Cyber Security Knowledge Graphs

Supervisor: Dr. Cogan Shimizu

Presented by: Rakesh Kandula

Contents

1) Developing an Ontology for Cyber Security Knowledge Graphs (CSKG)
    i. Motivation behind research
    ii. Related Work
    iii. Data Sources
    iv. Ontology Design and Implementation
    v. Future Work
2) A review of knowledge graph application scenarios in cyber security
    i. Motivation behind research
    ii. Different Cybersecurity Ontologies
    iii. Information Extraction (IE) Methods
    iv. Datasets to develop CSKGs
    v. Application Scenarios
    vi. Research Directions
3) Usage of current research in our research

Developing an Ontology for Cyber Security Knowledge Graphs (CSKG)

Motivation behind Research

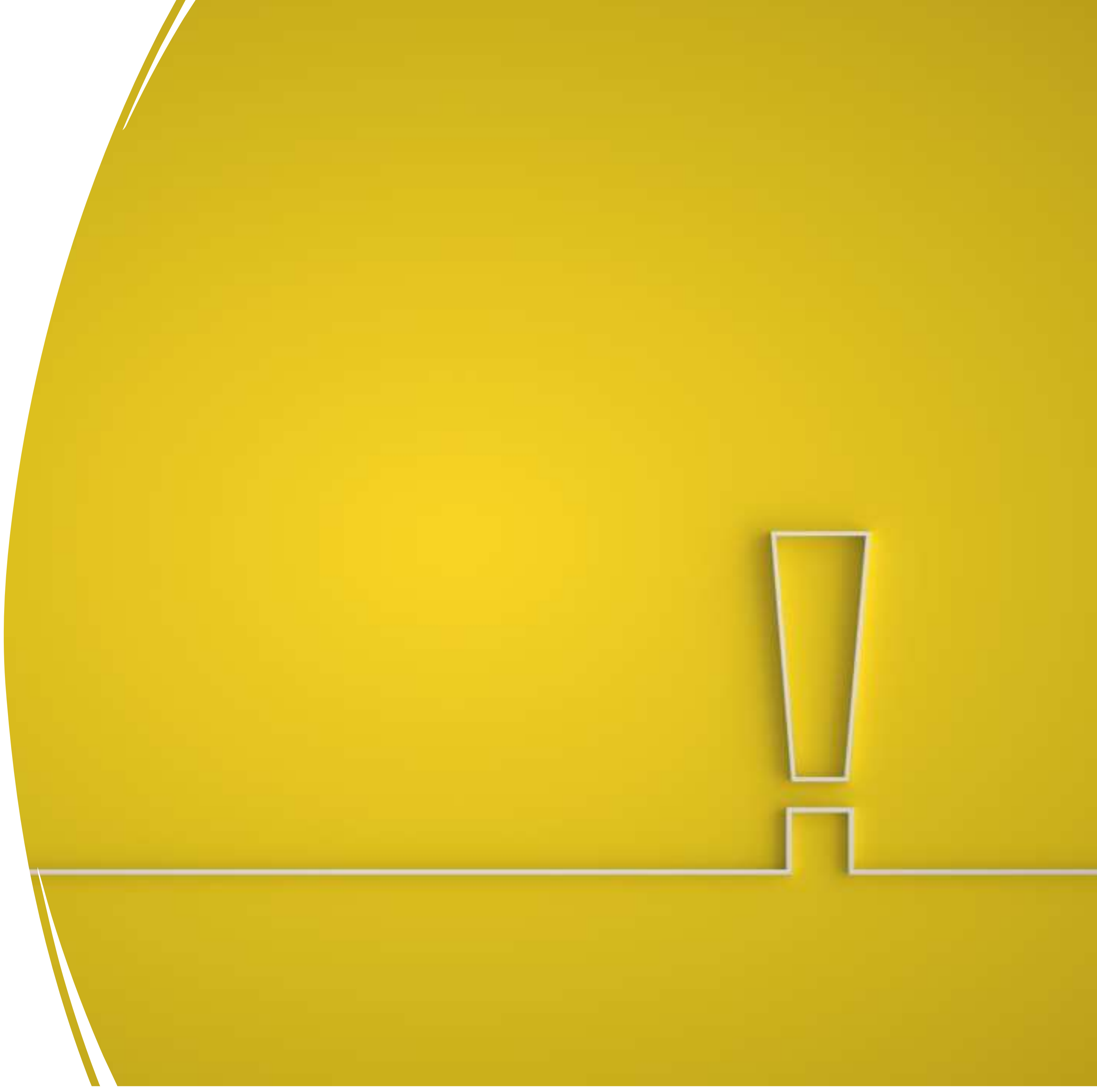
1. The main idea is to address the need of developing a standardized and structured representation of cyber security knowledge in the form of ontology.
2. Improving the Knowledge Representation in Cyber Security Knowledge Graphs (CSKG) Domain.
3. Integrating the Knowledge from different sources.
4. Trying to Standardize the ontology in CSKG.

Related Work

1. Before developing an Ontology, authors researched some of the previous works conducted in this field.
2. Authors mainly focused on two ongoing projects:
    i. Work done by group of authors from University of Maryland Baltimore County (UMBC)
    ii. Secondly, investigations done by Massachusetts Institute of Technology Research Establishment (MITRE) to develop ontology for cyber security domain.

Data Sources

1. In most cases, data is collected mainly from two types of sources:
   1. Structured Source
   2. Unstructured Source
2. This CSKG ontology is developed by integration of data from both structured and unstructured sources.
3. It states that data is collected from 13 structured sources and fed to a pipeline which collects the data and converts it to GraphSON format.
4. There is also ongoing work to develop the pipeline for Unstructured data, but there are some difficulties to develop.



Ontology Design and Implementation

1. The ontology developed contains 15 entity types and 115 properties.
2. It is not mentioned the methodology used in construction of ontology; authors provided the ontology in pictorial format.
3. It specifies the ontology using JSON Schema, which is compatible with GraphSON format, which makes validation of incoming data simpler.
4. Usage of JSON schema has also limitations, compared to OWL like it is simple to perform automatic reasoning and to infer new relationships.

Future Work

1. This paper describes the creation of ontology to represent the cyber security domain, integrating data from different sources.
2. In future, authors are studying the availability of inter-operate with Structured Threat Information Expression (STIX) by MITRE.
3. Interoperability will become increasingly important as this area develops

A Review of Knowledge Graph Application Scenarios in Cyber Security

Motivation behind research (paper2)

1. The main aim of this paper is to demonstrate multiple scenarios and applications of CSKG.
2. The paper discusses the different types of cyber security ontologies.
3. Extraction of cyber security entities and relation between them.
4. Understanding of multiple cybersecurity datasets.
5. This paper mainly focusses on how CSKGs can be utilized in industry to enhance cybersecurity analysis and operations.

Different Cybersecurity Ontologies

The figure describes the different types of cybersecurity ontologies.

There are Unified Security Ontologies and Scenario based Ontologies

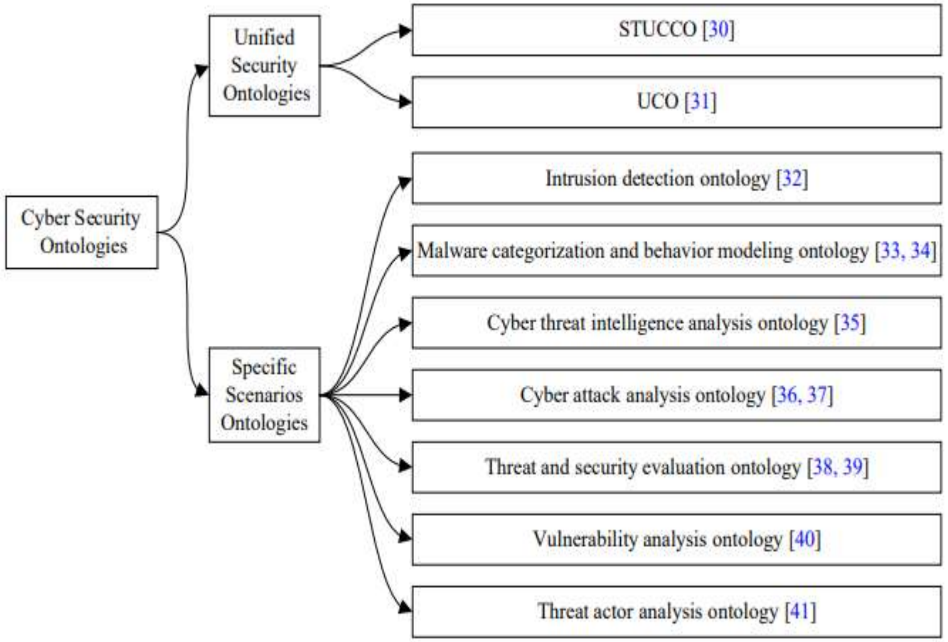

Cyber Security Ontologies
Unified Security Ontologies
STUCCO [30]
UCO [31]
Specific Scenarios Ontologies
Intrusion detection ontology [32]
Malware categorization and behavior modeling ontology [33, 34]
Cyber threat intelligence analysis ontology [35]
Cyber attack analysis ontology [36, 37]
Threat and security evaluation ontology [38, 39]
Vulnerability analysis ontology [40]
Threat actor analysis ontology [41]

Figure: Different Types of CSKG Ontologies

Information Extraction (IE) methods

1

Authors discussed the methodologies to extract Information to construct CSKGs. Mainly there are two types of IE:

- Named Entity Recognition (NER)
- Relation Extraction (RE)

2

While Extracting NER, deep learning methods are beneficial while compared to traditional approaches like rule-based methods, Maximum Entropy Model (MEM), Decision Trees etc.

3

The process to extract relations is not simple, the authors suggested RNN and LSTM methods.

Datasets to develop CSKGs

1. Development of KGs especially in the field of cyber security, it requires huge volume of data and latest one. The authors classified the data sets into three categories:
   1. Open-Source datasets
   2. The datasets for IE in cybersecurity domain
   3. The other datasets which helps new researchers to find new solutions

Application Scenarios

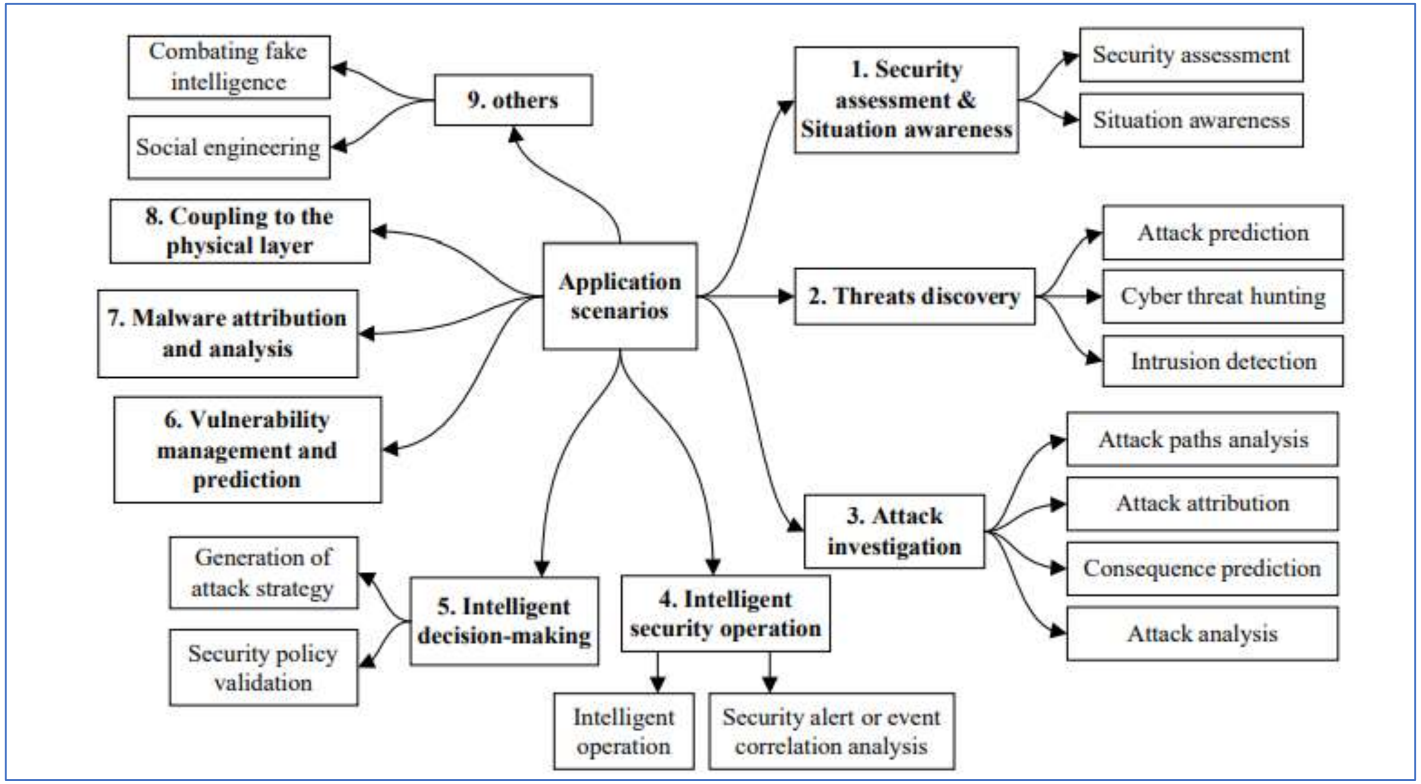

Application scenarios
1. Security assessment & Situation awareness
Security assessment
Situation awareness
2. Threats discovery
Attack prediction
Cyber threat hunting
Intrusion detection
3. Attack investigation
Attack paths analysis
Attack attribution
Consequence prediction
Attack analysis
4. Intelligent security operation
Intelligent operation
Security alert or event correlation analysis
5. Intelligent decision-making
Generation of attack strategy
Security policy validation
6. Vulnerability management and prediction
7. Malware attribution and analysis
8. Coupling to the physical layer
9. others
Combating fake intelligence
Social engineering

Figure: Application scenarios of CSKG

1. There are multiple application scenarios of CSKG.
2. We can see different scenarios in the figure.
3. It all describes about the software threat applications.

Research directions

1. The authors mentioned some research opportunities in CSKG:
   i. Open-source dataset construction
   ii. The construction of a dynamic cyber security knowledge graph
   iii. The application scenarios of the cyber security knowledge graph
   iv. The evaluation criterion of the cyber security knowledge graph

Usage of current research in our research

Research conducted in both papers are identical. One discusses about the Ontology creation and the other describes about the construction, IE, applications of CSKG.

As of now there is no unified ontology in the cyber security domain to construct the knowledge graphs on top of it.

But we can utilize the work done in this field and relate to our research, in areas of Information Extraction from structured or unstructured sources.

Utilize the Relation Extraction methods like deep learning method, RNN, LSTM mentioned in the research.

The methodology for continuous knowledge graph expansion can help to grow our knowledge graph with new data sources, enriched entities and relationships etc.

The classification framework can guide the entities and relationships to model our use case like hardware components, certifications, supply chain, vendor, vulnerabilities etc.

We can use the graph reasoning techniques can be applied to infer the trust scores of hardware components, supply chain, vendor based on their connections and properties in knowledge graph.

thank you