Extracting C&C from Malware

The Role of Malware Sample Analysis in Botnet Detection

By Gunter Ollmann, VP of Research, Damballa

Introduction
There often appears to be little or no difference between malware and botnets. If a computer system is infected with either then, as far as users and IT staff are concerned, it is compromised and can no longer be trusted for confidential business use. However, this distinction is important. Both are used by organized cyber criminals for financial gain, but botnets add another dimension to the threat – the ability to be remotely controlled and serve as a digital bridge into an organization.

Botnets add another dimension to the threat – the ability to be remotely controlled and serve as a digital bridge into an organization.

Modern botnet software typically ships with the full spectrum of malicious capabilities found in top-of-the-line malware. It becomes a ‘botnet’ if it contains features that allow it to communicate with a criminal Command-and-Control (C&C) infrastructure and can be remotely controlled. Even then, for a compromised host infected with a botnet agent to actually become a member of a botnet, rather than just another Remote Access Trojan (RAT), the agent must be able to associate itself with a C&C infrastructure capable of simultaneously managing multiple botnet agents.

Therefore, the network characteristics of malware and botnets provide a means of separating these two distinct threats – and a way of discerning the difference between a host ‘compromise’ and a network security ‘breach’. This paper reviews the pros and cons of extracting C&C information from malware samples obtained from the Internet and from within a corporate network, and the value of this information in identifying, differentiating and mitigating botnet activity within an enterprise organization.

C&C from Malware
Threat research organizations maintain huge collections of malware samples, typically obtained via a mix of paid-for commercial subscription services, shared vendor data and community malware feeds. They supplement these resources with vendor-specific honeynets, spam traps and sinkholes that capture additional samples. Subjecting these malware samples to automated analysis in an appropriate investigative environment can often reveal the presence of malicious network functionality. The goal is to extract key network behaviors that indicate botnet capabilities and might be used to enumerate possible C&C systems.

There are significant technological nuances associated with the extraction of network behaviors from malware. In addition, the malware itself is increasingly capable of defeating attempts to automatically extract these behaviors. That said, the automated processing of malware samples has proven to be a valuable source for botnet C&C information.

The Fallacy of Reactive C&C Extraction
One perceived limitation of using malware samples as a source for C&C information lies with the fact that samples must first be obtained before the critical information can be extracted. Therefore, any technology capable of detecting or protecting against botnet threats will be reactive – and dependent on the vendor having access to a sample.

The situation is further complicated by the way in which the vast majority of malware samples are obtained. Most samples are acquired via passive Internet monitoring technologies, which means that botnet visibility is largely limited to widely circulating threats or threats that tend to be indiscriminate in the systems they target. As such, malware and botnets specifically targeted at corporations are rarely captured using such systems, and rarely have their C&C information extracted.

It is true that the extraction of C&C information from captured malware samples is a reactive process. However, this reality does not necessarily mean that technologies utilizing this data are also reactive.

Botnet masters have adopted a highly organized and mechanized approach to the creation and distribution of their malware. Industrial-scale serial variant production lines – capable of churning out tens-of-thousands of new, unique and
“undetectable” malware per day – are de rigueur for most criminal operators. It costs practically nothing for the botnet controller to create and distribute new bot agents – and guarantee that they won’t be detected by antivirus products on the day of their release. However, the built-in C&C capabilities of a botnet are much less flexible or diverse. That vulnerability is critical for identifying and defeating botnet threats.

In order to make botnets robust against law enforcement takedown notices or hijack efforts by security researchers and criminal competitors, botnet controllers need to invest substantially more time and money on C&C infrastructure than they do on malware construction. Even though there are a growing number of techniques criminals can use to make C&C robust against loss, enumeration and operator disclosure, each of these alternatives also make the chosen C&C less flexible to change. Every serially generated malware sample relies upon the same C&C infrastructure to enable botnet functionality. Therefore, C&C becomes the chokepoint for detection and mitigation.

As time goes by, some botnet controllers eventually alter their C&C configuration, but at a pace far slower than that of the botnet’s malware components. The net effect of this limitation is that access to the latest malware samples (which the botnet distributes and installs upon victim systems) is not a prerequisite for detecting new compromises, or even for monitoring the growth of a botnet. Serial variant malware production systems are may bypass content inspection and signature-based detection systems, but it is not a viable technique for bypassing network-based botnet detection.

**Advances in Botnet C&C**

Malware authors and botnet controllers are creative criminals, always searching for new ways to bypass threat detection systems. While existing C&C infrastructures and topologies represent an Achilles heel in avoiding network detection and mitigation of botnet infections, organized cybercriminals have been able to build an increasingly diverse pallet of possible communication channels that can be hijacked for C&C.

There has been much discussion within the security industry about the exploitation of covert channels and micro-blogging sites for C&C communications. To date, actual field-deployed representations of these possible C&C vectors have been limited to what can best be described as “proof of concept” attempts. They have garnered interest from the news media and academic circles, but they have yet to be proven to be sufficiently robust for use by professional criminal organizations.

![Fig. 1. Micro-blogging C&C: Twitter accounts have been used in a proof-of-concept capacity for botnet C&C. Commands have been encrypted and represented as status updates in this example.](image-url)

The good news is that, while many of these techniques appear to provide a viable communication vehicle for botnet C&C, they are generally fragile once the algorithm has become public. In essence, although botnet masters can choose from a multitude of possible C&C techniques, many of their more sophisticated implementations require either expensive infrastructure investments or are quickly unraveled once a malware sample is obtained and its network communication behaviors understood.
For example, a botnet master may implement a custom C&C technology that leverages popular micro-blogging sites. This decision allows the botnet controller to create new (and multiple) accounts that host command instructions, which may in turn be encrypted or otherwise obfuscated. Botnet agents that retrieve the content from these free accounts then identify and follow any commands that may be present.

When a malware sample is obtained and decoded by a security researcher, the C&C algorithm can be easily deduced. It then becomes a relatively trivial matter to close or hijack malicious micro-blogging accounts. In short, botnet masters may choose to implement many different C&C technologies. However, very few are resistant to being unraveled at the network layer once a malware sample has been captured.

**Enterprise Botnet C&C Enumeration**

Due to the Internet-centric focus on malware and the relative ease in obtaining samples of these broad-spectrum threats, the automated analysis of C&C information extraction from malware samples is rarely indicative of the targeted threats encountered within enterprise networks. As such, enterprise-focused botnets represent a blind-spot for these techniques.

The primary method for overcoming this limitation is to acquire suspicious files and malware samples from within the enterprise itself – for example, by requiring transparent proxies, mail gateways and other network monitoring technologies to surrender copies of intercepted files to a specific internal repository. Suspicious files gathered in this way can be automatically studied and their network behaviors extracted from within the enterprise network.

There are, however, severe limitations with attempting to extract botnet network behaviors from within an enterprise network:

- Most malware identifies that it is being executed within a sandbox or other virtualization technology. Depending upon the malware author’s intent, malware that discovers itself being executed in such an analysis environment may appear to be benign (i.e. not execute any malicious routines) or may select to exploit the sandbox technology in an attempt to gain control of the investigative system.
o In the former case, the suspicious sample will not exhibit any network C&C behaviors and not be identified as a threat.

o In the latter case, the malicious file may compromise the integrity of the analysis system, which could result in an enterprise network breach.

• Most botnet malware will "test" the network in order to discover whether Internet access is possible. These tests may range from attempts to reach a public Web site (to timestamp the malware infection), launch of a test spam message, or contact systems maintained by the botnet controller for tracing infection success. These tests are typically independent of the C&C channel that the botnet malware will rely upon for actual malicious activity.

o If the malware fails to pass its network tests, it will behave differently, e.g. initiate worming capabilities, hide C&C behaviors until it passes the test, or simply act benign.

o If the botnet malware passes its tests because Internet network connectivity is possible from within the sandbox/virtualization technology, then the enterprise network has been exposed.

  ▪ The botnet controller then knows that the enterprise is using an automated investigative technology (tests have been passed, but malicious activities are not possible) and can therefore fine-tune future targeted attacks.

  ▪ If malicious actions are possible, the third-party victims (and observers) of the attack may take actions against the enterprise network. Typical steps include degrading the network’s reputation, which adversely affects the organization from sending legitimate business traffic), or legal action against the unwitting host as an enabler of criminal activity.

• Several classes of botnet malware targeted at enterprise networks rely upon internal clocks and specific user behaviors before C&C can be established. The malware sample must be executed benignly on a host for hours before it attempts to reach out to the C&C server. As such, botnet malware executed within a virtual environment may not yield C&C information or malicious intent within a timeframe sustainable for automated analysis within an enterprise.

• In order to reduce the size of the malware component and to leverage more advanced communication channels for C&C, botnet malware increasingly requires specific applications to be installed on the victim system. For example, botnet malware may attempt to make use of BitTorrent or Skype applications for connecting to C&C. If these applications are not present and correctly configured on the sandbox or virtualization technology, no network C&C information can be harvested.

In general, the preferred method of automatically analyzing suspicious files gathered from within an enterprise network is to conduct the work outside of the enterprise. The ideal location for security analysis is within an anonymous cloud so that botnet creators and controllers obtain the least amount of information about the target (or analyzing technology). That way, the enterprise reputation is not tarnished, nor is the enterprise network subject to additional forms of breach.

**Beyond Extraction**

The process of extracting network behaviors and deriving C&C information is well understood, as are the limitations of these techniques. This process yields much of the actionable intelligence necessary for uncovering the presence of botnet malware within an enterprise environment. However, greater detection and mitigation efficiencies can be obtained with additional processing, correlation and clustering.

For example, a botnet master may produce 5,000 unique serial variant malware agents each day and associate each of them with 20 distinct C&C channels – of which each malware variant will be aware of and initially attempt to connect to 5 of them. Each week, half of these 20 core C&C channels are replaced with new ones because they have been shut down by law enforcement, hijacked, or otherwise removed from the botnet. Any previously compromised hosts will be issued an updated list of current C&C channels every hour – and have their malware component replaced or updated each evening. Over a period of three months, the botnet master will have released 450,000 unique pieces of malware and have utilized 130 different C&C channels – but at the end of the day they all belong to the same botnet master, doing the same thing.
The ability to associate all these malware variants and C&C channels together in this simple example, and to manage the detection and mitigation of this particular botnet threat as a single entity, greatly reduces the burden on enterprise teams responsible for managing botnet protection. This ability to cluster botnet C&C information derived from a broad spectrum of malware samples into distinct botnets also allows an organization to track and assign a risk to the overall threat, rather than individual malicious agents seeking individual network components deployed as part of the broader botnet.

**Conclusions**

While Internet malware and botnet malware may appear to be similar at a cursory level, they represent very different threats to the enterprise. The differentiating factor between these two threats lies with botnet malware’s ability to circumvent enterprise network defenses and be remotely controlled in a collective manner by a cybercriminal controller. These network characteristics provide a means of discerning the difference between a host ‘compromise’ and a network security ‘breach’.

Because of the way in which malware samples are typically acquired by security vendors and research organizations, detection and mitigation of botnet threats are traditionally focused what can best be described as “broad-spectrum” Internet malware. Since enterprise targeted botnet malware is distributed in a different way and is far less likely to be encountered using standard sample gathering technologies, it is typically not well represented in existing perimeter and host-based protection technologies.

Enterprise botnet malware samples must be gathered from within the enterprise if they are to be useful in detecting and mitigating targeted threats. However, there are pros and cons to automatically extracting botnet C&C information from the samples – particularly if this extraction is conducted within the corporate network. When botnet C&C information is extracted from malware samples, it is important that due care is applied to the clustering and correlation of this information in order to track the threat as it evolves – and to provide an efficient means for managing mitigation of the threat.

Botnet threats evolve, with new features and tactics appearing constantly. The challenge in keeping up with these rapidly changing criminal technologies is to seek the metaphorical forest itself, rather than becoming distracted by large numbers of trees. Proper handling of botnet malware samples, including where and how to capture and decode them, provides critical insight into these threats without revealing sensitive information to online criminals or opening enterprise networks to attack.

**Further Reading**

- “Serial Variant Evasion Tactics: Techniques Used to Automatically Bypass Antivirus Technologies”, Damballa, October 2009
- “Botnet Communication Topologies: Understanding the intricacies of botnet Command-and-Control”, Damballa, June 2009

**About Damballa**

Damballa is a pioneer in the fight against cybercrime. Damballa provides the only network security solution that detects the remote control communication that criminals use to breach networks to steal corporate data and intellectual property, and conduct espionage or other fraudulent transactions. Patent-pending solutions from Damballa protect networks with any type of server or endpoint device including PCs, Macs, Unix, smartphones, mobile and embedded systems. Damballa customers include mid-size and large enterprises that represent every major market, telecommunications and Internet service providers, universities, and government agencies. Privately held, Damballa is headquartered in Atlanta.  
[http://www.damballa.com](http://www.damballa.com)

Prepared by:
Damballa Inc.
[http://www.damballa.com](http://www.damballa.com)
Copyright 2009. All rights reserved worldwide.