Spatiotemporal Risk-Averse Routing

Farabi Iqbal and Fernando Kuipers

Network Architectures and Services, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands
{M.A.F.Iqbal, F.A.Kuipers}@tudelft.nl

Abstract—A cyber-physical system is often designed as a network in which critical information is transmitted. However, network links may fail, possibly as the result of a disaster. Disasters tend to display spatiotemporal characteristics, and consequently link availabilities may vary in time. Yet, the requested connection availability of traffic must be satisfied at all times, even under disasters. In this paper, we argue that often the spatiotemporal impact of disasters can be predicted, such that suitable actions can be taken, before the disaster manifests, to ensure the availability of connections. Our main contributions are three-fold: (1) we propose a generic grid-based model to represent the risk profile of a network area and relate the risk profile to the availability of links and connections, (2) we propose a polynomial-time algorithm to identify connections that are vulnerable to an emerging disaster risk, and (3) we consider the predicted spatiotemporal disaster impact, and propose a polynomial-time algorithm based on an auxiliary graph to find the most risk-averse path under a time constraint.

I. INTRODUCTION

A Cyber-Physical System (CPS) is a system of physical elements that are managed and controlled by intelligent computational elements. The computational elements detect potential issues with the physical systems and react accordingly by reconfiguring the physical systems. One of the important aspects in managing the network of a CPS is that the availability of network services, e.g., network connectivity, is ensured at all times. A network connection between two network nodes is often provided via an end-to-end path (a sequence of network links) between the nodes. Network clients often care only about their connection availability (the probability that the connection is functioning at a random time in the future), and are often oblivious to how the end-to-end path is assigned. Different network clients may request different connection availability and the assigned end-to-end path must satisfy that requested availability, even under the failure of network links.

The availability of a connection depends on the availability of the links constituting its assigned end-to-end path. Although links are designed to be as robust as possible, link failures are still a recurring problem, especially due to natural disasters (adverse events due to the force of nature, e.g., earthquakes, hurricanes and floods) and human-based disasters (adverse events due to intentional or accidental actions of humans, e.g., construction works, nuclear explosions and sabotage). Safeguarding connections against disaster risks is important for satisfying the requested connection availability.

Certain disaster risks may be anticipated beforehand, e.g., by disaster early warning systems (e.g., hurricanes can be anticipated hours in advance [1]) or by predicting near-future disaster occurrences from earlier statistics. For instance, Donnellan et al. [2] conducted a study to estimate the probability that an earthquake of certain magnitude occurs near Los Angeles between May 2015 and May 2018, based on the earlier March 2014 earthquake. When the geospatial impact of disaster risks on the network area can be foreseen, CPS network operators can configure new connections with safer end-to-end paths or reroute vulnerable existing connections through safer network areas. Disaster risks may also display spatiotemporal behavior by moving around in the network area, affecting different parts of the network area at different times. Hence, the spatiotemporal nature of disaster risks needs also to be considered in ensuring the availability of connections.

Our main contributions can be summarized as follows:

- We develop a generic grid-based model to represent the risk profile of a network area and relate the risk profile to the availability of links and connections.
- We propose a polynomial-time algorithm to identify connections that are vulnerable to a disaster risk.
- We propose a polynomial-time algorithm, based on the generation of a flexible auxiliary graph, for finding the most risk-averse end-to-end path under a time constraint, when disaster risks are spatiotemporal.

The remainder of this paper is organized as follows. In Section II, we introduce our proposed grid-based model, discuss possible approaches for assigning the risk profiles, and relate the risk profiles to the availability of links and connections. We propose an approach for identifying connections that are vulnerable to an emerging disaster risk in Section III and analyze the effect of different disaster sizes on the number of vulnerable connections and for different network utilization levels. Section IV explains our approach for finding the most risk-averse path under a time constraint. We discuss related work in Section V and conclude in Section VI.

II. GRID-BASED MODEL

A. Availability of Grid Rectangles

We propose a grid-based model of equally-sized rectangles for representing the network area (e.g., a terrestrial network area, an underwater network area, an urban network area or any combination of them). 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 $F$ of $|F|$ equally-sized rectangles. Each grid rectangle $f \in F$ is assigned with a risk in the form of an availability value.
Fig. 1. An example of a grid of risk profiles.

Fig. 2. Different risk boundaries.

Fig. 3. Overlapping disaster risks.

$A_f$ between zero to one, which represents the probability that the network area bounded by rectangle $f$ is free from the impact of disasters during a specific time period. The risk that is assigned to each grid rectangle depends on the geospatial attributes of the network area bounded by the grid rectangle. Adjacent grid rectangles may or may not be assigned with equal risk value. For instance, almost ninety percent of the world’s earthquakes occur along the Pacific Ring of Fire [3]. Link failures also occur more frequently in areas with higher populations, such that a grid rectangle in a city should be assigned with a grid availability that is lower than a grid rectangle in a rural area. Figure 1 shows an example of a network area modeled by 104 grid rectangles. The accuracy of the grid in representing a network area can always be tuned by adjusting the granularity of the grid ($|F|$).

The availability of a grid rectangle can also be determined by the risk of disasters in the grid rectangle. A disaster risk $r$ is characterized by its occurrence probability $Pr(r^o)$ and impact probability $Pr(r^i)$. Both probabilities are important since although natural disasters have less occurrence probability than human-based disasters, natural disasters often have higher impact probability than human-based disasters [4]. Disasters can also occur without enough impact to damage their area-of-effect, e.g., an earthquake of magnitude below 2.5 poses no harm to buildings. The probability $Pr(r)$ of a disaster $r$ occurring and damaging its area-of-effect is

$$Pr(r) = Pr(r^o) \times Pr(r^i)$$ (1)

Our grid-based model also eases the representation of various risk boundaries (e.g., confined risks, scattered risks and risks with unaffected interior gaps as shown in Figure 2). Confined risks, e.g., controlled demolitions and electromagnetic pulse attacks, have contained area-of-effect with regular or irregular boundaries. Scattered risks, e.g., heat waves and thunderstorms, have scattered area-of-effects. A grid rectangle $f$ can be affected by a set of disasters $R$ as shown in Figure 3, with each disaster $r \in R$ occurring independently of one another, but can occur simultaneously. The availability $A_f$ of a grid rectangle $f$ is

$$A_f = \prod_{r \in R} (1 - Pr(r_f))$$ (2)

where $Pr(r_f)$ is the probability of disaster risk $r \in R$ occurring and damaging grid rectangle $f$.

B. Availability of Links and Paths

A network $G$ consists of a set $N$ of $|N|$ network nodes and a set $L$ of $|L|$ network links. We focus on link availability, since link failures are more frequent than node failures [5]. Each link $(u, v) \in L$ can be represented as a straight line between nodes $u$ and $v$, or as non-straight concatenations of multiple straight line segments of irregular lengths between nodes $u$ and $v$ [6]. Each link $(u, v) \in L$ overlaps a set of grid rectangles $O_{uv} \subseteq F$. The failure of any grid rectangle $f \in O_{uv}$ causes the failure of link $(u, v)$, irrespective of the other grid rectangles in $O_{uv}$ that do not fail. We consider the availability $A_{uv}$ of each link $(u, v) \in L$ as the product of the availability of all the grid rectangles that link $(u, v)$ crosses.
detects nodes ∈ f.

Algorithm 1 Detecting Vulnerable Connections
1: populate an R-tree Y with all the grid rectangles f ∈ F
2: for each link (u, v) ∈ L
3: compute its minimum bounding rectangle MBR_{uv}
4: find the set O_{uv} ∈ Y that overlaps MBR_{uv}
5: for each grid rectangle f ∈ O_{uv}
6: if f does not overlap link (u, v)
7: remove f from O_{uv}
8: compute the projected availability A'_{uv} of link (u, v)
9: for each connection c ∈ C
10: compute its projected path availability A'_{P_c}
11: if A'_{P_c} < A_c
12: add c into the vulnerable connection set C'

A_{uv} = \prod_{f \in O_{uv}} A_f \tag{3}

The availability of a connection equals the availability of its assigned end-to-end path. Since a path P consists of a number of links, the availability of a path A_P is the product of the availability of its links.

A_P = \prod_{(u, v) \in P} A_{uv} \tag{4}

III. DETECTION OF VULNERABLE CONNECTIONS

In the emergence of a risk of disaster to parts of the network area at a point in time, vulnerable existing connections (connections that cannot satisfy their requested availability once the disaster manifests) need to be detected and properly rerouted to safer paths. Only then can the availability of connections be ensured.

A. Problem Definition

Detection of Vulnerable Connections (DVC) problem: Given a network G of a set N of |N| nodes and a set L of |L| links, a grid F of |F| grid rectangles representing the area into which G is embedded, a set C of |C| existing connections, and a set F' ⊆ F of |F'| grid rectangles that are vulnerable to disaster risk r. Each grid rectangle f ∈ F is characterized by a grid availability A_f, and each grid rectangle f ∈ F' is characterized by a projected worst-case reduced grid availability A'_f due to disaster risk r. Each link (u, v) ∈ L connects nodes u and v, and overlaps a set O_{uv} ⊆ F of |O_{uv}| grid rectangles. Each connection c ∈ C is characterized by a requested connection availability A_c and an end-to-end path P_c. Identify the set C' ⊆ C of connections that are vulnerable to disaster risk r.

The DVC problem is polynomially solvable when the grid-based model of Section II is considered.

B. Our Approach

We propose Algorithm 1 for solving the DVC problem. In line 1 of Algorithm 1, an R-tree [7] (a depth-balanced data structure for organizing objects using bounded rectangles) is populated with all the grid rectangles. Lines 3-4 use the minimum bounding rectangle (MBR) of each link for performing a window query on the R-tree Y, by recursively checking the R-tree nodes for grid rectangles that overlap the MBR of the link. Lines 5-7 confirm that the grid rectangles overlap the link and not just the MBR of the link. The R-tree eliminates the need for checking pairwise overlap between all possible link and grid rectangle pairs, by identifying beforehand the grid rectangles that may overlap each link. The projected availability of links is computed using Equation 3 in line 8, and the projected availability of existing connections is computed using Equation 4 in line 10. If the projected path availability A'_{P_c} of a connection c is less than its requested connection availability A_c, c is vulnerable to disaster risk r.

Populating the R-tree takes at most O(|F| \log |F|) time [8]. Finding the grid rectangles that overlap each link takes at most O(|F|) time, since in the worst case, a link can overlap all grid rectangles. Computing the availability of connections and identifying vulnerable connections takes at most O(|C||L|) time. Summing up all contributions, the worst-case time complexity of Algorithm 1 is O(|F| \log |F| + |L||F| + |C|).

C. Analysis

We analyze the effect of the size of the disaster risk to the number of vulnerable connections for different levels of network utilization. We generate the network as a Waxman graph [9] in a grid F of |F| grid rectangles. The Waxman network is frequently used for representing spatial networks, e.g., the Internet topology [10]. |N| nodes are placed uniformly at random coordinates in the grid, and the link existence is reflected by ie, where ℓ_{uv} is the Euclidean distance between nodes u and v, and a is the maximum Euclidean distance between any nodes. Higher i leads to higher link densities, and lower j leads to shorter links. We consider only connected graphs, such that there is at least one path between each node. Each grid rectangle f ∈ F is assigned with a random availability A_f between 0.9999 and 1.0000. Simulations are conducted on an Intel(R) Core i7-4600U 2.1GHz machine of 16GB RAM memory, with |F| = 2500, |N| = 20, i = 0.6 and j = 0.6. All results are averaged over five thousand runs.

We generate a random set |C| of C existing connections according to the network utilization level for each simulation run. The network utilization is the average utilization of all links, with each link having |W| = 50 capacity. In an iterative manner (until the network utilization is reached), a connection c is assigned with a random source-destination node pair (x_c, y_c), and an end-to-end path P_c with the highest possible availability A_{P_c} (using Dijkstra’s algorithm [11] with \log A_{uv} as the link weight of each link (u, v) ∈ L). Each connection c ∈ C is then assigned with a random requested connection availability A_c between 0.7000 and A_{P_c}.

We consider both confined and scattered risks in our analysis. We generate a confined emerging risk by randomly selecting a grid rectangle as the epicenter, and randomly expanding set F' with one of the adjacent grid rectangles until the required |F'| is achieved. We ensure that confined
Emerging risks of different sizes have the same epicenter for each simulation run for a fair analysis. We generate a scattered emerging risk by randomly selecting \(|F'|\) grid rectangles from \(F\). The projected reduced availability (once the risk manifests) of each grid rectangle \(f \in F'\) is assumed to be half of its original value.

Figure 4 shows the effect of the size of the emerging risk on the number of vulnerable connections. As the size of the emerging risk increases, more connections are vulnerable to the emerging risk. A scattered emerging risk is more detrimental to connections than a confined emerging risk. More connections are also vulnerable to the emerging risk as the network utilization level increases, as shown in Figure 5.

Vulnerable connections need to be rerouted through safer network areas, such that the connection availability can be satisfied when the risk manifests. Figure 6 categorizes vulnerable connections into reroutable and unreroutable connections. A connection is reroutable if there is at least an alternate path in the network that can satisfy the requested connection availability. Else, the connection is unreroutable. Although the number of reroutable connections increases with network utilization, the ratio between the number of reroutable connections and the number of unreroutable connections increases as well.

IV. Spatiotemporal Risk-Averse Routing

Disasters may also travel within or pass through the network area, such that its impact on the risk profiles of the network area differs in time. For example, consider the network area shown in Figure 7. The network consists of four nodes and four links, and the network area is represented as a grid of 36 grid rectangles. It takes one time slot to traverse links (1, 3) and (3, 4), and two time slots to traverse links (1, 2) and (2, 4). At time slot \(t_0\), a hurricane manifests at the upper right part of the network area, reducing the availability of the grid rectangles in its area-of-effect. After a time slot, the hurricane moves towards the lower middle of the network area with stronger impact, while affecting links (1, 3), (2, 4) and (3, 4). After another time slot, the hurricane grows stronger and moves towards the upper left part of the network area, while affecting links (1, 2), (1, 3) and (3, 4). Hence, some grid rectangles have different availabilities at different time slots.

A. Problem Definition

Spatiotemporal Risk-Averse Routing (SRR) problem: Given a network \(G\) of a set \(N\) of \(|N|\) nodes and a set \(L\) of \(|L|\) links, a grid \(F\) of \(|F|\) grid rectangles representing the area of \(G\), a source node \(x \in N\), a destination node \(y \in N\), and a time window \(T\) of \(|T|\) time slots. Each grid rectangle \(f \in F\) is characterized by a grid availability \(A_{ft}\), for each time slot \(t \in T\). Each link \((u, v) \in L\) connects nodes \(u\) and \(v\), is characterized by a link delay \(\delta_{uv}\), in the unit of time slots) and overlaps a set of grid rectangles \(O_{uv} \subseteq F\) of \(|O_{uv}|\) grid rectangles. Find a path \(P\) from node \(x\) to node \(y\), between the time period \(t_{\Delta_1} \in T\) and \(t_{\Delta_2} \in T\), such that the path availability \(A_P\) is maximized.

The time window is assumed to be discretized into discrete time slots (e.g., by using the common divisor among all link delays as the unit of the time slots). Each link delay represents the number of time slots required to traverse the link. A routing decision is made before the travel commences, and traffic follows the assigned end-to-end path irrespective of any further network state change. Waiting may be allowed at certain or all nodes, such that traffic can stay for a duration of time slots at the nodes before leaving the nodes. By waiting at a node, the
(b) risk profiles at $t_0$

(a) risk profiles at $t_0$

(c) risk profiles at $t_2$

Fig. 7. An example of a grid with spatiotemporal risk profiles.

Algorithm 2 Auxilary Graph Generation

1: initialize an empty graph $H = (V, E)$
2: for each node node $n \in N$
3:  for each time slot $t \in T$ where $t < t_{\text{max}}$
4:    add nodes $n_t$ and $n_{t+1}$ into $V$
5:    if waiting is allowed at node $n$ at time slot $t$
6:      add link $(n_t, n_{t+1})$ into $E$ where $A_{n_t, n_{t+1}} = 1$
7:  for each link $(u, v) \in L$
8:    for each time slot $t$ where $t + \ell_{uv} \leq t_{\text{max}}$
9:    insert link $(u_t, v_{t+\ell_{uv}})$ into $E$
10: for each link $(u, v) \in E$
11: compute the worst-case availability $A_{uv}$ of link $(u, v)$
12: $\ell_{uv} = - \log A_{uv}$

The SRR problem finds the most risk-averse path (the path with the highest possible availability), while ensuring that the traffic reaches the destination node at least at $t_{\Delta_2}$. In the SRR problem, links have fixed delay but spatiotemporal availability (since the availabilities of grid rectangles are spatiotemporal). The SRR problem is thus a multi-criteria problem that maximizes the path availability under a path delay constraint. The SRR problem is polynomially solvable when the grid model of Section II and the notion of time slots are considered.

B. Our Approach

We propose a polynomial-time graph transformation algorithm (shown in Algorithm 2) that uses an auxiliary graph to reflect the notion of time slots. Using our auxiliary graph, the SRR problem can be solved by a polynomial-time min-cost routing algorithm (e.g., Dijkstra’s algorithm [11]). For instance, the auxiliary graph for the network in Figure 7 is shown in Figure 8. In lines 2-6, each node $n \in N$ is represented by $|T|$ auxiliary nodes. Vertical unidirectional auxiliary links with perfect availability are added between the different time slots of an auxiliary node when waiting is allowed at the node during that time slot. In lines 7-9, each link $(u, v) \in L$ is represented by at most $|T|$ unidirectional auxiliary links, which also reflect the time slots needed to traverse the link. We consider the availability of an auxiliary link to be the worst-case availability of the link during the time slots spent to traverse the link. The auxiliary graph contains at most $|N||T|$ auxiliary nodes and $(|N| + |L|)(|T| - 1)$ auxiliary links. In line 12, the auxiliary link weight $\ell_{uv}$ of each auxiliary link $(u, v) \in E$ is set to the negative logarithmic value of its availability. The worst-case time complexity of Algorithm 2 is $O(|L||F| + |T||(|N| + |L|))$.

The most risk-averse path from node $x \in N$ to node $y \in N$ between $t_{\Delta_1}$ and $t_{\Delta_2}$ can be acquired by using an appropriate min-cost routing algorithm (e.g., Dijkstra’s algorithm [11]) to find the min-cost path (using $\ell_{uv}$ as the cost of each link $(u, v) \in E$) from node $x_{\Delta_1} \in V$ to a temporarily created node $y' \in V$ that is connected from nodes $(y_{\Delta_1}, y_{\Delta_1+1}, \ldots, y_{\Delta_2}) \in V$ via directed links with zero link cost, in the auxiliary graph $H$. $y'$ is temporarily created because the traffic may arrive at the destination node earlier than $t_{\Delta_2}$, while waiting is not allowed at the destination node. When disjoint risk-averse paths are needed, an appropriate min-cost disjoint paths algorithm (e.g., Suurballe’s algorithm [12]) can be used instead.

C. Analysis

We analyze the effect of the size of the time window on the time required to generate the auxiliary graph and find the most risk-averse path. We again use a connected Waxman graph with the properties mentioned earlier in Section III, for each
simulation run. Each link \((u, v) \in L\) is randomly assigned with a delay \(\ell_{uv}\) between one to four time slots. Each grid rectangle \(f \in F\) is assigned with a random availability \(A_f\) between 0.00 and 1.00 for each time slot. Waiting is allowed indefinitely at all nodes.

The time required to generate the auxiliary graph increases with the increase of the size of the time window, as shown in Figure 9, with line 11 of Algorithm 2 dominating the running time. It is worth noting that the auxiliary graph need only be created once for a specific time window, and can be reused to find the most risk-averse path for any other node pair under time constraints that are part of the time window. The time required to find the most risk-averse path, if it exist, in the auxiliary graph is substantially less than the time required to generate the auxiliary graph in all the tested cases.

V. RELATED WORK

Kuipers provides an overview of survivability algorithms [13]. Dikbiyik et al. [14] propose risk-aware provisioning of connections to minimize the loss for network operators when a disaster occurs. They also consider a post-disaster reprovisioning scheme to recover disrupted connections. We, however, aim to reduce the number of disrupted connections by detecting vulnerable connections before the disaster, and route connections using the most risk-averse paths. We also consider link availability as a function of the spatiotemporal risk profile, instead of link component availabilities [15]. Conventional temporal routing, e.g., [16], [17], often aims to minimize the expected end-to-end path delay under temporal link delays. On the other hand, we maximize the path availability under spatiotemporal link availabilities, while also considering a time constraint under fixed link delays. In addition, our grid-based model enables the representation of more complex disaster boundaries, complementing earlier work that assumes specific geometric shapes of disaster boundaries, e.g., circular [18], ellipses [19], general polygons [19] or half-planes [20].

VI. CONCLUSION

In this paper, we propose a generic grid-based model to represent the risk profile of a network area, a polynomial-time algorithm to identify connections that are vulnerable under the risk of a disaster, and a polynomial-time algorithm to find the most risk-averse end-to-end path under a time constraint when disaster risks are spatiotemporal. We also show that larger disaster size leads to more vulnerable connections, and scattered disasters are more detrimental to network connections than confined disasters. The number of vulnerable connections increases with the increase in network utilization, and the possibility of rerouting vulnerable connections using alternative paths decreases with the increase in network utilization.

Possible future directions that can be derived from this paper are finding the minimum delay path that satisfies an availability constraint, using a probability density function to represent the risk profile, and extending the grid-based model for use in a three-dimensional Cartesian plane.

REFERENCES