

RAPID RECOGNITION OF BLACKLISTED FILES AND FRAGMENTS

MICHAEL MCCARRIN
BRUCE ALLEN



MANY THANKS TO:

- OSDFCCon and Basis
- Bruce Allen
- Scott Young
- Joel Young
- Simson Garfinkel

All of whom have helped with this project immensely.

PROBLEM STATEMENT:
GIVEN A LIST OF INTERESTING FILES, FIND ALL
DEVICES THAT CONTAIN AT LEAST ONE.



A COMMON SOLUTION: HASH EVERYTHING. COMPARE HASHES.

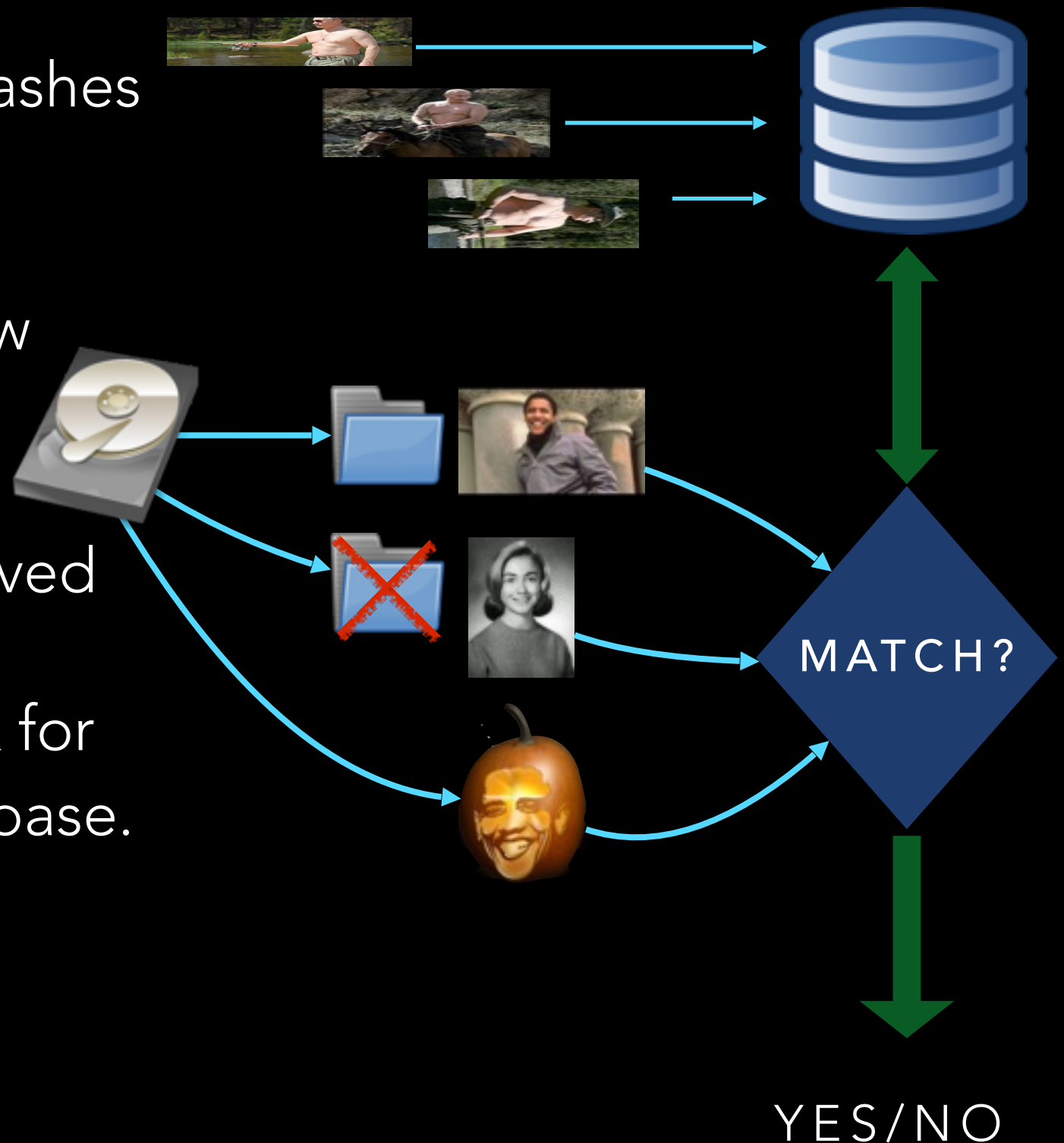
1. Keep a database of the hashes of your interesting files.

2. Extract files from each new device you see:

- Overt + undeleted + carved

3. Hash extracted files. Look for matches against the database.

4. Solved!



THERE ARE MANY COMMON SCENARIOS WHERE FILE HASHING DOESN'T WORK.

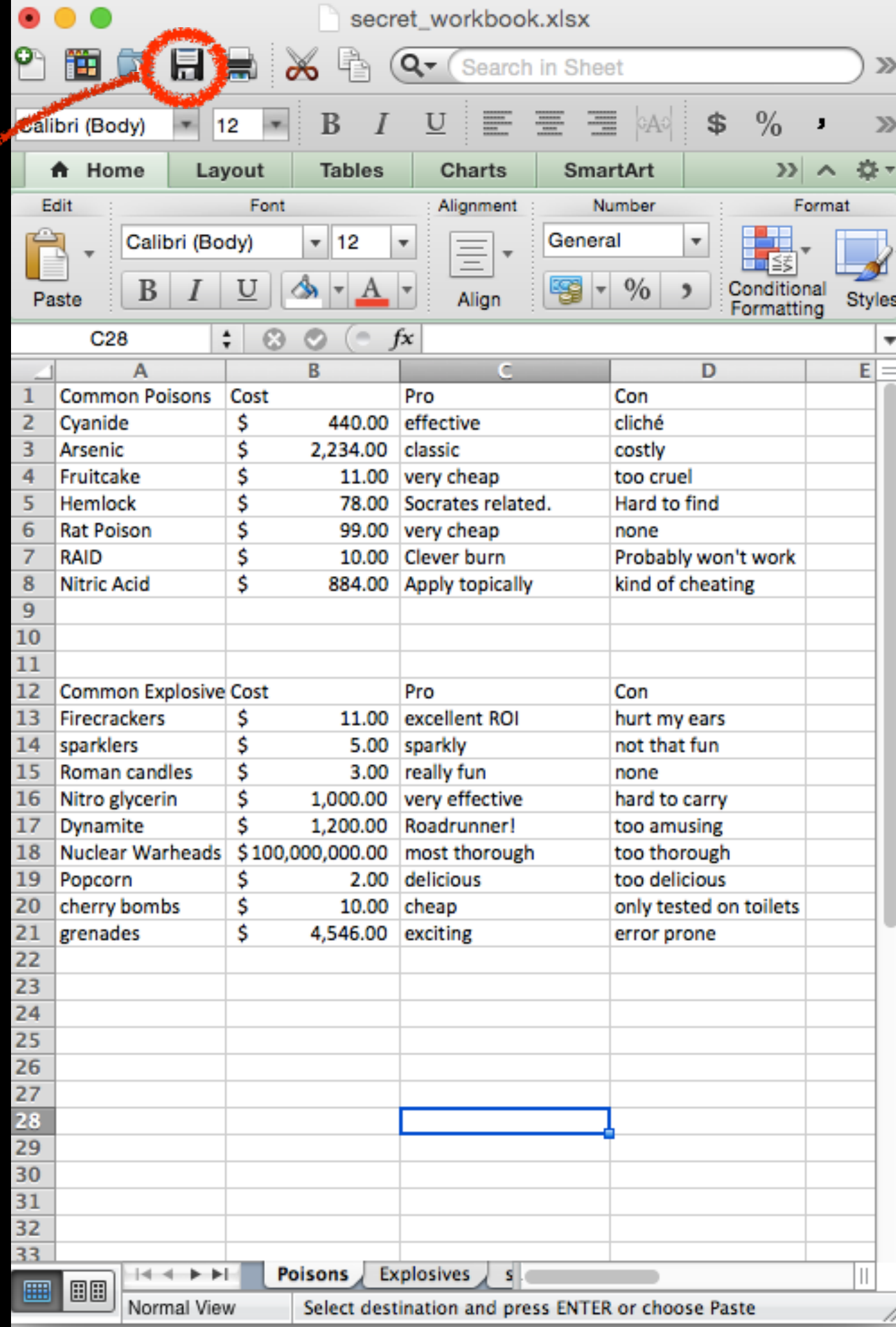
- Comparing cryptographic hashes only finds **exact** matches.
- Flip a bit, the hash changes completely.
- Everybody knows this already.
- Usually it comes up when people are worrying about mildly clever adversaries.
- **But...** there are many more mundane reasons to worry.

EXAMPLE 1: EXCEL CHANGES FILE CONTENT WHEN YOU SAVE

"GET NEW MD5"

How to get a new MD5 for
your Excel file:

- Open your file.
- Click save.
- Complete!



HERE'S THE HEX VIEW OF THE CHANGES

sw.hex (~Google Drive/NPS/gatherings/OSDFCon) - VIM3

BEFORE SAVING

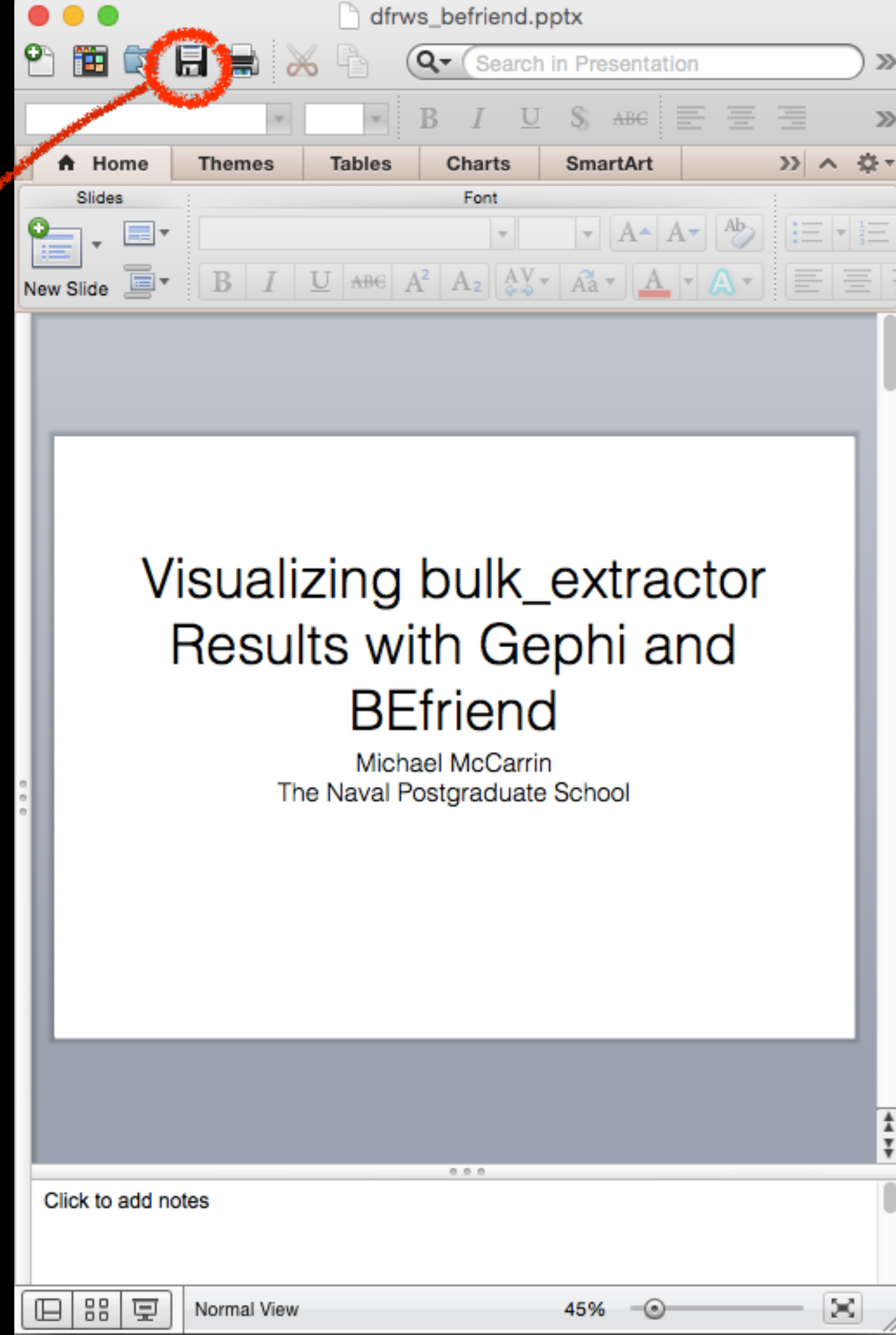
AFTER SAVING

SHIFTED FROM
HERE TO THE END
(~1.5K)

EXAMPLE 2: POWERPOINT EXHIBITS SIMILAR BEHAVIOR

"GET NEW MD5"

- Not as many changes, but one byte is enough.
- Note that the user content is 100% unchanged!



DIFF OF POWERPOINT HEX DUMPS BEFORE AND AFTER SAVING:

be_1.hex (~ /Google Drive/NPS/gatherings/OSDFCon) - VIM3																																							
+ 1 +---77421 lines: 00000000: 504b 0304 1400 0600 0800 0000 2100 ab70 P										+ 1 +---77421 lines: 00000000: 504b 0304 1400 0600 0800 0000 2100 ab70 P																													
77422	0012e6d0:	6e74	68ef	9393	f43a	1bf0	0e20	2df8	de80	nth....:....-...	77422	0012e6d0:	6e74	68ef	9393	f43a	1bf0	0e20	2df8	de80	nth....:....-...																		
77423	0012e6e0:	6ed7	913d	d645	e225	6de3	17c7	0ecf	23d5	n...=.E.%m.....#.	77423	0012e6e0:	6ed7	913d	d645	e225	6de3	17c7	0ecf	23d5	n...=.E.%m.....#.																		
77424	0012e6f0:	addc	13b2	8fd2	a775	6312	a23f	8fb6	578cuc..?.W.	77424	0012e6f0:	addc	13b2	8fd2	a775	6312	a23f	8fb6	578cuc..?.W.																		
77425	0012e700:	e613	142a	dc68	e195	c657	161a	a64d	f113	...*.h...W...M..	77425	0012e700:	e613	142a	dc68	e195	c657	161a	a64d	f113	...*.h...W...M..																		
77426	0012e710:	00a0	81d0	3f7b	df36	fde7	f900	0000	ffff?{.6.....	77426	0012e710:	00a0	81d0	3f7b	df36	fde7	f900	0000	ffff?{.6.....																		
77427	0012e720:	0300	504b	0304	1400	0600	0800	0000	2100	..PK.....!	77427	0012e720:	0300	504b	0304	1400	0600	0800	0000	2100	..PK.....!																		
77428	0012e730:	0390	fab4	6601	0000	5902	0000	1100	0801	...f...Y.....	77428	0012e730:	70d9	0c9a	6601	0000	5902	0000	1100	0801	p...f...Y.....																		
77429	0012e740:	646f	6350	726f	7073	2f63	6f72	652e	786d	docProps/core.xml	77429	0012e740:	646f	6350	726f	7073	2f63	6f72	652e	786d	docProps/core.xml																		
77430	0012e750:	6c20	a204	0128	a000	0100	0000	0000	0000	l...{(.....	77430	0012e750:	6c20	a204	0128	a000	0100	0000	0000	0000	l...{(.....																		
77431	0012e760:	0000	0000	0000	0000	0000	0000	0000	0000	77431	0012e760:	0000	0000	0000	0000	0000	0000	0000	0000																		
77432	0012e770:	0000	0000	0000	0000	0000	0000	0000	0000	77432	0012e770:	0000	0000	0000	0000	0000	0000	0000	0000																		
77433	0012e780:	0000	0000	0000	0000	0000	0000	0000	0000	77433	0012e780:	0000	0000	0000	0000	0000	0000	0000	0000																		
77434	0012e790:	0000	0000	0000	0000	0000	0000	0000	0000	77434	0012e790:	0000	0000	0000	0000	0000	0000	0000	0000																		
+ 77435	+--- 5 lines: 0012e7a0: 0000 0000 0000 0000 0000 0000 0000 0000 0000 ...																			+ 77435	+--- 5 lines: 0012e7a0: 0000 0000 0000 0000 0000 0000 0000 0000 0000 ...																		
77440	0012e7f0:	0000	0000	0000	0000	0000	0000	0000	0000	77440	0012e7f0:	0000	0000	0000	0000	0000	0000	0000	0000																		
77441	0012e800:	0000	0000	0000	0000	0000	0000	0000	0000	77441	0012e800:	0000	0000	0000	0000	0000	0000	0000	0000																		
77442	0012e810:	0000	0000	0000	0000	0000	0000	0000	0000	77442	0012e810:	0000	0000	0000	0000	0000	0000	0000	0000																		
77443	0012e820:	0000	0000	0000	0000	0000	0000	0000	0000	77443	0012e820:	0000	0000	0000	0000	0000	0000	0000	0000																		
77444	0012e830:	0000	0000	0000	0000	0000	0000	0000	0000	77444	0012e830:	0000	0000	0000	0000	0000	0000	0000	0000																		
77445	0012e840:	0000	0000	0000	0000	0000	0000	0000	0000	77445	0012e840:	0000	0000	0000	0000	0000	0000	0000	0000																		
77446	0012e850:	0000	0000	0000	0000	006c	924b	4fc3	3010L.KO.0.	77446	0012e850:	0000	0000	0000	0000	006c	925f	4fc2	3014L..0.0.																		
77447	0012e860:	84ef	48fc	07cb	f7d4	49a1	3ca2	2448	3cca	..H.....I.<.\$H<.	77447	0012e860:	c5df	4dfc	0e4d	df47	3708	4a96	3112	547c	..M..M.G7.J.1.T																		
77448	0012e870:	854a	08ca	435c	90b1	b78d	85e3	58f6	b669	.J..C\.....X..i	77448	0012e870:	91c4	28fe	892f	a6b6	17d6	d075	4d7b	61e0	..(.. /.....uM{a.																		
77449	0012e880:	f9f5	3869	09ad	e068	cdf8	f3ec	acb3	8b55	..8i...h.....U	77449	0012e880:	a7b7	1b38	31fa	d89c	d35f	cf3d	b7d9	6457	...81....._.=.dW																		
77450	0012e890:	a5c9	129c	57b5	c969	3288	2901	236a	a9ccW..i2.)#j..	77450	0012e890:	6ab2	05e7	5565	c634	e9c5	9480	1195	5466	j...Ue.4.....Tf																		
77451	0012e8a0:	3ca7	4fd3	7174	4689	476e	24d7	b581	9cae	<.0.qtF.Gn\$.....	77451	0012e8a0:	35a6	4f8b	5934	a2c4	2337	92eb	cac0	98ee	5.0.Y4...#7.....																		
77452	0012e8b0:	c1d3	8be2	f020	1336	15b5	837b	575b	70a86...{W[p.	77452	0012e8b0:	c1d3	497e	7e96	099b	8aca	c1bd	ab2c	3854	..I~.....,8T																		
77453	0012e8c0:	c093	4032	3e15	36a7	25a2	4d19	f3a2	848a	..@2>.6.%M.....	77453	0012e8c0:	e049	2019	9f0a	3ba6	05a2	4d19	f3a2	8092	.I...;...M.....																		
77454	0012e8d0:	fb41	7098	20ce	6a57	710c	4737	6796	8b4f	.Ap. .jWq.G7g..0	77454	0012e8d0:	fb5e	7098	202e	2b57	720c	47b7	6296	8b35	^p. .+Wr.G.b..5																		
77455	0012e8e0:	3e07	368c	e313	5601	72c9	91b3	1618	d99e	>.6...V.r.....	77455	0012e8e0:	5f01	ebc7	f105	2b01	b9e4	c859	038c	6c47+....Y..lG																		
77456	0012e8f0:	48b7	4829	7aa4	5d38	dd01	a460	a0a1	0283	H.H)z.]8...`....	77456	0012e8f0:	a447	a414	1dd2	6e9c	6e01	5230	d050	8241	.G....n.n.R0.P.A																		
77457	0012e900:	9e25	8384	fd7a	115c	e5ff	bdd0	293b	ce4a	..%...z.\.....);J	77457	0012e900:	cf92	5ec2	7ebc	08ae	f4ff	5e68	9513	67a9	..^~.....^h..g.																		
77458	0012e910:	e1da	8699	b671	77d9	526c	c4de	bdf2	aa37qw.RL.....7	77458	0012e910:	706f	c34c	c7b8	a76c	290e	62e7	de79	d519	po.L...l).b..y..																		
77459	0012e920:	364d	3368	8eba	1821	7fc2	5e27	77f8	dda8	6M3h...!...^'w...	77459	0012e920:	ebba	eed5	8336	46c8	9fb0	d7f9	dd63	3b6a6F.....c;j																		
77460	0012e930:	9132	6d57	0268	9149	91a2	420d	c5b3	f20b	.2mW.h.I..B.....	77460	0012e930:	a44c	d395	009a	6752	a4a8	5043	feac	fc86	.L....gR...PC....																		
77461	0012e940:	aed5	5768	947c	2cf4	e73b	acd0	7181	b523	..Wh. ,.,.;.q..#	77461	0012e940:	6bf5	191a	251f	1bbd	7e87	1d3a	2eb0	72e4	k...%.....~...r.																		
77462	0012e950:	0fe0	171a	3d69	1496	e416	6ca9	4868	9a5c=i.....l.Hh.\	77462	0012e950:	01fc	46a3	27b5	c282	dc82	2d14	094d	93e9	..F.'.....-..M..																		
77463	0012e960:	decc	9c02	2333	d643	daba	35f7	3809	9b99#3.C..5.8...	77463	0012e960:	cdd2	2930	3263	1da4	a95b	738f	f3b0	99a5	..)02c...[s.....																		
77464	0012e970:	2990	97eb	62a2	44c9	4167	ecaf	d4ba	1d2c)...b.D.Ag.....,	77464	0012e970:																											

My Clever Plan

Jo Smith

1 USE MATH

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

$$\begin{aligned}(x+y)^3 &= (x+y)^2(x+y) \\ &= (x^2+2xy+y^2)(x+y) \\ &= (x^3+2x^2y+xy^2)+(x^2y+2xy^2+y^3) \\ &= x^3+3x^2y+3xy^2+y^3\end{aligned}\tag{1.1}$$

Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies

1.1 NOW FOR THE CLEVER BIT

Lorem ipsum dolor sit amet, consectetur adipiscing elit.

$$A = \begin{bmatrix} A_{11} & A_{21} \\ A_{21} & A_{22} \end{bmatrix}\tag{1.2}$$

Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem.

My Clever Plan

Jo Smith

1 USE MATH

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

$$\begin{aligned}(x+y)^3 &= (x+y)^2(x+y) \\ &= (x^2+2xy+y^2)(x+y) \\ &= (x^3+2x^2y+xy^2)+(x^2y+2xy^2+y^3) \\ &= x^3+3x^2y+3xy^2+y^3\end{aligned}\tag{1.1}$$

Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies

1.1 NOW FOR THE CLEVER BIT

Lorem ipsum dolor sit amet, consectetur adipiscing elit.

$$A = \begin{bmatrix} A_{11} & A_{21} \\ A_{21} & A_{22} \end{bmatrix}\tag{1.2}$$

Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem.

AGAIN, HERE IS THE HEX VIEW.

IF YOU LOOK CLOSELY YOU CAN SEE THE TIME STAMP HAS CHANGED

The image shows a side-by-side comparison of two hex dump files, likely generated by a tool like VIM2. The files are named 'plan1.hex' and another file (partially visible as 'plan1.hex'). The hex dump displays data in hexadecimal format, with each line representing a memory address and its corresponding data. The data is organized into columns, with the first column showing the address and the subsequent columns showing the data in hexadecimal. The text 'plan1.hex (~/.Google Drive/NPS/gatherings/OSDFCon) - VIM2' is visible at the top of the left pane. The text 'plan1.hex (~/.Google Drive/NPS/gatherings/OSDFCon) - VIM2' is visible at the top of the right pane. The hex dump shows a comparison of two files, with the left pane showing the original file and the right pane showing a modified version. The modified version shows changes in the timestamp and the file ID. Red circles highlight the changes in the timestamp and the file ID. The timestamp is located at the end of the file, and the file ID is located at the beginning of the file. The timestamp is '4817-04'00'' and the file ID is '7094817-04'00''. The modified version shows a timestamp of '5245-04'00'' and a file ID of '7095245-04'00''. The text 'FILE ID?' is visible at the bottom of the image.

```
1 +---2338 lines: 00000000: 2550 4446 2d31 2e35 0a25 d0d4 c5d8 0a33 %P + 1 +---2338 lines: 00000000: 2550 4446 2d31 2e35 0a25 d0d4 c5d8 0a33
339 00009220: 6473 7472 6561 6d0a 656e 646f 626a 0a33 dstream.endobj.3 2339 00009220: 6473 7472 6561 6d0a 656e 646f 626a 0a33 dstream.endobj.
340 00009230: 3120 3020 6f62 6a0a 3c3c 0a2f 5072 6f64 1 0 obj.<<./Prod 2340 00009230: 3120 3020 6f62 6a0a 3c3c 0a2f 5072 6f64 1 0 obj.<<./Pro
341 00009240: 7563 6572 2028 7064 6654 6558 2d31 2e34 ucer (pdfTeX-1.4 2341 00009240: 7563 6572 2028 7064 6654 6558 2d31 2e34 ucer (pdfTeX-1.
342 00009250: 302e 3136 290a 2f43 7265 6174 6f72 2028 0.16)./Creator ( 2342 00009250: 302e 3136 290a 2f43 7265 6174 6f72 2028 0.16)./Creator
343 00009260: 5465 5829 0a2f 4372 6561 7469 6f6e 4461 TeX)./CreationDa 2343 00009260: 5465 5829 0a2f 4372 6561 7469 6f6e 4461 TeX)./CreationD
344 00009270: 7465 2028 443a 3230 3135 3130 3237 3039 te (D:20151027 2344 00009270: 7465 2028 443a 3230 3135 3130 3237 3039 te (D:20151027
345 00009280: 3438 3137 2d30 3427 3030 2729 0a2f 4d6f 4817-04'00')./Mo 2345 00009280: 3532 3435 2d30 3427 3030 2729 0a2f 4d6f 5245-04'00')./Mo
346 00009290: 6444 6174 6520 2844 3a32 3031 3531 3032 dDate (D:2015102 2346 00009290: 6444 6174 6520 2844 3a32 3031 3531 3032 dDate (D:201510
347 000092a0: 3730 3934 3831 372d 3034 2730 3027 290a 7094817-04'00'). 2347 000092a0: 3730 3935 3234 352d 3034 2730 3027 290a 7095245-04'00'
348 000092b0: 2f54 7261 7070 6564 202f 4661 6c73 650a /mapped./False. 2348 000092b0: 2f54 7261 7070 6564 202f 4661 6c73 650a /mapped./False
349 000092c0: 2f50 5445 582e 4675 6c6c 6261 6e6e 6572 /PTeX.Fullbanner 2349 000092c0: 2f50 5445 582e 4675 6c6c 6261 6e6e 6572 /PTeX.Fullbanne
350 000092d0: 2028 5468 6973 2069 7320 7064 6654 6558 (This is pdfTeX 2350 000092d0: 2028 5468 6973 2069 7320 7064 6654 6558 (This is pdfTeX
351 000092e0: 2c20 5665 7273 696f 6e20 332e 3134 3135 , Version 3.1415 2351 000092e0: 2c20 5665 7273 696f 6e20 332e 3134 3135 , Version 3.14
352 000092f0: 3932 3635 2d32 2e36 2d31 2e34 302e 3136 9265-2.6-1.40.16 2352 000092f0: 3932 3635 2d32 2e36 2d31 2e34 302e 3136 9265-2.6-1.40.
353 00009300: 2028 5465 5820 4c69 7665 2032 3031 3529 (TeX Live 2015) 2353 00009300: 2028 5465 5820 4c69 7665 2032 3031 3529 (TeX Live 2015
354 +--- 95 lines: 00009310: 206b 7061 7468 7365 6120 7665 7273 696f kp + 2354 +--- 95 lines: 00009310: 206b 7061 7468 7365 6120 7665 7273 696f
449 00009900: 2b0a 656e 6473 7472 6561 6d0a 656e 646f +.endstream.endo 2449 00009900: 2b0a 656e 6473 7472 6561 6d0a 656e 646f +.endstream.endo
450 00009910: 626a 0a33 3220 3020 6f62 6a0a 3c3c 0a2f bj.32 0 obj.<<./ 2450 00009910: 626a 0a33 3220 3020 6f62 6a0a 3c3c 0a2f bj.32 0 obj.<<./
451 00009920: 5479 7065 202f 5852 6566 0a2f 496e 6465 Type /XRef./Inde 2451 00009920: 5479 7065 202f 5852 6566 0a2f 496e 6465 Type /XRef./Inde
452 00009930: 7820 5b30 2033 335d 0a2f 5369 7a65 2033 x [0 33]./Size 3 2452 00009930: 7820 5b30 2033 335d 0a2f 5369 7a65 2033 x [0 33]./Size
453 00009940: 330a 2f57 205b 3120 3220 315d 0a2f 526f 3./W [1 2 1]./Ro 2453 00009940: 330a 2f57 205b 3120 3220 315d 0a2f 526f 3./W [1 2 1]./Ro
454 00009950: 6f74 2033 3020 3020 520a 2f49 6e66 6f20 st 30 0 R./Intr 2454 00009950: 6f74 2033 3020 3020 520a 2f49 6e66 6f20 st 30 0 R./Intr
455 00009960: 3331 2030 2052 0a2f 4944 205b 3c30 4137 31 0 R./ID [<0 2455 00009960: 3331 2030 2052 0a2f 4944 205b 3c30 4137 31 0 R./ID [<0
456 00009970: 4542 3845 3846 3132 3546 3644 4533 4536 EB8E8F125F6DE3E6 2456 00009970: 3932 3130 4337 4333 3439 4436 4645 4341 9210C7C349D6FE
457 00009980: 3943 3744 3431 3830 4542 3042 393e 2031 9C7D4180EB0B9> < 2457 00009980: 3834 4142 4644 4237 3136 3130 423e 2031 84ABFDB71610B>
458 00009990: 3137 4345 4238 4538 4631 3235 4636 4445 17CEB8E8F125F6DE 2458 00009990: 3041 3739 3231 3043 3743 3334 3944 3641 0A79210C7C349D
459 000099a0: 3345 3639 4337 4434 3138 3045 4230 4239 3E69C7D4180EB0B9 2459 000099a0: 4543 4238 3441 4246 4442 3731 3631 3042 ECB84ABFDB7161
460 000099b0: 3e5d 0a2f 4c65 6e67 7468 2039 3320 2020 >1./Length 93 2460 000099b0: 3e5d 0a2f 4c65 6e67 7468 2039 3320 2020 >1./Length 93
461 000099c0: 2020 2020 200a 2f46 696c 7465 7220 2f46 7465 7220 2f46 2461 000099c0: 2020 2020 200a 2f46 696c 7465 7220 2f46 7465 7220 2f46
462 000099d0: 6c61 7465 4465 636f 6465 0a3e 3e0a 7374 lateDecode.>>.st 2462 000099d0: 6c61 7465 4465 636f 6465 0a3e 3e0a 7374 lateDecode.>>.st
463 000099e0: 7265 616d 0a78 da15 c9b9 0d80 5010 4341 ream.x.....P.CA 2463 000099e0: 7265 616d 0a78 da15 c9b9 0d80 5010 4341 ream.x.....P.C
464 000099f0: 9bfb 73df 1219 2109 8280 ba68 8302 20a2 ..s.....h... 2464 000099f0: 9bfb 73df 1219 2109 8280 ba68 8302 20a2 ..s.....h...
465 00009a00: 3e0a 016f 307a 5a2f 80cf 83a3 80a8 a0b6 >..o0zZ/..... 2465 00009a00: 3e0a 016f 307a 5a2f 80cf 83a3 80a8 a0b6 >..o0zZ/.....
466 +--- 6 lines: 00009a10: d249 2d15 efdd b65e 7cf1 e81a bb1b 2925 .I- + 2466 +--- 6 lines: 00009a10: d249 2d15 efdd b65e 7cf1 e81a bb1b 2925
```

FILE ID?

CONCLUSION: MANY APPLICATIONS CHANGE FILE-LAYER CONTENT.

- The point is not that your investigation relies on Excel files or finding PowerPoint presentations.
- The point is: **applications don't care** about forensic integrity.
- The changes above are **transparent** to the user.

COROLLARY: WE MAY BE MISSING EVIDENCE.

Other things you won't find with file hashes:

- Files that are **deleted** and partially overwritten.
- Slightly **edited** files.
- **Corrupted** files in overt space.
- Badly **carved** files.
- **Fragments** of files in unallocated space.

WORSE: THESE ARE CASES WHERE THE SUSPECT IS NOT EVEN TRYING.

- At most, they require someone to click save or empty the recycling bin.
- The state produced by **default behavior** is much more concerning than any single clever person.
- Many reasons to improve this process, but the most compelling is **economic:**
 - **Solving the problem should be cost less than creating it.**

AN ARGUMENT FOR "APPROXIMATE MATCHING" STRATEGIES:

- Not a fancy thing: we often just want to find files with the **exact same** user-level content.
- Fundamentally, the approach is not that different:
 1. Cut the object into pieces.
 2. Hash the pieces.
 3. Count matching hashes.

THIS PROBLEM HAS MANY KNOWN SOLUTIONS.

Numerous sophisticated matching schemes:

- Rabin fingerprinting, ssdeep, sdhash, frag_find, sector hashing...
- Lots of good stuff—we try to build on it.
- A skilled examiner could find everything we mentioned.
- Many groups have in-house solutions.
- **Why write a new tool?**

REASONS FOR A NEW TOOL

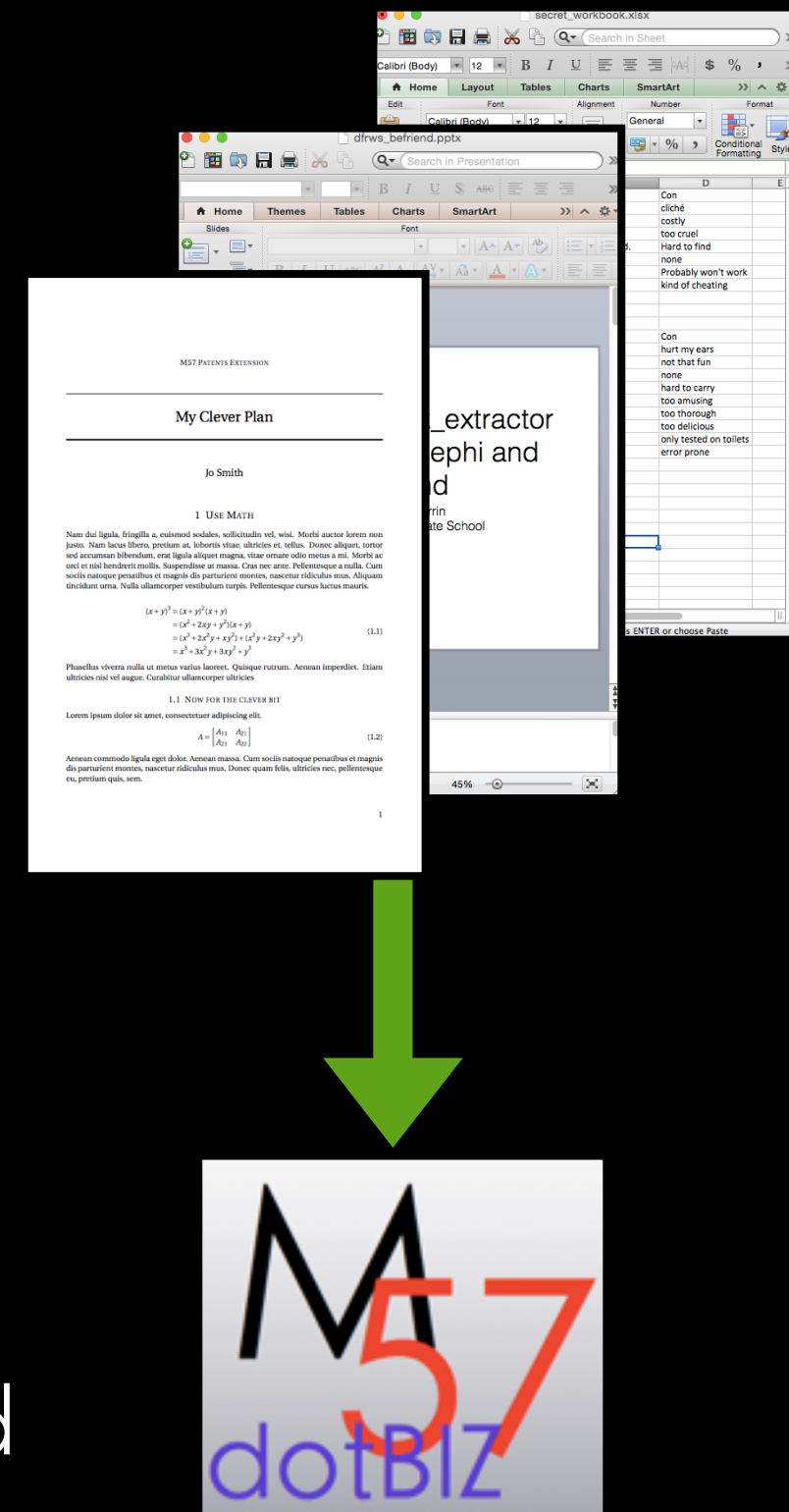
1. Nothing on the previous slide will change the economic problem.
2. Make already developed techniques more available.
3. Getting it right is harder than it first seems.
 - In particular, the feature selection problem has not been solved.

OUR GOALS: TRIVIAL, SCALABLE, COMPREHENSIBLE.

- **Trivial:** It has to be easier to discover targets than it is to obscure them.
- **Scalable:** Should be able to search for millions of files at once.
- **Comprehensible:** operation of tool and interpretation of results must be intuitive.

SECTORSCOPE DEMO

- Target: the 3 files shown above.
- Media: a modified disk image from the m57 patents scenario (exFAT).
- Files were added to a subfolder on the image then deleted.
- Subfolder was deleted then replaced so it is not trivially easy to recover through Autopsy / TSK.



STEP 1: CREATE A BLACKLIST

- We store our blacklist in hashdb: a fast, lightweight key-value store we designed for this project.
- Blacklist files are hashed in 512-byte chunks and imported into the database using the hashdb bulk_extractor scanner.
- At the moment, this requires a command line.

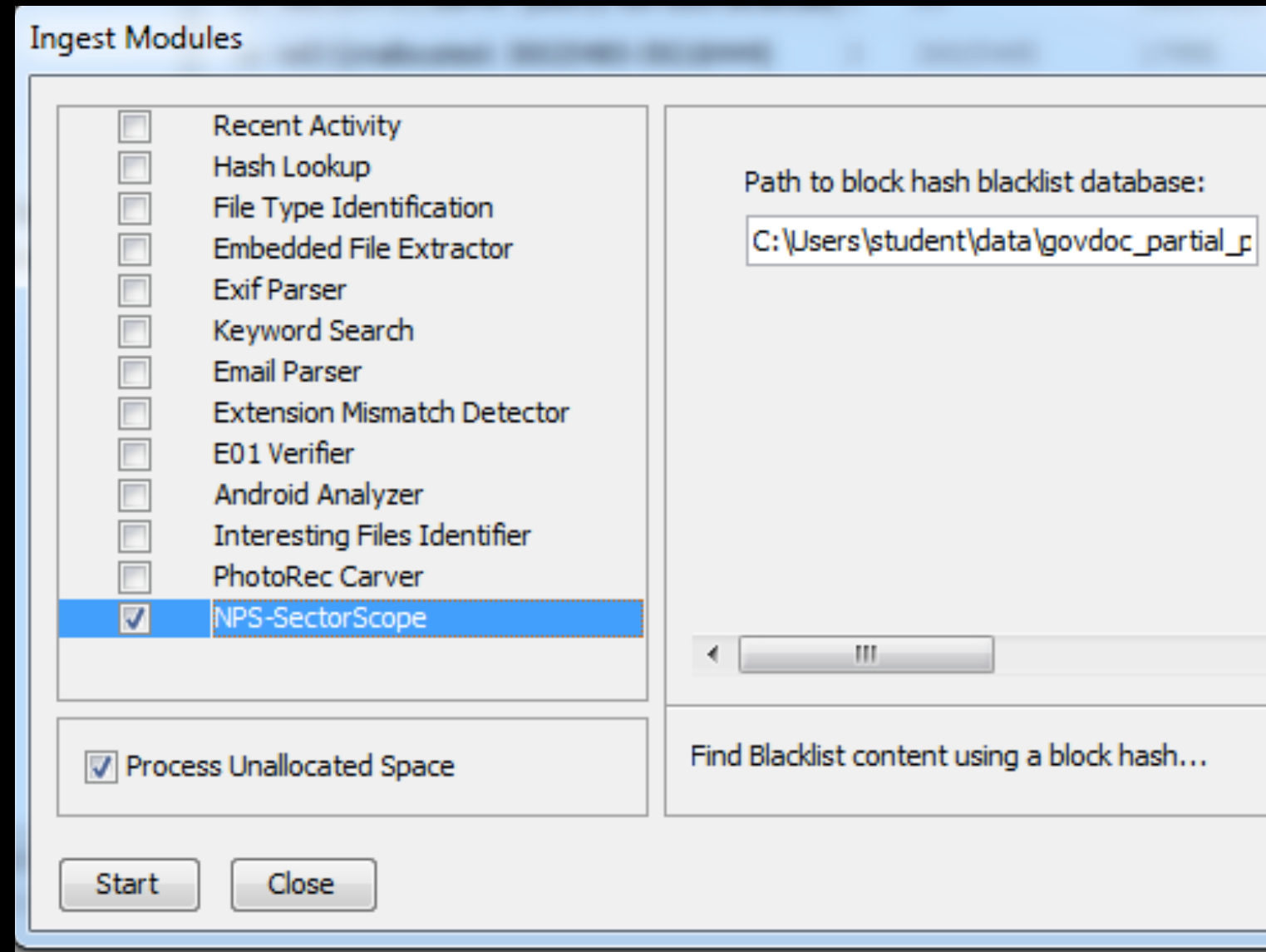


STEP 2: SCAN MEDIA

- Could do this at the command line...

- Let's use Autopsy.
- Install SectorScope and the SectorScope plugin.

- Add your disk image to the case.



- Run SectorScope ingest module.

STEP 3: ANALYZE THE RESULTS.

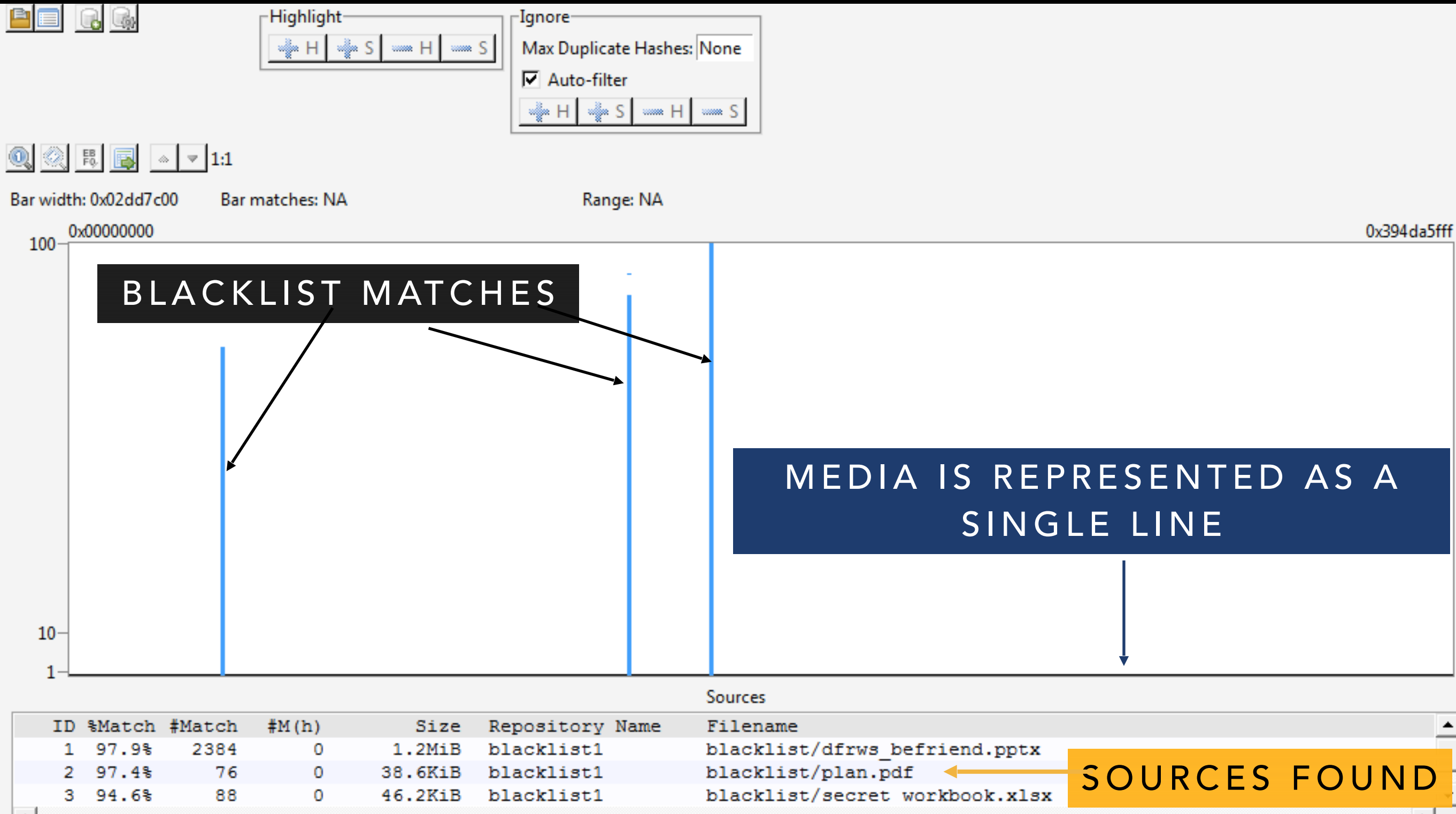
Open SectorScope from the Reports listing:

The screenshot shows the Autopsy 3.1.3 interface. The left sidebar contains a tree view with the following categories: Data Sources, Views, Results, Hashset Hits, E-Mail Messages, Interesting Items, Tags, and Reports. The Reports category is selected and highlighted in blue. The main pane displays a 'Directory Listing' of reports. It includes a table with the following data:

Source Module Name	Report Name	Created Time	Report File Path
NPS-SectorScope	jo-2009-12-10.E01	2015-10-28 01:40:30 PDT	C:\Users\student\autopsy\osdfcon_1\ModuleOutput\NPS-Sector

Below the table, there are tabs for 'Hex', 'Strings', 'Metadata', 'Results', 'Text', and 'Media'. The 'Results' tab is currently selected.

RESULTS FROM OUR 3 FILE BLACKLIST.



PAN AND ZOOM TO NAVIGATE THE MEDIA.

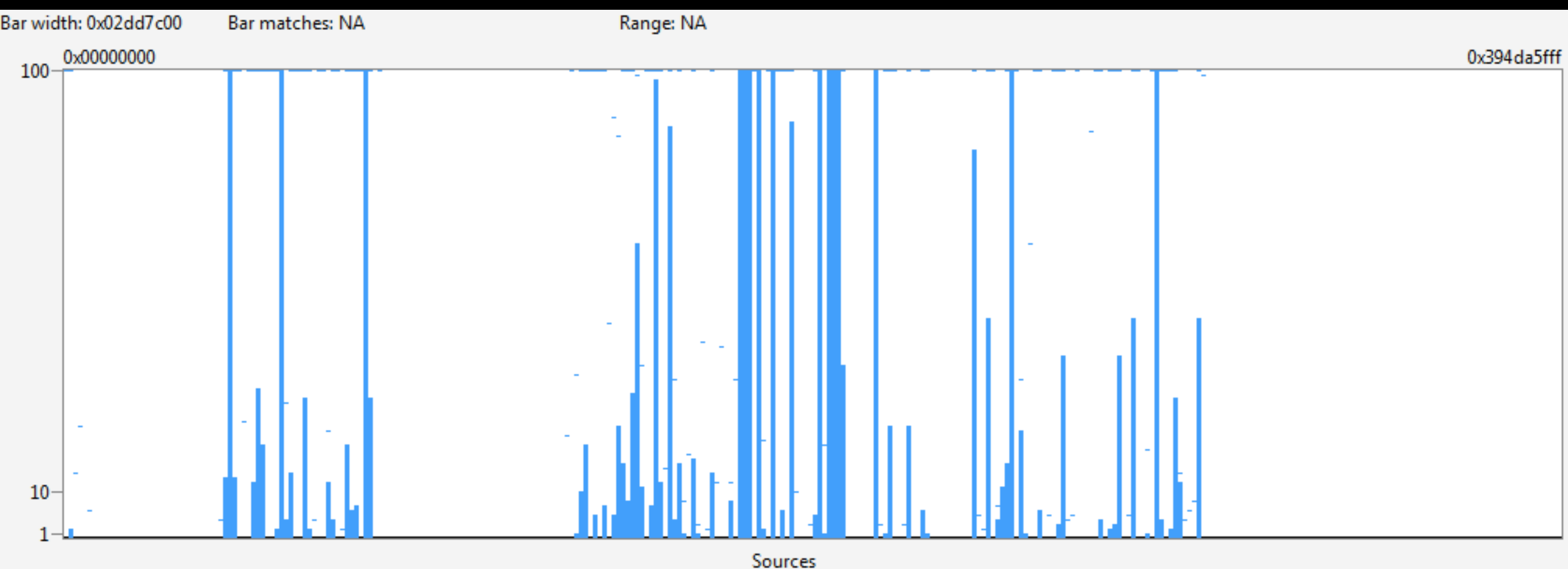
- The list is sorted by percent of file found.
- Highlight a file to see where it is.
- Ignore it to get it off the screen.
- Select a range to see which sources are there.
- Adjust the resolution.
- Open a hex-dump of the data.

LET'S LOOK AT A NON-
TRIVIAL EXAMPLE.

THE SCAN RUNS BULK_EXTRACTOR'S HASHDB MODULE TO QUERY THE BLACKLIST.

- Scanning took 109 seconds to scan against 20,772,713 hashes
- Lenovo Thinkpad with 8G of RAM, and a Dual Core i7-5600U CPU (2.6 GHz)
- You can beat this speed with faster hardware
- Compare to 99 seconds to scan against 2,607 hashes: a factor of a thousand takes 10 seconds longer.

THINGS GET MORE COMPLICATED WITH 20 MILLION HASHES.

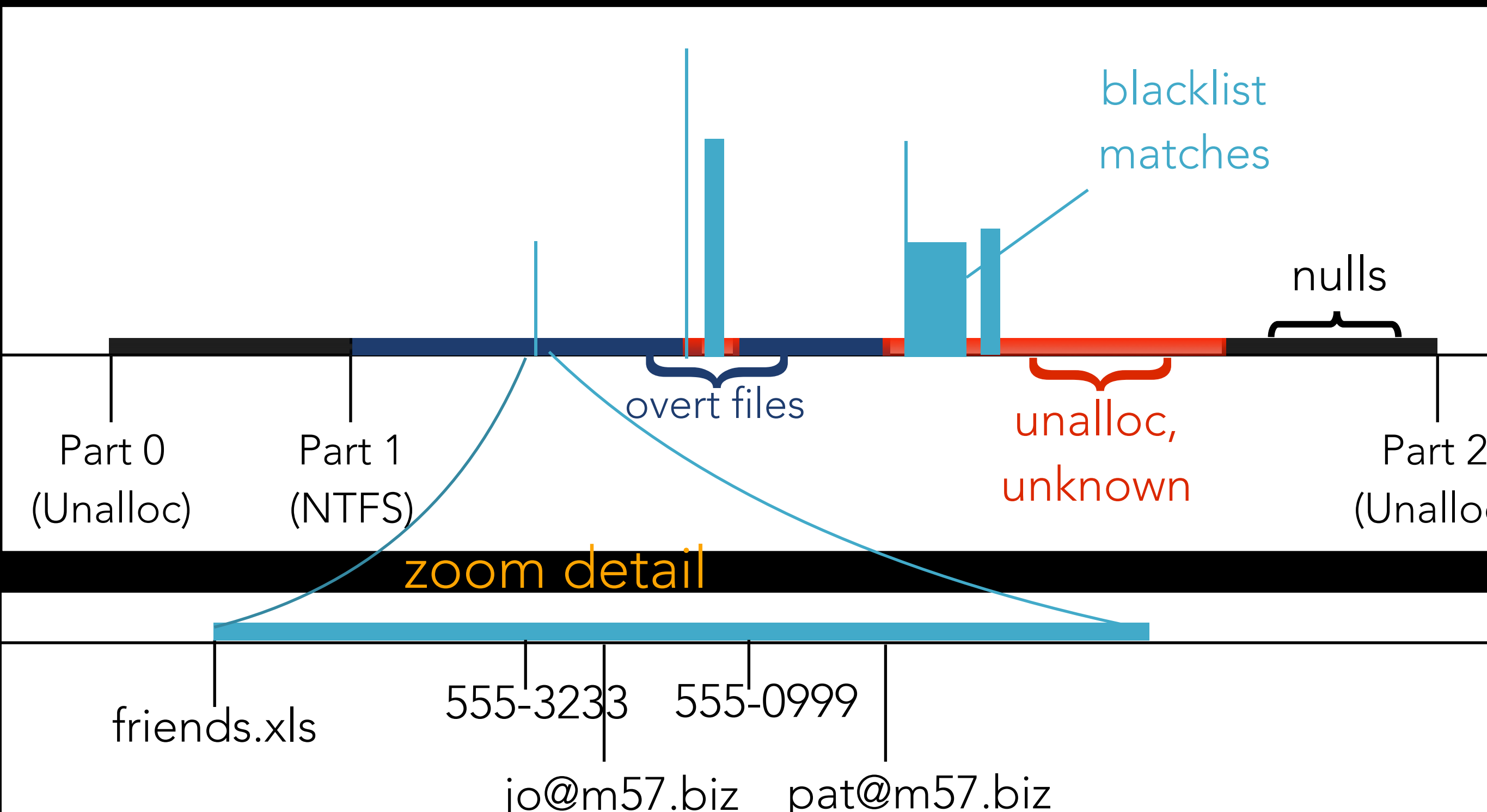


- Added from govdocs—not all high entropy.
- Need to use noise reduction to filter out false positives.

NOISE REDUCTION IS TRICKY

- Down at the level of a 512-byte block, it's not really clear what a "match" tells us.
- Some blocks are extremely rare. (2^{4096} is a lot of possibilities).
- Some blocks show up all over the place.
- We use an algorithm published in a recent DFRWS paper: Garfinkel, S. L., & McCarrin, M. (2015). Hash-based carving: Searching media for complete files and file fragments with sector hashing and hashdb. Digital Investigation, 14, S95-S105.

FUTURE WORK: ANNOTATE WITH TSK AND BULK_EXTRACTOR DATA



YOU NEED TO KNOW WHAT YOU DON'T KNOW.

- Few forensics tools will tell you this.
- How much of the image contains data that you know nothing about at all?
- Our end goal is to bring available tools together to map out every sector.

Questions?

Contact:

Michael McCarrin
mrmccarr@nps.edu

Bruce Allen
bdallen@nps.edu

SectorScope on GitHub:

<https://github.com/NPS-DEEP/NPS-SectorScope>