https://blog.crysys.hu/2013/02/miniduke/

February 27, 2013

Miniduke

Earlier in February 2013, FireEye announced the discovery of a new malware that exploited a 0-day vulnerability in Adobe Reader. Now, we announce another, as yet unknown malware that exploits the same Adobe Reader vulnerability (CVE-2013-0640).

This new malware was named Miniduke by Kaspersky Labs with whom we carried out its first analysis. Our participation in this research was justified by a detected Hungarian incident. A detailed report on the results of our joint efforts has been published by Kaspersky Labs on their Securelist blog site. That report describes what we currently know about the operation of Miniduke including its stages, and also information on the C&C infrastructure and communications. We have published another report from CrySys Lab that contains information on the indicators of Miniduke infections and gives specific hints on its detection. This blog entry is a brief excerpt of our report.

The available malware samples are highly obfuscated, and compiled by a polymorphic compiler. The attackers were able to produce new variants with only a few minutes difference between compile times. Therefore, the number of distinct samples could be very large. Hashes of known samples are published in our detailed report on indicators.

Due to a large number of compiled samples, there is a high chance that the current version is difficult to detect by signatures. Yet, there are common features in the samples that can be used to identify the malware components.

In every case we encountered, the “Program Files/Startup” contains a file with .lnk extension after installation. This is used to start up the malware after the computer is rebooted.

A not fully cross-checked information is that, during installation, the malware will be copied in two copies on the system and the two executables differ. This might mean that the executable modifies itself. For example, we recovered the following two files:

md5sum base.cat :113e6fc85317fdd135e3f5f19e6c7a58 *base.cat

md5sum ~6rld.tmp : c786a4cdfe08dbe7c64972a14669c4d1 *~6rld.tmp

where base.cat is the startup file, which is created based on ~6lrd.tmp. base.cat is stored in the “All users” directory, whereas ~6lrd.tmp is stored in a user’s directory, e.g., in the guest user directory as “C:\Documents and Settings\guest\Local Settings\Application Data\~6rld.tmp”. This user directory contains at least one more file, update.cmd, with a specific content that could be used for detection.

As for stage 3 of the attack, it is important to note that it is not yet analyzed deeply. So once a victim downloads the ~300k long piece of stage 3 code, we don’t know what happens with the previous stages, and we have no information about detections once this stage is reached, except the usage of the C&C server news.grouptumbler.com. Another variant of the stage 3 code is much smaller, only 14k long, and connects to a server in Turkey.

We have identified the following servers delivering stage 2 and stage 3 code to victims:

arabooks.ch 194.38.160.153 / Switzerland

artas.org 95.128.72.24 / France

tsoftonline.com 72.34.47.186 / United States

www.eamtm.com 188.40.99.143 / Germany

The C&C server used by stage 3 of the malware is news.grouptumbler.com (IP 200.63.46.23) and it is located in Panama.

There are multiple layers of C&C communications in the malware. First, the malware uses Google search to receive information from its master. Then, it uses the Twitter messaging service looking for the twits of a specific Twitter user. Commands received via this channel trigger the download of stage 2 and stage 3 code.

Basic detection can be based on the queries that are initiated by the victim computer within seconds:

www.google.com – port TCP/80 - HTTP

twitter.com –port TCP/443 - SSL

www.geoiptool.com –port TCP/80 - HTTP

Known search strings in Google search can also be used to detect the malware:

lUFEfiHKljfLKWPR

HkyeiIDKiroLaKYr

lUFEfiHKDroLaKYr

Unfortunately, these strings are most likely unique to each C&C server or victim, thus unknown samples might use other strings, but possibly with the same length.

Examples for twits containing the URL of the C&C server are shown below:

The weather is good today. Sunny! uri!wp07VkkxYt3Mne5uiDkz4Il/Iw48Ge/EWg==

Albert, my cousin. He is working hard. uri!wp07VkkxYmfNkwN2nBmx4ch/Iu2c+GJow39HbphL

My native town was ruined by tornado. uri!wp07VkkxYt3Md/JOnLhzRL2FJjY8l2It

The malware also sends a query to the geoiptool. An example is shown below:

GET / HTTP/1.1

User-Agent: Mozilla/5.0 (compatible; MSIE 7.0; Windows NT 6.0; en-US; Trident/5.0)

Host: www.geoiptool.com

The malware retrieves the URL of the stage 2/3 delivery C&C server from Twitter messages as described above. Then, we can observe the first query from the victim towards the server. This query contains pure HTTP traffic on port 80 to the server following the template below.

GET /original/path/shortname/index.php?e=aaaaaaaaa

where:

shortname can be a number of strings, generally human readable (e.g. lib, engine, forum, forumengine etc.)

“e=” is not constant, can be anything, but generally 1-2 letters long

aaaaaaaa stands for some Base64-like text (see details below)

the servers used are assumed to be legitimate sites, just hacked by the attackers.

Based on this format, we can detect a valid query as follows:

The name of the first GET parameter should be discarded

this means “e=” is not important

we saw only one GET parameter, queries with multiple parameters are likely not used

For detection, the Base64-like string “aaa…” should be first modified as follows:

“-” should be replaced by “+”

“_” should be replaced by “/”

This results in correct Base64 encoding, which can be decoded with library functions such as base64_decode. After decoding, a string of data, partially binary, will be available. Parts are separated by the delimiter character “|”. The format and a numerical example are below:

binary data (~100 bytes)|numerical ID (~10 digits)|version number

e.g.,

binary data|5551115551|1.13

As the binary data itself may contain the ”|” character, parsing should start from the end (i.e., the numerical ID starts from the second “|” character from the end). In additional, the ID length may vary (not fully confirmed), but it seems to be around 10 digits. Finally, the version number always follows the pattern “one digit.two digits”, e.g., 1.1X 3.1X.

gif-image

The C&C server’s response – if it sends encrypted files – is a GIF file containing a small icon, and after that, the malware. For stage 3, the file downloaded has a larger size (~300KB). It also begins with a GIF header, but that header is only 13 bytes long, and then starts the encrypted executable (see picture above).

ID	Name	Description
T1027	Obfuscated Files or Information	The available malware samples are highly obfuscated, and compiled by a polymorphic compiler.
T1314	Host-based hiding techniques	The available malware samples are highly obfuscated, and compiled by a polymorphic compiler.
T1037.001	Boot or Logon Initialization Scripts: Logon Script (Windows)	In every case we encountered, the “Program Files/Startup” contains a file with .lnk extension after installation. This is used to start up the malware after the computer is rebooted.
T1102.002	Web Service: Bidirectional Communication	There are multiple layers of C&C communications in the malware. First, the malware uses Google search to receive information from its master. Then, it uses the Twitter messaging service looking for the twits of a specific Twitter user. Commands received via this channel trigger the download of stage 2 and stage 3 code.
T1071.001	Application Layer Protocol: Web Protocols	There are multiple layers of C&C communications in the malware. First, the malware uses Google search to receive information from its master. Then, it uses the Twitter messaging service looking for the twits of a specific Twitter user. Commands received via this channel trigger the download of stage 2 and stage 3 code.
T1334	Compromise 3rd party infrastructure to support delivery	the servers used are assumed to be legitimate sites, just hacked by the attackers.
T1312	Compromise 3rd party infrastructure to support delivery	the servers used are assumed to be legitimate sites, just hacked by the attackers.
T1001	Data Obfuscation	The C&C server’s response – if it sends encrypted files – is a GIF file containing a small icon, and after that, the malware. For stage 3, the file downloaded has a larger size (~300KB). It also begins with a GIF header, but that header is only 13 bytes long, and then starts the encrypted executable (see picture above).
T1027	Obfuscated Files or Information	The C&C server’s response – if it sends encrypted files – is a GIF file containing a small icon, and after that, the malware. For stage 3, the file downloaded has a larger size (~300KB). It also begins with a GIF header, but that header is only 13 bytes long, and then starts the encrypted executable (see picture above).