Scientific Report: Short Term Scientific Mission COST Action CA15127

Fernando Kuipers

May 24, 2017

1 STSM Details

STSM Title: Calculating Risk Profiles for Large-Scale Disasters

STSM Applicant: Dr. Fernando Kuipers, Technische Universiteit Delft, Delft, the Netherlands

Host: János Tapolcai, Budapest University of Technology and Economics, Budapest, Hungary

Period: May 14-20, 2017

Working group: WG1

2 Purpose of the STSM

Through this STSM, our purpose was to start a collaboration to investigate how to compute the survivability of a network to geographically correlated challenges. Typically, the survivability is given as a single value, most often the expected value or worst-case value of the measure, which limits the amount of information that can be obtained from such a computation. Our plan was to focus on calculating the whole probability distribution of a vulnerability measure. Single-value properties like the expected value, or the worst-case value, can subsequently be easily derived from the distribution itself.

3 Description of the work carried out during the STSM

During the STSM visit, Fernando has given a presentation, entitled "Spatiotemporal network resilience." The talk indicated that disasters typically affect a large geographical region and could cause extensive and long-lasting network outages that are not in the scope of existing protection schemes. Moreover, disasters tend to display spatiotemporal characteristics, and consequently link availabilities may vary in time. In the talk, Fernando presented his work on analyzing the vulnerability of networks to (spatiotemporal) disasters and computing alternative routes that minimize the potential damage from a pending disaster.

In addition to the presentation, the bulk of the time was spent on discussing the topic of the STSM. Discussions were held with János Tapolcai, Balázs Vass, Zalan Heszberger, and Gábor Rétvári. During our discussions, it became clear that we could take two paths in parallel: (1) trying to compute the (likelihood of) possible failure states in the network, given a list of possible types of disaster, and (2) trying to compute the (likelihood of) possible probabilistic shared risk link groups (pSRLGs and their corresponding failure states) in the network, based on a given failure model that gives the probability of a link damage as a function of distance, if a disaster hits at a place in the vicinity. The second approach closely related to the work that David Hay had conducted during his STSM at Budapest University of Technology and Economics. We therefore have decided to join forces.

Our results for both activities are described in the following section.

4 Description of the main research results obtained

Research on (the vulnerability of a network to) so-called geographically correlated challenges has mostly reflected the vulnerability of the network by a single value of a single measure.

We take a different approach and consider (1) based on a set of possible disasters (of varying shapes), the distribution of the measure after one of these disasters randomly occurs, and (2) based on a failure model, all possible pSRLGs.

4.1 Given set of disasters

The following work is done together with Jorik Oostenbrink, an MSc student under Fernando's supervision.

We assume the network $G = \{V, E\}$, consisting of nodes V connected by links E, is embedded in a plane, and lies completely in a bounded convex region $R \subseteq \mathbb{R}^2$. The network can either be directed or undirected.

Nodes $v_i \in V$ are modeled as points $p_i \in R$. Instead of modeling them as straight line segments, each link is modeled as a finite sequence of line segments connecting their nodes.

We model disasters deterministically, i.e., we assume all links intersecting a disaster area fail. If a node lies within a disaster area, all of its links must have at least one endpoint in the disaster area and therefore would fail. We do not explicitly model node failures, as the failure of all incident links is equivalent to the node failing.

We assume that we are given a finite set of possible disasters D. We further assume that exactly one of these disasters will manifest at a time. The probability of multiple (independent) disasters occurring simultaneously is generally very small and thus is ignored. Each disaster $d \in D$ has a disaster area $A(d) \subseteq \mathbb{R}^2$ and an occurrence probability $P(d)$. Note that $\sum_{d \in D} P(d) = 1$.

Our model and methods can be used with any shape of disaster area, as long as it is possible to calculate if a line segment intersects it.

There are multiple ways to obtain the set D. For example, in the case of earthquakes, one could generate potential earthquakes in a Monte Carlo approach based on fault parameters. The United States Geological Survey (USGS) provides tools to compute such earthquake scenarios. Another approach could to take a historic set of the last N earthquakes above a certain magnitude. Finally, one could use a given set of earthquake scenarios as input.

As an intermediate step towards computing measure distributions, we first consider the probability distribution over the state of the network after a random disaster.

Let a failure state s be defined as a set $s \subseteq E$, where $e_i \in s$ if and only if e_i is down.

Let S be the random value indicating the failure state after the disaster and let $S(d)$ be the failure state after disaster $d \in D$.

Because we assume exactly one disaster occurs, we have

$$P(S = s) = \sum_{d \in D | S(d) = s} P(d) \tag{1}$$

The value of a measure only depends on the state of the network, and thus it only needs to be computed once per possible failure state, instead of once for each $d \in D$. By iterating over possible failure states instead of disasters, we can potentially significantly reduce computation time when computing the distribution over a measure.

Consider a measure M. Let $M(d)$ be the value of the measure after disaster d, and $M(s)$ the value of the measure in failure state s. Note that $M(d) = M(S(d))$.

Similarly to equation 1, we have

$$\begin{aligned} P(M = m) &= \sum_{d \in D | M(d) = m} P(d) \\ &= \sum_{s \in S[D] | M(s) = m} \left(\sum_{d \in D | S(d) = s} P(d) \right) \\ &= \sum_{s \in S[D] | M(s) = m} P(S = s) \end{aligned} \tag{2}$$

The disadvantage of computing a distribution instead of a single value is that one may be overwhelmed by the amount of data. Thus it is important to properly visualize the results in a useful fashion, for example via a cumulative distribution function (CDF).

4.2 Computing pSRLGs

In this line of work, the network is modeled as an undirected connected geometric graph $G = (V, E)$ with $n = |V|$ nodes and $m = |E|$ links. The nodes of the graph are embedded as points in the Euclidean plane. Assuming that the network area can be projected onto a two-dimensional Cartesian plane, the grid can be generated by partitioning the Cartesian plane into a set $x_{max} \times y_{max}$ equally-sized squares. We refer to each square by its horizontal and vertical position in the grid, i.e. (x, y) refers to the square in the x^{th} column and y^{th} row. For ease of presentation, we place the network in a coordinate system where the (x, y) coordinate is exactly the center point of square (x, y).

Each grid rectangle (x, y) is assigned with a **risk** $g(x, y)$ in the form of a non-negative value of at most one, which represents the probability that the area bounded by square (x, y) contains the epicenter of a disaster during a specific time period. The risk that is assigned to each grid square depends on the geospatial attributes of the network area bounded by the grid square. Adjacent grid rectangles may or may not be assigned with equal risk value.

Let $\delta(x, y, e)$ be the conditional probability that link e fails, given that there is a disaster with epicenter (x, y). Then our goal is to compute the probability of all possible failure states, e.g. the probability that two links close to each other fail if a disaster strikes close to them. Since the two links do not necessarily fail independently from each other, e.g. consider two links on the same position in a conduit, the mathematics becomes challenging and we are still in the process of getting to a solution.

5 Future collaboration with the Host institution

During the STSM, it became obvious that Fernando Kuipers and János Tapolcai share many research interests, and therefore we plan on collaborating well beyond the scope of this STSM. Of course, our present and near-future collaboration will revolve around following up on the work (e.g., on pSRLG) that was initiated in this STSM. Moreover, our collaboration also includes David Hay, who previously had spent an STSM with János.

6 Foreseen publications/articles resulting from the STSM

Fernando has been invited to give a talk (+ corresponding paper) at the first ACM international workshop on Critical Infrastructure Network Security (CINS), to be held in Urbana-Champaign, Illinois, USA on June 5, 2017, in conjunction with ACM SIGMETRICS. Fernando will work together with Jorik Oostenbrink to present there the work on computing network vulnerability given a set of possible disasters as input.

In addition, we (Fernando Kuipers, János Tapolcai, Balázs Vass, Zalan Heszberger, and David Hay) are currently in the process of working out the pSRLG model, with the plan to submit it as a paper to a conference. The work might also include a section related to computing network vulnerability given a set of possible disasters as input.