Value Creation Plan

51MSR615-AA: Clustering Cyber Incidents: Cyber MO

|  |  |  |
| --- | --- | --- |
| Milestone Date | Description | Type of Deliverable |
| November 15, 2013 | Java code which takes as input a corpus of text documents and produces a dissimilarity matrix (based on cosine distance of tfidf vectors (including stemming and a knob to filter off “stop-word” terms). Initial Demonstration on NYTimes articles. | Software (Chris) DONE |
| December 1, 2013 | Construction of R script(s) for taking dissimilarity matrix as input and which outputs cluster descriptions and cluster assignments with confidence AWAITING stress test of Orange | Software (Chris) |
| December 1, 2013 | Java code which implements several varieties of Spectral Clustering. The code will be tested and verified to work in reasonable time on a moderate number of data objects (a few thousand) on my laptop. The Spectral Clustering varieties that will be implemented:  a.            The basic Spectral Clustering algorithm where the user specifies K, the number of clusters.  b.            Three different variations of the basic Spectral Clustering algorithm that eliminates the need for the user to choose K. | Software (Chris) |
| December 1, 2013 | Neo4j bundle demo: example attacks represented as csv, quickly loaded into neo4j + graph query | Software (Adam) |
| December 15, 2013 | Clustering of sample of existing campaigns based on similarity of CRITS listed indicators | Analysis report (Eric) |
| December 15, 2013 | OpenStack VM setup (a VM set up with all of the software and demo data we need to demonstrate tools are working) | Configuration (Adam) |
| Jan 1, 2014 | STIX->CSV parser. Current distance matrix constructor takes csv as input so in order to accept STIX we can write a new data importer or just convert STIX to csv. | Software (Adam) |
| January 15, 2013 | Clustering of existing campaigns based on similarity of CRITS listed TTP | Analysis report (Eric) |
| Feb 1, 2014 | Code which ingests STIX encoded events and supports graphical queries against them such as ‘find all incidents that use are associated with an IP at MITRE also associated with other incidents and which use Malware with a specific MD5 code | Software (Adam) |
| Feb 1, 2014 | Development of spectral clustering code (possibly multiple parts) that takes vectors of cyber incidents in STIX (described by text and/or structured data) and outputs suggested groupings with cluster descriptions and confidence | Software (Chris) |
| Feb 1, 2014 | Establish specific tech transfer opportunities with his DHS Cybersecurity and Communications sponsor | Tech Transfer candidates (Eric working with Sean Barnum) |