LESSONS LEARNED WHILE SINKHOLING BOTNETS – NOT AS EASY AS IT LOOKS!

Rainer Link, David Sancho
Trend Micro

Email Rainer_Link@trendmicro.de,
David_Sancho@trendmicro.es

ABSTRACT

Botnets are a well-known security threat for businesses and end-users alike. They are made up of many infected computers under the control of a criminal or criminal gang. The main power of a botnet is in its numbers: the bigger it is, the more it can do because of the compounded bandwidth and computing power of its members. However, small botnets are also often used in order to stay beneath the radar. Sinkholing is a technique that aims to redirect the traffic meant for the malicious server to an analysis server owned by the researchers. In this way, the malicious traffic coming from each of the botnet clients goes straight to the research box, ready to be analysed.

This paper talks about the lessons we have learned from our previous experience of sinkholing botnets, as well as suggestions for researchers on how to realize this endeavour. We will discuss sinkholing as a vehicle for information gathering, and show how it is only of limited use in shutting down botnets. It is not the technical aspects of sinkholing that are interesting, as these are well known among researchers. Instead, the real-world difficulties involved in carrying out these operations will be covered. Some examples include the difficulty of working with certain ISPs or Registrars, what to do when you are suddenly receiving a large volume of Personal Identifiable Information (PII) and problems such as sinkholing a C&C server that is hosted on a compromised domain. We'll also cover best practices, things to avoid, areas where researchers should tread carefully and why a few drinks at the bar with an ISP technician are worth much more than years of experience with IP tables!

INTRODUCTION

Sinkholing is a researcher's technique to redirect/hijack traffic from a malicious machine to an ad-hoc box controlled by the researcher so that it can later be analysed. By looking at a botnet's traffic, the researcher can gather clues and obtain a lot of intelligence about the botnet and its authors/owners.

It is not clear who coined the expression 'sinkholing' but it seems to allude to the fact that the malicious server is not responding any more and becomes a hole that absorbs the traffic.

The earliest reference to the term 'sinkholing' that we know of comes from a 2006 presentation by Marco Cova, now at the University of Birmingham [1] – although the term 'Internet sink' has been around longer, for example as used in a paper from RAID 2004 [2].

There are three main ways of sinkholing a botnet command and control server:

- The registrar can redirect a domain name to the researcher's box. This only works when the connection is based on a DNS name.
- Ask the ISP to redirect an existing IP to the researcher's box. This can only work when the researcher's box is located in the IP range of the same provider.
- The 'walled garden' approach. Ask the ISP to redirect all traffic destined for an IP to the researcher's box.

Another approach that can yield similar results is to contact the hosting provider of a C&C and ask them to take down the server, but also to pass a copy of the hard disk on to the researcher. While this can yield similar results to what we describe below, we do not think of this as sinkholing in the classic sense.

There is yet another method that may be useful to gauge the current impact of old and deceased botnets. Registering expired domains that used to point to C&C servers allows the researcher to measure how many live bots are still trying to contact that old C&C server.

1. SINKHOLING USES – HOW EFFECTIVE IS IT?

Sinkholing can provide a great deal of information for the researcher. When properly performed, the researcher's box may receive all the incoming communication that the original C&C server would have received. There are two sides to this information:

Metadata:

- Who is trying to contact the server (IP address)?
- When are they doing it (timestamp)?
- The URLs and paths they use.
- The countries where they are located (usually based on the IP address and geo-location, but sometimes this may be included as a parameter in the bot communication).

Usually, the different URLs contacted can help ascertain the intention of the botnet's owner, i.e. 'drop.php' (to report stolen data); 'checkin.php' (to connect for the first time), etc. This can be considered metadata as well.

A source-IP analysis can also yield interesting results. For example, if within the IP ranges of the infected PC base there is a significant number of computers from a specific company, this could mean a mass-compromise or a targeted attack. This can be deduced by looking at this metadata information as well.

The data proper:

- This changes with each botnet but there usually is a data-stealing component in almost every case. At the very least, botnets report machine name, operating system and some such data. This can include all sorts of personally identifiable information – from here on, referred to as 'PII' – which we'll discuss in section 3.

Note, however, that sinkholing is most effective when there is a single server. In multiple C&C server configurations, sinkholing won't be as useful. The same applies to P2P botnets, those using different or rotating domain names or those that use Domain Generating Algorithms to locate the C&C server. In these cases, sinkholing will only produce a subset of the botnet traffic, which is generally not much more

useful than the traffic captured by monitoring a single node. The real value of the sinkholing technique is in seeing all of the botnet traffic.

Moreover, having an existing C&C server redirected to a researcher's machine requires a great deal of coordination with the registrar, the country's CERT and their law enforcement team. This process is usually not lightning fast, and often depends on the researcher's relationships with each of the stakeholders in each step of this process. Sinkholing is not an activity that generates money or any other direct benefit to the registrar/ISP.

There is also the possibility of not just redirecting the botnet traffic to a researcher's box but instead leaving the attackers' resources unresolved, be it domain names, IPs or even the physical machines. These actions don't gather any information about the botnet but obliterate it or a portion of it, which may be valuable in itself.

In addition, there are so many C&C servers in any given moment that sinkholing can be ineffective as a neutralizing tool. In other words, sinkholing is not a point-and-shoot kind of tool to remove every botnet. It is unfortunate that high-profile botnets that should be neutralized via sinkholing tend to use one or more of the techniques that make it ineffective (e.g. DGA, P2P etc.).

There have been a few high-profile botnet-neutralizing operations that have resulted in the demise of botnets. Those examples have been carefully timed and very well researched ahead of time. This is especially complicated the more technically sophisticated the botnet is.

For instance, Waledac [3] was neutralized in February 2010 and it involved taking down servers and seizing domain names with careful timing because it used a decentralized peer-to-peer communication system to control the whole botnet. The botnet was distributed in layers so the researchers had to take down all of the control servers at the same time to prevent any of them from acting as backup to the others or allowing the criminal to add new backup servers to the list. Rustock [4], taken down in March 2011, was a similar case that needed court orders to seize criminal property.

An example of a botnet that was difficult to take down was Conficker [5] which used a Domain-Generating Algorithm to call home. Every day, the worm would generate some 250 predictable random-looking URLs to connect to. If the bad guys wanted to send a new command, they just had to register one of the domains to be used on a particular day and serve the content. Neutering this system sparked the collaboration of the whole security industry in what was called the Conficker Working Group. With the help of registrars, they were able to block the registration of all those possible 250 daily domains that the algorithm was able to generate. The criminals' response to this was to create a new algorithm that was able to generate many orders of magnitude more domains that its predecessor, about 50,000 domains a day. This kept researchers at bay for a while. That's how complex it can get to target a specific botnet to redirect/neutralize/sinkhole. In the case of DGA-enabled botnets, it's not enough to register a domain and receive the requests. The coverage would only be partial. In these cases, you need to target every possible domain that the DGA is able to generate. Given these kinds of challenges, the industry has spawned special working groups to help with these complex situations.

There is also a factor to take into account. Targeting a botnet such as Zeus or SpyEye brings to the community a much lower return on investment (ROI) than sinkholing a more specific botnet. This is because Zeus botnets tend to appear and disappear very often as the software to create them is readily available in the underground. A botnet such as Waledac or Conficker is unique to the criminal group that created it and therefore eliminating it is a bigger deal and usually a much bigger challenge for any researcher. The amount of research that has to be undertaken before taking on a very specific botnet is much greater than an off-the-shelf botnet. Knowing the networking side as well as the application/protocol layer is a must when dealing with specific botnets.

An ROI analysis is necessary before sinkholing any particular botnet. The dilemma is that the easier a botnet is to sinkhole, the less useful this activity is. Regardless, the simplest scenario still requires considerable resources and synchronization among different parties. Sinkholing is not necessarily worthwhile in every situation.

2. REAL-WORLD DIFFICULTIES OF SINKHOLING

Sinkholing a botnet generally requires making a request to the registrar to redirect the malicious domain name to the analysis box. Unless there is a court order that compels them to comply with such a request, without the explicit consent of the owner/end-user of the domain, the registrar is unable to grant such requests.

Moreover, asking a registrar for help with sinkholing always means additional work for them and for some registrars/ISPs the abuse/security team is most likely understaffed ('it only costs money...'). In addition, depending on how the botnet works, it may not be as simple as changing the name resolution – a simple DNS change – and if you need them to redirect traffic at the router level, that would mean a much bigger change in their network configuration. Therefore, such a request is more likely to be refused than a simple change in DNS.

Working with CERTs, law enforcement units, registrars and ISPs can be very complicated when dealing with many different countries with different laws. In certain countries, there are so-called bulletproof hosting providers, which will immediately ignore any requests. This can hinder the sinkholing project considerably. Moreover, if sinkholing becomes more popular or successful, it is to be expected that the bad guys will start using more bulletproof hosting providers. They will also stop using 'simple botnets' and instead will favour more complex ones, like spreading C&C servers across regions, P2P-style botnets or more advanced DGAs.

Another point to consider is the amount of coordination sinkholing requires. There is no single entity that currently coordinates all sinkholing efforts or activities undertaken throughout the world. Such lack of coordination is likely to result in conflicting activities.

If sinkholing is to be an effective tool to fight botnets, there probably needs to be a global organization responsible for coordination of sinkholing activities. Any such organization would need to involve security researchers, law enforcement and ISPs/registrars. To ensure that the targets of sinkholing

efforts are not informed in advance, access to the activities of the organization would need to be closely vetted. Of course, even the existence of such an organization raises a lot of unanswered issues:

- Who would be responsible for coordination?
- How would such an organization obtain authorization for such activities?
- What happens if several organizations want to sinkhole the same botnet?
- What will the obligations of the organization be in the event it discovers a criminal activity?

Even supposing that the redirection takes place, the rest is not exactly a bed of roses. A researcher's analysis machine will be flooded with requests and will often be DDoSed in retaliation. Utilizing cloud providers can help by allowing a wider broadband connection to withstand the network and CPU load. The challenge in those cases is to make sure you comply with the provider's terms of service or risk your cloud account being removed. Even then, using a cloud provider may still be a risk, as we all know from the *Amazon EC2* outage back in April 2011 [6]. A large-scale DDoS attack might also affect other customers of the cloud provider – causing the cloud provider to stop your server in order to alleviate issues for its other customers. The botnet herder might also try to hack the researcher's analysis machine, allowing it to be used for other attacks or illegal activities. Therefore very close 24-hour monitoring of the analysis machine is highly recommended. A direct emergency line to the cloud provider/ISP is also helpful should anything go wrong with the analysis box.

Other very important considerations are the legal consequences and issues regarding sinkholing. Since this document is only intended to address some of the practical issues revolving around the activity, we would recommend that a legal advisor be consulted before undertaking any sinkholing projects.

As previously mentioned, one of the aims of sinkholing is to study the botnet and to gain a better understanding of how big the botnet is. However, some sinkholing efforts have gone much further than simply monitoring the traffic received. There have been several cases where the sinkholing server has been used to actively send commands or clean-up tools to the infected bots in order to remove infection. While this may appear to be good idea, gaining consent from the owner of the affected machines may be necessary since, as the researcher, you are now executing code on such machines. Also if the botnet has a 'remove bot' command, what happens if this command inadvertently crashes the machine – who is liable here, the researcher or the botnet owner? One mitigation concept was demonstrated at LEET 2008 for the Storm worm [7]. Another more recent example is the takedown of Coreflood [8], with the help of *Microsoft MSRT* [9].

There are many, many things to consider when sinkholing a botnet – but in particular we would advise other researchers to make decisions on the following when entering into such activities:

- How long should you continue to run the sinkhole before turning it off?
- What is your plan of action should the botnet owner manage to reactivate the botnet?
- What do you do in the situation where the bot has an in-built kill switch (i.e. after X days without receiving a command from the C&C, it damages the machine in some way)?

To summarize, many issues and factors need to be considered before commencing any sinkholing project. Moreover, running a sinkhole may not ultimately be the right solution. This document presents only some ideas and opinions regarding those issues and factors. Nonetheless, we highly recommend that you consider the points we have raised if you decide to become involved in any sinkholing activities.

3. WHAT TO DO WITH PERSONALLY IDENTIFIABLE INFORMATION (PII)?

A thorn in every sinkholing researcher's side has always been the issue of how to deal with personally identifiable information (or PII, for short). When you sinkhole a C&C server successfully, you are likely to collect significant amounts of data and among it, personal data or PII. Since laws and regulations differ from country to country as to the processing, collection and storage of PII, you should consult your legal advisor before doing any of these.

If you use a cloud provider to perform sinkholing operations, it is also important to check their policies regarding PII collection and storage and determine whether their systems can properly secure such information. Your organization may be responsible for any lost data in the event of a leak or security breach. This is especially important for credit card data and login/password databases. If you suspect that a sinkholing exercise may capture PII data, it may be best to forward it to a secure offline data repository within your control. This ensures that in case of a shutdown of the cloud instance, the data is still available to be analysed at any moment.

Unfortunately, not collecting any PII would make it difficult to identify the target. Without such information, one would not be able to inform the targeted organization that the attack is taking place against them. This information is of significant value to understand the whole picture and to put together all the pieces of the puzzle: who is behind the attack, what their motives are and where they may be located. The nature of the PII received can often offer some theories about possible answers to those questions. Not collecting such data would prevent the researcher from getting a glimpse of the bigger picture.

Regardless, there should be no need to decrypt or store any network traffic that potentially contains PII. Use of this information is limited so it is advisable to process such data as it is being collected without storing the information.

In any event, having a set of principles on data collection from the outset, is a very useful way of knowing what to do once unexpected things come up. In this regard, the Torpig sinkholing paper from 2009 [10] set two simple and efficient principles that can be safely applied to any white hat sinkholing project. We recommend using them and abiding by them at all times as a clear ethical guideline.

Given all the various complex regulations or laws relating to the processing, collection and storage of personal data, it would be advisable to consult a legal advisor before doing any of these things.

4. STRANGE POSSIBLE CASES

The goal here is to raise awareness of situations which need to be considered before attempting a sinkholing project. There are several unusual cases that may arise from your sinkholing practice. As researchers accumulate more experience on these types of activities, they will stumble upon their fair share of unusual situations.

What if the bots connect to IP addresses, not DNS names?

Generally speaking, redirecting an IP address to your machine can only be done by the ISP at the routing level. ISPs are unlikely to comply without some sort of court order. But as everybody knows, it is technically possible to do this without the ISP's assistance. The redirection would probably need to target a local IP address within the ISP's network and then be tunnelled to the researcher's location.

What if the bots connect to specific folders within a compromised server?

If this is the case, you have two options:

a) The more intrusive one is to redirect the whole server to the analysis box and then configure the machine's web server to redirect all legitimate traffic to the real server. This may result in all sorts of problems and can impact the site's quality of service. However, this is not recommended as it would intercept non-malicious traffic.

b) The customary way of dealing with these cases is to obtain a court order to redirect the specific folders' URLs to the researcher's box. The configuration is not as straightforward as redirecting the whole domain but it might be possible with the help of the hosting company/ISP and a court order. In fact, an L4 switch/proxy/load-balancer may be necessary and the hosting company/ISP is unlikely to buy this infrastructure for a specific request; so keep in mind that even with a court order, this kind of redirection might not be possible in smaller ISPs.

What if the bot clients are configured to act in a special way when the server is not available/acting strangely?

The most extreme of these cases involves a hypothetical client that damages the machine in this situation. In real-life malware, though, this means accessing a backup C&C server or causing some delay in the machine that could affect its performance. For instance, if the analysis machine always responds with an empty 200 OK HTTP response, the bot clients might not be ready for such a response and they may crash or fall in low performance loops. A complete analysis of the bot code should be performed before sinkholing takes place. If the full reversing cannot be achieved in time, blackboxing tests can help determine how a regular bot client reacts to the sinkholing server as C&C, if possible.

Your sinkhole can be recognized as the real C&C server

Cases like the *Damballa* mistake [11] show how a sinkhole may appear to be the real C&C server on a cursory inspection. Many research boxes looking for C&C servers operate from automated scripts and a change of server only flags your research box as a malicious new black spot on the Internet. This not only happened with *Damballa* but it occurs often. In extreme cases, the IP space you're operating from might be flagged as malicious and this might lead to measures being taken by your service provider. You should make sure that any sinkholing activities you undertake are in compliance with any terms of any contract with your service provider.

5. RECOMMENDATIONS

Sinkholing is not a silver bullet. It is a very useful tool in certain specific situations, but is also a tool that takes considerable time and co-ordination to be effective and useful. It is always important to think long and hard about the ROI of any sinkholing activity. There are plenty of things that can go wrong but when used effectively, it can neutralize the whole botnet and provide very valuable stats for the researcher about those behind it.

It would be desirable to have a global organization that could coordinate sinkholing efforts or activities and that could lend legitimacy to and help in obtaining the requisite authorization to perform such activities. That would make the process a lot smoother and it would help eliminate one of the hurdles that sinkholing has today: coordination among several parties.

The researchers need to be aware of the various limitations and think carefully about them before embarking on these activities, especially considering the restrictions on processing, collection and storage of PII. Sinkholing can be a very rewarding experience and the information that can be gleaned from such activities may well justify the efforts and time put into the sinkholing project.

REFERENCES

[1] Cova, M. A presentation of 'Modeling Botnet Propagation using Time Zones'. 21 Feb 2007. http://www.cs.ucsb.edu/~kemm/courses/cs595G/marco.pdf.

[2] Yegneswaran, V.; Bedford, P.; Ploka, D. On the design and use of Internet sinks for network abuse monitoring. RAID 2004. http://pages.cs.wisc.edu/~vinod/raid-paper.pdf.

[3] With legal nod, Microsoft ambushes Waledac botnet. 25 Feb 2010. http://news.cnet.com/8301-1009_3-10459558-83.html.

[4] Prolific Spam Network Is Unplugged. 17 Mar 2011. http://blogs.wsj.com/digits/2011/03/17/prolific-spam-network-is-unplugged/.

[5] Conficker Working Group: Lessons Learned. 7 Jun 2010. http://www.confickerworkinggroup.org/wiki/uploads/Conficker_Working_Group_Lessons_Learned_17_June_2010_final.pdf.

[6] AWS outage timeline and recovery strategy. 25 Apr 2011. http://aws.amazon.com/message/65648/ and http://www.randomhacks.net/articles/2011/04/25/aws-outage-timeline-and-recovery-strategy-downtimes.

[7] Holz, T.; Steiner, M.; Dahl, F.; Biersack, E.; Freiling, F. Measurements and Mitigation of Peer-to-Peer-based Botnets: A Case Study on Storm Worm. LEET 2008. http://www.honeyblog.org/junkyard/paper/storm-leet08.pdf.

[8] Feds to remotely uninstall Coreflood bot from some PCs. 27 Apr 2011. http://www.computerworld.com/s/article/9216199/Feds_to_remotely_uninstall_Coreflood_bot_from_some_PCs.

[9] A second MSRT release in April. 26 Apr 2011. http://blogs.technet.com/b/mmpc/archive/2011/04/26/a-second-msrt-release-in-april.aspx.

[10] http://www.cs.ucsb.edu/~seclab/projects/torpig/.

[11] http://www.h-online.com/security/news/item/Damballa-s-analysis-of-botnet-C-C-servers-criticised-1126699.html (Oct 28 2010).