1. A deeper look into ScanBox

Please e-mail us at threatintelligence@uk.pwc.com for a version of this report with additional indicators that you are welcome to distribute so long as it is not on public channels (TLP-GREEN).

We have observed actors amending the ScanBox framework to evade existing public signatures, detailed below.

Overview

Security researchers have often made the mistake of assuming that when a specific tool was observed being used in espionage attacks, it was representative of activity of a single actor. More frequently, however, many are now identifying that distinct groups of attackers are sharing their toolsets, just as in the cybercrime world.

One such toolset, the ScanBox framework, is now shared between a number of groups who conduct espionage attacks. Evidence suggests that these groups include those behind the recent Forbes and Anthem attacks. This short paper outlines our current perspectives on the previously discussed espionage groups currently using the framework and a hint that a 5th player is getting in on the game.

ScanBox performs keylogging of users when they visit a compromised website, without requiring malware to be deployed, and can collect a great deal of information which can be used to tailor future attacks.

In October we published some details of the ScanBox tool set. Since then we have encountered 24 additional sites compromised with the framework. Over this time we have observed changes to the code and novel techniques for executing.

We have also received a number of tip offs from other researchers, as well as queries from victims who were directly targeted by those using the framework. We would like to extend our thanks to these individuals for their contributions towards this research.
Who’s using it, and who’s being targeted?

The following diagram shows the links in tools and targets between the groups discussed in our previous blog, but newer information has since come to light which allows us to more accurately associate these groups with known threat actors:

Between these clusters, we've seen strategic web compromises designed to target users in the following countries:
Variations on the framework

Since our last post there have been several alterations to the ScanBox code base, including new modules, changes to avoid signature based detection, as well as extra techniques to try to identify whether those being scanned are real machines or researchers.

Fears of proliferation

In some cases we have been able to watch developers update and test variants of the framework, and even come across server-side code being tested by budding hackers.

Our findings are detailed below.
2. Updates to the ScanBox Framework

Following on from our previous post on ScanBox\(^1\), we have watched the clusters of activity outlined with close interest, as well as keeping an eye on new adopters of the ScanBox framework.

For those who didn’t read our last entry on the framework, nor the excellent work by Jaime Blasco\(^2\) which preceded it, ScanBox is a framework written in JavaScript and PHP which allows an attacker to perform reconnaissance and key logging of visitors to compromised websites without requiring any malware to be downloaded or installed.

The framework has remained in use since initial analyses were published, and further analysis of the code, public reporting, as well as the infrastructure used to host ScanBox infections has given us a better picture of some of the clusters of activity we identified in our earlier blog.

Technical Updates

In addition to the four websites we previously identified hosting the malicious code, we’ve now identified a further 24 websites hosting the framework. Anonymised data relating the countries and sectors affected are given in appendix A.

Broadly, the ScanBox framework codebase has remained the same, however there are slight nuances in some aspects of the code, or in the software attackers choose to search for.

Software checks

In cases where the attackers have included software checks within their ScanBox code, it tends to be for the same original list of filenames (Appendix C). In some cases the attackers customise the list, presumably based on the things they’re expecting to find. For example, adding or removing additional security products based on the predominant software providers in their target region.

It’s also worth noting that the standard list includes quite a lot of software which is less relevant to security (examples include WinRAR, iTunes and WinZip). Some of these may be included in order to help the attackers to try and identify real victims vs researchers/sandboxes/honeypots. An example list is shown in the following screenshot:

---


In one cluster of ScanBox activity, instead of checking for specific files as per the method above, the attackers have implemented a different method of tracking files and drives present on the victim machine.

The first piece of new functionality is that the attackers build a possible list of drive names (A-Z) and scan for the existence of each drive:
function checkDrives(filepath)
{
    var txt = String.fromCharCode(69,63,120,109,108,32,118,101,114,115,105,111,110,61,34,49,46)
    var xmlDoc = new ActiveXObject("Microsoft.XMLDOM");
    xmlDoc.async = true;
    try
    {
        xmlDoc.loadXML(txt);
    } catch (e)
    {
        var t=filepath;
        ie_drives.push(t);
    }
}

checkFolders

They do not check files directly either, instead, checking for the presence of a list of folder names:

function checkFolders(filepath)
{
    var txt = String.fromCharCode(69,63,120,109,108,32,118,101,114,115,105,111,110,61,34,49,46)
    var xmlDoc = new ActiveXObject("Microsoft.XMLDOM");
    xmlDoc.async = true;
    try
    {
        xmlDoc.loadXML(txt);
    } catch (e)
    {
        var t=filepath;
        folders_res.push(t);
    }
}
In this case, rather than determining what security software is present, this check would mainly be useful in assisting the attacker with identifying the victim’s operating system. This could then be used to tailor future attacks (i.e. should I deploy malware which can bypass UAC? Should I send malicious documents targeting CVE-2012-0158 or not?).

Also, bizarrely the attackers check the path ‘Program Files (x80236)’ – if anyone knows what this corresponds to, please get in touch and let us know.

**Avoiding analysis?**

As we stated earlier, some of the features of the original ScanBox code were probably designed to help those analysing results distinguish between honeypots/analysis environments and real-world victims. The newly added variables include:

- colorDepth – This may help to identify virtual machines which are typically configured with specific graphics options.
- Local Time on the machine – checking that the local time on the machine matches the expected time given the geo-location of the infected IP address – in many cases analysis environments are not configured with the correct time.

**Updates to evade signatures**

Possibly in response to our previous report, we’ve seen some of the groups using ScanBox alter the content of the modules to evade detection. Attackers do read reports, both to help them attack and also to evade detection.

One change has been the URI formats used to deliver stolen key log data from obvious URLs such as:

/k.php?data=[KeyloggerData]

To more subtle URLs such as:

/[KeyloggerData].jpg

We also note that following our release of a signature to detect the phrase “No Java or Disable”, which was present in a number of related frameworks, some of the attackers have now changed this to “No or Disable”.

---

3 For example as in http://pwcblogs.pwc.com/cyber_security_updates/2015/01/destructive-malware.html
3. Previously... on ScanBox...

Clusters

In our previous entry on ScanBox we described four clusters of activity – this section includes updates on those clusters.

Previously, we were only able to cluster activity based on the infrastructure used and the associated malware, we now have a sufficient number of samples to cluster based on differences in implementation between the code used for ScanBox in each case.

Flash Cluster (aka Cluster 1)

The recently published FBI flash alert #A-000049-MW⁴, cited domain names previously referenced in our blog as being related to an actor known as Deep Panda, which we’ll get to in a minute. The link appears to have been made from the use of the DerUsbi malware family though, which we know to be used by several espionage actors.

What is interesting is the potential overlap between the target selection of this cluster and the targets of those behind the recently reported Forbes compromise⁵,⁶.

Checking the Google SafeBrowsing results for *.googlecaches.com, shows that the domain was used for distribution of malicious code for a significant period after our previous blog entry:

```
Diagnostic page for googlecaches.com

What happened when Google visited this site?
Of the 3 pages we tested on the site over the past 90 days, 0 page(s) resulted in malicious software being downloaded and installed without user consent. The last time Google visited this site was on 2014-12-12, and the last time suspicious content was found on this site was on 2014-12-12.
Malicious software includes 213 trojan(s), 11 exploit(s).
This site was hosted on 2 network(s) including AS32097 (WIL-KC), AS58879 (ANCHNET).

Has this site hosted malware?
Yes, this site has hosted malicious software over the past 90 days. It infected 15 domain(s), including gokbayrak.com, sefc.com.hk, turkkorseyi.com.
```

This includes distribution via gokbayrak.com. Looking at whether any other domains or IP addresses were observed delivering malicious software via gokbayrak.com shows that it was also observed delivering malware via 88.80.190[.]133. This is the same IP address that was cited in iSight’s reporting of the Forbes breach.

---

In fact, the IP address 88.80.190.133 was involved in the compromise of the same 3 websites as SafeBrowsing shows were affected by googlecaches.com.

The same three sites targeted with the 0-day used in the Forbes attack were also observed distributing malware via googlecaches[.]com - we therefore believe it’s likely that the group we previously described as ‘Cluster 1’ was behind the Forbes compromise.

We can use similar techniques to explore other actors who have the same tasking as Cluster 1. In the screenshot of websites delivering malware via gokbayrak.com, we saw that the domain name macanna.com[.]tw was also used to distribute malware from this page.
Whilst macanna.com.tw appears to be a legitimate site, it has also been observed as being a command and control destination for malware. The malware sample in question has a hash of 3b8d7732de3b3e8823d241e7cd3185c4. The same sample also communicates with happynewyear.dns04.com, which in turn resolves to the IP address 115.23.172.151, which hosts a large number of other malicious host names:
These following hostnames are activity associated with the actor best known as TH3Bug⁷ - named after their choice of Poison Ivy password. Their malware samples are present in the same cluster:

Deep Cluster (aka Cluster 2, centred on news.foundationssl[.]com)

This cluster relates to the threat actor referred to as Deep Panda by CrowdStrike, as was confirmed in a recent blog post\(^8\). In turn, this is widely believed to relate to the incident at Anthem, as described in a Krebs post\(^9\).

The graph below shows the links between the we11point.com domain name, and news.foundationssl[.]com as seen on the CrowdStrike blogpost:

![Graph showing links between domains](image)

The Krebs article also points toward other possible (although not explicit) links between the domain allegedly used in the Anthem hack (we11point.com) to Cluster 2 through shared WHOIS details, as we11point.com was registered by domain re-seller ‘li2384826402@yahoo.com’

On its own, this would not be sufficient to associate the two clusters, but it is useful to note as a ‘softer’ overlap.

---


Mystery Cluster 3 (aka Cluster 3, centered on qooq1e[.]com):

Cluster 3 remains a mystery, unfortunately the code used in this instance is the most slim line version, and has since not been widely re-used – it is unclear who was behind the compromise using this domain name.

Evil Cluster (aka Cluster 4, centred on webmailgoogle[.]com):

We’d previously missed the link between Cluster 4 – and malware widely known as ‘EvilGrab’ or ‘Vidgrab’. From our view point, this malware is exclusively used by one group, known by CrowdStrike as Stone Panda\(^\text{10}\).

In addition to the four clusters outlined above, within the 24 additional compromises identified, we believe there are at least 3 other distinct groups using the framework.

**Digital Quartermastering**

In their 2013 paper ‘From Quartermaster to Sunshop’\(^\text{11}\), FireEye described the concept of a Digital Quartermaster, a kind of malware supply chain for intrusions, where a skilled team would develop toolsets for a range of attackers who deploy them. The shared use of ScanBox may match up quite well to this hypothesis, and indeed even to some extent the naming schema overlaps, as iSight refer to the actor behind the Forbes breach as Codoso, but suggest they are publically known as ‘Sunshop’. In all likelihood this relates to a series of blogs by FireEye which refer to a series of web compromises in 2013 as being the SunShop\(^\text{12}\) campaign.

Although we did not notice the correlation immediately, there is good overlap between the groups we’ve described above, and the clusters of activity described as sharing a Flash 0-day in early 2014 by Symantec\(^\text{13}\). Although other groups have since begun using the framework, the first groups to use the framework (clusters 1-4) correspond quite nicely to the existing Symantec blog. For reference, we’ve overlaid our ScanBox clusters against the likely clusters Symantec created behind the scenes for their blog, as well as other popular names:

<table>
<thead>
<tr>
<th>ScanBox</th>
<th>CrowdStrike</th>
<th>Symantec</th>
<th>Other Names</th>
<th>Vulnerabilities &amp; Frameworks</th>
<th>Malware Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>???</td>
<td>Sunshop (iSight)</td>
<td></td>
<td>ScanBox CVE-2014-6332</td>
<td>Briba, Poison Ivy</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>Deep Panda</td>
<td>Sakurel</td>
<td></td>
<td>ScanBox CVE-2014-0322 (Internet Explorer)</td>
<td>Sakurel, DerUsbi, many others</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CVE-2012-4792 (Internet Explorer)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CVE-2014-0502 (Adobe Flash)</td>
</tr>
</tbody>
</table>

\(^{10}\) [http://www.slideshare.net/CrowdStrike/crowd-casts-monthly-you-have-an-adversary-problem](http://www.slideshare.net/CrowdStrike/crowd-casts-monthly-you-have-an-adversary-problem)

\(^{11}\) [https://www.fireeye.com/content/dam/legacy/resources/pdfs/fireeye-malware-supply-chain.pdf](https://www.fireeye.com/content/dam/legacy/resources/pdfs/fireeye-malware-supply-chain.pdf)

\(^{12}\) [https://www.fireeye.com/blog/threat-research/2013/05/ready-for-summer-the-sunshop-campaign.html](https://www.fireeye.com/blog/threat-research/2013/05/ready-for-summer-the-sunshop-campaign.html)

Please note that each vendor has their own way of grouping activity together, so these mappings are given on a best efforts basis.

**Player 5 has entered the game....**

In all examples of ScanBox deployments discussed so far, we believe that the scripts were deployed to anyone who visited websites of interest to a given sector which the attacker was able to compromise—and that the attackers waited for victims to visit the compromised site. However one group of attackers using the ScanBox framework are now actively sending e-mails to potential victims, where the e-mails contain links to websites hosting ScanBox. We believe these attackers are not covered by the existing clusters 1 through 4, as the code differs from that used elsewhere.

This method of sending links rather than waiting for visiting a specific compromised website has two main advantages:

- The advantage that the attacker doesn’t have to compromise sites that are relevant to the sector they wish to perform reconnaissance against, so generally this will make things easier for the attacker; and,
- The attacker has to deal with fewer false positives in terms of data received from victims. Even good IP whitelisting techniques will result in some false positives, by controlling the visitors however they can ensure only those they want to scan are scanned.

This group differs from the others based on the following characteristics:

- They send e-mails with links to compromised websites, rather than compromising sites of interest to their targets. Attackers send links to victims using a similar technique to that described in our Sofacy phishing blog\(^1\), where multiple redirects are deployed, one being a decoy, the other in this case loading ScanBox:

- Uses the ‘checkFolders’ function within the ScanBox code, rather than explicitly checking for files; and,
- Hosts the ScanBox code on the same page they have compromised rather than on a 3rd party IP address or domain name owned by the attacker.

So far we have identified four low key websites, all belonging to small companies based in the United States or Canada which are being abused in this fashion.

### 3.1. Going behind enemy lines – fears of proliferation and upcoming attacks

At one point during our investigation into infrastructure hosting ScanBox code, we identified a server which appeared to be used for development and testing purposes. On this occasion, the server side code was publicly accessible, allowing us to gain insights into the development and testing phase of an attack using ScanBox. This also included the ScanBox framework’s own detailed reconnaissance against the developer themselves.

We noted the developer repeatedly uploading the modified versions to VirusTotal, presumably in an attempt to improve evasion of anti-virus.

We do not believe this developer is part of the core group that has access to the original implementation, but is instead another actor, who is likely rebuilding ScanBox from samples they find online.
The screenshot on the left is from a version currently in development by a possible attacker, the corresponding screenshot on the right is from a public article15.

15 http://www.cnxhacker.com/2015/01/19/6412.html
4. Conclusion

The publication of threat information allows us to draw links between different campaigns, tools and malware but we need to be careful about which links we consider to be significant and ensure we’re confident in how information that’s publically available was derived. The summary above is just our view of the overlaps in web based tools/exploits and targets between different threat actors, but those with different datasets may be able to draw different conclusions.

Last time, we identified three possible hypotheses to explain the overlap between the ScanBox users, in this blog, based on the data we have available, we can settle on just one of these conclusions:

“2. Selections of actors share some resources, as per previous observations with similar kits by some security vendors.”

Specifically, our key conclusions are:

- **[High Confidence]** - The DQM theory presented by FireEye and later explored by Symantec in 2014 about likely tool and exploit sharing between a specific set of groups continues to hold true, we can only speculate as to the nature of the relationships organisationally between these groups.

- **[Medium Confidence]** - We currently believe that the activity represented by Cluster 1 and the activity related to Th3Bug are distinct, but that there is overlap in who they are tasked to target.

- **[Low confidence]** – Th3Bug is one of the other actors who is in receipt of the same shared resource pool as those clusters already identified in this blog.

What is not clear is why specific resources (web-based exploits) appear to be shared, and why others (primarily malware families) are sometimes kept within a specific cluster.
5. Signatures

Snort Signatures – TLP WHITE


alert tcp $EXTERNAL_NET any -> $HOME_NET any (msg:"--[PwC CTD] -- MultiGroup - ScanBox and Targetted Watering Holes Content (var version\;var ax\;var e\;try{axo=new ActiveXObject)"); flow:established,from_server; file_data; content:"var version\;var ax\;var e\;try(axo=new ActiveXObject”; classtype:trojan-activity; reference:url,pwc.blogs.com/cyber_security_updates/2014/10/scanbox-framework-whos-affected-and-whos-using-it-1.html; metadata:tlp WHITE,author CDD; sid:xxxxxx; rev:2015021901;

alert tcp $EXTERNAL_NET any -> $HOME_NET any (msg:"--[PwC CTD] -- MultiGroup - ScanBox and Targetted Watering Holes Content (document.getElementsByTagName('head').item(0).appendChild(form_tag)"); flow:established,from_server; file_data; content:"document.getElementsByTagName('head').item(0).appendChild(form_tag)"; classtype:trojan-activity; reference:url,pwc.blogs.com/cyber_security_updates/2014/10/scanbox-framework-whos-affected-and-whos-using-it-1.html; metadata:tlp WHITE,author CDD; sid:xxxxxx; rev:2015021901;

alert tcp $EXTERNAL_NET any -> $HOME_NET any (msg:"--[PwC CTD] -- MultiGroup - ScanBox and Targetted Groups Watering Hole Code (Chr(CInt(ns(i)) Xor n))"); flow:established,from_server; file_data; content:"Chr(CInt(ns(i)) Xor n)"; classtype:trojan-activity; reference:url,pwc.blogs.com/cyber_security_updates/2014/10/scanbox-framework-whos-affected-and-whos-using-it-1.html; metadata:tlp WHITE,author CDD; sid:xxxxxx; rev:2015021901;

6. Appendices

Appendix A – ScanBox Sites

Where the site was referenced via phishing as opposed to ‘Watering Hole’ based activity, it has been excluded from the following table.

Where we have an assigned cluster, but have not discussed it in this document, we have given ‘Cluster [Letter]’ these are not intended as names for groups. Where we do not have an associated group we have listed ‘unknown’ under this field.

<table>
<thead>
<tr>
<th>Country</th>
<th>Sector/Target</th>
<th>Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN</td>
<td>Uyghur</td>
<td>Cluster 1</td>
</tr>
<tr>
<td>US</td>
<td>Think Tank</td>
<td>Cluster 2</td>
</tr>
<tr>
<td>US</td>
<td>Think Tank</td>
<td>Cluster 2</td>
</tr>
<tr>
<td>US</td>
<td>Think Tank</td>
<td>Cluster 2</td>
</tr>
<tr>
<td>KR</td>
<td>Hospitality</td>
<td>Cluster 3</td>
</tr>
<tr>
<td>JP</td>
<td>Industrial Sector</td>
<td>Cluster 4</td>
</tr>
<tr>
<td>GB</td>
<td>Chemicals</td>
<td>Cluster 4</td>
</tr>
<tr>
<td>JP</td>
<td>Geological Surveying</td>
<td>Cluster 4</td>
</tr>
<tr>
<td>VN</td>
<td>Government</td>
<td>Cluster A</td>
</tr>
<tr>
<td>JP</td>
<td>Education</td>
<td>Cluster A</td>
</tr>
<tr>
<td>JP</td>
<td>Geological Surveying</td>
<td>Cluster A</td>
</tr>
<tr>
<td>MN</td>
<td>Government</td>
<td>Cluster B</td>
</tr>
<tr>
<td>MM</td>
<td>Government</td>
<td>Cluster C</td>
</tr>
<tr>
<td>MN</td>
<td>Media</td>
<td>Unknown</td>
</tr>
<tr>
<td>MN</td>
<td>Media</td>
<td>Unknown</td>
</tr>
<tr>
<td>CN</td>
<td>NGO</td>
<td>Unknown</td>
</tr>
<tr>
<td>VN</td>
<td>Government</td>
<td>Unknown</td>
</tr>
<tr>
<td>AU</td>
<td>Government</td>
<td>Unknown</td>
</tr>
<tr>
<td>CN</td>
<td>Technology</td>
<td>Unknown</td>
</tr>
</tbody>
</table>
### Appendix B – Indicators of Compromise – TLP:WHITE

This table only includes related new single value IoCs which were not already published in our previous blog, which we are happy to share at TLP:WHITE

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Value</th>
<th>Artefact type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>1.9.5.38</td>
<td>IP Address</td>
</tr>
<tr>
<td>Cluster 1</td>
<td>103.255.61.227</td>
<td>IP Address</td>
</tr>
<tr>
<td>Cluster 1</td>
<td>118.193.153.221</td>
<td>IP Address</td>
</tr>
<tr>
<td>Cluster 1</td>
<td>118.193.153.227</td>
<td>IP Address</td>
</tr>
<tr>
<td>Cluster 1</td>
<td>174.121.122.73</td>
<td>IP Address</td>
</tr>
<tr>
<td>Cluster 1</td>
<td>4639c30b3666eb1b3927d5579790a88bff68e8137f18241f4693e0d4539c608</td>
<td>Malware Hash</td>
</tr>
<tr>
<td>Cluster 1</td>
<td>809959f390d5a49c8999ad6ff27fd92f1b2bo</td>
<td>Malware Hash</td>
</tr>
<tr>
<td>Cluster 1</td>
<td>ab58b6a7de25d8f6e4b70a24e0ccede0d5f6129df02a9e61293e1d17649a2</td>
<td>Malware Hash</td>
</tr>
<tr>
<td>Cluster 1</td>
<td>e8a8f8e39040fe3695217b4e4f1316177d675ed</td>
<td>Malware Hash</td>
</tr>
<tr>
<td>Cluster 1</td>
<td>file.googlecaches.com</td>
<td>Hostname</td>
</tr>
<tr>
<td>Cluster 1</td>
<td>gtm.googlecaches.com</td>
<td>Hostname</td>
</tr>
<tr>
<td>Cluster 1</td>
<td>js.googleapiscache.com</td>
<td>Hostname</td>
</tr>
<tr>
<td>Cluster 1</td>
<td>owa.outlookssl.com</td>
<td>Hostname</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>122.10.10.161</td>
<td>IP Address</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>204.152.199.43</td>
<td>IP Address</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>50.2.24.211</td>
<td>IP Address</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>bak.mailaunch.com</td>
<td>Hostname</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>f1890cc9d6dc8402142683406339453941468d8</td>
<td>Malware Hash</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>us-mg6.mail.yahoo.mailaunch.com</td>
<td>Hostname</td>
</tr>
</tbody>
</table>

### Appendix C – Standard software list detected by ScanBox

7z
AhnLab_V3
antifyfx
a-squared
avg2012
avira
Bit9
bitdefender_2013
BkavHome
COMODO
Dr.Web
eemet4.1
eemet5.0
eScan
eset_nod32
ESET-SMART
ESTsoft
Further information

For more in-depth coverage, including full details of the analysis behind this blog as well as additional indicators which can be used to detect similar samples, or if you have any other queries, please give us a shout at threatintelligence@uk.pwc.com.