Developing an Ontology of the Cyber Security Domain

Semantic Technologies for Intelligence, Defense, and Security (STIDS) 2012
October 23-26, 2012

Dr. Leo Obrst

Penny Chase

Dr. Richard Markeloff

MITRE

“A Spectacular Failure for the Antivirus Industry”

June 1, 2012 http://www.wired.com/threatlevel/2012/06/internet-security-fail/

- Mikko Hypponen, Chief Research Officer of F-Secure, reports:
  - Stuxnet and Duqu went undetected for more than a year
  - Flame went undetected for more than two years
  - Stuxnet, Duqu and Flame “hid in plain sight”
    - Digitally signed components to mimic trustworthy applications
    - Based on standard libraries that do not arouse suspicion
  - Attackers tested them against all of the relevant antivirus products on the market
  - Zero-day exploits used in these attacks are unknown to antivirus companies by definition
- Commercial antivirus products “can't protect against targeted malware created by well-resourced nation-states”

Combating the Malware Threat

- Malware is one of the most serious threats to cyber security
- Malware may pose as ordinary software
- Progress on malware detection hampered by proprietary solutions
- MITRE-supported standards counteract proprietary solutions
- With these standards and semantic technologies we can bring malware defense to a new level

Standards Supported by MITRE

- MAEC – Malware Attribute Enumeration and Classification
- CCE – Common Configuration Enumeration
  - 11000 entries in CCE list
- CAPEC – Common Attack Pattern Enumeration and Classification
  - 400 attack patterns in 68 categories in CAPEC dictionary
- CVE – Common Vulnerabilities and Exposures
  - 53000 vulnerabilites in CVE dictionary
- OVAL – Open Vulnerability and Assessment Language
  - 14000 definitions in MITRE OVAL repository (other repositories exist)
- More at http://makingsecuritymeasurable.mitre.org

Enabling Automated Active Defense

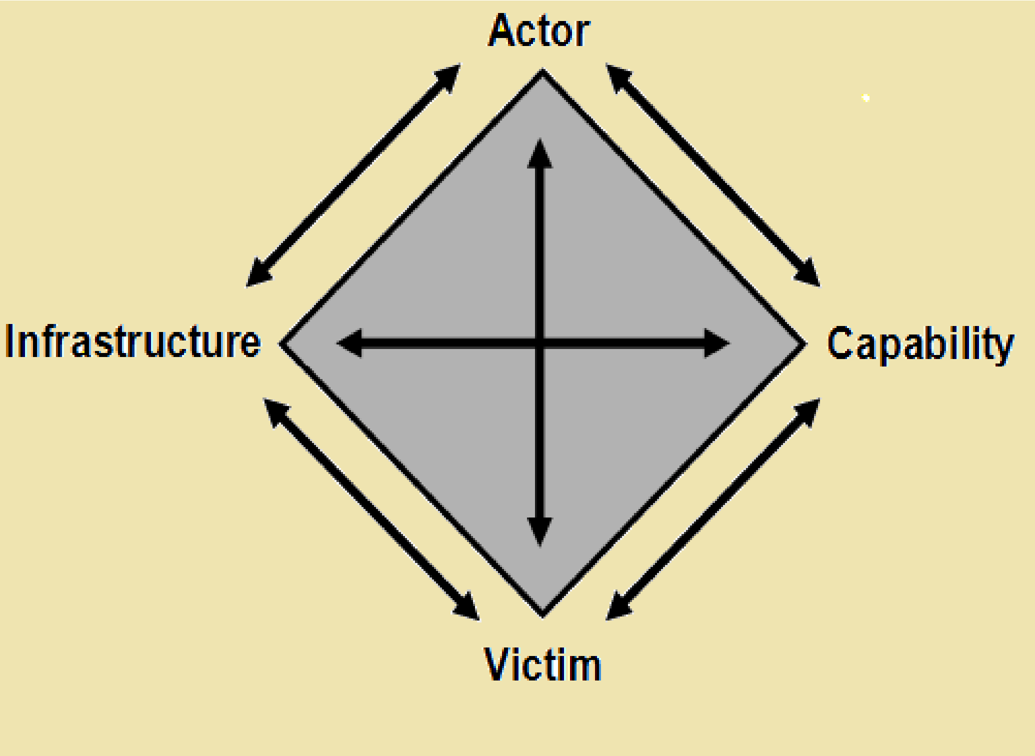
- Existing standards are descriptive languages implemented in XML
  - XML lacks formal semantics
- Semantic models of these standards would enable:
  - Integrating existing data silos
  - Bringing automated reasoning to bear on malware detection
- Would this make it possible to find Flame, Stuxnet?
  - Probably not today
- Could potentially apply the 80-20 rule to malware defense
  - 80% of incursions handled automatically
  - 20% require human intervention

Goals of the Cyber Ontology Effort

- Ultimate goal: Develop an ontology of the cyber security domain expressed in OWL
  - To enable integration across disparate data sources
  - To support automated cyber defense
- Initial focus is on malware
- Explain the process followed in developing the Cyber ontology and catalog the sources upon which it is based
- Provide a compilation of resources useful for constructing semantic models in the cyber security domain

The Diamond Model of Malicous Activity

- Provides the overarching conceptual framework
- The four corners account for all the major dimensions of a malicious cyber threat
  - Infrastructure: networks, software, hardware
  - Actor: the one threatening the victim
  - Capability: The tools available to the actor
    - Exploits
    - Infection vectors
    - C2 tools

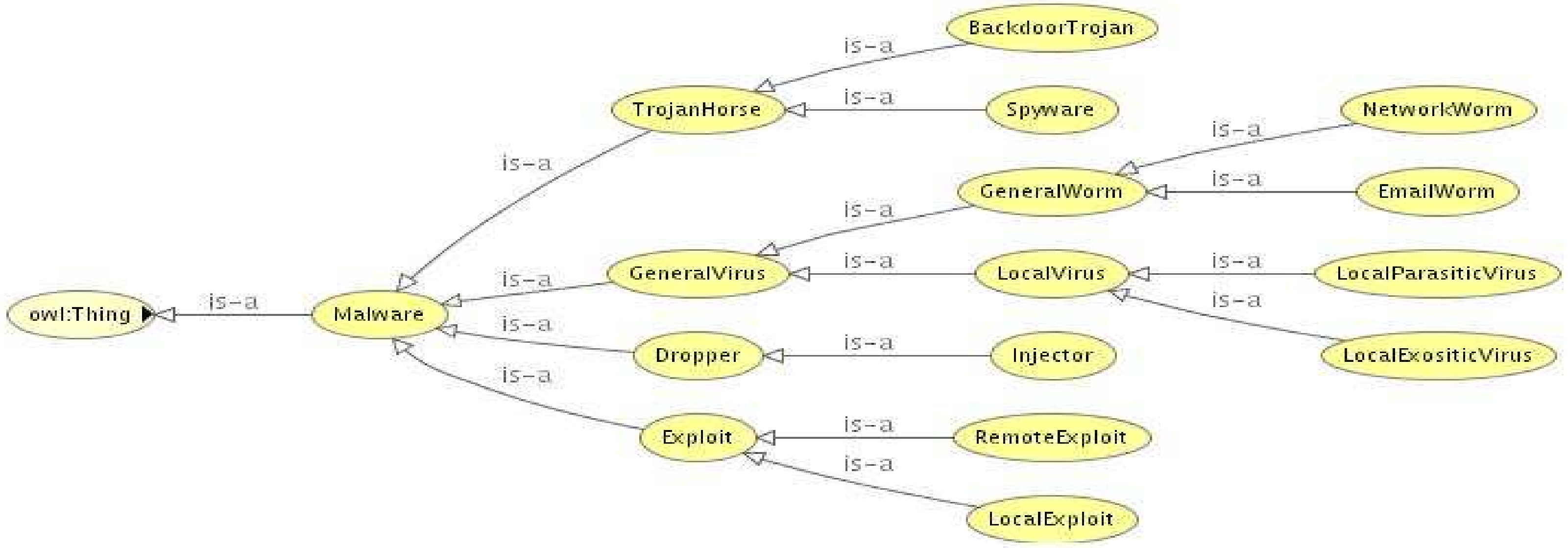


Ontology Development Methodology

- "Middle-out" analysis
  - Aspects of both top-down and bottom-up analysis
  - Bottom-up analysis requires understanding the semantics of the underlying data sources
  - Top-down analysis requires understanding the semantics of the end-users
- Enumerate competency questions
  - Questions the ontology needs to answer
- Reuse of existing ontologies
  - Including foundational, mid-level, utility, and reference ontologies
- Harvest existing schemas, data dictionaries, glossaries, and standards
  - Can provide entities, relationships, properties, attributes, and value ranges

Existing Cyber Security Ontologies

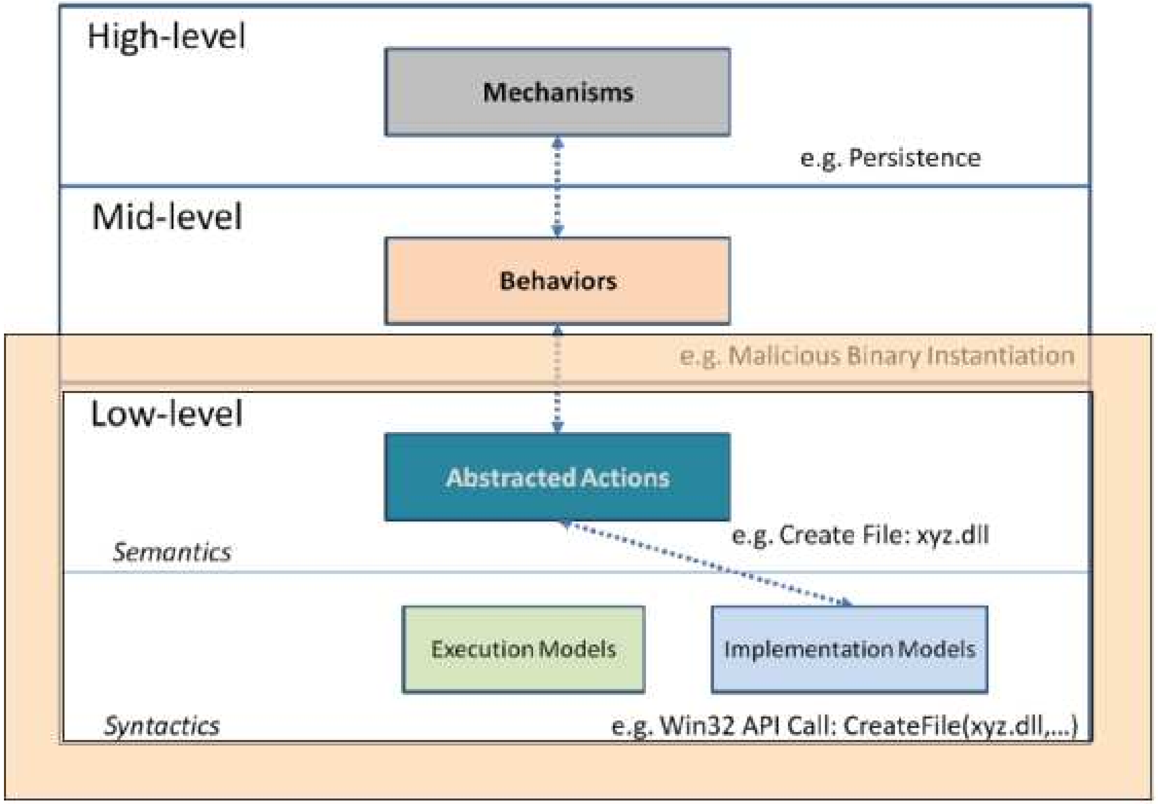
- NetOps Ontology
  - Domain: Government network management
  - Developed by MITRE to support the the Network Operations Community of Interest (COI)
- Swimmer's Malware Ontology (2008)
  - Only non-trivial malware ontology we could find
- Main source for malware domain knowledge: MAEC



MITRE

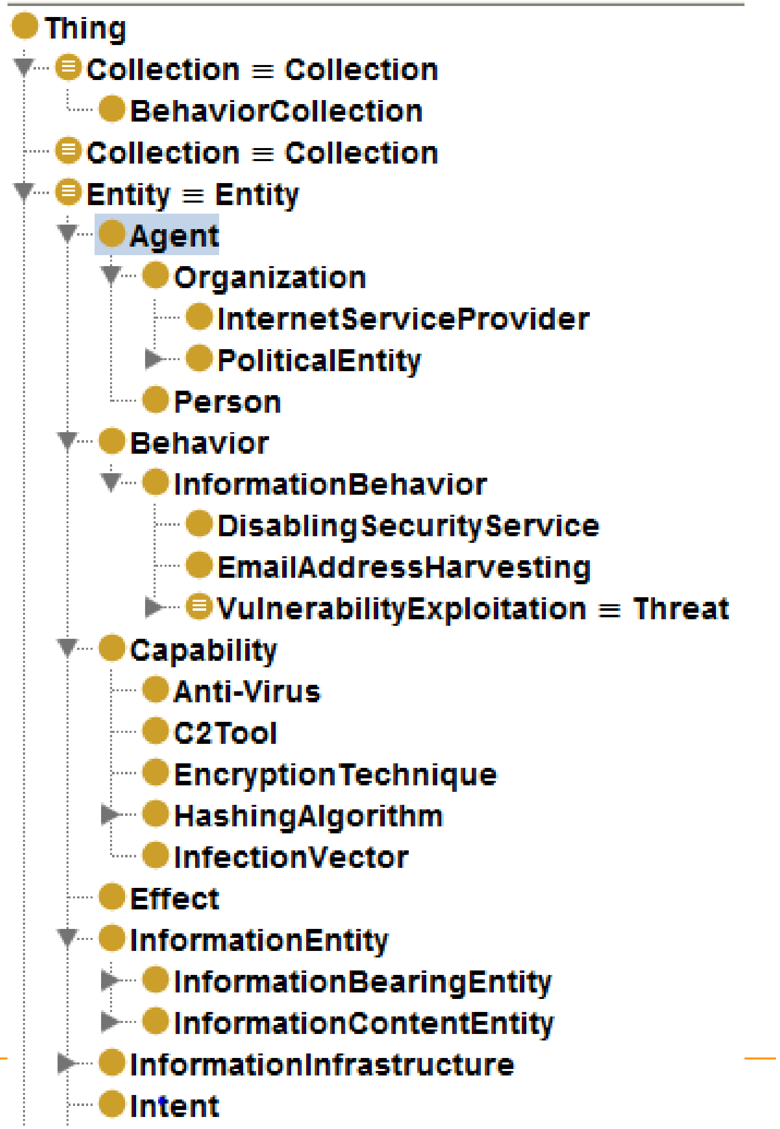
MAEC Tiered Architecture

- Lowest level: Actions such as hardware accesses and system state changes
  - Abstracted away from their implementations
- Middle Level: Discrete components of malware functionality
- Top Level: Organized groups of behaviors
  - Propagation
  - Insertion
  - Self-defense



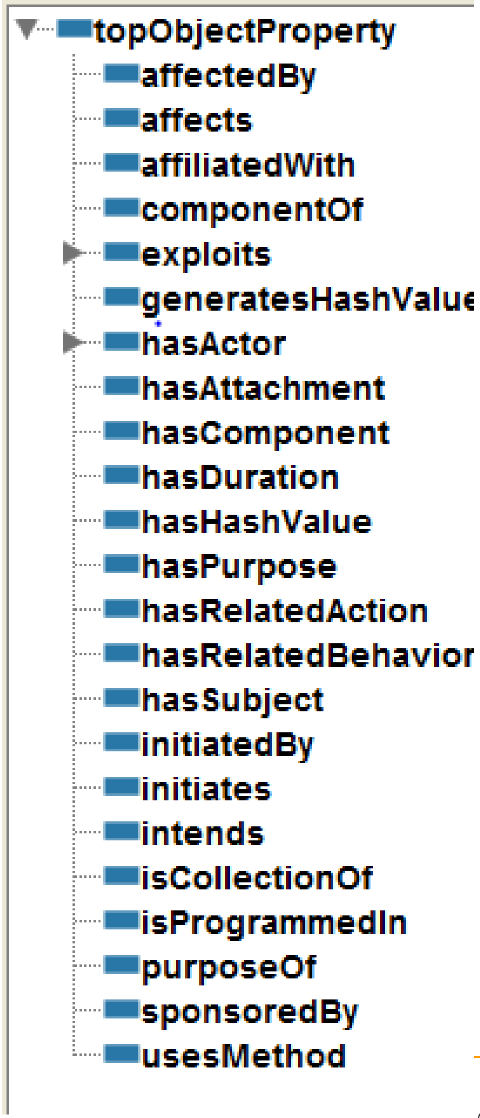
Current Malware Ontology

Classes (partial)



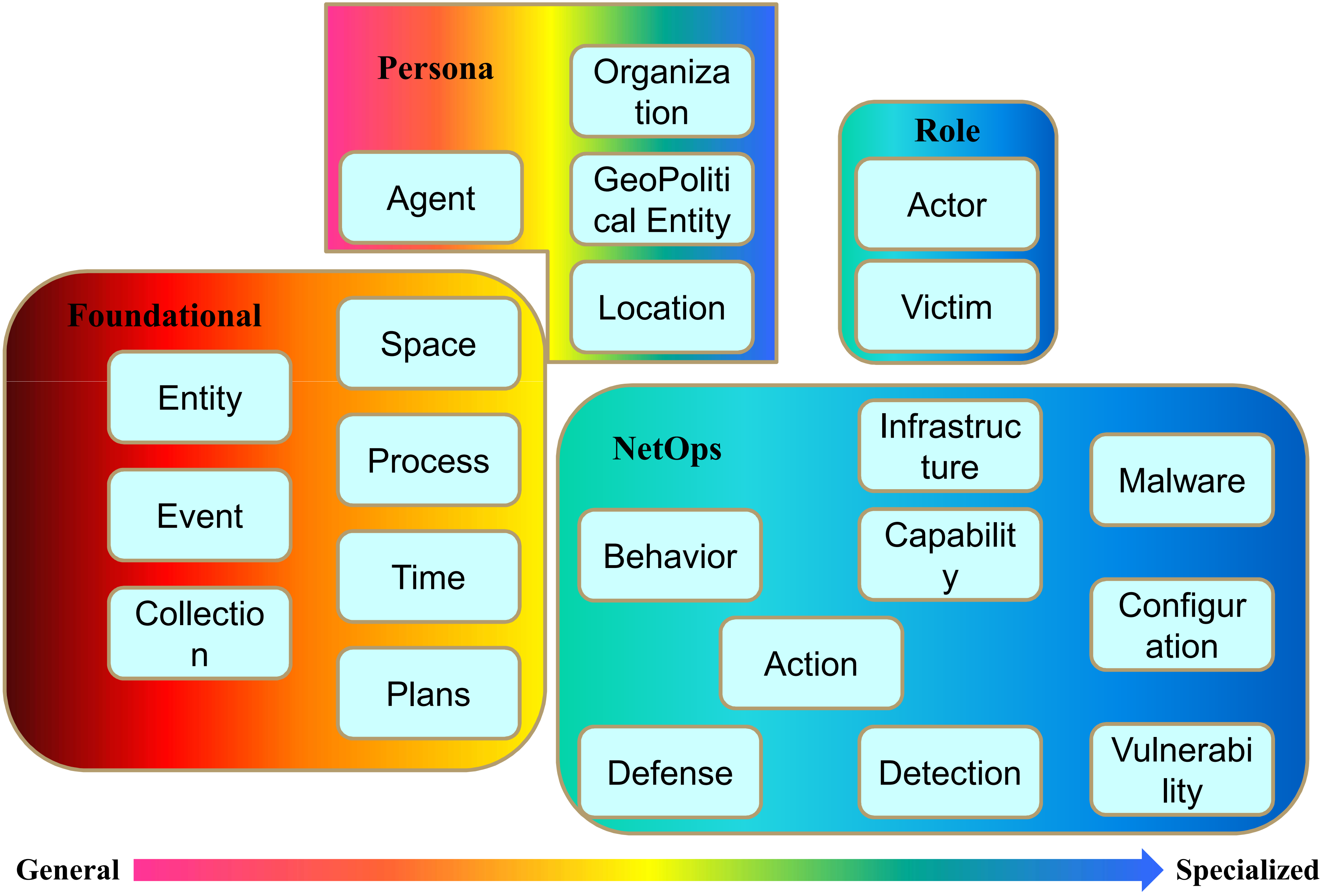
- Thing
  - Collection ≡ Collection
    - BehaviorCollection
  - Collection ≡ Collection
  - Entity ≡ Entity
    - Agent
      - Organization
        - InternetServiceProvider
        - PoliticalEntity
      - Person
    - Behavior
      - InformationBehavior
        - DisablingSecurityService
        - EmailAddressHarvesting
        - VulnerabilityExploitation ≡ Threat
    - Capability
      - Anti-Virus
      - C2Tool
      - EncryptionTechnique
      - HashingAlgorithm
      - InfectionVector
    - Effect
    - InformationEntity
      - InformationBearingEntity
      - InformationContentEntity
    - InformationInfrastructure
    - Intent

Properties



- topObjectProperty
  - affectedBy
  - affects
  - affiliatedWith
  - componentOf
  - exploits
  - generatesHashValue
  - hasActor
  - hasAttachment
  - hasComponent
  - hasDuration
  - hasHashValue
  - hasPurpose
  - hasRelatedAction
  - hasRelatedBehavior
  - hasSubject
  - initiatedBy
  - initiates
  - intends
  - isCollectionOf
  - isProgrammedIn
  - purposeOf
  - sponsoredBy
  - usesMethod

Cyber Ontology Architecture



Persona

Organization

Agent

GeoPolitical Entity

Location

Role

Actor

Victim

Foundational

Space

Entity

Process

Event

Time

Collection

Plans

NetOps

Infrastructure

Malware

Behavior

Capability

Configuration

Action

Defense

Detection

Vulnerability

General

Specialized

Other Cybersecurity Resources

- Incident Object Description and Exchange Format (IODEF)
  - Data format for describing and exchanging incident information
  - From IETF
- OpenIOC
  - XML format for sharing intelligence related to Indicators of Compromise (IOCs)
  - From Mandiant
- Web Application Security Consortium (WASC) Threat Classification
  - Similar to CAPEC
- Verizon Enterprise Risk and Incident Sharing (VERIS) framework
  - Used to collect security incident data
- Many other resources available

Next Steps

- The current Cyber ontology is focused primarily on malware and some “diamond model” aspects
- Need more infrastructure and capabilities
- Expand behavioral aspects and events
- Signatures, complex cyber command & control (C2), obfuscation, encryption support
- Rules & automated reasoning support using Rule Interchange Format (RIF) & Logic Programming
  - Detect prospective malware
  - Provide alerts and rule-based recommendations to human malware analysts

Thanks!