOT.ACG	Unknown		fb904725283ddb5ddf134a07431441c1	Windows Server
☑ 89	ENVELOPE.DLL	Match	Dynamic Link Library	322bf8e46a4395b52a8f1a4d3e234007	Microsoft Office 2000 - Sma
☑ 90	EXCEL.EXE	Match	Windows Executable	a969724206760c7a02de8363d641a3fc	
☑ 91	EXCEL.PIP	Unknown		7234f35e7df648c9da60b7f4b54239a1	Microsoft Office 2000 - Sma
☑ 92	EXCEL9.OLB	Match	OLE Object Library	2be3ab9beeefcf85e6b872e794b30247	Microsoft Office 2000 - Sma
☑ 93	F1.ACG	Unknown		305224f5d702f51b57823089dd61da7c	Windows Server
☑ 94	FILTERS.TXT	Match	Text	02a91bcfaa85efc2bd7676688e3f8b22	Microsoft Office 2000 - Sma
☑ 95	FINDER.EXE	Match	Windows Executable	2658c5058bf2a1c51ccd4519dff227a4	Microsoft Office 2000 - Sma
☑ 96	GENIUS.ACG	Unknown		0d071b84895ecdec42156bece09ce745	Microsoft Office 2000 - Sma
☑ 97	GRAPH9.EXE	Match	Windows Executable	ee5e12e366e0b65f06f69b37b9fec3c6	
☑ 98	GRAPH9.HLP	Match	Help	5c5cde7dc3086f6207ae7f28090da018	Excel
☑ 99	GRAPH9.OLB	Match	OLE Object Library	802472f054175a425e509e87ea4b46d7	Microsoft Office 2000 - Sma
☑ 100	HLP95EN.DLL	Match	Dynamic Link Library	64af4fc64cb04c371c6330203c362bb4	
☑ 101	IMPMAIL.DLL	Match	Dynamic Link Library	8da58da27b9bd24c0425ba4c0012721d	Microsoft Office 2000 - Sma
☑ 102	INTLBAND.HTM	Match	Web Page	c88169ceea4875883a6c6fb139a93149	Microsoft Office 2000 - Sma
☑ 103	LOGO.ACG	Unknown		25f1b8da0cdce429d7e312ac061dad39	Windows Server

☑ 196/230 Demodisk: PS 10845 LS 10845 CL 2609 SO 000 FO 0 LE 1

```
000000 ÐÏ·à¡±·á·················>···þÿ ················ã···········å·······þÿÿÿ····ß···à···á···â···ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
000118 ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
000236 ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
000354 ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿ
000472 ÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿÿì¥Á·7  ···ø·¿··············Õ ····bjbjU·U··················  ····"2··7|··7|·· ·
000590 ·······················´·······ÿÿ··········ÿÿ··········ÿÿ··························l····□·······□··□·······□·········r·······r·
000708 ······r················□·······□·······□·······□·······□·······¤···4···□········ò2··h···ä·······ä·······ä·······ä·······
000826 ú·······)·······)·······)·······q2······s2······s2······s2······s2······s2······s2··$···Z4··  ···z6··□···□2··········
```

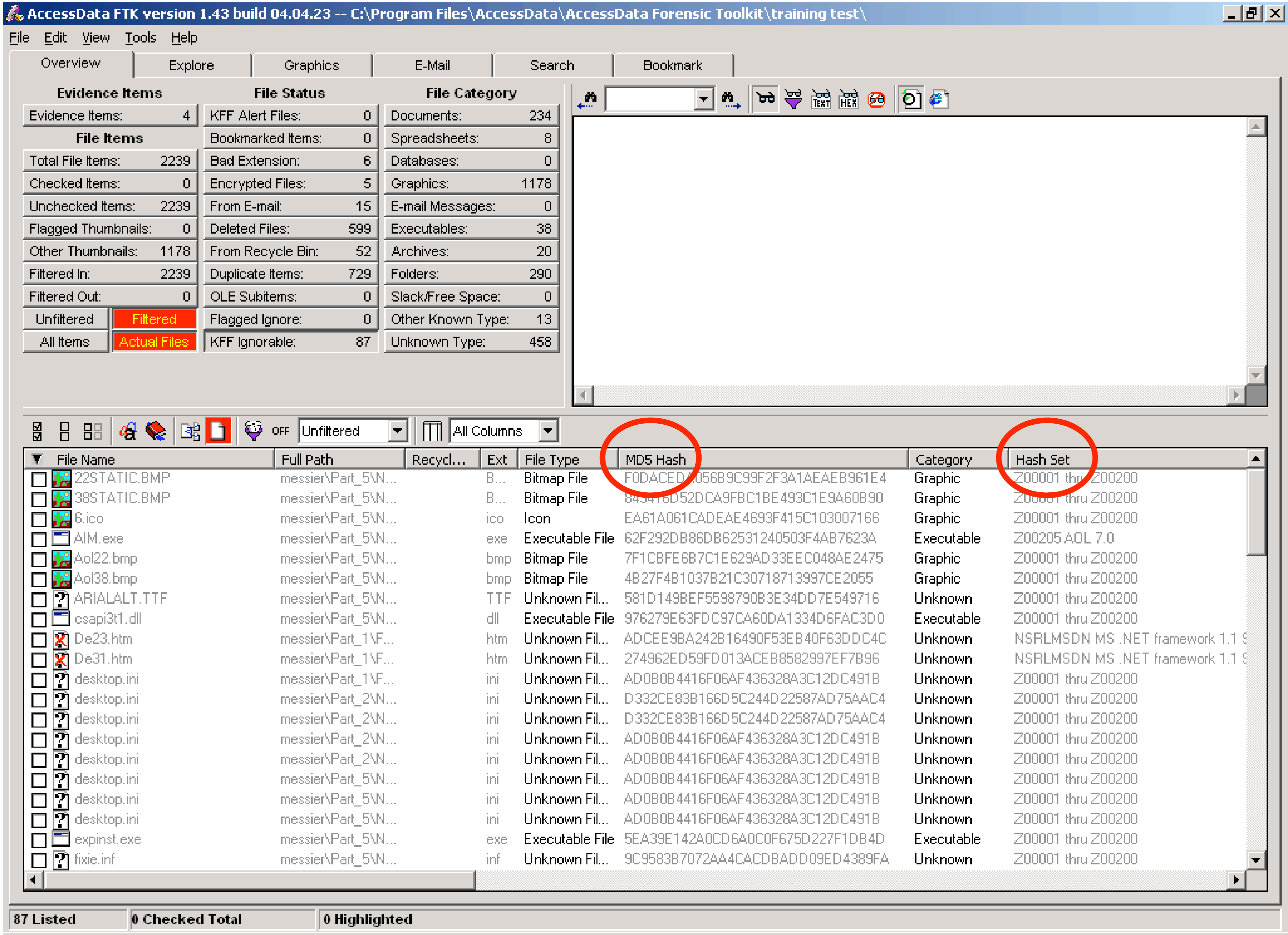
demo data + nsrl\Demodisk\NSRL meeting July 191.doc


AccessData FTK version 1.43 build 04.04.23 -- C:\Program Files\AccessData\AccessData Forensic Toolkit\training test\
File
Edit
View
Tools
Help
Overview
Explore
Graphics
E-Mail
Search
Bookmark

Evidence Items		File Status		File Category	
Evidence Items:	4	KFF Alert Files:	0	Documents:	234
File Items		Bookmarked Items:	0	Spreadsheets:	8
Total File Items:	2239	Bad Extension:	6	Databases:	0
Checked Items:	0	Encrypted Files:	5	Graphics:	1178
Unchecked Items:	2239	From E-mail:	15	E-mail Messages:	0
Flagged Thumbnails:	0	Deleted Files:	599	Executables:	38
Other Thumbnails:	1178	From Recycle Bin:	52	Archives:	20
Filtered In:	2239	Duplicate Items:	729	Folders:	290
Filtered Out:	0	OLE Subitems:	0	Slack/Free Space:	0
Unfiltered	Filtered	Flagged Ignore:	0	Other Known Type:	13
All Items	Actual Files	KFF Ignorable:	87	Unknown Type:	458

	File Name	Full Path	Recycl...	Ext	File Type	MD5 Hash	Category	Hash Set
□	22STATIC.BMP	messier\Part_5\N...		B...	Bitmap File	F0DACEDA056B9C99F2F3A1AEAEB961E4	Graphic	Z00001 thru Z00200
□	38STATIC.BMP	messier\Part_5\N...		B...	Bitmap File	843416D52DCA9FBC1BE493C1E9A60B90	Graphic	Z00001 thru Z00200
□	6.ico	messier\Part_5\N...		ico	Icon	EA61A061CADEAE4693F415C103007166	Graphic	Z00001 thru Z00200
□	AIM.exe	messier\Part_5\N...		exe	Executable File	62F292DB86DB62531240503F4AB7623A	Executable	Z00205 AOL 7.0
□	Aol22.bmp	messier\Part_5\N...		bmp	Bitmap File	7F1CBFE6B7C1E629AD33EEC048AE2475	Graphic	Z00001 thru Z00200
□	Aol38.bmp	messier\Part_5\N...		bmp	Bitmap File	4B27F4B1037B21C30718713997CE2055	Graphic	Z00001 thru Z00200
□	ARIALALT.TTF	messier\Part_5\N...		TTF	Unknown Fil...	581D149BEF5598790B3E34DD7E549716	Unknown	Z00001 thru Z00200
□	csapi3t1.dll	messier\Part_5\N...		dll	Executable File	976279E63FDC97CA60DA1334D6FAC3D0	Executable	Z00001 thru Z00200
□	De23.htm	messier\Part_1\F...		htm	Unknown Fil...	ADCEE9BA242B16490F53EB40F63DDC4C	Unknown	NSRLMSDN MS .NET framework 1.1 S
□	De31.htm	messier\Part_1\F...		htm	Unknown Fil...	274962ED59FD013ACEB8582997EF7B96	Unknown	NSRLMSDN MS .NET framework 1.1 S
□	desktop.ini	messier\Part_1\F...		ini	Unknown Fil...	AD0B0B4416F06AF436328A3C12DC491B	Unknown	Z00001 thru Z00200
□	desktop.ini	messier\Part_2\N...		ini	Unknown Fil...	D332CE83B166D5C244D22587AD75AAC4	Unknown	Z00001 thru Z00200
□	desktop.ini	messier\Part_2\N...		ini	Unknown Fil...	D332CE83B166D5C244D22587AD75AAC4	Unknown	Z00001 thru Z00200
□	desktop.ini	messier\Part_2\N...		ini	Unknown Fil...	AD0B0B4416F06AF436328A3C12DC491B	Unknown	Z00001 thru Z00200
□	desktop.ini	messier\Part_2\N...		ini	Unknown Fil...	AD0B0B4416F06AF436328A3C12DC491B	Unknown	Z00001 thru Z00200
□	desktop.ini	messier\Part_5\N...		ini	Unknown Fil...	AD0B0B4416F06AF436328A3C12DC491B	Unknown	Z00001 thru Z00200
□	desktop.ini	messier\Part_5\N...		ini	Unknown Fil...	AD0B0B4416F06AF436328A3C12DC491B	Unknown	Z00001 thru Z00200
□	desktop.ini	messier\Part_5\N...		ini	Unknown Fil...	AD0B0B4416F06AF436328A3C12DC491B	Unknown	Z00001 thru Z00200
□	expinst.exe	messier\Part_5\N...		exe	Executable File	5EA39E142A0CD6A0C0F675D227F1DB4D	Executable	Z00001 thru Z00200
□	fixie.inf	messier\Part_5\N...		inf	Unknown Fil...	9C9583B7072AA4CACDBADD09ED4389FA	Unknown	Z00001 thru Z00200

87 Listed | 0 Checked Total | 0 Highlighted



AccessData FTK version 1.43 build 04.04.23 -- C:\Program Files\AccessData\AccessData Forensic Toolkit\training test\
File
Edit
View
Tools
Help
Overview
Explore
Graphics
E-Mail
Search
Bookmark
OFF
Unfiltered
All Columns

Evidence Items		File Status		File Category	
Evidence Items:	4	KFF Alert Files:	0	Documents:	234
File Items		Bookmarked Items:	0	Spreadsheets:	8
Total File Items:	2239	Bad Extension:	6	Databases:	0
Checked Items:	0	Encrypted Files:	5	Graphics:	1178
Unchecked Items:	2239	From E-mail:	15	E-mail Messages:	0
Flagged Thumbnails:	0	Deleted Files:	599	Executables:	38
Other Thumbnails:	1178	From Recycle Bin:	52	Archives:	20
Filtered In:	2239	Duplicate Items:	729	Folders:	290
Filtered Out:	0	OLE Subitems:	0	Slack/Free Space:	0
Unfiltered	Filtered	Flagged Ignore:	0	Other Known Type:	13
All Items	Actual Files	KFF Ignorable:	87	Unknown Type:	458

	File Name	Full Path	Recycl...	Ext	File Type	MD5 Hash	Category	Hash Set
□	22STATIC.BMP	messier\Part_5\N...		B...	Bitmap File	F0DACEDA056B9C99F2F3A1AEAEB961E4	Graphic	Z00001 thru Z00200
□	38STATIC.BMP	messier\Part_5\N...		B...	Bitmap File	843416D52DCA9FBC1BE493C1E9A60B90	Graphic	Z00001 thru Z00200
□	6.ico	messier\Part_5\N...		ico	Icon	EA61A061CADEAE4693F415C103007166	Graphic	Z00001 thru Z00200
□	AIM.exe	messier\Part_5\N...		exe	Executable File	62F292DB86DB62531240503F4AB7623A	Executable	Z00205 AOL 7.0
□	Aol22.bmp	messier\Part_5\N...		bmp	Bitmap File	7F1CBFE6B7C1E629AD33EEC048AE2475	Graphic	Z00001 thru Z00200
□	Aol38.bmp	messier\Part_5\N...		bmp	Bitmap File	4B27F4B1037B21C30718713997CE2055	Graphic	Z00001 thru Z00200
□	ARIALALT.TTF	messier\Part_5\N...		TTF	Unknown Fil...	581D149BEF5598790B3E34DD7E549716	Unknown	Z00001 thru Z00200
□	csapi3t1.dll	messier\Part_5\N...		dll	Executable File	976279E63FDC97CA60DA1334D6FAC3D0	Executable	Z00001 thru Z00200
□	De23.htm	messier\Part_1\F...		htm	Unknown Fil...	ADCEE9BA242B16490F53EB40F63DDC4C	Unknown	NSRLMSDN MS .NET framework 1.1 S
□	De31.htm	messier\Part_1\F...		htm	Unknown Fil...	274962ED59FD013ACEB8582997EF7B96	Unknown	NSRLMSDN MS .NET framework 1.1 S
□	desktop.ini	messier\Part_1\F...		ini	Unknown Fil...	AD0B0B4416F06AF436328A3C12DC491B	Unknown	Z00001 thru Z00200
□	desktop.ini	messier\Part_2\N...		ini	Unknown Fil...	D332CE83B166D5C244D22587AD75AAC4	Unknown	Z00001 thru Z00200
□	desktop.ini	messier\Part_2\N...		ini	Unknown Fil...	D332CE83B166D5C244D22587AD75AAC4	Unknown	Z00001 thru Z00200
□	desktop.ini	messier\Part_2\N...		ini	Unknown Fil...	AD0B0B4416F06AF436328A3C12DC491B	Unknown	Z00001 thru Z00200
□	desktop.ini	messier\Part_2\N...		ini	Unknown Fil...	AD0B0B4416F06AF436328A3C12DC491B	Unknown	Z00001 thru Z00200
□	desktop.ini	messier\Part_5\N...		ini	Unknown Fil...	AD0B0B4416F06AF436328A3C12DC491B	Unknown	Z00001 thru Z00200
□	desktop.ini	messier\Part_5\N...		ini	Unknown Fil...	AD0B0B4416F06AF436328A3C12DC491B	Unknown	Z00001 thru Z00200
□	desktop.ini	messier\Part_5\N...		ini	Unknown Fil...	AD0B0B4416F06AF436328A3C12DC491B	Unknown	Z00001 thru Z00200
□	expinst.exe	messier\Part_5\N...		exe	Executable File	5EA39E142A0CD6A0C0F675D227F1DB4D	Executable	Z00001 thru Z00200
□	fixie.inf	messier\Part_5\N...		inf	Unknown Fil...	9C9583B7072AA4CACDBADD09ED4389FA	Unknown	Z00001 thru Z00200

87 Listed | 0 Checked Total | 0 Highlighted

Identification Metrics

Operating System	Files Installed	Percent Identified	Files Unknown	Files in Distribution
Win 98	4,266	93%	297	18,662
Win ME	5,169	93%	383	11,512
Win NT WS	1,659	86%	239	17,904
Win 2KPro	5,963	86%	839	16,539
Win XPPro	9,404	86%	1,293	19,546

Compare hashes from known OS media to hashes of installation of that OS; best case scenario

Identification Metrics

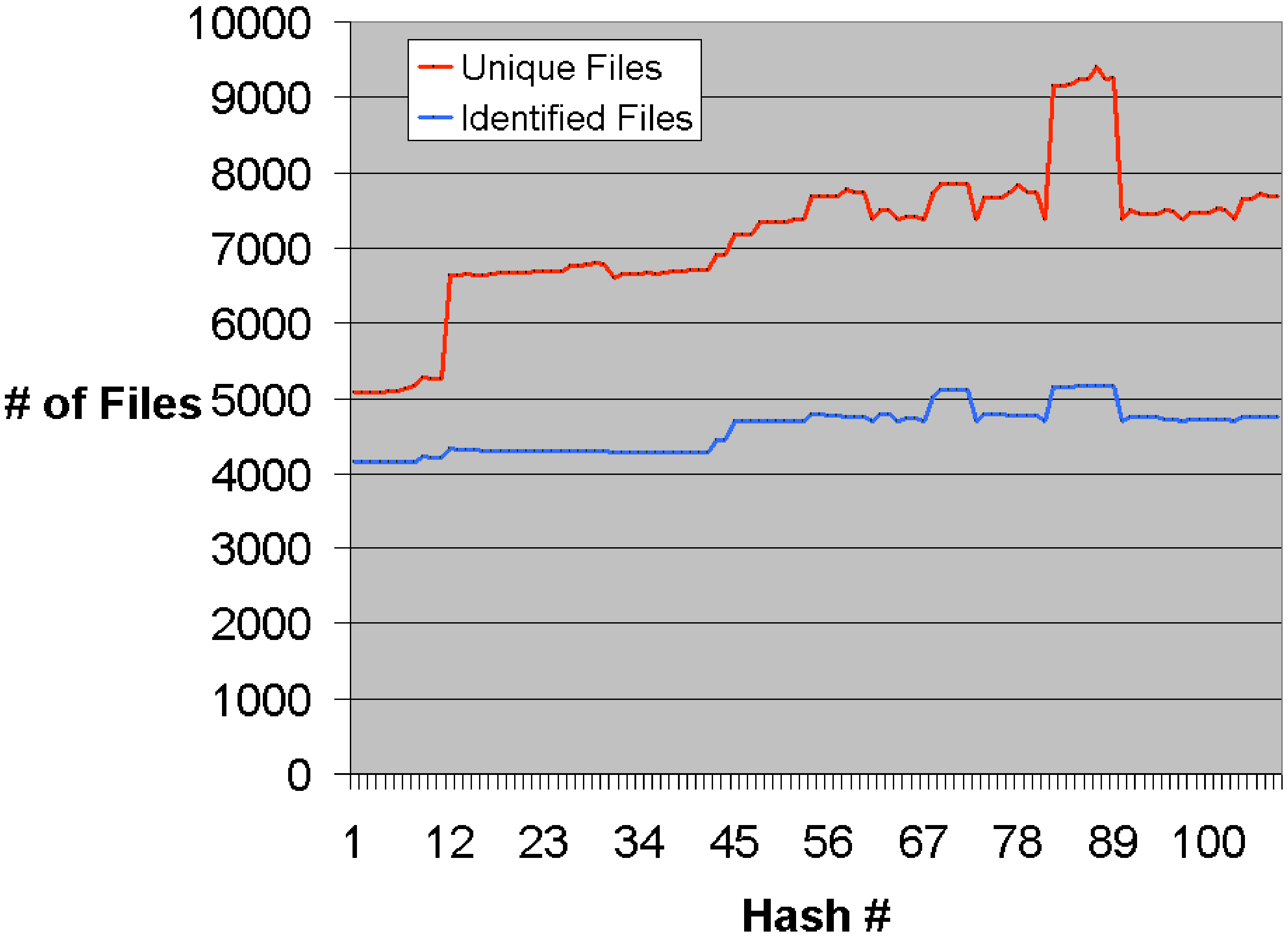
Operating System	Files Installed	Percent Identified	Files Unknown	Files in Distribution
Win 98 + Office 2K	23,464	98%	596	43,327
Win ME + Office 2K	24,112	98%	526	32,758

Compare hashes from known media to hashes of installations; best case scenario

Identification in Practice

Operating System	Files Installed	Percent Identified	Files Unknown	Notes
NIST PC #2 W2K	59,135	20%	47,124	Manager's PC email, memos
NIST PC #1 W2K	18,048	35%	11,839	"Normal" use Email, writing
NIST PC #3 WNT	14,186	54%	6,618	Researcher, Several apps
NIST PC #4 W98	16,397	55%	7,404	Researcher, Several apps
NIST PC #5 W98	34,220	75%	8,667	Project development

File Identification on a Changing Windows 2000 System



Hashing Limitations

- Eliminate known files on seized machine
- Only as good as the hashed collection
- Applicable feedback from installations
- Dynamic files - may use block size hashes
- Audio, images easily changed

NARA Research

- Use hashing process on non-classified Presidential materials
- Identify application files
- Identify duplicate files
- Access to older installed software

NARA Statistics

- 93 computer systems
  - Pre-filtered to contain only software
- 51,146 individual files
- 11,118 distinct files (SHA-1)
- 8,077 files originating in specific application(s)
- 7,610 file names
- 4,326 of 8,077 exactly match application file names
- Able to trace system “pedigree”

Contacts

Doug White

Michael Ogata

www.nsrl.nist.gov

nsrl@nist.gov

Barbara Guttman

barbara.guttman@nist.gov

Sue Ballou, Office of Law Enforcement Standards

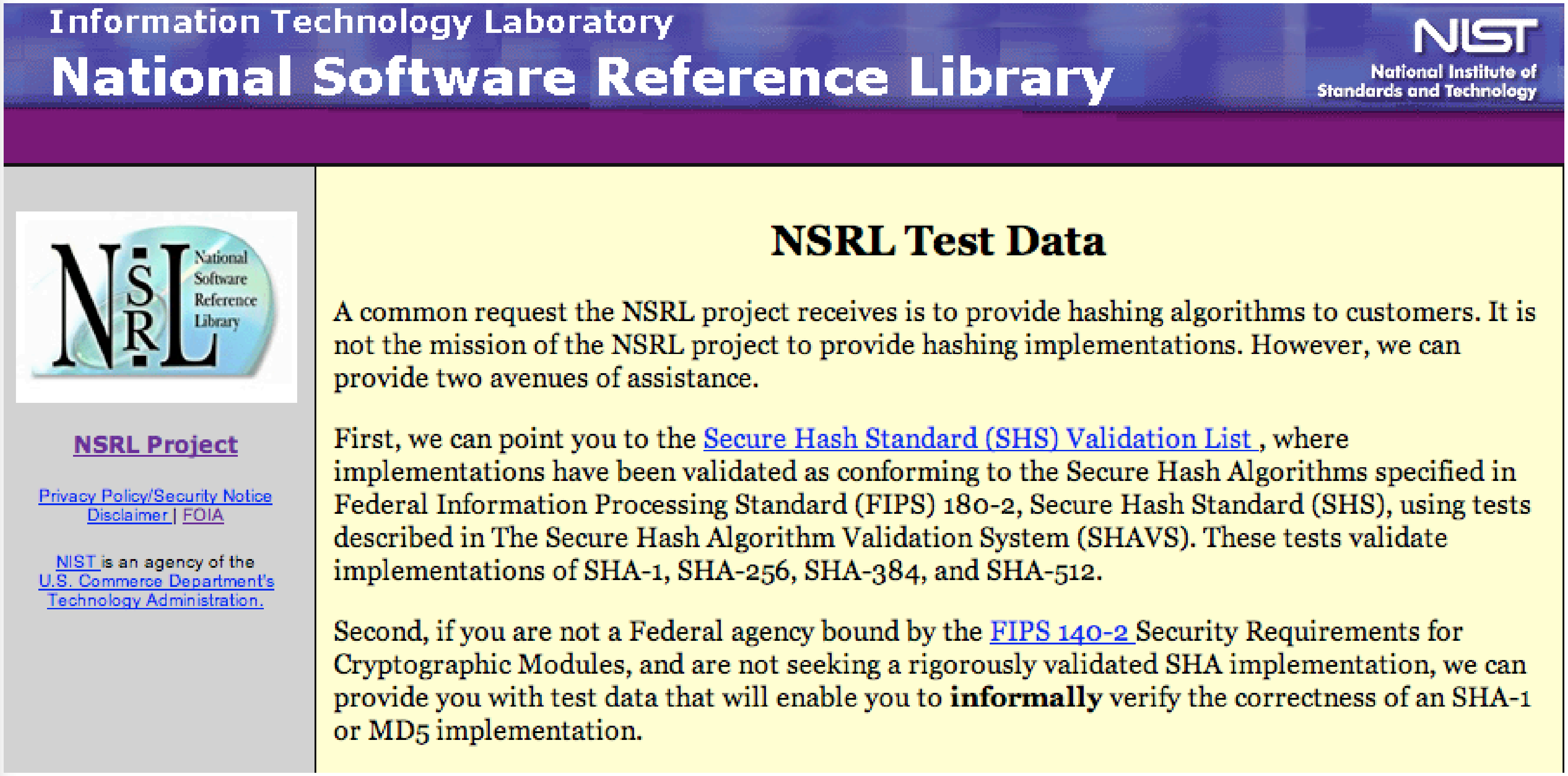
Rep. For State/Local Law Enforcement

susan.ballou@nist.gov

NSRL Software Collection

- Media in format as available to the public
- Consumer products available in stores
- Developer products available as vendor services
- Malicious software
- “Cracked” software

Hash Verification



Information Technology Laboratory

National Software Reference Library

NIST National Institute of Standards and Technology

NSRL Project

Privacy Policy/Security Notice
Disclaimer | FOIA

NIST is an agency of the U.S. Commerce Department's Technology Administration.

NSRL Test Data

A common request the NSRL project receives is to provide hashing algorithms to customers. It is not the mission of the NSRL project to provide hashing implementations. However, we can provide two avenues of assistance.

First, we can point you to the Secure Hash Standard (SHS) Validation List , where implementations have been validated as conforming to the Secure Hash Algorithms specified in Federal Information Processing Standard (FIPS) 180-2, Secure Hash Standard (SHS), using tests described in The Secure Hash Algorithm Validation System (SHAVS). These tests validate implementations of SHA-1, SHA-256, SHA-384, and SHA-512.

Second, if you are not a Federal agency bound by the FIPS 140-2 Security Requirements for Cryptographic Modules, and are not seeking a rigorously validated SHA implementation, we can provide you with test data that will enable you to **informally** verify the correctness of an SHA-1 or MD5 implementation.

www.nsrl.nist.gov/testdata

Hash Collision News

- **The NSRL project does not see any fatal ramifications from the collision announcements**.
- Details posted at http://www.nsrl.nist.gov/collision.html within 2 days
- This was not a "pre-image" attack; that is, the researchers did not identify a known file in the NSRL and attempt to generate a different file with a matching hash value.
- Nothing presented at Crypto 2004 indicated that SHA-1 has been broken
- There are known MD5 collisions and weaknesses; the NSRL data provides an MD5 to SHA-1 mapping to facilitate the migration away from MD5.
- SHA-1 will be superceded in 2010 by FIPS 180-2, Secure Hash Standard (SHA-224, 256, 384,512). The NSRL will provide a SHA-1 to SHA-256 mapping.
- The NSRL provides several hash values and the file size, and it is highly improbable that a pre-image attack will be found soon that can generate a combination of hash collisions.

Hashes

- Like a person’s fingerprint
- Uniquely identifies the file based on contents
- You can’t create the file from the hash
- Primary hash value used is Secure Hash Algorithm (SHA-1) specified in FIPS 180-1, a 160-bit hashing algorithm
  - 10^{45} combinations of 160-bit values
- “Computationally infeasible” to find two different files less than 2^{64} bits in size producing the same SHA-1
  - 2^{64} bits is one million terabytes

SHA-1 Mathematics

- Bit sequence is padded to a multiple of 512
- Messages of 16 32-bit words, n*512, n>0
- 80 logic functions are defined that accept 3 32-bit words and produce 1 32-bit word
- 80 constants defined, 5 32-bit buffers initialized
- 80 step loop:
  - Manipulate message into 80 32-bit words
  - Use shifts, functions, addition on buffers
- 160-bit SHA is string in the 5 32-bit buffers