INTRODUCTION TO OMNISOC
The situation prior to formation

• The Big10 Academic Alliance (1993)
  – Sharing of best practices,
  – Building relationships and trust

• Security organizations spread thin
  – Dedicated professionals, competing priorities, outnumbered
    • Protecting Infrastructure
    • Conducting assessments
    • Passive defense
    • Incident response

• Increasing reliance on infrastructure by society
  – redefines impact of some security incidents
Ambition -> improve security posture

- Dedicated teams
  - Identify and protect
  - Detection
  - Response and recovery
- 24x7x365 staffing
- Improved threat sharing
- Reduced detection time
- Controlled cost
Challenges to address

- How to share sensitive data responsibly
- Issues local to every institution
- Sector specific tailoring for Higher Education
- Evolving thread landscape
The approach

• Figure out how to share data
  – Rely heavily on existing relationships and agreements within BTAA
  – Use policy to build trust, use technology to enforce
    • NDAs, Service agreements + 2fa, RBAC

• Created a shared Security Operations Center
  – Provide 24x7 coverage
  – Consume shared data
  – Detect and coordinate

• Members refine their own local capabilities
  – Leveraged SOC service frees up time
What is a Security Operations Center (SOC)?

“a team primarily composed of security analysts organized to detect, analyze, respond to, report on, and prevent cybersecurity incidents.”

Source: http://www.mitre.org/
Creation of the OmniSOC

Founding Members (no particular order)
- Rutgers University
- Purdue University
- Northwestern University
- University Nebraska - Lincoln
- Indiana University

- Located at Indiana University
- Thee teams
  - Tier 1 (2 FTE + Pool)
  - Security Analysts (3 FTE)
  - Platform Engineering (4.5 FTE)
- Leverage GlobalNOC
  - Tools
  - IT Infrastructure
  - Ops procedures
  - Facilities
Objectives

• **Gather** security event data from participating members
• **Perform** analysis to identify actionable threats
• **Engage** members when needed
• **Meet** with members regularly to **refine** service and processes
Objectives

- Provide **hosted SIEM** service for interested customers
- **Automate** analysis where possible
- Use **Machine Learning** for detection of pattern based indicators

10/15/18

https://omnisoc.iu.edu/
Devilish Details

• Estimated per member data volume a concern
  – 1.5TB / day for 30 days.
  – We are going to need a scalable approach

• Data sources vary by member
  – Need to be flexible and vendor agnostic
  – Make the most of what already exists
  – Normalize data to support cross data source analysis

• This is a production system
  – Need to design for survivability and robustness
The Platform

• Site local appliances + central clusters
• ELK for storage and analysis
  – Elastic, Logstash, Kibana
• Member Data
  – 30 days of data retention
  – Some > 100k events / sec, avg 50k
  – Low volume events
    • Intrusion Detection System detected specific action
  – High volume events
    • Record of all network activity, every flow

10/15/18 https://omnisoc.iu.edu/
Aggregation Appliances

• Operated by OmniSOC
  – 2 per member
• Collects relevant event data
• Encrypted communications
  – SSH from bastions inbound
  – VPN outbound
• Store and forward design
  – withstand multi-hour communications outage without data loss
• Monitored 24x7
Data Flow

- Collect data from existing sources using Logstash
- IDS / IDS
  - Bro
  - Suricata
  - Cisco Firepower
  - Palo Alto
- Netflow
Data Flow

- Data is buffered in the aggregators local to each member site
- Central systems pull data in
• Using Logstash, convert data to standardize form
• Consistent formats and names for 5-tuple info, etc
  – Source IP represented as native CIDR not string
  – Timestamps represented as time not unsigned integer
Data Flow

- Add GeoIP information for IP addresses
- Provide origin AS info
- Tag critically important resources
- Tag known bad resources using available threat feeds
  - REN-ISAC, etc

10/15/18
https://omnisoc.iu.edu/
Data Flow

- insert data into Elasticsearch database
- 1 elastic cluster per member
- data is retained for most recent N days
What might we do with it? An example.

- Ingest network flow and IDS data
- Store it in a way we can search on it
- Create displays to explore and share between SOC engineers
- Alert on events of interest, apply ML techniques to detect anomalies, etc.
We have data now what?

**Identify**
- Understand Mission
- Establish Governance
- Develop Risk Management Strategy and Processes
- Manage Assets

**Protect**
- Manage Access
- Provide Awareness and Training
- Manage Data Security
- Perform Maintenance

**Detect**
- Manage Protection Technology
- Log - network, service, endpoints
- Establish and Maintain Monitoring Technology
- Maintain and Use Information Protection Process and Procedures
- Incident Response, Policies, Plans, Procedures, Techniques, etc.

**Respond**
- Maintain Situational Awareness
- Understand Monitored Environment
- Provide Call Center Services
- Conduct Real-Time Monitoring and Triage
- Logs, Alerts, Signs, Clusters, Prioritize

**Recover**
- Notify IT Team
- Incident Details; Indicators of Compromise
- Implement Countermeasures (Contain)
- Coordinate Incident Response Process
- Analyze Incident

- Confirm
- Determine Impact/Scope
- Gather and Handle Evidence
- Communicate and Share Information
- Advise, Bulletins, Alerts, Senior Leadership
- Improve Response Process and Capabilities

Doc.: 10/15/18 | https://omnisoc.iu.edu/
OmniSOC Structure

• Service Desk
  – Initial assessment and response
  – Coordination with members and engineering
  – Drive process

• Security Engineering
  – Analysis of security data
  – Assessment of risk and threat

• Platform Engineering
  – Data collection
  – DevOps analytics environment
  – data curation, modeling and tuning
Generalized Workflow

- Tier 1 triage possible incident
  - Automated alert
  - Staff identified
- Tier 2 confirm incident and scope
  - Guide next action
- Tier 1 engage with member institution
- Member provides feedback
Tools Used

• Member Communication
  – Phone
  – Slack
  – Email

• Incident Tracking
  – Service Now

• Data Analysis
  – Kibana
Incident Response

- Service Desk Monitors alertmon for OmniSOC alarms 24/7
- Alarms are based on “watchers” created by platform and security teams
- Watchers trigger on new SIDs (Security Identifiers) which are representative of something new or unexpected occurring on the members network
- Alarms trigger service desk to create Incident in ServiceNOW and spawn a task for Security Engineering to review
- Service desk monitors Incident in ServiceNOW for entire lifecycle of event
Incident Response

- OmniSOC Security Engineering receives incident task from Service desk.
- Security Engineering uses Kibana to analyze data provided by member institution to investigate the event.
- May also use other systems for threat intel (e.g. VirutsTotal or SES from REN-ISAC), or scapegoat VM to connect to remote site in the event of malicious web sites.
- Depending on what the security engineers find, they provide instruction to OmniSOC SD to either close incident as false positive or escalate to member for review and response.
Incident Response

Service Desk Response and Escalation → Security Engineering Review → OmniSOC Escalation to Member → Member Response and Feedback

- OmniSOC security engineer creates a message using standard template to be sent to member
- Standard template used for message includes all investigation details such as timestamps, IPs, incident type, source of data (Bro, Snort, Suricata, Palo Alto, etc)
- Message is forwarded to SD, who escalates incident to the member institution
- Key here: Message is used by OmniSOC to explain Indicators of Compromise (IoC) discovered to members by being as specific as possible so that they may interpret that data and take appropriate action (hopefully without need for clarification from OmniSOC)
Incident Response

- Member Institution receives notification from OmniSOC regarding indication of compromise
- Member performs analysis of event, takes action dependent on their internal procedures
- Member provides feedback to OmniSOC regarding usefulness of data received
- OmniSOC takes feedback from customer to improve service (reduce level of false positives, remove unnecessary monitoring rules, improve response time, etc.)
Future

• Improved technical capabilities
  – Make it easier to track events across members and data sources
  – Increase level of automated detection using Machine Learning
• Increase number of R1 schools
  – More schools == better detection
• Figure out how to support smaller schools and nets
• Grow staff
Misc Questions
How many hosts? How do you manage?

- 335 hosts total (300 in the mothership)
- Puppet for base builds
- Ansible for application configuration and deploy
- Foreman for orchestration
  - [https://www.theforeman.org](https://www.theforeman.org)
  - VM and physical box lifecycle management
    - GUI+DNS+DHCP+PXE+Puppet+Nagios
- Management ELK cluster for performance data on the others
Why ELK?

• Evaluation zeroed in on Elastic and Splunk
• Closed source vs open source with commercial add ons
• Better alignment with our long term vision
• Familiarity with ELK
  – Been using it for internal logs for several years after migrating off Splunk
  – flow analysis -> NetSage etc
• Check out ElastiFlow
  – [https://github.com/robcowart/elastiflow](https://github.com/robcowart/elastiflow)
Are you using Kuberdockonetes?

- Not yet, enough risk for this phase.
- Our testing suggests there are some performance reasons to do so:
  - Memory tests
    - Containers @ 99.6% of host
    - **VMs @ 66-77%** of host (read vs write)
  - Disk tests
    - Containers @ 99 – 93% of host (random vs seq)
    - **VMs @ 93-85%** of host (random vs seq)
How are you using ELK?

• We have platinum support from Elastic
• X-Pack features in use
  – RBAC -> large set of users, distinct roles
  – Watchers
  – Monitoring
  – Machine Learning
• Much of this can be done in purely open source with additional work with possible exception of RBAC.
How many X,Y,Z are you dealing with?

• Week before school starts @ IU
  – 5.2 billion documents per day
  – 13.6TB of data per day
  – Peaks ~ 100k events / second
  – 680 bytes per flow for connection logs, there are others
  – 6 logstash nodes
  – 18 host nodes
  – 36 warm nodes
What are your growth plans?

• 5 schools now, looking to grow this year
  - Improve our field of view
• Figure out how we can support regionals and other network providers
• Long term -> support smaller schools without well established / dedicated security resources (not sure how yet)
Questions or Comments

Edward Balas
ebalas@iu.edu