DGAs and Cyber-Criminals: A Case Study

by Manos Antonakakis, Jeremy Demar, Christopher Elisan, John Jerrim

In recent years, Domain Generation Algorithms (DGAs) have evolved from a proof-of-concept technique, capable of bypassing legacy static reputation systems (e.g. Domain Blacklists), into full-featured stealth modules embedded within an increasing number of advanced and evasive commercial crimeware toolkits today. DGAs are also referred to as a form of “domain fluxing.”

This case study details how Damballa Labs uncovered criminal DGA activity long before the malware using the DGA technique was ever identified by the security community. This discovery was accomplished using patent-pending machine learning technology and years of passive DNS data collection and analysis. In addition, this case study describes how Damballa Labs, starting only with the identified DGA behavior, tied the DGA behavior to the criminal command-and-control (C&C) infrastructure and then to the malware, infection vectors and campaigns. The identified malware is a Zeus version 3 variant that uses peer-to-peer as its primary C&C channel and only resorts to the DGA-generated domains if it fails.

Identifying DGA Behavior Using Machine Learning Technology

Utilizing Damballa's extensive global visibility of DNS traffic, Damballa Labs can identify key characteristics of DGA-based crimeware deployments. One of the key detection attributes for crimeware that employs a DGA to find live C&C servers, rests in its failure - in particular, its daily production of unsuccessful DNS resolutions for nonexistent domain names. These nonexistent domain name responses (NXDomains) have proven to be a reliable detection feature for DGA usage.

Linking a list of domains to a specific criminal campaign can be difficult. Starting with a long list of only NXDomains makes the problem even more difficult. For this case study, we started with a list of NXDomains that were not yet associated with any known malware. We then clustered DGA-generated NXDomains, and then we mapped the NXDomains to the registered domains that were used as the criminals’ C&C infrastructure. We then continued to work backwards to identify the malware and then find the exploit sites, redirector domains, and the original spam messages that lead to the infection.

One of the patent-pending machine learning technologies used in cyber threat research at Damballa Labs is called the DGA Identification System, which is used to analyze passive DNS data to automatically identify clusters of NXDomains and tag those clusters as:

- Known malicious - with the appropriate attribution to malware family and criminal operator.
- Unknown - identified as DGA activity, but not yet known whether it is malicious or benign.

The DGA Identification System can correlate a large number of NXDomains for any identified DGA-generated domain cluster. The cluster that we focused on for this case study generated over 1900 domains in four days - a massive amount of domains to manually inspect and classify. Searching for a sampling of these domains in various data sources and online revealed nothing – not surprising as they are NXDomains. There were no malware reports that listed any of these domains or linked them to any sort of DGA. The following table provides examples of NXDomains generated for this cluster.

mzezeuetfrawm69ptezj56jwczewo41l58l48.com	izm39bsftgthrgte11j26hrh24etj26hvmvg33.org
f12gsbskslrm19a47bvd50jzn40cviri65n30c49.biz	h24m29dxfqp42fyhya47fwbwjritfzj46ktcz.info
c19eujxjwdzatjupsb18m29m49dzmuiwd20fq.ru	ixa27gti65bznxm29fxkwc39hvnxnuf22lzg43.com
jxkxgzo41o21j26a67dwo51fxe21o11g63g33fqcz.info	f52lzdyavnzdvjwc59g33hxeqiwmqe31h34ly.info
kzoxh24g13oza47punsm29lxprjwo21f42cwar.com	nxp62b48kveyjwfrh64h64gtkvawk17g33fqps.com
p32gta67i15bta27c69aze31fqjrn60atavpudy.net	iqpzdtfzoqftp42mrn20n30dxpxiwk47h54c69.org
fymsazetkwhxm29jqd50o51pzhud40aratf12.com	kvbtltjwoxf52b68nslulzgvevh44bvatey.biz

kum19g33l58eravl28o41gxnyprdwj36mrhrmz.org	kviuf62gqe61cyotowbzg53f42f32bwe11p52py.org
l58d40evk47p22lrj36h14gtayaycudve21n60iy.biz	e51ntn50ivgsltoqhyfsd30b18cti15jwmtly.net
mrmsoql48kxhrmrfvkuaun50cwbtezktl48.com	hwbsk17lqjudvj66n60bvk47p12d40hqkzc29ht.net
g43a37iyk67hzishsnug33o61ksm29cybzmybw.biz	gro61jwg43f12etdqpxewkto61a57e11f52nvbx.com
kzosbzl28ltiyi25c29nup42f52nzk47a67j46h44.com	l28oqbsmxmvkxmtmua47nroyktgxc59frou.ru
ira47i45d50hyd20mrh24fze31c29h14a47jto51nr.info	o51p62csm19kvhuorewc59l38dtdsb38axlxjz.ru
j66hvbrn20esn20msi35nwhupzg23p42mxp62f62.org	ixn30i45ktcqa57axp52b48bxi55pvpxlskwar.com
byiyjujrk27cvdulvkrh14kwdymvd10lycv.com	lrlsh34jyo61azeti35kyitiqj26ctf12ntj16.org
i15psfxfxgto41kyitp12c19o31k27d30dzlzo11.ru	ftn30hud30owoqjvayfsdtl28d20kvhvaqp12.com
gvowftfzo41f62l58k37fyazezdriqmwcxi45.info	aua67pvhqcxdtovjsatntl18cza67fujyb58.biz
brdwgzp32o61jtl18ezazhuo61lqnwlqdsjr.ru	b68nyn10myjulqgyovjxgugqb38fzbqbqc39.com
a27g13fvpwnzhthrduhql68nze41k27mtlsnu.org	p42k27cxo51hwm39ctc29n60n40dwiwfumvlxar.org
jrb18dyi35jxmrexiyo61k37owd50oue51ntb58.net	lri25j26msl18nymskvdyo51jqfwjwa67d50kr.biz

Timeliness of the data was another challenge we had to address when trying to map the domains to malware and criminal campaigns. The clusters of NXDomains generated by a DGA normally do not appear in any public reporting. When anti-virus (AV) or sandboxing vendors generate reports on samples that we know are related to malware samples that utilize DGAs, the only domains that might be listed are the ones that were generated at that snapshot in time. If the malware sample was run in a sandbox even one day off from when we collected the NXDomains there is rarely an overlap. Trying to find any evidence of what these domains were used for or what malware was using them proved difficult. When we first started looking at these clusters we had some clusters that were over six months old, and it was still very rare to find any public information that referenced the domains.

Discovery of Domains and Distribution

What were the real C&C domains?

Our first task was to take the list of NXDomains identified by the DGA Identification System and try to find the domains that resolved to an IP address. Utilizing our passive DNS data, and newly automated techniques, we were able to locate nine domains that matched the DGA patterns that were resolving in a three day window (one day before and one day after the cluster was detected). These domains resolved to two IP addresses.

Domains:

- c69l18byjzo31hwk67b68eqd10mug43o41lzpsdt.biz
- e21p12fsh24a57frj66lrgvhqjzb58pzpshwdt.org
- ernubri35gyc59l58ovbxaym39m49fzmuctb48.info
- guh54htm59k57kvksiqe41m69axbqcwbuf22g23.org
- h54lzm29g43csptazc59k57fqf22fqpvn20itgt.biz
- kwe41duevivixhsbtfvc19h24grjzbzi25pz.net
- mvhsexctb28ctawpqbsn30hzp42mrm29g43cu.org
- n40cqd20m29fujtm69e41c39nubyivjvasiymt.org
- ounyk47fyg43gvf52gwpvjso51lthveqfsow.info

IPs:

- 64.186.144.31
- 66.228.60.7

These IP addresses also hosted one other domain that appeared to be related to the criminals. The domain lsdjbfsb9u4b2jjdjkjdj(dot)com has historically been on several IP addresses with domains that match this DGA pattern. It appeared that this domain was related to a fake AV campaign run by the criminals involved in the DGA campaign we were tracking.

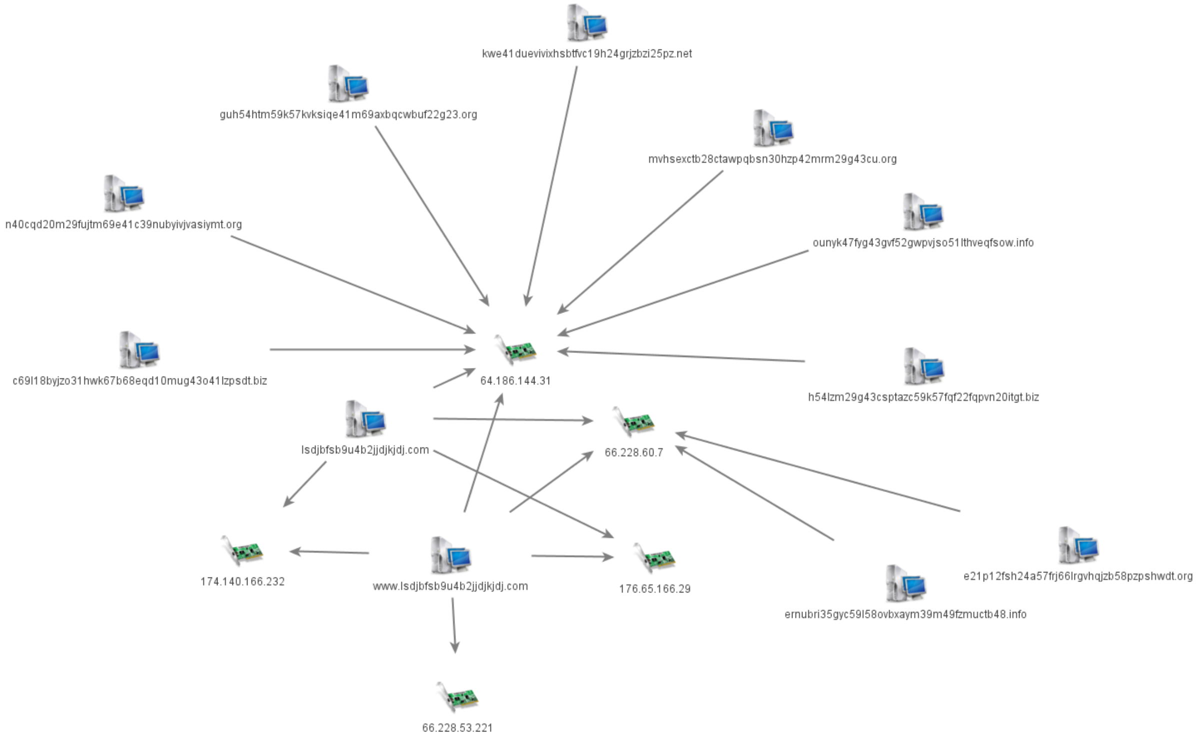

kwe41duevivixhsbtfvc19h24grjzbzi25pz.net
guh54htm59k57kvksiqe41m69axbqcwbuf22g23.org
mvhsexctb28ctawpqbsn30hzp42mrm29g43cu.org
n40cqd20m29fujtm69e41c39nubyivjvasiymt.org
ounyk47fyg43gvf52gwpvjso51lthveqfsow.info
c69l18byjzo31hwk67b68eqd10mug43o41lzpsdt.biz
64.186.144.31
h54lzm29g43csptazc59k57fqf22fqpvn20itgt.biz
lsdjbfsb9u4b2jjdjkjdj.com
66.228.60.7
e21p12fsh24a57frj66lrgvhqjzb58pzpshwdt.org
174.140.166.232
www.lsdjbfsb9u4b2jjdjkjdj.com
176.65.166.29
ernubri35gyc59l58ovbxaym39m49fzmuctb48.info
66.228.53.221

Figure 1: A View of the C&C Infrastructure

Using passive DNS data for these domains, we were able to create a link map that shows our understanding of the campaign using this DGA at the point in time that the domains were generated.

Searching for the domains in our malware database we were not able to find any malware using these domains.

From C&C to Infection

Now that we had the C&C domains, what could we find out about the malware and how it was distributed?

Looking at the WHOIS data for the domains, we found that most used a variety of privacy services but the criminals behind this campaign made a mistake and used a real email address (kiles(at)island.net) for four of them. Looking at WHOIS data, we were able to find two other domains that shared the same email address.

Looking for other information on the email address, it looked like a legitimate user that is unlikely to have any ties to this malware campaign. We did not follow this any further and currently believe this account may have been hijacked and used without his knowledge.

The exploit sites were hosting mostly Blackhole Exploit kits but at least one was running a Sakura kit. We found a pattern in the naming and were able to find a list of exploit sites.

- chillecash.com
- chillegraph.com
- sulusium.com
- sulusality.com
- sulusius.com
- sulusize.com
- vellaband.com
- vellastation.com
- vellalink.com
- vellaline.com

Now that we had the domains used to infect victims with the malware, it was easy to get a copy of the malware for analysis. We discovered that this malware was a Zeus version 3 variant that uses peer-to-peer (P2P) as its primary C&C channel and only resorts to the DGA-generated domains if it fails to connect to its P2P network. This explains why we couldn’t find any reporting on these domains linked to malware - all of the sandbox systems/data we had access to did not restrict access to the P2P channel and the malware never had to resort to its backup plan. More information on the malware can be found in the Malware Analysis Section.

From Infection to Distribution

We did some digging into the infector/exploit domains and found that they were linked to the popular Better Business Bureau (BBB) and NACHA spamming campaigns. The NACHA spam has been around for a while and we are not sure exactly when they switched to this version of Zeus. BBB spam is newer, having started in late November and appears to have used this version since the beginning.

Closing the Loop

The first reference we were able to find to this version of Zeus was in early October 2011. Our system was able to find DGA domains at least as far back as mid-September 2011. This means that we were able to see the malware activity before any other reporting existed. We were able to detect this as a Zeus/Murofet variant weeks before anyone knew about the malware or had a copy of it to analyze. At this point, we only had one more step to prove that our system properly grouped the NXDomains for this particular piece of malware. We ran the malware in our lab on a computer with the date set to the same day that we collected the domains that we were using to map them to malware. We found that the domains generated by the sample were a correct match for the domains in our passive DNS database.

Malware Analysis

Host Behavior

Malware Infection Routine

Upon execution of Contacts.EXE (MD5 = ccec69613c71d66f98abe9cc7e2e20ef) in a Windows 7 64 bit environment, the malware installs itself under a randomly named folder using a random filename in the following location:

C:\Users\<username>\AppData\Roaming\

For this specific case study, the malware that was dropped is named LEEGU.EXE (MD5 = 8f60afa9ea1e761edd49dfe012c22cbf) and the folder's name is AHAVW. The location where the malware is dropped is therefore:

C:\Users\<username>\AppData\Roaming\ahavw\leegu.exe

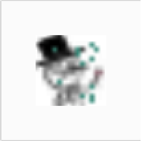


Figure 2: Leegu.EXE's icon sporting Lulzsec's logo

The significance of having the malware installed in the AppData\Roaming folder is to make it machine independent and roam with the user. This means that if the infection is within an enterprise environment, every time an infected user logs into a different machine, the malware becomes active on that machine.

To autostart, the malware added the following registry entry:

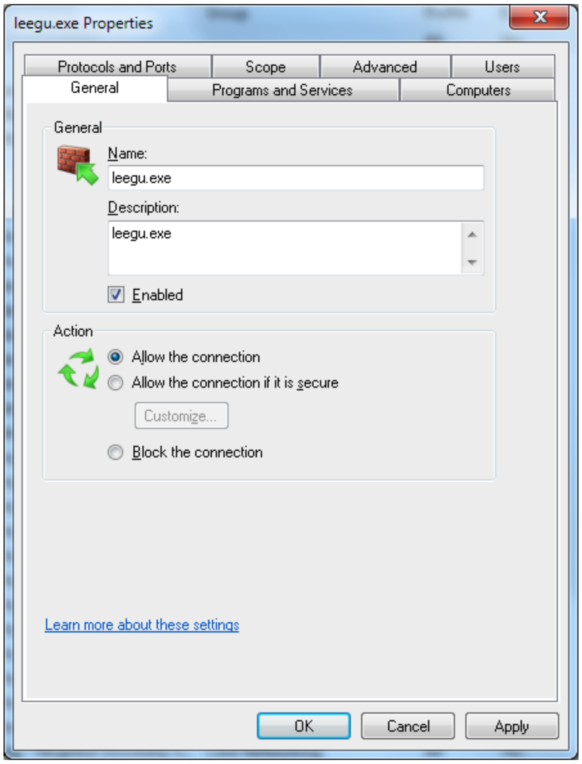
Key = HKCU\Software\Microsoft\Windows\CurrentVersion\Run

Name = {297DCA1C-2305-AD40-6DDB-F23F45FCD122}

Data = C:\Users\<username>\AppData\Roaming\ahavw\leegu.exe

The use of HKCU key makes the added startup entry applicable only to the currently logged-in user. Based on this registry modification and the location where the malware is dropped, the malware has geared itself to be a user specific infection rather than system specific.

The malware also creates new firewall rules to allow UDP and TCP inbound connection to the malicious application as seen in the figures below.



leegu.exe Properties

General

Name:
leegu.exe

Description:
leegu.exe

☑ Enabled

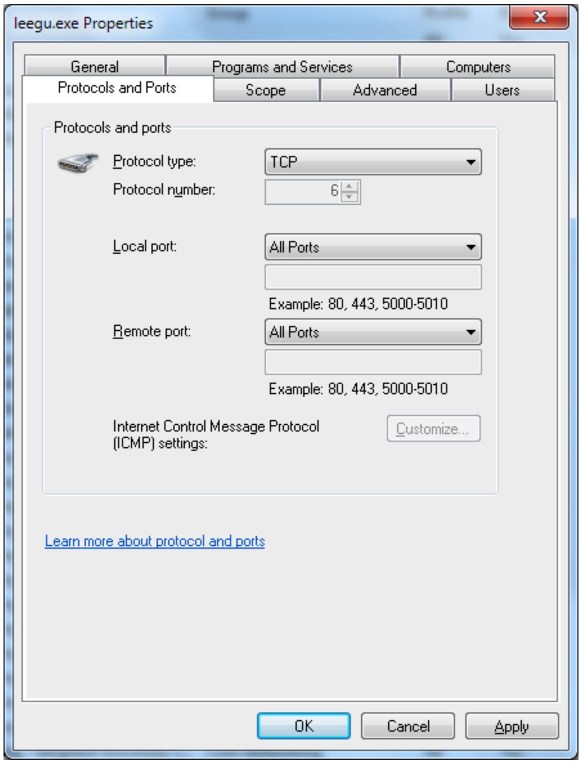
Action

● Allow the connection

○ Allow the connection if it is secure

○ Block the connection

Figure 3: Enabled Firewall Rule



leegu.exe Properties

Protocols and ports

Protocol type: TCP

Protocol number: 6

Local port: All Ports

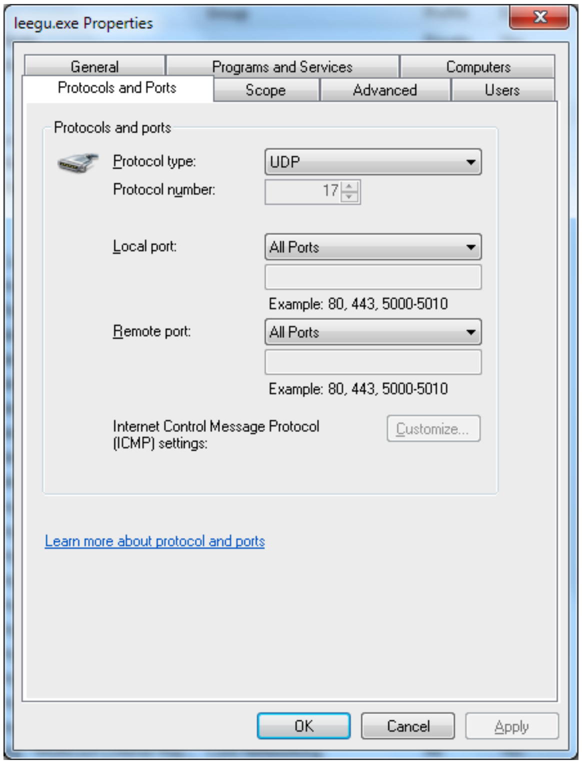
Example: 80, 443, 5000-5010

Remote port: All Ports

Example: 80, 443, 5000-5010

Internet Control Message Protocol (ICMP) settings:

Figure 4: Allow TCP / All Ports connections



Protocols and ports

Protocol type: UDP

Protocol number: 17

Local port: All Ports

Example: 80, 443, 5000-5010

Remote port: All Ports

Example: 80, 443, 5000-5010

Internet Control Message Protocol (ICMP) settings: Customize...

Figure 5: Allow UDP / All Ports connections

After setting up Leegu.EXE, Contacts.EXE deletes itself and passes control to Leegu.EXE.

AV Evasion Technique

It's easy to think that Contacts.EXE is a traditional dropper based on the malware infection routine. But looking closely into the samples reveals that Leegu.EXE is actually a polymorphic generation of Contacts.EXE. For every infection, a new polymorphic generation is produced. This results in no two dropped malware samples being alike. But as with other polymorphs, the form is different but the function remains the same.

Malware Intent

Further inspection of the malware reveals it to be a variant of Zeus. And as with other Zeus variants, it steals banking information and then exfiltrates the stolen information via encrypted HTTP POST.

Network Behavior

After the successful installation of the malware, it then reaches out to other infected hosts using P2P. If connection is successful, it gets updates from and shares information with these other bots. If no other infected hosts are found or if there is no P2P connection, the malware goes into plan B. This is revealed by blocking the appropriate UDP and TCP ports to disable the malware's P2P communication, thus, preventing it from communicating and exchanging information with its malware peers.

The malware's plan B is to connect to a centralized C&C using domains generated by a DGA. This ensures that even without peers, the malware would still be able to reach out and communicate with its criminal operator. One thing to note about plan B is its utility in enterprise environments where P2P is often blocked. But before the malware attempts connecting to any of these domains, which is oftentimes revealing, it first checks whether the compromised system has Internet connectivity by initiating connection to the following domains:

www.bing.com
www.google.com

If there is Internet connection, it tries to connect to DGA-generated domains. It cycles through the different domains until it finds a valid connection. Figure 6 summarizes the communication flow of the malware.

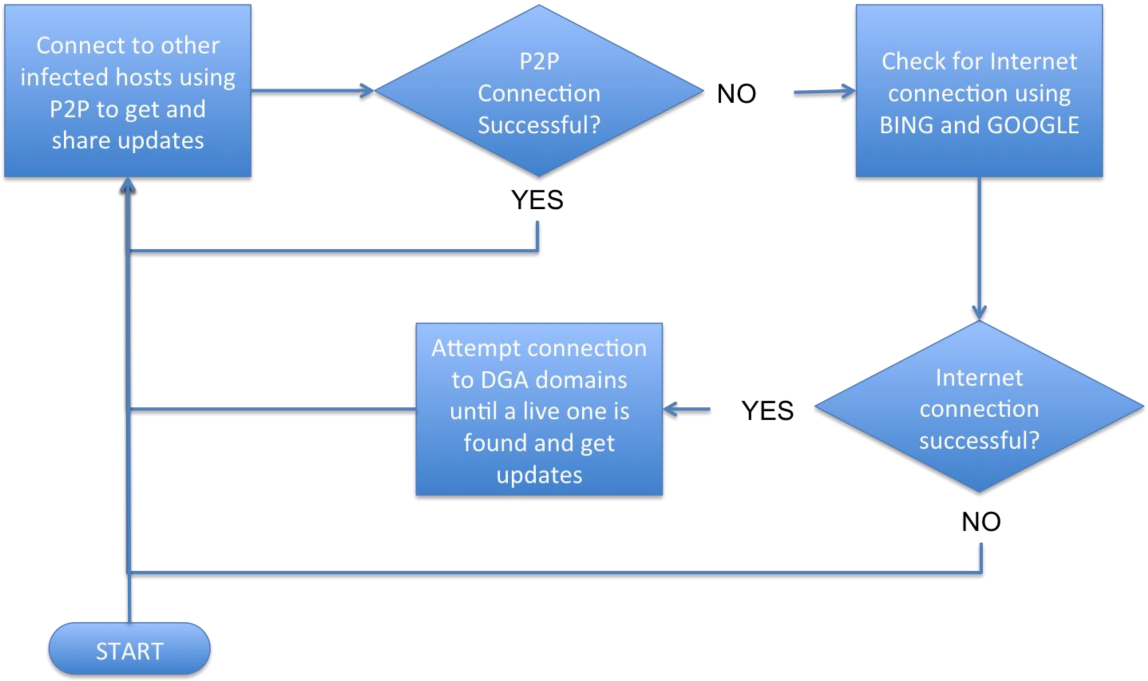

Connect to other infected hosts using P2P to get and share updates
P2P Connection Successful?
NO
Check for Internet connection using BING and GOOGLE
YES
Attempt connection to DGA domains until a live one is found and get updates
YES
Internet connection successful?
NO
START

Figure 6: Malware Communication Flow

P2P

Network traffic flow analysis was performed on the malware installed on both Windows XP and Windows 7. Looking initially at Windows 7, the malware makes an initial connection utilizing a P2P protocol to establish connections with other infected hosts. The initial connection is made to 118.171.x.x (Host A) using UDP port 11229. Host A subsequently sends 284K of updates to the local malware. A total of 50 P2P connections are then initiated by the malware to external hosts.

One of the connections to 118.233.x.x (Host B) results in an additional 390K download of updates to the malware.

A series of “high port to high port” TCP connections is attempted between the malware and three external hosts. One of these attempts is successful, resulting in several flows to TCP port 16359 to a second host in 118.233.x.x (Host C). The packet analysis shows HTTP headers with binary payloads.

The Windows XP analysis showed a number of similarities to Windows 7. This time, the malware downloaded 72 fewer octets of updates from previously mentioned Host A using the same UDP port 11229. A total of 88 P2P connections are then initiated by the malware, with one of connections resulting in an additional 389K download from 59.116.x.x, almost identical in size to that received from Host B in the Windows 7 test. As before, there was a high-port to high-port TCP connection between the malware and an external host (75.130.x.x, port 11968) that included HTTP headers and binary payloads.

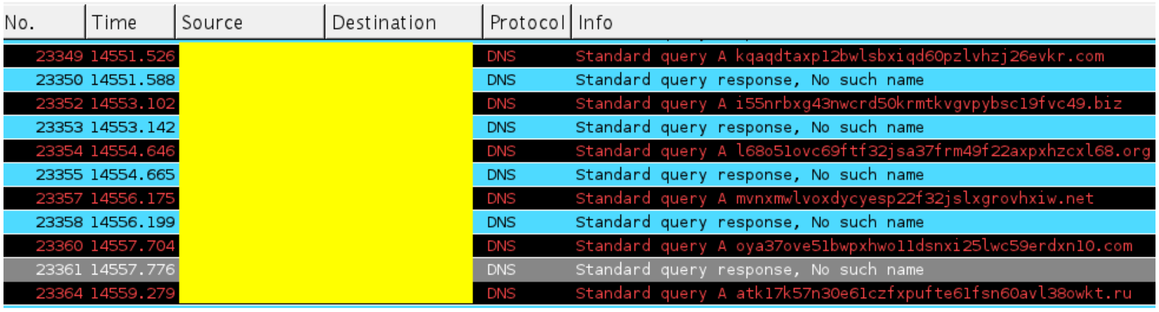
Examining both sets of data, nearly all of the first 15 connection attempts made by both tests were to the same destination IP addresses. Thereafter, the connection attempts to the same IP addresses occur in regular clusters, with connection attempts to other IPs breaking the groups.

Analysis of the details of the P2P protocol implementation is ongoing.

DGA

As a back up to the P2P component, the malware utilizes a centralized form of command structure. But instead of having a finite set of hard-coded domains to reach out to the C&C that can easily be identified and blocked, the malware utilizes a DGA to generate domains.

The DGA algorithm utilizes the current date as its ‘seed’ and generates approximately a thousand domains for a given day. The generated domains are made up of purely alphanumeric characters and utilize one of the following TLDs: com, biz, org, net, ru, and info.



No.	Time	Source	Destination	Protocol	Info
23349	14551.526	[redacted]	[redacted]	DNS	Standard query A kqaqdtaxp12bwlsbxiqd60pzlvhzj26evkr.com
23350	14551.588	[redacted]	[redacted]	DNS	Standard query response, No such name
23352	14553.102	[redacted]	[redacted]	DNS	Standard query A i55nrbxg43nwcrd50krmtkvgvpybsc19fvc49.biz
23353	14553.142	[redacted]	[redacted]	DNS	Standard query response, No such name
23354	14554.646	[redacted]	[redacted]	DNS	Standard query A l68o51ovc69ftf32jsa37frm49f22axpxhzcxl68.org
23355	14554.665	[redacted]	[redacted]	DNS	Standard query response, No such name
23357	14556.175	[redacted]	[redacted]	DNS	Standard query A mvnxmwlvoxdycyesp22f32jslxgrovhxiw.net
23358	14556.199	[redacted]	[redacted]	DNS	Standard query response, No such name
23360	14557.704	[redacted]	[redacted]	DNS	Standard query A oya37ove51bwpxhwo11dsnxi25lwc59erdxn10.com
23361	14557.776	[redacted]	[redacted]	DNS	Standard query response, No such name
23364	14559.279	[redacted]	[redacted]	DNS	Standard query A atk17k57n30e61czfxpufte61fsn60avl38owkt.ru

Figure 7: Malware reaching out to DGA domains

Summary

Damballa Labs was able to find and identify a cluster of NXDomains that led us down a path where we were able to not only find the working C&C domains and malware but also the activities that led to the infections. We were able to use the system and process we created to find everything all the way back to the spam campaign that was distributing it. And we were able to find the domains related to the criminal activity weeks before the first malware sample was found and analyzed. The example we used in this case study was typical in many ways to the behavior seen in other DGA-based campaigns but the use of the DGA only as its plan B made these domains a little more elusive.

The automated systems and techniques used for this case study are now in production operation in Damballa Labs. Damballa provides the earliest possible detection of DGA-based threats, and the detection is not dependant on actually having malware samples. This is now an integral part of Damballa FirstAlert, the cyber threat intelligence system that powers the Damballa Failsafe and Damballa CSP products. Damballa FirstAlert provides enterprise, ISP and telco security teams the earliest possible protection from advanced malware, botnets and persistent threats.

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

Prepared by:
Damballa Inc.
http://www.damballa.com