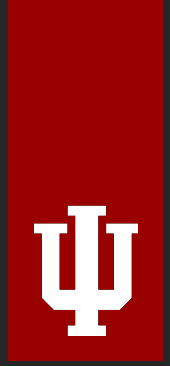


University Data Can Fuel AI/ML Cybersecurity Research

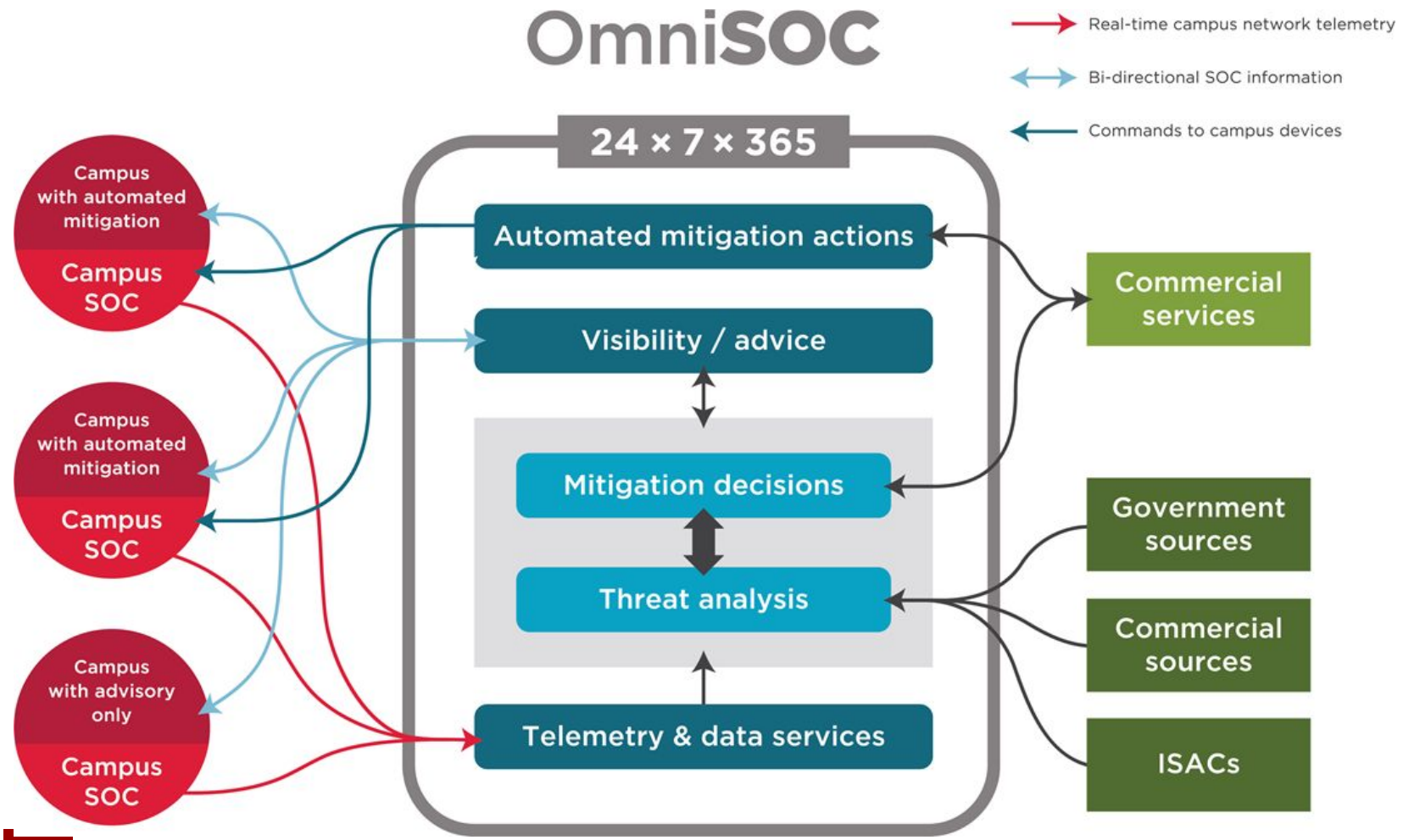
A Vision with an OmniSOC Proofpoint

Von Welch, Brad Wheeler, & Ryan Kiser

INDIANA UNIVERSITY

Cybersecurity ML Needs Data!

1. Relevant
2. Enormous
3. Context
4. Current


OmniSOC
Real-time campus network telemetry
Bi-directional SOC information
Commands to campus devices
24 × 7 × 365
Campus with automated mitigation
Campus SOC
Campus with automated mitigation
Campus SOC
Campus with advisory only
Campus SOC
Automated mitigation actions
Visibility / advice
Mitigation decisions
Threat analysis
Telemetry & data services
Commercial services
Government sources
Commercial sources
ISACs

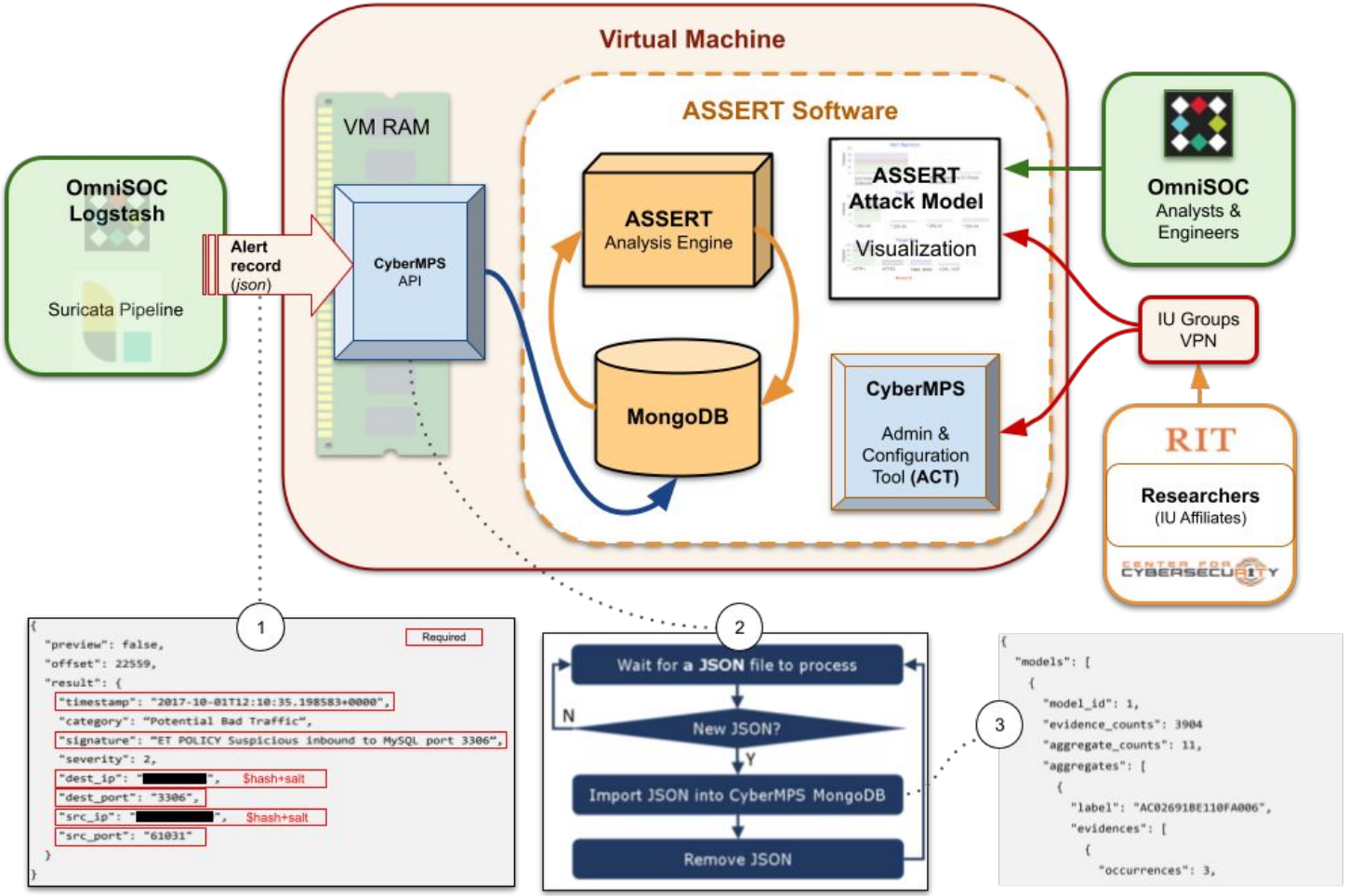
The data that enables this is...

1. Relevant
2. Enormous
3. Contextualized
4. Current

An example: ASSERT

ASSERT is an unsupervised learning system which categorizes related events into attack models which can be used by analysts to understand attacker behavior.

Okutan, A., Yang, S.J. ASSERT: attack synthesis and separation with entropy redistribution towards predictive cyber defense. *Cybersecur* **2,** 15 (2019). https://doi.org/10.1186/s42400-019-0032-0


Virtual Machine
ASSERT Software
VM RAM
OmniSOC Logstash
Suricata Pipeline
Alert record (json)
CyberMPS API
ASSERT Analysis Engine
MongoDB
ASSERT Attack Model Visualization
CyberMPS Admin & Configuration Tool (ACT)
OmniSOC Analysts & Engineers
IU Groups VPN
RIT
Researchers (IU Affiliates)
CENTER FOR CYBERSECURITY
1
Required
{
"preview": false,
"offset": 22559,
"result": {
"timestamp": "2017-10-01T12:10:35.198583+0000",
"category": "Potential Bad Traffic",
"signature": "ET POLICY Suspicious inbound to MySQL port 3306",
"severity": 2,
"dest_ip": "[redacted]", $hash+salt
"dest_port": "3306",
"src_ip": "[redacted]", $hash+salt
"src_port": "61031"
}
}
2
Wait for a JSON file to process
N
New JSON?
Y
Import JSON into CyberMPS MongoDB
Remove JSON
3
{
"models": [
{
"model_id": 1,
"evidence_counts": 3904
"aggregate_counts": 11,
"aggregates": [
{
"label": "AC02691BE110FA006",
"evidences": [
{
"occurrences": 3,

Outcomes

1. Case studies about different models
2. New functionality for the prototype
3. New ideas about how to use ASSERT
4. De-identified data approved for follow-up work

Yang, S., Okutan, A., Werner, G., Su, S., Goel, A., & Cahill, N. (2021). Near Real-time Learning and Extraction of Attack Models from Intrusion Alerts. ArXiv, abs/2103.13902.

University Operational Data

1. Untapped resource
2. There are many challenges
3. We think we can overcome them

Work with us!

ResearchSOC - rsoc@iu.edu
OmniSOC - soc@omnisoc.iu.edu
IU CACR - cacr@iu.edu