Entropy Measures in Critical Infrastructure Graphs

Markus Klemetti, Samir Puuska, Jouko Vankka

National Defence University, Finland

markus.o.klemetti@gmail.com, samir.puuska@mil.fi, jouko.vankka@mil.fi

Critical infrastructure (CI) consists of assets and systems which are essential in maintaining vital societal functions. Disruption or destruction of CI has significant effects on safety, health and security of people. Faults in the infrastructure may quickly affect other systems and cause cascading failures across the interconnected network. Real-time awareness of CI health and performance is a necessity for both everyday usage and efficient incident response and disaster mitigation. Critical infrastructure has become a noteworthy field of contemporary research, where various differing modelling formalisms and analysis methods have been studied. These approaches include graphs, bayesian belief networks, neural networks and fuzzy logic, among other more esoteric ones [1].

In information theory, entropy is understood as a measure for uncertainty or information. The concept of entropy was first proposed in 1948 by Claude E. Shannon in his paper "A Mathematical Theory of Communication" [3]. For a random variable X we define its entropy to be

$$H(X) \equiv -\sum_{x} P(x) \log_2 [P(x)]$$

where x goes through all possible states of X. In other words, entropy is the expected value of information associated to a single event. By examining entropy associated with events, we can deduct how reliable the information actually is.

The information is usually measured in bits (shannons). One bit contains information that is the equivalent of a single coin flip. Entropy is high in cases where the end result is hard to predict (e.g. uniform distribution). Inversely, heavily biased probability distributions have lower entropy, because results do not usually contain much information. The end result is guessable from the distribution alone.

The research described in this paper is part of a larger research project, called Digital Security of Critical Infrastructures (DiSCI). The DiSCI project aims to find solutions for estimating and minimising threats facing the CI at national level. During the DiSCI project, the Situational Awareness of Critical Infrastructure and Networks (SACIN) software framework was developed for evaluating CI monitoring concepts [5]. In the framework, the JDL data fusion model [4] was used for CI sensor integration. The research described in this paper covers the JDL levels 2 and 3 in the SACIN framework. Levels 2 and 3 are responsible for situation analysis and future risk estimation. We have previously presented a concept implementation for critical infrastructure monitoring, as well as modelling and analysis framework suitable for national scale operations [5].

In this paper we expand the previously proposed theoretical framework with time dependent stochastic elements. We accomplish this by exploring the possibility of using the concept of Shannon's entropy as a support tool for CI monitoring and analysis. We also research the possibility of software implementation as a part of our existing software framework.

We presented our proposition for critical infrastructure system (CIS) model in [2]. In our model, critical infrastructure is presented as a directed graph, where each vertex is a CI system. Edge directions represent dependency relations between the systems. Each node is associated with a finite state machine (FSM) which represents the status (health, capability etc.) of the system in question. In this paper we expand the model by associating a probability distribution to each FSM, which accounts for the flow of time and previous confirmed sensor reading. When time passes, it is increasingly more likely that the current status of one particular system does not reflect the last sensor reading. As time passes, the uncertainty about the state of the sensor increases. By relying on statistical probabilities that have been previously observed or are known, it is possible to make predictions about its current state. The probability distributions may have been collected by observing the operation of the sensor for a longer time period, or they may have been defined by the sensor operator.

Attached to each FSM is a *status function* S, where $S: Q \rightarrow [0,1]$ maps each state q of the automaton to a number that represents its state, 0 being not operational and 1 being fully operational. By observing the expected value $E(S(X))$, it is possible to estimate the status of the system in question. The entropy of the random variable $S(X)$ informs us of the reliability of the estimate. The interlinked FSM structure guarantees that the dependencies between sensors are taken into account.

We believe that the proposed model improves CI situational awareness by taking into account the increasing uncertainty created by the passage of time. The changes we have made also help clarify the user interface of SACIN, since in addition to numeral representation it is easy to visualize the expected values and entropies of the sensors using for example different colors and tones on the sensor icons. This allows the operator to quickly obtain an overview of the current situation in the CI network.

References:

[1] Ouyang, Min. "*Review on modeling and simulation of interdependent critical infrastructure systems*." Reliability engineering & System safety 121 (2014): 43-60.

[2] Puuska, Samir et al. *"Modelling and Real-time Analysis of Critical Infrastructure Using Discrete Event Systems on Graphs"*, 2015 IEEE International Conference on Technologies for Homeland Security, April 14 - 16, 2015, Boston, USA.

[3] Shannon, C. E., "*A Mathematical Theory of Communication*." The Bell System Technical Journal, Vol. 27 (1948).

[4] Steinberg, Alan N., Christopher L. Bowman, and Franklin E. White. "*Revisions to the JDL data fusion model*." AeroSense'99. International Society for Optics and Photonics, 1999.

[5] Timonen, Jussi et al. "*Situational awareness and information collection from critical infrastructure*." IEEE 2014 6th International Conference On Cyber Conflict (CyCon 2014), 2014: 157-173.