Correlating Cyber Incident Information to Establish Situational Awareness in Critical Infrastructures

Giuseppe Settanni
Austrian Institute of Technology, Digital Safety & Security Department - A-1220 Vienna, Austria

PROPOSED SOLUTION

- Three methods for security information correlation,
  - Artifact-based
  - Word-based
  - Dictionary-based
- Discover implicit relations among relevant documents,
- Support the establishment of cyber situational awareness,
- Based on Vector-Space-Models (VSM),
- Using Cosine Similarity to compare security documents.
- Preliminary Evaluation
  - Accuracy and efficiency assessment performed.
- Application

Features Extraction

Document Dataset \( D = \{d_1, d_2, \ldots, d_m\} \)

Feature Set \( \mathcal{F} = \{f_1, f_2, \ldots, f_n\} \)

Feature Vector \( \mathbf{v}_d = (v_{d_1}, v_{d_2}, \ldots, v_{d_n}) \)

Feature Frequency

Term Frequency \( TF_{f,d} = \frac{d_f}{\sum_{f 
\in \mathcal{F}} d_f} \)

Inverse Document Frequency \( IDF_f = \ln \left( \frac{\# \text{of documents}}{\# \text{of documents containing } f} \right) \)

Binary Frequency \( b_{f,d} = \begin{cases} 1, & \text{if } f, d \in \mathcal{F} \\ 0, & \text{otherwise} \end{cases} \)

Document Similarity Metric

\( s(d_x, d_y) = \cos(\theta) = \frac{\mathbf{v}_x \cdot \mathbf{v}_y}{\|\mathbf{v}_x\| \|\mathbf{v}_y\|} \)

Document Correlation Methods

**Artifact-based**

**Word-based**

**Dictionary-based**

**Evaluation**

Evaluation Dataset Generation

\( s_{ij} \): similarity score between document \( d_i \) and \( d_j \)

\[ s_{ij} = \frac{|t_i \cap t_j|}{|t_i \cup t_j|}, \quad t: \text{tolerance} \]

\( d_{xy} \): # of MDs mutually present in \( d_x \) and \( d_y \)

\[ l_{ij} = l(d_x, d_y) = \max_{(i,j)} \left( \frac{d_{ij}}{\max(d_{ij}, d_{xy})} \right) \]

\( l_{xy} \): linking score

System Implementation

Open Research Questions & Future Work

- Application of semantics to identify related documents
  - Characterize documents according to their meaning
- Definition of new (tailored) correlation metrics
  - How to measure if two documents are related to one another?
- Consider human feedback to improve correlation accuracy
- Use graph theory to analyze and visualize correlation results
- Implement translation of artifacts to be language-independent
- Evaluation with real incident data

Relevant Publications