The big picture

• **Objective:**
  Data discovery
  Specifically, the speed of discovery

• **Approach:**
  Iterate at scale within seconds to minutes over billions of data points

• **Method:**
  Explain the BRO installation
  Behavioral baselining
  Validate – triage – remediate
Architecture

• **Big data:**
  BRO at 8Gb/s at multiple sites

• **Super-computer:**
  Spark, Scala, and graph engine
  Cray Urika-GX

• **Analytics:**
  Behavioral analytics (the rest of this presentation)
Enterprise multi-sensor BRO installation
Or
“Twere that it were so simple”

Deployment Scenario

• Multiple systems, multiple sites.
• Limited physical access and only SSH-based access to the systems.
• Some systems deployed outside of firewall.
• Largely unsupervised.
• Long retention period for logs.
• Different port configurations for each site.
R-Scope: Enterprise-Ready Bro

Requirements for Bro Sensor*

• Fault tolerant.
• Access to diagnostic information.
• Secure. (sensor + data)
• Easily configured.
• Visibility into system performance.

* In addition to all the things we love about bro
Fault Tolerance

• Light-weight Process Monitoring
  • Proc connector and a Netlink socket to obtain process events from the kernel
  • BPF filter to accept only FORK, EXEC and EXIT events.
  • EXIT event return code used to detect crashes.
  • Processes to be monitored register with the watchdog and provide API to restart.
  • To detect bro crashes
    
    WIFSIGNALED(exit_code) || !WIFEXITED(exit_code)
  • To detect splunkd crashes
    * WIFSIGNALED(exit_code) && WTERMSIG(exit_code) != 15

• Reboot on kernel panic
  • Add the line `kernel.panic = 10` to `/etc/sysctl.conf` to reboot after 10 seconds
Fault Tolerance

• Customized handling of system memory pressure. (Kernel OOMs)
• Prevent disk full scenarios by using both date based and disk size based expiration.
• Log all interactions
  • Goal is not to fail.
  • Definitely not fail the same way twice.
• Serial and console access.
• DVD based re-provision for catastrophic failures.
Diagnostic Information

• SNMP for monitoring the sensor.
  • Status and notifications for process state changes and ipmi events.
  • Implemented using net-snmp via agent-X protocol
    • [http://www.net-snmp.org](http://www.net-snmp.org)
    • Agent-X protocol allows customizations to be in a separate process.
  • Custom MIB for extensions
    • Start with standard MIBs and create your own.
    • Use mib2c along with custom MIB to auto generate net-snmp interface code.

```
mib2c -c /usr/share/snmp/mib2c.scalar.conf <scalarNameFromMIB>
mib2c -c /usr/share/snmp/mib2c.mfd.conf <tableNameFromMIB>
```
  • Integrate the auto-generated interface code into your C/C++ application that provides custom data.
  • DISMAN monitoring to generate traps/notifications.

```
proc splunkd 2 2
proc rsyslogd 1 1
monitor -r 180 -S -o prNames -i sysName.0 -o prErrMessage "process table" prErrorFlag != 0
```
Appliance Security

- Limited set of open ports and accessible services.
- Controlled access to ssh and snmp.
  - Explicit configuration of allowed hosts/networks.
  - Iptables configured via https://github.com/ldx/python-iptables
- No privileged or native shell access.
- Curated and patched software components.
- Bro processes run without elevated privilege.
  - http://linux.die.net/man/8/setcap
  - setcap cap_setuid,cap_net_admin,cap_net_raw,cap_ipc_lock+eip bro
- Encrypted drives tied to physical HW
  - LUKS encryption for data volume.
  - LVM partitions on top of LUKS.
  - File systems on LVM partitions.
  - LUKS password tied to physical HW. Disks are random data unless married to physical machine
  - Very little performance hit on modern intel chip
Easy Configurability

• Ansible for sensor configuration
  • Automation engine for configuration management
    • https://www.ansible.com/how-ansible-works
  • “raw” mode can be used with CLI frameworks (i.e R-Scope, broctl)
    • Only need ssh access to system.
    • $ ansible rscope -m raw -a 'system status'
      • For two hosts the output may look similar to this:
        ---------------------
        rscope-1 | SUCCESS | rc=0 >>
        16:32:00 up 10:05, 1 user, load average: 6.89, 4.59, 4.97
        rscope-2 | SUCCESS | rc=0 >>
        13:32:01 up 8 days, 21:07, 3 users, load average: 0.00, 0.00, 0.15
  • Tasks and playbooks for setting and getting system configuration.
    • Tasks are collection of commands
    • Playbooks are collection of tasks.
Easy Configurability

• Ansible for sensor configuration
  • Tasks
    - name: "apps-production"
      raw: "apps list"
      register: apps_prod_out
      failed_when: "'Failed' in apps_prod_out.stdout or 'Error' in apps_prod_out.stdout"
  • Playbooks
    - hosts: rscope
      connection: ssh
      gather_facts: False
      tasks:
        - name: "apps-configuration"
          include: tasks/showconfig_apps.yml
        - name: "files-configuration"
          include: tasks/showconfig_files.yml
        - name: "intel-configuration"
          include: tasks/showconfig_intel.yml
        - name: "logs-configuration"
          include: tasks/showconfig_logs.yml
System Performance Analysis

• Performance metrics captured as bro logs.
  • Port metrics – ingest rates and drop stats
    • #fields
      • timestamp
      • monport
      • bps_rx_port
      • pckts_rx_port
      • pckts_dropped_port
    • 1469527201.417345 mon2 457909256 76386 0
    • 1469527201.613132 mon0 581533648 99921 0
    • 1469527202.417429 mon2 436484712 74734 0
    • 1469527202.853441 mon0 495422648 89723 0
  • Worker metrics – processing rates and packet size histograms
  • CPU metrics – utility rates
• Offline analysis/graphing performed using Pandas.
  • Allows programmatic manipulation of time series data.
  • Calculate statistics like mean, median, standard deviation, percentiles etc.
  • Integrated with matplotlib for plot generation.
• Data ingest

  • # Read the log file, skip comments and headers, use the 'timestamp' field as the timestamp
    # and parse it using the datetime_parser function.
    data = pd.read_table(log_file, comment='#', skiprows=8, names=field_names, sep='s+',
                        parse_dates=['timestamp'], date_parser=datetime_parser)

  • Create an index based on the configured index fields.
    data = data.set_index(log_stream['index_fields'])

  • def datetime_parser(x):
    dt = datetime.utcfromtimestamp(self.__timestamp_seconds(x))
    dt = dt.replace(tzinfo=pytz.utc)
    required_time_zone = pytz.timezone(self.__analysis_config['time_zone'])
    dt = dt.astimezone(required_time_zone)
    return dt

• Data Aggregation ( Combine two time series )

  • aggregated_series = port1_series.add(port2_series, fill_value=0)

• Juxtapose two time series

  • juxtaposed_series = field_values.unstack(level=0).fillna(0)
Additional Features

• Scripts for filtering log types.
  
  https://github.com/reservoirlabs/bro-scripts/tree/master/logging

• Quick self test to verify system health.
  
  • Useful for spot checks
  
  • Useful during upgrades

• `ethtool -S` to gather port metrics and provide real time stats.

• `Rsync` for data offload framework.

  • Best performance with latest version and options
    
    `--havL --info=progress2 --no-inc-recursive --progress -e`
Performing at-scale analysis
Deloitte Advisory’s Approach

- Lead with a broad analysis of transactional data
- Focus on simple analytical questions
- Identify behaviors and candidate IOCs rather than signatures
- Combine transactional data with enrichment/contextual information
- Iterate

Use sophisticated data science to advance analytic maturity and speed

http://detect-respond.blogspot.com/2013/03/the-pyramid-of-pain.html
Approach In Context

80-90% of all activities fall in this range

Deloitte Advisory’s methods enable us to identify and search for behaviors, which causes more pain to adversaries

http://detect-respond.blogspot.com/2013/03/the-pyramid-of-pain.html
Facts And Figures

• Record types
  • Conn
  • HTTP
  • SSL
  • Notice

• 254 billion records (60 billion / month)

• 244 bytes per record (average)

• 62TB on disk
Behaviors Under Analysis

• Infection 🦠

• Beaconing 📡

• Scanning 📦

• Data Exfiltration ➡️
Infection Detection

• Multiple methods
  • Redirection identification
  • Correlated HTTP requests to identify repeated patterns

• Define the expectation
  • Redirection depth is low
  • Noise is going to be high

• Observations in practice
  • Outlier analysis is key
  • Context is also key
  • Noise is high
  • Least investigated area in Deloitte Advisory’s engagements
  • Combinatorial complexity

• Success stories
  • Zero-sized pixel with high request depth identified and correlated with threat intel
Beaconing

• Multiple methods
  • Flocking identification based on Protographs by Janies and Collins, 2011
  • Top talkers
  • Entropy-based methods
    • Inter-session timing
    • Outbound bytes

• Define the expectation
  • Look for repetition over time
  • Look for deviations from diurnal cycles

• Observations in practice
  • Aggregation of methods
  • Dynamically adjust thresholds
  • Use heuristics and counting over parametric methods
  • Keep the humans in the loop

• Success stories
  • Identified IOT devices phoning home
  • Identified malware infections

https://zoomapps.files.wordpress.com/2015/04/entropy-is-disorderliness.png
Scanning

• Start with Threshold Random Walk relationship described by Jung, Paxson, et al
  • http://www.icir.org/vern/papers/portscan-oak04.pdf

• Define the expectation
  • Most connections are successful
  • Clusters of unsuccessful connections are useful and unusual

• Observations in practice
  • Count number of unsuccessful attempts
  • Bin by time period
  • Group by client IPs
  • Focus on specific ports (services of interest)

• Success stories
  • Identified scanning nodes used by the internal security team
Data Uploads

• Start with Producer / Consumer relationship described by Bullard and Gerth
  • https://resources.sei.cmu.edu/asset_files/Presentation/2014_017_001_90063.pdf
  • https://github.com/reservoirlabs/bro-producer-consumer-ratio

• Define the expectation
  • Numerator = number of uploaded bytes
  • Denominator = number of downloaded bytes
  • Clients usually download

• Observations in practice
  • Keep a ratio rather than normalize [0, inf)
  • Use a logarithm to bin

• Success stories
  • Multi-function devices beaconing out of the network, uploading megabytes a day
  • Site-to-site backups
What Are Deloitte Advisory’s Teams Finding?

- Protocols running on off ports
- Tunneling protocols going to unexpected places
- Anomalous uploads
- “TOR bridge” nodes
- Candidate C2 nodes
- Beaconsing detection
- Drive-by downloads

Deloitte Advisory is providing ongoing behavioral analytics and correlation to supplement existing SIEM and search tools.
Contact Information
BroCon ‘16

Eric Dull
Deloitte Advisory
edull@Deloitte.com
202-306-4469

Dilip Madathil
Reservoir Labs
madathil@reservoir.com
This presentation contains general information only and Deloitte Advisory is not, by means of this presentation, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This presentation is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your business. Before making any decision or taking any action that may affect your business, you should consult a qualified professional advisor.

Deloitte Advisory shall not be responsible for any loss sustained by any person who relies on this presentation.

As used in this document, “Deloitte Advisory” means Deloitte & Touche LLP, which provides audit and enterprise risk services; Deloitte Financial Advisory Services LLP, which provides forensic, dispute, and other consulting services; and its affiliate, Deloitte Transactions and Business Analytics LLP, which provides a wide range of advisory and analytics services. Deloitte Transactions and Business Analytics LLP is not a certified public accounting firm. These entities are separate subsidiaries of Deloitte LLP. Please see www.deloitte.com/us/about for a detailed description of the legal structure of Deloitte LLP and its subsidiaries. Certain services may not be available to attest clients under the rules and regulations of public accounting.