



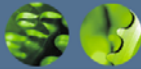
Advanced File Carving With FTimes

CEIC 2007
May 8, 2007

KoreLogic, Inc:

Andy Bair
pab-ceic@korelogic.com

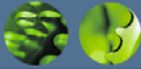
Jay Smith
jsmith-ceic@korelogic.com



Overall Agenda

- Basic Section
 - Introduction - File Carving Overview
 - Background – Terminology & Methodology
 - FTimes – Introduction to FTimes Map Mode
 - Lab #1 – Extract JPEG Image From A Tar Ball
- Advanced Section
 - XMagic – Fundamentals
 - XMagic & Carving
 - Lab #2 – Extract Word Doc From A Packet Capture
 - Conclusions

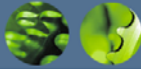
© Copyright 2003-2007 KoreLogic, Inc. All Rights Reserved 2



Advanced Data Carving

- How can you carve out a file when you do not know the SOF and EOF?
- Define and using XMagic to assist in carving out blocks of data.

© Copyright 2003-2007 KoreLogic, Inc. All Rights Reserved 3



Agenda

- **File Carving Review**
- XMagic – Fundamentals
- XMagic & Carving
- Lab #2 – Extract Word Doc From A Packet Capture
- Conclusions


© Copyright 2003-2007 KoreLogic, Inc. All Rights Reserved 4



File Carving Review

- Many file types have well-known values or Magic numbers in the first bytes of the file header
- Typical file carvers
 - Identify specific types of file headers and/or footers
 - Carve out blocks between these two boundaries
 - Stop carving after a user-specified or set limit has been reached
- Unfortunately, not all file types have a standard footer signature, so determining the EOF can be difficult -- thus the need for limits
- What can be leveraged to help us XMagic

© Copyright 2003-2007 KoreLogic, Inc. All Rights Reserved 5




Reality Check! #3

Johnny Badguy was reading copies of the CIO's email messages through his email client. He "thought" he deleted the files, but the deleted files were still on the drive and had partially been overwritten.

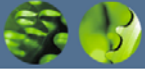
....hmmm....

Using data carving, you can extract any remaining portions of the email message as evidence of the theft.

Not being able to carve complete files is not always a bad thing.



© Copyright 2003-2007 KoreLogic, Inc. All Rights Reserved 6



Agenda

- File Carving Review
- **XMagic – Fundamentals**
- XMagic & Carving
- Lab #2 – Extract Word Doc From A Packet Capture
- Conclusions

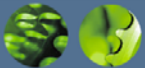


XMagic Fundamentals – Magic Review

<http://ftimes.sourceforge.net/FTimes/XMagic.shtml>

- Use XMagic to develop statistics (entropy, averages, %-ctypes, ...)
- To understand XMagic, requires knowledge of the file(1) command and magic(5)
- Magic number – special constant (traditionally) used to identify a particular type of file (e.g., tcpdump magic is 0xa1b2c3d4)
- file(1) command – determines file types using magic numbers
- Typical file(1) command usage:

```
$ file ftimes.zip
ftimes.zip: Zip archive data, at least v2.0 to extract
```



XMagic Fundamentals – File and Magic Example

```
$ file ftimes.zip
```

```
magic:
0 string PK\003\004 Zip archive data
>4 byte 0x09 \b, at least v0.9 to extract
>4 byte 0x0a \b, at least v1.0 to extract
>4 byte 0x0b \b, at least v1.1 to extract
>4 byte 0x14 \b, at least v2.0 to extract
```

```
ftimes.zip:
0 1 2 3 4 5 6 7
50 4b 03 04 14 00 00 00 |PK.....|
```

```
ftimes.zip: Zip archive data, at least v2.0 to extract
```



XMagic Fundamentals – Comparison to Magic

- Split operator/value pair into separate fields
- Supports
 - Regular expression Magic via Perl Compatible Regular Expressions (PCRE)
 - Block-based entropy calculations
 - Block-based average calculations
 - Block-based percent calculations for ctype(3) character classes
 - Block-based hash calculations (MD5, SHA1, and SHA256)
 - Several different test operators for all of its block-based tests



XMagic Fundamentals – Comparison to Magic

- Test operator/value (if test operator not supplied in standard Magic, the implied operator is '=')
- ```
Magic: 0 string \037\235 compress'd data
XMagic: 0 string = \037\235 compress'd data
```
- Place holder when the test value is to be ignored:
- ```
Magic: >6 byte x type %c
XMagic: >6 byte x - type %c
```



Extending XMagic

- Convert a series of string/[Bbc] tests to the equivalent regexp test:

```
Magic: 0 string/B = \=pod\n Perl POD document
Magic: 0 string/B = \n\=pod\n Perl POD document
Magic: 0 string/B = \=head1\ Perl POD document
Magic: 0 string/B = \n\=head1\ Perl POD document
Magic: 0 string/B = \=head2\ Perl POD document
Magic: 0 string/B = \n\=head2\ Perl POD document
XMagic: 0 regexp =~ ^\n?(?:pod\n|head[12]) Perl POD document
```



Extending XMagic (2)

- Convert a search/<number> test to an equivalent regexp:<number> test

```

Magic: 0 search/20 = foo The venerable %s document
XMagic: 0 regexp:20 == foo The venerable %s document

```
- Block-based test types to harvest various topographical information:

```

XMagic: 0 byte x - 512
XMagic: >40 row_entropy_1:512 x - \b,%f
XMagic: >40 row_average_1:512 x - \b,%f
XMagic: >40 percent_ctype_alnum:512 x - \b,%f
XMagic: >40 sha1:512 x - \b,%s

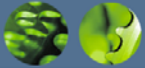
```



Reality Check! #4

We have a proprietary file format or database. The subject of our investigation has taken a copy and stashed it in the slack space of the drive.

Using FTimes along with XMagic you can create a custom XMagic signature that digs through the subject image looking for the stashed files.



Agenda

- File Carving Review
- XMagic – Basics
- XMagic & Carving**
- Lab #2 – Extract Word Doc From A Packet Capture
- Conclusions

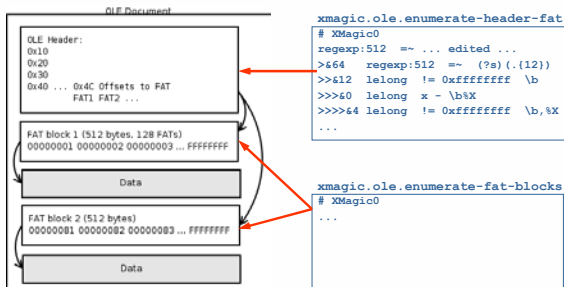


Background - Terminology

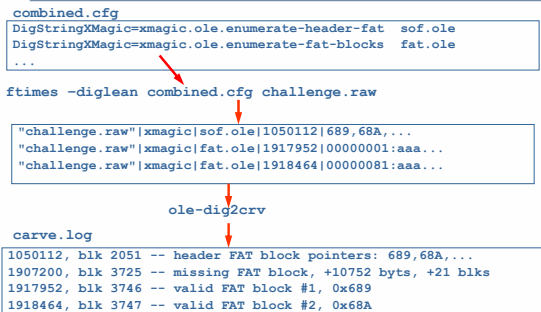
- SOF/EOF** - start/end of file
- SOI/EOI** - start/end of image
- FAT** - file allocation table (also referred to as SAT blocks)
- OLE** - Object Linking and Embedding, Microsoft's compound documents
- Magic** - constant used to identify file or data type (also known as file typing)
- XMagic** - Extended Magic
- Zero-storage** - Zero additional space needed to carve out files. Information about sector information is used to keep track of location of data to be carved



XMagic & Carving - OLE Documents



XMagic & Carving: Enumerate File Structure





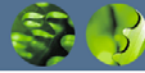
XMagic & Carving: Collect Entropy, Averages, % Character Type

```
stats-512.cfg.xmagic
# XMagic
0 byte x - 512
>%0 row_entropy_1:512 x - \b|\f
>%0 row_entropy_2:512 x - \b|\f
...
```

```
stats-512.cfg
Basename=-
DigStringXMagic=stats-512.cfg.xmagic stats-512
...
```

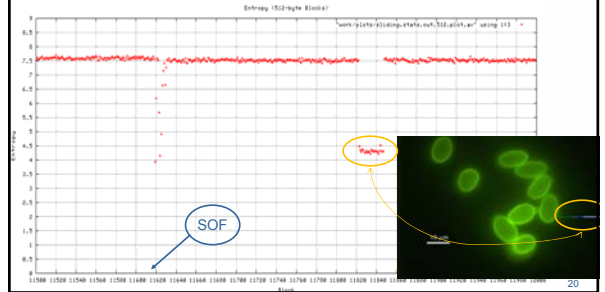
```
ftimes -diglean stats-512.cfg challenge.raw
```

```
name|type|tag|offset|string
"challenge.raw"|xmagic|stats-512|0|512|4.656387|7.282739|...
"challenge.raw"|xmagic|stats-512|512|512|4.667385|7.244524|...
...
```



XMagic & Carving: Plot Sliding Statistics

This sliding entropy graph shows the start of the JPEG image at block 11619. The graph also reveals a drop in entropy at block 11820.



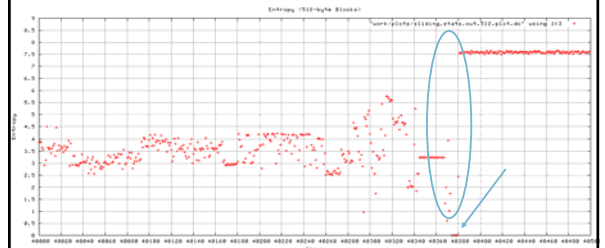
XMagic & Carving: Statistics

- Sliding entropy and average
 - Detect data stream edges
 - Usually block boundary
- Sliding entropy classifies data types
 - Entropy 4-6 = Likely TEXT and HTML blocks
 - Entropy 7-8 = Likely compressed or encrypted such as ZIP and JPEG blocks



Entropy Example

- An entropy graph showing drastic changes in 1-byte entropy around block 40380. It's easy to see where the entropy changes and that it typically occurs on block boundaries.



Reality Check! #5

A subject has been hiding his data within unused blocks, but not necessarily on sector boundaries.

You can use XMagic with sliding entropy and the graphs to determine if there are patterns in the data.



Agenda

- File Carving Review
- XMagic – Basics
- XMagic & Carving
- Lab #2 – Extract Word Doc From A Packet Capture
- Conclusions



Agenda

- File Carving Review
- XMagic – Basics
- XMagic & Carving
- Lab #2 – Extract Word Doc From A Tar Ball
- **Conclusions**



Conclusions

- There is “no one tool to rule them all”
- Using multiple tools can lead to more information that needs to be analyzed.
- Using a carved data block validator can help to reduce false positives.
- If that final piece of the case is eluding you, consider using techniques such as sliding entropy to help look for irregular fluctuations -- these may be indicators of file boundaries.
- The attacker or subject could be using data hiding techniques that have the data residing fragmented or in odd locations on the disk.
- Not all situations can be solved with a point and click.
- You need to know the fundamentals of data carving before you adopt a tool that does it for you.



Links and Info

<http://www.korelogic.com>

<http://ftimes.sourceforge.net>

http://www.korelogic.com/Resources/Projects/dfrws_challenge_2006/

Email: pab_ceic@korelogic.com
jsmith_ceic@korelogic.com