Doxing and Anti-Doxing – Part I

By Jason Andress

For those of us following or taking part in the various hacktivist activities happening around the globe on a regular basis, doxing is a regular feature. We wake up in the morning to find the personal lives of businessmen, hackers who have made target of themselves for one reason or another, government employees, and a host of others spilled out onto the Internet for the entire world to see. Doxing can be a tool for use in security testing, investigation, or research on the positive side. But it can also be a tool for humiliation, harassment, and worse on the negative side.

In the Part I of this article, we will discuss what exactly doxing is and the tools and techniques we might use to carry out such an attack. In the Part II of this article we will talk about the steps we can take to at least lessen its impact, should we find ourselves on the receiving end of such efforts.

Discuss in Forums {mos_smf_discuss:Andress}

What is Doxing?

The word doxing is a simple word involving a bit of mangling of the English language in order to communicate a somewhat more complex concept. We arrive at doxing by starting with documents, shortening it to docs, applying a bit of leetness to make it dox, then transforming it into a verb: documents -> docs -> dox -> doxing. But what is it? Doxing is the process of locating, to the greatest extent possible, all of the information available on an individual, this being generally followed by the exposure of the information discovered to a group or the general public. Those following along might also realize that doxing, information reconnaissance, OSINT, and a number of other similar concepts are all very closely related, so much so that we might successfully argue that they are slight variations on describing the exact same concept.

We commonly see doxing used by hacking groups such as Anonymous, LulzSec, AntiSec, and so on. An excellent example is the large scale doxing of law enforcement-related personnel by Anonymous in December of 2011, an act which was reportedly carried out in revenge for the close attention being given to hacktivist groups by various law enforcement agencies. In this particular attack the information on over 7000 people was exposed, including names, addresses, social security numbers, email contents, passwords to sensitive systems, and a great deal of other information.

On the white hat side of the fence, a somewhat more restrained form of doxing is also used, although generally much more limited in the set of techniques available, and generally lacking the public exposure of information. Doxing techniques may be used by penetration testers, security researchers, incident responders, and investigators to collect information on potential targets, track down information regarding the origins of tools used in attacks or malware, or to locate the originator of an attack. In March of 2012, the FBI is said to have used information gained from doxing and turned over by another hacking group to arrest Hector Monsegur, a.k.a Sabu, who is widely supposed to have been the founder and/or leader of LulzSec.

Ultimately, doxing is searching for information on an individual, although usually taken to a much greater length than the typical light cyberstalking or ego surfing many of us engage in on a daily basis.

Why Would Anyone Want To Do This?

The motivations behind doxing, whether originated by the good guys or the bad guys, are generally not positive for the person who is the target of such activity. As we mentioned, doxing is used by those who are considered to be on the dark side, hackers (in the bad sense) and hacktivist groups, and also by others such as stalkers, identity thieves, internet trolls, and so on. Usually people in this group seek out said information to attack or harass their targets in some fashion. The specific motivations here may vary somewhat, but we can quickly come up with specific cases in which the entirety of the available data on an individual might be used including name, address and social security number (the identity theft trifecta), account credentials, telephone numbers, and so on.

Even on the light side, where we might find an investigator, incident responder, law enforcement, or other similar personnel using such techniques, the consequences for the target will likely be at least unpleasant, if not targeted at the same ultimate goals. In actuality, much of the set of techniques we would call doxing is simply referred to as investigation in such communities.

What are the Consequences of Being Doxed?

The consequences of doxed information being exposed can range from slight irritation to serious threat to health, livelihood, or potentially life. We can very quickly see where exposing information on a person’s social activities, sexual preference, medical history, and other such interesting bits of information may be seriously damaging. This type of exposure could easily result in public embarrassment, severe reputational damage, loss of employment, identity theft, and worse.

A fairly serious example of doxing and subsequent attacks can be seen in the actions taken by Anonymous against Aaron Barr, then the CEO of HBGary Federal, a defense contracting company. In February of 2011, Barr announced his infiltration of Anonymous and said he would expose the information he had found in a talk at a security conference that year. As with most cases of poking a wasp nest with a stick, this ended badly. Anonymous doxed Barr and attacked the HBGary Federal servers, later posting tens of thousands of emails from the HBGary Federal systems on the Internet, as well as the body of personal information on Barr himself. They subsequently took over social networking accounts, compromising servers, and generally causing quite a bit of havoc. Ultimately, this series of attacks resulted in Barr resigning, reputation damage to Barr and HBGary Federal, some level of investigation by the US House Armed Services Subcommittee on Emerging Threats and Capabilities, and the US Congress.

Doxing Techniques

A number of different sources can provide information for doxing efforts. We can, for instance, collect from social networking sites and tools, people-oriented search tools, pay search sites, public records, and any of a number of other places. Some such sources or tools are generic and will show data on nearly any name we care to enter, and others are very specific and pertain to a particular company, city, or the like.

Typically when we start to dox someone, we will have some small amount of information to start with. A name, depending on how common the name is, is a good starting point; a name and an email address are better. If we have an email address on which to search, we may immediately turn up other sources of information, particularly if the target in question uses the email address as a common account name, posts online frequently, etc. Given a name, email address, and a city, we may be able to turn up a home address, employer, professional organizations, local sports teams or hobby groups, and so on. Each additional piece of information we turn up adds to the body of information we have and makes validating the next piece of information along the path much easier. The more information we have to begin with, the easier our job is.

If we are starting with a weak set of information, or the target has a very common name (James Smith would be a problem), we may have to do a bit of inference at the beginning. One of the most likely starting points would be to pin down a physical location to a smaller area. If we know James Smith had a particular IP address at some point, we look this IP up to find which ISP or company this IP belongs to. Based on this, we might narrow down our search parameters to the areas this ISP serves or this company operates in. Of course the danger exists that our inference is entirely wrong, and we have just gone down the wrong path entirely. But we do have to start somewhere.

Social Networking Sites

Social networking sites can provide a virtual gold mine of information for doxing purposes. A typical person who is at all active online will typically have accounts on at least two or three social networking sites. Depending on the site in question, we may find all manner of personal information (some not fit for public consumption) including current and past employers, education, physical location data, and a plethora of other items.

Of benefit to those using such sites to collect information is the myriad of privacy and sharing settings, each being entirely unique to a particular site. Those of us who are security professionals may have a good grip on handling such settings and may be aware of the need to restrict our personal information, putting us in a better position than the common user to properly safeguard our information. In some cases, this may not be good enough to completely protect us. The companies who run these social networking tools regularly update their privacy settings, and many of them allow “friends” to take actions like tagging pictures with names and re-sharing information, thus circumventing our security efforts.

Additionally, friends, family, co-workers, et al may provide another avenue for gathering information, even if the target has a properly secured account and is extremely careful. Given a particularly chatty friend on a social networking site, we may not even need information directly from the target. This is one of the main reasons that doxing efforts often extend outside of the target in question.

We may also be able to bypass the security measures on a social networking account by simply asking for access to the information through whatever friending mechanism available for the service in question. Such an approach will often enjoy success if we create an account impersonating someone who the target already knows or has some history with, such as a friend, co-worker, classmate etc. We can see an excellent example of this in the Robin Sage incident in which a security researcher impersonating a woman managed to friend over 300 people and gain access to all manner of sensitive information, including classified military information on troop movements.

General Online Content

Although social networking sites provide us with one of the richest sources we may find, they are by no means the only sources available. We can find all types of information by looking for online resumes, blog postings, postings to newsgroups, archives of local newspapers, newsletters from professional organizations, records of births, deaths, and marriages, any of a number of public records and other data. The problem we encounter when digging for such data is finding what we actually want in the massive volume we might need to sift through.

People-oriented Search Tools

Given this enormous body of information available to search, it is helpful to filter some of this through sets of tools that will do some of the work for us. Fortunately, there are a number of services that will conduct searches oriented around individuals and will often give us at least some portion of the information we seek for free. Some of the more common

tools include:

• Pipl.com

• Spokeo.com

• Zabasearch.com

• Mylife.com

• Peekyou.com

and many, many, others. These sites will commonly turn up names, addresses, birth dates, family members, pictures, documents, employers, and quite a bit of other information. Such sites do not exist out of pure altruism, so they will often display a certain amount of information as a hook and then ask for a payment to access the remainder. We can usually get enough free information from such search engines on which to base further searches or general digging, making these sites worth a visit.

In addition there are a number of pay sites that exist for the purpose of performing “background checks,” allowing us direct access to databases of information collected on individuals. These sites will likely include the same set of information as the people-oriented search engines (in fact they may be the same company), but the better sites will also have access to more difficult to reach records such as criminal proceedings, court documents, mortgage documents, and other similar items that are public or semi-public records. Most of this data is available to the individual in general, should we choose to look for it, but it often requires considerably more legwork and expense to obtain. A few background check sites include:

• Intelius.com

• Ussearch.com

• Peoplefinders.com

Information about Domains and Networks

Considering the connectedness of the average computer-savvy person these days, chances are we will be able to turn up an IP address or domain name connected to them in some fashion. Given a small amount of such information, unless the person has been particularly careful, we will often be able to quickly find a good deal more with a few simple searches.

Whois searches and searches of DNS records can often give us contacts for the domain or IP in question, sometimes

even being directly connected to the individual who is our target. While this seems unlikely and entirely too easy, such information is often present. Aggregation tool sites such as Netcraft.com, IPinfoDB.com, and yougetsignal.com can also provide us with additional information such as where the system on the other end of the domain name or IP might be physically located, what software it is running, and any of a number of other useful bits.

Lastly, we would be remiss to not mention the Wayback Machine at web.archive.org. The Wayback Machine archives the content of a huge number of web servers on a regular basis, and looking at changes to a website over time can be extremely instructive. It may be that the system on the other end of a domain name contains no interesting information now, but it might have a month ago, or a year ago, or five years ago. The Wayback Machine can be en extremely helpful tool for many research efforts.

Our Friend Google

Google can be the doxers best friend. We all know what Google is and how easy it is to type "firstname lastname" into the search field and get a few hits. There are, however, considerably more advanced ways of searching Google that will get us better results.

Google hacking is the use of advanced operators in search engine queries (not necessarily just Google), in order to enable more targeted searches. As we mentioned, this is not specific to Google and similar search parameters can be used with most any search engine. Lists of advances operators can generally be found on the page for the search engine in question. For Google, the advanced operators can be found here http://support.google.com/websearch/bin/answer.py?hl=en&answer=136861 and for Bing here http://msdn.microsoft.com/en-us/library/ff795620.aspx. For most any search engine, we can find the advanced operator listing by searching for the engine name and “advanced operators”. While we will find some variation in query construction between different engines, the construction is usually fairly similar.

A large body of work exists for using advance operators to perform very specific searches, alongowith a few books. The book Google Hacking for Penetration Testers by Johnny Long (available from Syngress) is an entire volume dedicated to this specific subject. Although it is a bit long in the tooth at this point, it is still a good resource. We can also look to the Google Hacking Database (GHDB) at http://www.hackersforcharity.org/ghdb/ or http://www.exploit-db.com/google-dorks/ for a database of specific searches. These databases contain a wide variety of security specific searches and are available to the public through a few simple clicks.

Metadata

Metadata is data about data, and we can find such data associated with almost any file with the exceptions to this being vanishingly small. We can see a common example of metadata in the creation and modification timestamps associated with almost all files. Metadata can provide us with another excellent source for our doxing efforts. We can find this data in word processing documents, presentations, image files, videos, and any of a number of similar artifacts. In this metatdata, we can often locate various interesting items such as usernames, hostnames, network paths, various dates, hardware information, and a variety of other interesting bits. For files created on hardware containing GPS capabilities, we may also find embedded GPS coordinates for the location where the file was created, databases or temporary files containing a history of physical locations, and the like. Overall, metadata is definitely worth taking a look at if we find files in the course of our search.

There are number of tools that can provide us with the capability to sift through metadata. For general usage, we can use Metagoofil (although this is now a bit aged and requires some effort to get working properly), FOCA which is in the same general vein as Metagoofil, ExifTool which ostensibly handles image formats but actually does a great number of other formats as well, and several others. Becoming familiar with metadata tools can lead to all sorts of interesting information.

Maltego

Any of a variety of other tools might assist us in our reconnaissance efforts whilst doxing. While it would be nearly impossible to develop an exhaustive list, there is at least one that deserves a special mention, namely Maltego. Maltego is “an open source intelligence and forensics application”. Maltego enables us to conduct, in many cases, a certain portion of our doxing in an automated fashion. Maltego, given a starting place, such as an IP address, hostname, name, etc… will attempt to ferret out other related items of information.

Results from Maltego can be hit or miss, depending on the information available to find. It is absolutely fantastic at tasks like combing through data for an organization. We may also discover a larger set of information from Maltego than we can comfortably cope with, is we are not careful to limit its scope.

Keeping Track

Last, but certainly not least, we will want some method for keeping track of the information we find. We could absolutely use a simple text editing tool like Notepad, Gedit, or the like. Such tools are very useful for taking notes as we go along, but ultimately not a good tool in the long run. In the case where our research has been thorough enough to include multiple individuals, a plain text document will quickly become a difficult task to keep up with and will not lend itself well to searching or correlating information.

As with many larger sets of data, or data we might need to manipulate in various ways, spreadsheets and/or databases are a very handy tool. We will typically want to develop a common template for our doxing efforts, so we do not miss particular items by oversight, i.e.:

• Maiden name

• Facebook account

• Twitter account

• IP addresses

• System names

• Domain names

• Blog URL
• Name

• Address

• DoB

• SSN

• Email addresses

• Phone numbers

• Employer

this, of course, is a very small sample and our template would need to cover considerably more. In an exhaustive doxing effort, we would likely end up with a stack of such collections of data.

There are also a few commercial tools are purpose-built for just such a use. CaseFile is one such tool, and was created by the makers of Maltego. This provides us with a much more tailored solution, but may be overkill for some smaller efforts.

Part II

In the next part of this article, we will be discussing the opposite side of the doxing equation, namely anti-doxing. Now that we have covered what doxing is and how it is performed, we will talk about how we can protect ourselves and help to mitigate such an attack when we are on the receiving end.

Dr. Jason Andress (ISSAP, CISSP, GPEN, CEH) is a seasoned security professional with a depth of experience in both the academic and business worlds. In his present and previous roles, he has provided information security expertise to a variety of companies operating globally. He has taught undergraduate and graduate security courses since 2005 and conducts research in the area of data protection. He has written several books and publications covering topics including data security, network security, penetration testing, and digital forensics.