OSINT with FOCA 2.6

By Russ McRee – ISSA member, Puget Sound (Seattle), USA Chapter

Join the Discussion
Connect

Prerequisites

Windows host or *nix with Wine

Current .NET framework, 3.5 or higher

It's been an interesting few months. First, we learned of vast quantities of diplomatic cables exposed via Wikileaks, and then we saw HBGary compromised by Anonymous, revealing some flagrant missteps on the part of the HBGary Federal division. I'll leave the endless debate over the rights and wrongs of both these incidents alone entirely and instead direct your attention to a tool useful for Open Source Intelligence or OSINT. Given exposures of the scale and scope of the above mentioned headline grabbers, I propose two sound tactical reasons to practice OSINT: offense and defense. It's simple. Forgive me this one use of well-referenced Sun Tzu wisdom: "If ignorant both of your enemy and yourself, you are certain to be in peril."

I use OSINT tactics in an approved manner most often on behalf of those for whom I am conducting penetration testing. I also strongly recommend it as a defensive tactic on behalf of your organizations and yourselves to ensure you aren't leaking data unnecessarily.

There is some concern and risk for practitioners regarding the possible consequences, legal or otherwise, of conducting OSINT. I propose treating it like other potentially controversial methods. Much like vulnerability research, coordinated disclosure is likely the best approach. Others might disagree with me on the grounds that if the information is available to search crawlers, it is automatically "open source." Information tagged as *confidential* or *secret* by an organization who unknowingly or errantly leaks it due to misconfiguration or policy violation would still consider that information "for their eyes only" even if accidently exposed. Enough preaching; this is a tools column.

I've previously discussed, both here and on my blog, Maltego,[1] coupled with a variety of Googledorks/Bingdorks as useful for OSINT.

FOCA (Fingerprinting Organizations with Collected Archives) 2.6 is an interesting tool that focuses heavily on document metadata extraction while incorporating other extreme search capabilities. Version 2.6 went live literally as I was writing this.

During the reconnaissance phase of any penetration testing engagement, a tool such as FOCA is extremely useful. Rather than depending on a variety of recon methods, FOCA will provide many related services for you. The FOCA project leads have indicated that for more than the last year and a half FOCA has been a primary tool in their own engagements.

I reached out to Alejandro Martin Bailon and Chema Alonso of Informatica 64,[2] part of the FOCA development team, who pointed me to their DEF CON 18 presentation, "FOCA2: The Foca Strikes Back."[3] This 38 minute video is well worth your time and is quite entertaining; Chema and Jose are quite funny (it gets a bit racy at 06:30 minutes in). They ran FOCA against whitehouse.gov, but stopped themselves just short of army.mil, then hightailed it offstage (remember the above mentioned risks). When asked when they would release FOCA for Linux they replied only that a foca eats penguins. FOCA is "seal" in Spanish ;-)

Not sure if was also intended this as an analogy for the Navy SEALs, but "special warfare" tactics against the wrong target can only bring one thing if you're not careful. So sayeth Robot: "Danger, Will Robinson!"

Just trust me on this, and watch their presentation.

The newest version of FOCA has been enhanced with capabilities that include network discovery, recursive analysis of URLs, information gathering, software recognition/integration (Burp Suite, EvilGrade), DNS cache snooping, PTR record scanning, and reporting. They also incorporated Exalead[4] search to supplement Google and Bing as both search engines often miss certain file types during crawls. You may run into some query limits when including Exalead results.

Remember potential legal issues, particularly around activity such as DNS zone transfers; it's illegal in some countries.

Finally, the FOCA project leads asked me to point out that, after you've discovered just how much information you might be leaking from your organization via metadata, they offer an IIS module, the IIS MetaShield Protector (commercial), that cleans document metadata as files are served but leaves them intact on the local file system. You can also use OOMetaExtractor[5] to clean individual OpenOffice files on your local system (freely available on CodePlex). I'll offer you one other defensive tactic to consider in the conclusion.

[1] http://www.paterva.com/web5.

[2] http://www.informatica64.com.

[3] http://www.securitytube.net/video/1353.

[4] http://www.exalead.com/search.

[5] http://oometaextractor.codeplex.com.

There is also an online version[6] of FOCA that you can utilize to examine documents, but it's limited to one file at a time. FOCA Online offers indication that they will not store uploaded files or their content, but I contend that discretion is the better part of valor.

Installing FOCA 2.6

FOCA installation is as simple as any other standalone Windows application. Download the compressed installation package, unzip it, and execute `Setup.msi`, assuming you have a current installation of the .NET framework.

While the FOCA team provided me FOCA Pro 2.5.6, which is typically provided to folks who attend their online training seminars[7] (100 €), the functionality in FOCA Free is identical with the exception of reporting. The Pro version makes use of the Crystal Reports runtime and offers. If you're using 64bit Windows, be sure to download the x64 version of the CR runtime or you will receive FOCA initialization errors. FOCA Pro 2.6 also includes a vulnerabilities panel.

I tested both Pro 2.5.6 and Free 2.6, but the remainder of this discussion will incorporate only elements available via Foca Free 2.6.

Conducting OSINT exercises with FOCA 2.6

I used three targets to test FOCA. For metadata functionality I used a couple personal, local documents and also focused on ISSA.org (those that are readily available to search engines). For the more aggressive network elements I targeted my own domain, holisticinfosec.org.

Assume no permission to either as you test this tool; please target only domains for which you've gained approval to do so.

Your first step as you begin any work with FOCA is to create a new project. Click *File*, then *New Project*, give it a name and provide associated domain(s). Remember to save the project as well.

I first tested local document analysis; you can drag and drop local files to FOCA, or right-click and *Add file* or *Add folder*, as seen in Figure 1.

I then chose *Extract Metadata* or *Extract All Metadata*.

The *Metadata Summary* will then present you with findings. As seen in Figure 2, when I ran FOCA against all my locally stored copies of NIST documents, it was quickly able to identify all software used to create the document. You'll have to believe me when I say that FOCA clearly identified who (user name) created the documents, and on what operating system.

For a remote retrieval effort, establish a new project, and then click *Custom Search* to expand the search form. The domain you indicated during project creation should be populated along with filetype operators you may be accustomed to using per Google hacking methodology. You can also right-click and *Add URLs* from files where you've built or acquired a TXT or CSV list. Click *Search All* and sit back and wait while FOCA crawls the domain for documents and enumerates additional information. Once the initial search is complete and all available documents are identified in the UI, right-click in the UI and select *Download* or *Download All*. You can only examine metadata for documents you have downloaded. Once all downloads are complete, again right-click and *Extract All Metadata*, then click *Metadata* from the toolbar, followed by *Analyze Metadata*. You'll note that this step then populates client data as viewed in *Network Data*, then *PCs/Servers*.

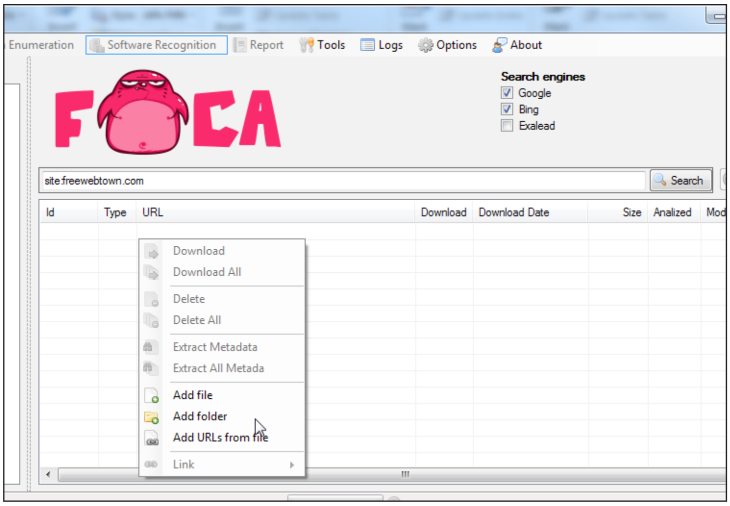
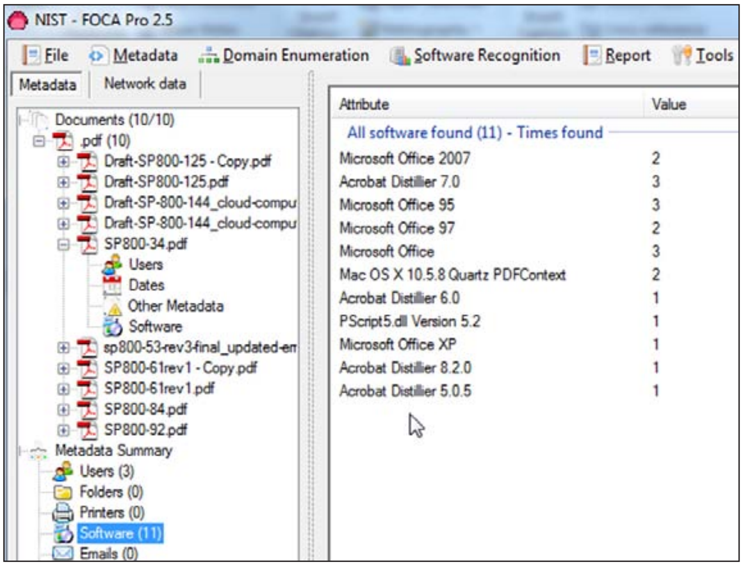


Figure 1 – Drag and drop or add a local file for FOCA metadata analysis



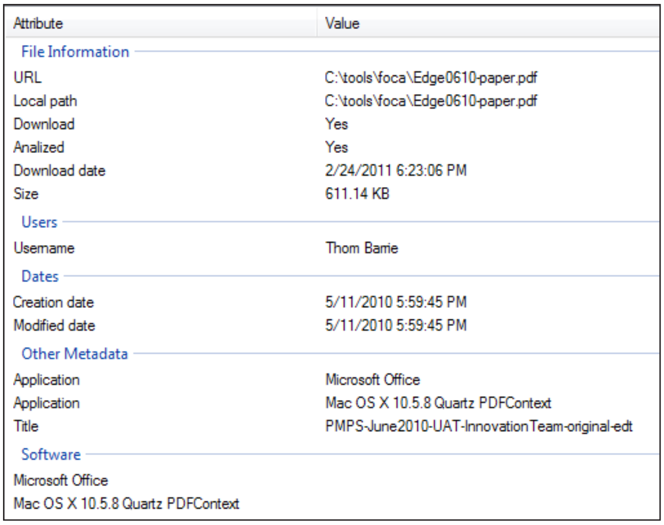
Attribute	Value
All software found (11) - Times found	
Microsoft Office 2007	2
Acrobat Distiller 7.0	3
Microsoft Office 95	3
Microsoft Office 97	2
Microsoft Office	3
Mac OS X 10.5.8 Quartz PDFContext	2
Acrobat Distiller 6.0	1
PScript5.dll Version 5.2	1
Microsoft Office XP	1
Acrobat Distiller 8.2.0	1
Acrobat Distiller 5.0.5	1

Figure 2 – FOCA identifies software used to create documents

As I had spoken to our esteemed editor Thom about this, he is the only one of the 41 unique users identified in the metadata culled from documents downloaded from ISSA.org whom I will discuss. The PCs/Servers view will show all documents found as related to each PC identified. Using the document Edge610-paper.pdf as reference, I switched back to the document metadata view for that PDF as seen in Figure 3. Now we

[6] http://www.informatica64.com/foca.

[7] http://www.informatica64.com/downloadfoca/Trainings.aspx.



Attribute	Value
File Information	
URL	C:\tools\foca\Edge0610-paper.pdf
Local path	C:\tools\foca\Edge0610-paper.pdf
Download	Yes
Analized	Yes
Download date	2/24/2011 6:23:06 PM
Size	611.14 KB
Users	
Username	Thom Barrie
Dates	
Creation date	5/11/2010 5:59:45 PM
Modified date	5/11/2010 5:59:45 PM
Other Metadata	
Application	Microsoft Office
Application	Mac OS X 10.5.8 Quartz PDFContext
Title	PMPS-June2010-UAT-InnovationTeam-original-edt
Software	
Microsoft Office	
Mac OS X 10.5.8 Quartz PDFContext	

Figure 3 – FOCA identifies system details per document

know that this very document you're reading right now was created on a Mac running OS X 10.5.8 with Microsoft Office and Quartz PDFContext.

Now you know the lengths the *ISSA Journal* editor will go to in the name of science and rich content on behalf of his reading constituency.

For our final set of FOCA-born magic tricks I focused only on my own resources; namely holisticinfosec.org. I checked all boxes under *Options* and saved them before this run including proxy searches, automatic directory listing methods to include PUT, DELETE, TRACE, and all fingerprinting options for HTTP, FTP, SMTP, and DNS. The exception, simply due to speed and redundancy, was that I did not select *Use all DNS* servers under the *DNS* tab. I clicked *Search All*, waited a few seconds then set to exploring. Under Network Data, then *Domains*, I right-clicked my domain as seen in Figure 4, and was presented with multiple options.

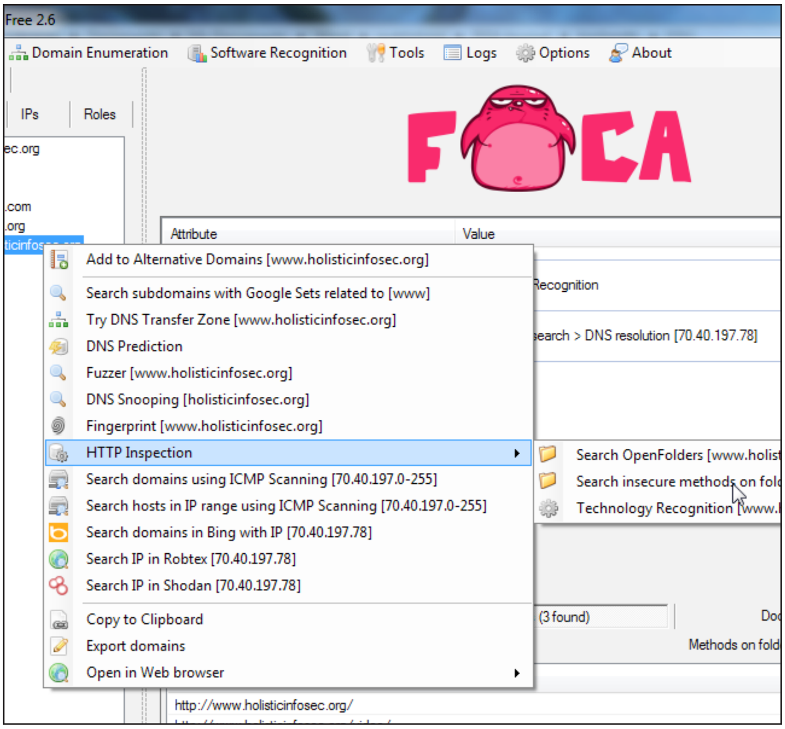

Free 2.6
Domain Enumeration
Software Recognition
Tools
Logs
Options
About
IPs
Roles
ec.org
.com
.org
ticinfos
FOCA
Attribute
Value
Add to Alternative Domains [www.holisticinfosec.org]
Search subdomains with Google Sets related to [www]
Try DNS Transfer Zone [www.holisticinfosec.org]
DNS Prediction
Fuzzer [www.holisticinfosec.org]
DNS Snooping [holisticinfosec.org]
Fingerprint [www.holisticinfosec.org]
HTTP Inspection
Search domains using ICMP Scanning [70.40.197.0-255]
Search hosts in IP range using ICMP Scanning [70.40.197.0-255]
Search domains in Bing with IP [70.40.197.78]
Search IP in Robtex [70.40.197.78]
Search IP in Shodan [70.40.197.78]
Copy to Clipboard
Export domains
Open in Web browser
Search OpenFolders [www.holis
Search insecure methods on fol
Technology Recognition [www.
Recognition
earch > DNS resolution [70.40.197.78]
(3 found)
Do
Methods on fold
http://www.holisticinfosec.org/

Figure 4 – FOCA options for deeper assessment

Check out *Fingerprinting* and *Technology Recognition* (expect a Captcha prompt here). If you choose to conduct *Domain Enumeration* with the DNS Search panel, you'll end up pulling related domains via IP Bing, which will include other domains hosted on your server if you're using a shared/collocated server. Again, be careful here. You could poke someone else's stuff too hard, or discover the nasty truth about shared servers. It's amazing what you'll discover; the good, the bad, and the ugly that may be hosted on the same server you're hosted on.

In conclusion

You can end up running down the proverbial rabbit hole for hours upon hours with FOCA; it is a deeply intriguing tool. Tools don't often scare me, but FOCA does; so I use it with the same respect I utilize for firearms. Don't point it without being fully aware of the consequences; keep your finger off the trigger; and don't let loose until the range master yells "Range hot!" And if you find something whack, report it responsibly, please.

Again, I must strongly recommend using FOCA against your own organizations. Explore for what you may be inadvertently and needlessly exposing. Better you find out before the bad guys use it against you. FOCA can truly help you better protect those you're charged with defending, so use it in good stead. Investigate OSINT in general via your preferred search provider. As a skillset and an undertaking, there is much value.

Finally, take a look at the Google Hack Honeypot (GHH);[8] it helps catch FOCA-like, search engine hacking behavior conducted against your servers. This tool will likely be the subject of a future column.

Ping me via email if you have questions (russ at holisticinfosec dot org).

Cheers…until next month.

Acknowledgements

—Alejandro Martin Bailon

—Chema Alonso

About the Author

Russ McRee, GCIH, GCFA, GPEN, CISSP, is team leader and senior security analyst for Microsoft's Online Services Security Incident Management team. As an advocate of a holistic approach to information security, Russ' website is holisticinfosec.org. Contact him at russ@holisticinfosec.org.

[8] http://ghh.sourceforge.net/index.php.