PRACTICAL EXPERIENCES
OF
BUILDING AN IPFIX BASED
OPEN SOURCE BOTNET DETECTOR

Mark Graham
OUTLINE

- RESEARCH PROBLEM: Botnet detection in Cloud Providers
- FLOW: IPFIX and NetFlow
- CONCEPTUAL FRAMEWORK: Build environment & challenges
- RESULTS
- CONCLUSIONS: Why do the results matter?
ABOUT ME...

- PhD Student at Anglia Ruskin University, Cambridge, UK
- Just passed my 2nd Year Confirmation on Candidature
- Supervisors: Adrian Winckles, Dr Erika Sanchez-Velazquez

Working PhD title:
“A Botnet Needle in a Virtual Haystack(s)”
RESEARCH PROBLEM

“To create a botnet detection mechanism for cloud service provider networks.”

Why a CSP environment?

- IOT - Ease of access for centralised data
- Attacks - From the cloud, on the cloud

Build Criteria:

- Privacy - Cannot site AV in tenant environment
- Isolation - A tenant cannot access another tenant
- IPv6 - must support this next gen protocol
RESEARCH PROBLEM... ATTACKS ON CSPs

1) HOST ESCAPE
   *Crisis Malware* (2012)

2) INTRA-VM ATTACKS
   *Ristenpart et al.* (2006)

3) VM ESCAPE
   *Cloudburst Malware* (2009)
   *Venom* (2015)
A BRIEF HISTORY OF FLOW

- 1980’s - SNMP
- 1990’s - Syslog
- 2002 - NetFlow v5 (Cisco)
- 2009 - NetFlow v9 (Cisco)
- 2013 - IPFIX (IETF Standard: RFC 7011-7015)
WHY FLOW?

- Flows based on PDU header information
  1) Publically available metadata (v’s PCAP does intrusive DPI)
  2) Data storage savings (3.1GB PCAP; 43KB in IPFIX)
  
  PCAP is a *phone call*; flow is the *phone bill* (who, when, how long)

- Traffic detection (v’s forensic/signature detection)
  
  Botnet takedown requires locating C&C

- NFv5 + PCAP used in botnet detection research since 2007 as a data capture method to feed into detection algorithms:
  
  Bothunter, Botsniffer, Botcop, Botzilla, BLINC, etc.
IPFIX v NetFlow

IPFIX was developed to address the drawbacks of NetFlow:

- **Standard:** Vendor Neutrality
- **Extensible:**
  - NFv5 – fixed template: 18 fields
  - NFv9 - 79 fields (104 if Cisco)
  - IPFIX - 433 fields (IANA)
- **Protocol:** NF is UDP; IPFIX supports UDP / TCP / SCTP (TLS)
- **Security:** IPFIX has C.I.A by design; including data obfuscation
- **Next Gen:** IPFIX supports IPv6, MPLS and multi-cast

*Caveat: Cisco NFv9 support some of these, but proprietary*
CRITERIA FOR “DATA CAPTURE” ELEMENT

- Cloud and IoT provider virtualised environments
- Respect cloud tenant data privacy (?)
- COTS & open source
- Feed into a neural network; which feeds into SDN / VM containment
- Ideally support NetFlow and IPFIX for experimental comparison
CONCEPTUAL DESIGN: VIRTUALISATION PLATFORM

1) Hypervisors:
   • **Xen (Citrix)**, **Hyper-V (Microsoft)**, **ESXi (VMWare)**
     Xen: Open Source
     Xen: Common in CSPs (AmazonAWS, OpenStack, Apache CloudStack)
     Xen: Full Para-virtualisation

2) Software Switches:
   • **OVS (Open vSwitch)**, **Hyper-V, Nexus (Cisco), vSwitch (VMWare)**
     OVS: Open Source
     OVS: Exports IPFIX, NetFlow v5/v9 and sFlow
     OVS: Sits well with Xen Hypervisor
EXPERIENCE OF BUILDING THE SYSTEM #1

Citrix XenServer 6.2.0:
- OS: Linux Centos v5.5 i386
- Hypervisor: Xen v4.1.5
- Hypervisor Mgmt: XenCentre (GUI)
- Virtual Switch: Open vSwitch v1.4.6

This worked fine for NetFlow v5, but...
EXPERIENCE OF BUILDING THE SYSTEM #1

1. XenCentre GUI does not support IPFIX
   OK - we could use command Line XAPI Toolstack

2. Open vSwitch only supports IPFIX on v1.10+
   We have OVS v1.4.6

3. Open vSwitch v1.10 requires Centos 5.5 i686
   We have Centos 5.5 i386 - ok, lets upgrade
   XenServer partitions DOM0 into 4GB, of which 3.8GB is used by Xen.
EXPERIENCE OF BUILDING THE SYSTEM #1

1. XenCentre GUI does not support IPFIX
   OK - we could use command Line XAPI Toolstack

2. Open vSwitch only supports IPFIX on v1.10+
   We have OVS v1.4.6

3. Open vSwitch v1.10 requires Centos 5.5 i686
   We have Centos 5.5 i386 - ok, lets upgrade
   XenServer partitions DOM0 into 4GB, of which 3.8GB is used by Xen.

I am not a Citrix expert...
Centos refused to be upgraded
EXPERIENCE OF BUILDING THE SYSTEM #2

Citrix XenServer Creedence v6.5:
- OS: Linux Centos v5.10 i686  ✔
- Hypervisor: Xen v4.4 (64 bit)  ✔
- Virtual Switch: Open vSwitch v2.1.2  ✔
  Supports IPFIX
  But, IPFIX would not export timestamps
  And IPFIX would not aggregate
EXPERIENCE OF BUILDING THE SYSTEM #2

Citrix XenServer Creedence v6.5:
- OS: Linux Centos v5.10 i686 ✔
- Hypervisor: Xen v4.4 (64 bit) ✔
- Virtual Switch: Open vSwitch v2.1.2 ✔
  Supports IPFIX
  But, IPFIX would not export timestamps
  And IPFIX would not aggregate

I am not an OVS expert...

OVS refused to work with IPFIX
EXPERIENCE OF BUILDING THE SYSTEM #3

Mark’s Bespoke Build v1.0:

- **OS**: Ubuntu 14.04  ➡️ NEW OS
- **Hypervisor**: Xen Project v4.4 (64 bit)  ➡️ STANDALONE
- **Hypervisor API**: XAPI Toolstack  ➡️ ADDITION
- **Virtual Switch**: Open vSwitch v2.0.2  ➡️ DOWNGRADED
- **IPFIX Exporter/Collector**: nProbe v6.15  ➡️ ADDITION

3 HOUR TOTAL INSTALL
## CLOUD STACK:

<table>
<thead>
<tr>
<th>Component</th>
<th>Version/Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DOM-0 OS</strong></td>
<td>Ubuntu 14.04</td>
</tr>
<tr>
<td><strong>Hypervisor</strong></td>
<td>Xen 4.4 (64 bit)</td>
</tr>
<tr>
<td><strong>Hypervisor API</strong></td>
<td>XAPI Toolstack</td>
</tr>
<tr>
<td><strong>Virtual Switch</strong></td>
<td>Open vSwitch v2.0.2</td>
</tr>
<tr>
<td><strong>Flow Exporter</strong></td>
<td>nProbe v6.15</td>
</tr>
<tr>
<td><strong>Flow Collector</strong></td>
<td>nProbe v6.15</td>
</tr>
<tr>
<td><strong>VM Management</strong></td>
<td>XenCentre v6.5</td>
</tr>
<tr>
<td><strong>Flow Protocol Support</strong></td>
<td>NetFlow v5  \ NetFlow v9 \ IPFIX</td>
</tr>
<tr>
<td><strong>Flow Traffic Presentation</strong></td>
<td>Neo4J</td>
</tr>
</tbody>
</table>
PROBE LOCATION
PROBE LOCATION
CURRENT WORK: IPFIX TEMPLATE

<table>
<thead>
<tr>
<th>Version: 10 (2)</th>
<th>Length (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Export Time Stamp = 2015-01-01 12:59:59 (4)</td>
<td></td>
</tr>
<tr>
<td>Sequence Number = 0 (4)</td>
<td></td>
</tr>
<tr>
<td>Observation Domain ID = 123456 (4)</td>
<td></td>
</tr>
<tr>
<td>Set ID = 1 (2)</td>
<td>Set Length (2)</td>
</tr>
<tr>
<td>Template ID = 456</td>
<td>Field Count = 8</td>
</tr>
<tr>
<td>Flow_Start_MilliSeconds = 152</td>
<td>Field Length (4)</td>
</tr>
<tr>
<td>Flow_End_MilliSeconds = 153</td>
<td>Field Length (4)</td>
</tr>
<tr>
<td>IN_Bytes = 1</td>
<td>Field Length (4)</td>
</tr>
<tr>
<td>IPv4_DST_Addr = 12</td>
<td>Field Length (4)</td>
</tr>
<tr>
<td>L4_SRC_Port = 7</td>
<td>Field Length (2)</td>
</tr>
<tr>
<td>L4_DST_Port = 11</td>
<td>Field Length (2)</td>
</tr>
<tr>
<td>Protocol = 4</td>
<td>Field Length (1)</td>
</tr>
<tr>
<td>BiFlow_Direction = 239</td>
<td>Field Length (1)</td>
</tr>
</tbody>
</table>
FUTURE WORK: CSP NEUTRALISATION ECO-SYSTEM

- Traffic Detector (IPFIX Template)
- Threat Intelligence (Honeypots)
- Neural Network (Bot Detection AI)
- Bot Neutralisation (SDN & VM placement)
FUTURE WORK: VISUALISATION

All Protocols

HTTP Only
LIMITATIONS OF THE SYSTEM

• Deep Packet Inspection
  Discarding the payload for privacy comes at a cost:
  – Can you detect botnets without DPI information? ... (Probably not)
  – IPFIX allows customisable Information Elements to capture DPI information
    We have developed a DPI template: HTTP, DNS, SMTP & IRC.
  – Where is the cross-over with tenant privacy? ... (Need to measure to detect)

• Encryption / VPN Traffic
  – Payload encryption wont impact traffic communication graphs
  – But encrypted PDU headers within a VPN will impact collection
**CURRENT WORK: EXTENDED IPFIX TEMPLATE**

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version: 10</td>
<td>Length (2)</td>
</tr>
<tr>
<td>Export Time Stamp</td>
<td>2015-01-01 12:59:59 (4)</td>
</tr>
<tr>
<td>Sequence Number</td>
<td>0 (4)</td>
</tr>
<tr>
<td>Observation Domain ID</td>
<td>123456 (4)</td>
</tr>
<tr>
<td>Set ID = 1</td>
<td>Set Length (2)</td>
</tr>
<tr>
<td>Template ID = 457</td>
<td>Field Count = 11</td>
</tr>
<tr>
<td>DNS ID</td>
<td>...</td>
</tr>
<tr>
<td>DNS TTL</td>
<td>...</td>
</tr>
<tr>
<td>DNS Query Name</td>
<td>...</td>
</tr>
<tr>
<td>DNS IP Address</td>
<td>...</td>
</tr>
<tr>
<td>HTTP GET</td>
<td>...</td>
</tr>
<tr>
<td>HTTP Referer</td>
<td>...</td>
</tr>
<tr>
<td>HTTP Location</td>
<td>...</td>
</tr>
<tr>
<td>HTTP Age</td>
<td>...</td>
</tr>
<tr>
<td>HTTP Cookie</td>
<td>...</td>
</tr>
<tr>
<td>HTTP Set Cookie</td>
<td>...</td>
</tr>
<tr>
<td>HTTP Via</td>
<td>...</td>
</tr>
</tbody>
</table>
BOTNETS FOR TESTING

C&С  P2P  IRC

WANTED

DEAD OR ALIVE
CONCLUSION

Using COTS technology we created a botnet traffic capture mechanism:

- 80%+ CSPs already collect flow traffic for network management
- Probe must be located on the hypervisor
- IPFIX template for botnet detection

Clouds will host IoT and Smart Cities:

- Cloud is an attack platform (ideal breeding ground of botnets)
- Cloud is an attack target (storage, other tenants, VE malware)
- Traditional AV is not suited for botnets
THANK YOU

mark.graham@anglia.ac.uk