ISOT HTTP Botnet Dataset

There are 2 different datasets: a botnet dataset consisting of malicious DNS traffic generated by different botnets, and a benign dataset consisting of DNS traffic generated by different known software applications.

Malicious Botnet DNS Traffic

The botnet dataset contains full DNS packets of 9 exploit kits that were collected in a virtual environment, as depicted in Figure 5. One of the bots (Atom)* did not generate traffic because it detects virtual machines. Each bot was deployed in a Windows XP virtual machine that ran for several days. The virtual environment was fully monitored from DNS server to the router. We configured and deployed an authoritative DNS server (in the lab) **ns1.botnet.isot**. with IP **192.168.50.88**. We configured each bot to specifically communicate with a command and control (C&C) server that we setup. For example, for the Zeus botnet, we setup a C&C server named zeus.botnet.isot.
The list of bots is given in Figure 1, where the following naming convention is used: *{**Exploit_kit_name**}.botnet.isot*

```
192.168.50.14   zyklon.botnet.isot
192.168.50.15   blue.botnet.isot
192.168.50.16   liphyra.botnet.isot
192.168.50.17   gaudox.botnet.isot gdox.botnet.isot dox.botnet.isot  **
192.168.50.18   blackout.botnet.isot
192.168.50.30   citadel.botnet.isot ***
192.168.50.31   citadel.botnet.isot ***
192.168.50.32   be.botnet.isot black energy
192.168.50.34   zeus.botnet.isot
```

Figure 1. List of botnets and their VMs

Notes:

* We have also deployed *ATOM* exploit kit, which is the third generation of Zeus, but we did not see any traffic going as we expected that it detects that it is running under a virtual machine.
** *dox, gdox, gaudox*, all are refereeing to the same exploit kit "*Gaudox*". We wrote three zone files *dox, gdox, gaudox* because at least three C&C channels were required to deploy the botnet. We used slightly different naming because we want to track how communication flows between the channels
*** citadel bot was installed twice in two different machine to check that it shows the same behaviour on different machines.

Dataset files:

The botnet data (malicious data) consists of 5 pcap files located under the subdirectory botnet_data; the 5 files are listed in Figure 2.

```
init.pcap
init2.pcap
init3.pcap
init5.pcap
init4.pcap
```

Figure 2. Botnet data files

Application DNS Dataset

The ISOT application dataset was collected from individual known (benign/normal) applications to profile their DNS behaviour. This allowed us to passively classify DNS traffic and differentiate malicious traffic vs. normal traffic. The data was collected in a virtual environment depicted in Figure 5. Each individual software application was installed on a virtual machine that was running windows 7. The DNS resolver of each machine was pointed to our DNS server **192.168.50.88**. The collected data is considered normal traffic since its coming from known applications. Figure 3 gives the list of applications and their VMs.

```
192.168.50.19   dropbox.com  Dropbox
192.168.50.50   Avast
192.168.50.51   Adobe Reader
192.168.50.52   Adobe Software Suite
192.168.50.54   Chrome
192.168.50.55   Firefox
192.168.50.56   Malwarebyte
192.168.50.57   WPS office
192.168.50.58   Windows update
192.168.50.59   utorrent.com bittorrent.com
192.168.50.60   fosshub.com audacity
192.168.50.61   Bytefence-com
192.168.50.63   Thunderbird Mozila
192.168.50.64   Avast
192.168.50.65   Skype
192.168.50.66   Facebook massager
192.168.50.67   CCleaner
192.168.50.68   Win update
192.168.50.69   Hitmanpro.com

-> background data from windows
> time.windows.com
> time.microsoft.akadns.net
> dns.msftncsi.com
```

Figure 3. List of applications and their VMs

Dataset files:

The application data (benign) consists of 3 pcap files located in the subdirectory application_data, and depicted in Figure 4.

```
- applicationDNS.pcap
- dns_application_2017_1.pcap
- dns_application_2017.pcap
```

Figure 4. Application data files

Timeline

The data was collected over the following period:

Start: 2017-06-14
End: 2017-06-21 18:31

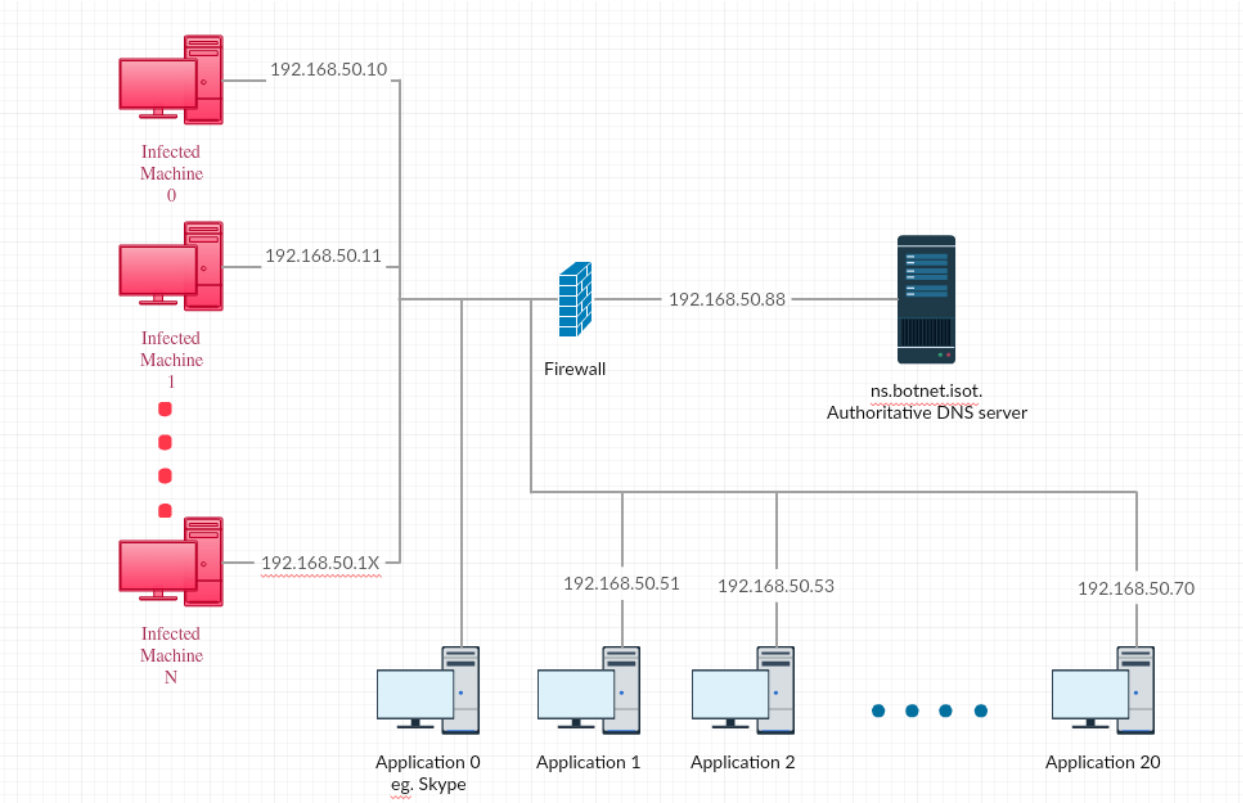

Figure 5: shows our virtual environment implementations

To Reference this dataset use:

Alenazi A., Traore I., Ganame K., Woungang I. (2017) Holistic Model for HTTP Botnet Detection Based on DNS Traffic Analysis. In: Traore I., Woungang I., Awad A. (eds) Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments. ISDDC 2017. Lecture Notes in Computer Science, vol 10618. Springer, Cham