Coping with Physical Attacks on Random Network Structures

Omer Gold, *Student Member, IEEE,*
Department of Computer Science
Bar-Ilan University, Ramat Gan, Israel 52900
goldgeo@cs.biu.ac.il

Reuven Cohen
Department of Mathematics
Bar-Ilan University, Ramat Gan, Israel 52900
reuven@math.biu.ac.il

***Abstract*—Communication networks are vulnerable to natural disasters, such as earthquakes or floods, as well as to physical attacks, such as an Electromagnetic Pulse (EMP) attack. Such real-world events happen at *specific geographical locations* and disrupt specific parts of the network. Therefore, the geographical layout of the network determines the impact of such events on the network's connectivity. Recent works focused on assessing the vulnerability of a deterministic (geographical) network to such events. Here, we focus on assessing the vulnerability of (geographical) *random networks* to such disasters and identifying the most vulnerable parts of a network where only partial (probabilistic) information about its geographical layout is given. We consider stochastic models in which nodes and links are probabilistically distributed geographically on a plane, and model the disaster event as a circular cut that destroys any node or link within or intersecting the circle. We develop algorithms for assessing the damage of such attacks and determining which attack locations have the most disruptive impact on the network. Our novel approach allows identifying locations which require additional protection efforts (e.g., equipment shielding). Overall, the paper demonstrates that using stochastic modeling and geometric probability techniques can significantly contribute to our understanding of network survivability and resilience.**

***Index Terms*—Network survivability, physical attacks, geographic networks, random networks, fiber-optic, network reliability, large scale failures, Electromagnetic Pulse (EMP).**

I. Introduction

In the last decades, telecommunication networks have been increasingly crucial for information distribution, control of infrastructure and technological services, as well as for economies in general. Large scale malfunctions and failures in these networks, due to natural disasters, operator errors or malicious attacks pose a considerable threat to the well being and health of individuals all over the industrialized world. It is therefore of considerable importance to investigate the robustness and vulnerabilities of such networks, and to find methods for improving their resilience and stability.

This emerging field of geographically correlated failures has started gaining attention only recently [2], [13]. However, unlike most of the recent work in this field, we focus on *random networks*. We consider a stochastic model in which nodes and links are probabilistically distributed geographically on a plane. The motivation behind it is to examine the reliability of a network where we possess only partial (probabilistic) information about its geographical layout. For example, a geographically hidden network where the adversary possesses only partial information about the network topology or no knowledge at all. For example, in densely populated areas the probability for stations (nodes) to exist is high compared to desolated areas, in which it is less likely to find many stations. Similarly, the probability for existence of a fiber (link) between two stations can be modeled as a function of the distance between the stations, the population density in the station's regions, and possibly other parameters relating to the endpoints and geography. Another example is the case where an adversary possess a 'noisy' map of the network's geographical layout.

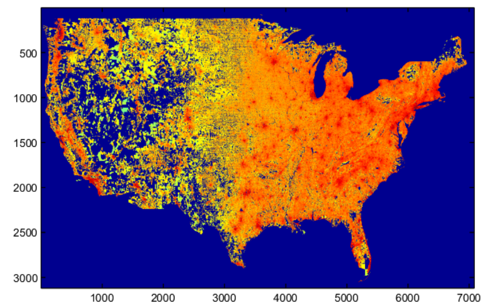

Fig. 1. Color map of the the USA population density in logarithmic scale. Data is taken from [7].

The ability to probabilistically model a network using information such as demography, topography, economy, can be used to produce an input to our algorithm for determining a location where an attack will cause maximum expected disruptive damage in terms of capacity or connectivity. This is important in assessing the expected damage from an attack by an adversary with limited knowledge. In order to design a more robust and well defended system one can consider the resilience of the actual network topology compared to the appropriate random model, and also consider the effect of hiding the actual physical location of fibers on the attack strategy and expected damage by an adversary.

II. Related Work

The issue of network survivability and resilience has been extensively studied in the past (e.g., [4], [6], [8]–[10], [16], [19], [20] and references therein). Most of these works concentrated on network topology, but did not consider the physical location of nodes and links.

The subject of random network models has been studied extensively. See, e.g., [3]. Some of the studied models also incorporated geometry into the random model. See, e.g., [5]. The properties of the real Internet topology and geometry have been studied experimentally in [14], [15], [18].

Recently, the subject of geographically correlated failures has been proposed [13], in which the effect of attacks or disasters span an extensive geographic area, affecting all network equipment within this area. In [13] an algorithm was devised for studying the effect of geographically correlated failures on a deterministic network, and a case study was presented about a fiber backbone in the USA. In [12] random geographical failures (cuts) to deterministic networks were studied, and in [11] the effect of cuts on flows has been studied.

Major contribution has been made recently in [1], [2], where runtime improvements and extensions were given, allowing for probabilistic failures (e.g., with Gaussian effect), light-path investigation and simultaneous attacks scenarios.

While various failure models were recently studied extensively in these works, they assume that the network's layout is deterministic. In this paper, we focus on spatial *random* networks. This applies to both the case where the network is derived from a random model, or to the case where the physical topology of the network is unknown to the attacker, who only possesses some statistical information such as the population density or the probability of having link between various locations. To the best of our knowledge, this paper is the first to study such geographical network failures in the context of spatial *random* networks.

III. Problem Statement

We study the model consisting of a random network immersed within a bounded convex set $B \subseteq \mathbb{R}^2$. We consider nodes as stations and links as cables that connect stations. Stations are represented through their coordinates in the plane $\mathbb{R}^2$, and links are represented by straight line segments defined by their end-points. The network is formed by a stochastic process in which the location of stations is determined by a stochastic point process. The distribution of the Poisson process is determined by the intensity function $f(u)$ which represents the mean density of nodes in the neighborhood of u. The number of nodes in a Borel set B follows a Poisson distribution with the parameter $f(B)$, i.e., the integral over the intensity of all points in the Borel set. Furthermore, the number of nodes in disjoint Borel sets are independent. An introduction to Poisson Point Processes can be found in [17].

In our model we consider a network $N = \langle p \sim \Pi(f), link \sim \omega, c \sim H, Rec \rangle$ where nodes are distributed in the rectangle $Rec = [a,b] \times [c,d]$ through a *Spatial Non-Homogeneous Poisson Point Process* $\Pi(f)$ where $f(p)$ is the *intensity function* of the Poisson Point Process Π at the point $p = (x,y)$. Let $\omega(p_i, p_j)$ be the probability for the event of existence of a link between two nodes located at p_i and p_j in Rec. $H(c, p_i, p_j)$ is the cumulative distribution function of the link capacity, C_{ij}, between two connected nodes at p_i and p_j, i.e. $P(C_{ij} \leq c) = H(c, p_i, p_j)$ where p_i and p_j are the locations of nodes i and j, respectively. It is reasonable to assume that $\omega(p_i, p_j)$ and $H(c, p_i, p_j)$ can be computed easily as a function of the distance from p_i to p_j and that the possible capacity between them is bounded (denoted by $max\{capacity\}$). We assume the following: the intensity function f of the Poisson Point Process, ω, and the probability density function h (the derivative of H), are functions of constant description complexity. They are continuously differentiable and Riemann-integrable over Rec, which also implies that our probability functions are of bounded variation over Rec, as their derivatives receive a maximum over the compact set $Rec \subseteq \mathbb{R}^2$.

We note that Poisson process is *memoryless* and *independent*. In particular, if we look on a specific region with high intensity values then it is likely to have multiple stations in this region. There are various point process models that can be used to model a random network, for example, Matérn hard-core process and Gibbs point processes class [17], which can be used where it is wanted that a random point is less likely to occur at the location u if there are many points in the neighborhood of u or for the hard-core process where a point u is either "permitted" or "not permitted" depending whether it satisfies the hard-core requirement (e.g. far enough from all other points). Methods similar to those presented here for the Poisson process can be applied to these models, as well.

Definition 1 (Circular Cut): A circular cut D, is a circle determined by its center point $p = (x_k, y_k)$ and by it's radius r, where $p \in Rec_r = [a+r, b-r] \times [c+r, d-r]$. [1] We sometimes denote the cut as $\mathrm{cut}(p,r)$ and Rec_r as Rec (depending on the context). Such a cut removes all nodes and links that intersect it (including the interior of the circle).

Our main goal is to find an attack location (or a set of locations) which has the highest expected disruptive impact on the network. We consider a fiber to be destroyed (failed fiber) if it intersects the cut, namely, the attack's influence region. The impact is measured by the total capacity of failed fibers, or by the total number of failed fibers, which is equivalent to the previous measure when all fibers have capacity 1.

IV. Damage Evaluation Scheme

In order to find the attacks that have the highest expected impact on the network, we develop a scheme to evaluate the total expected capacity of the intersected links (TEC) of the network with a circular attack in a *specific location*. First, we present the general idea behind it. We divide the intersection of a cut D (denoted also by $\mathrm{cut}(p,r)$) with a graph's edges into 3 independent types:

- $\alpha - link$, is the case where the entire edge is inside $\mathrm{cut}(p,r)$, which means both endpoints of the edge are inside $\mathrm{cut}(p,r)$ (see illustration in Fig. 2).
- $\beta - link$, is the case where one endpoint of the edge is inside $\mathrm{cut}(p,r)$ and the other endpoint is outside of $\mathrm{cut}(p,r)$ (see illustration in Fig. 3).
- $\gamma - link$, is the case where both endpoints are outside of $\mathrm{cut}(p,r)$ and the edge which connects the endpoints intersects $\mathrm{cut}(p,r)$ (see illustration in Fig. 4).

[1] For simplicity we assume that D can only appear in whole within Rec.

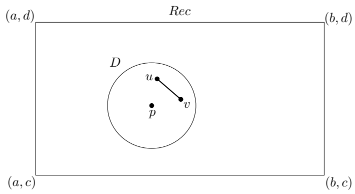

Fig. 2. α-link edge.

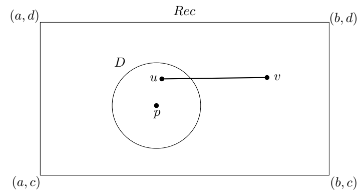

Fig. 3. β-link edge.

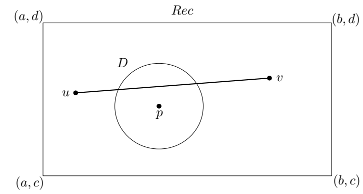

Fig. 4. γ-link edge.

Note that any intersected link with cut $D = \text{cut}(p, r)$ belongs to exactly one of the above types. For $\sigma \in \{\alpha, \beta, \gamma\}$, let X_σ be the total capacity of all the intersected $\sigma - link$ type edges with cut D, namely the damage determined by $\sigma - links$. Thus, it holds that the expected capacity of the intersected links of types α, β and γ is determined by:

$$E[X_\alpha] = \frac{1}{2} \iint\limits_{D} \iint\limits_{D} f(u)f(v)g(u,v)dudv \quad (1)$$

$$E[X_\beta] = \iint\limits_{Rec-D} \iint\limits_{D} f(u)f(v)g(u,v)dudv \quad (2)$$

$$E[X_\gamma] = \frac{1}{2} \iint\limits_{Rec-D} \iint\limits_{Rec-D} f(u)f(v)g(u,v)\mathcal{I}(u,v,D)dudv \quad (3)$$

where $g(u,v) = \omega(u,v) \int_0^{max\{capacity\}} h(c,u,v)c\,dc$ is the expected capacity between two nodes at points u and v (determined by the probability of having a link between them, times the expected capacity of this link). $\mathcal{I}(u,v,D)$ is the indicator function, giving one if the segment (u,v) intersects the circle D and zero otherwise.

Denote by $X = X_\alpha + X_\beta + X_\gamma$, the total damage determined by all the intersected links with cut D. Due to the linearity of expectation, we get that the total expected capacity of the intersected links (TEC) is $E[X] = E[X_\alpha] + E[X_\beta] + E[X_\gamma]$. Hence, it is sufficient to evaluate the expected damage caused by each of the 3 types of the intersected links separately. Summing them all together is the total expected damage caused by $D = cut(p, r)$.

A. Evaluating the Damage of Circular Cut Algorithm

We present an approximation algorithm $EDCC$ (see pseudocode in Algorithm 1) for evaluating the total expected capacity of the intersected links (TEC) of the network with a circular attack (cut) D in a *specific location*. Later, we use this algorithm to find attack locations with the (approximately) highest expected impact on the network. We give two different approximation analyses for algorithm $EDCC$ output. One is an additive approximation, and one is multiplicative. Although additive approximations in general are better than multiplicative, our analysis of the additive approximation depends on the maximum value of the functions $f(\cdot)$ and $g(\cdot,\cdot)$ over Rec, where $g(\cdot,\cdot)$ stands for the expected capacity between two points $u, v \in Rec$. While the maximum of $f(\cdot)$ and $g(\cdot,\cdot)$ over Rec can be high and affect the running time of the algorithm, for practical uses in "real-world" scenarios, it is usually low enough to make the running time reasonable. The multiplicative analysis which does not depend on the maximum of $f(\cdot)$ and $g(\cdot,\cdot)$ over Rec, depends on the variation bound of $f(p_i)f(p_j)g(p_i,p_j)$, namely a constant M which is an upper bound for the derivative of $f(\cdot)f(\cdot)g(\cdot,\cdot)$ over Rec. Define an additive ϵ-approximation to the TEC as a quantity $\tilde{C}$ satisfying $C - \epsilon \le \tilde{C} \le C + \epsilon$ for $\epsilon > 0$, where C is the actual expected capacity intersecting the cut. Similarly, define a multiplicative ϵ, ε-approximation to the cut capacity as a quantity $\tilde{C}$ satisfying $(1-\varepsilon)C - \epsilon \le \tilde{C} \le (1+\varepsilon)C + \epsilon$.

The algorithm uses numerical integration based on the discretization, where Rec is divided into a grid of squares with edge length Δ (we refer to Δ as the *"grid constant"*). The different approximations are pronounced in the $ComputeGrid(Rec, r, \epsilon)$ function in the algorithm, which determines the grid of constant Δ. Let $Grid$ be the set of these squares center-points. The algorithm evaluates the integrals numerically, using the points in $Grid$. Intuitively: the denser is the grid, the more accurate the results, at the price of requiring additional time to compute. Later, we examine the relation between the accuracy parameter ϵ and the grid constant Δ, this relation determines the implementation of $ComputeGrid()$ and the running time of our algorithm.

When computing the expectation of the γ-links damage caused by a cut D, we use the following lemma:

Lemma 1: For every node $u \in Rec \setminus D$, the set K_u (see lines 11-12 in Algorithm 1 and illustration in Fig. 5) satisfies: for every $v \in K_u$, (u,v) intersects D, and for every $w \in Rec \setminus (K_u \cup D)$, (u,w) does not intersect with D.
Thus, we run over every point u in $Grid$ which is outside the cut, and compute all possible γ-link damage emanating from u, using procedure $evaluateGamma(u, D, Grid)$. Summing them all together and dividing by 2, due to double-counting, we get the total expected γ-link damage.

B. Numerical Accuracy and Running Time Analysis

We give in this section two theorems regarding the accuracy of the algorithm. The first theorem gives an additive bound on the error in the calculation of the damage caused by a cut of radius r using numerical integration with grid constant Δ. We restrict our results to the case where $\Delta < r/2$, as otherwise the approximation is too crude to consider. The proof requires some technical lemmas, given in the Appendix.

Theorem 1: For a grid of constant Δ, a point $p \in Rec$, and the result $\tilde{C}$ for $\text{cut}(p,r)$ obtained by Algorithm 1, it holds that $C - \epsilon \le \tilde{C} \le C + \epsilon$, where C is the actual TEC value for $\text{cut}(p,r)$, and $\epsilon = c_0 \cdot \sqrt{\Delta}$ for some constant c_0 that depends on the maximum values of $f(\cdot)$ and $g(\cdot,\cdot)$, their variation bound, the sides of Rec and the radius r.

Algorithm 1 EvaluateDamageCircularCut (EDCC): Approximation algorithm for evaluating the damage of a circular cut.

capacity **procedure** EDCC($N, \mathrm{cut}(p, r), \epsilon$)
$D \leftarrow Circle(\mathrm{cut}(p, r))$ //D is a representation of the circular cut.
$Grid \leftarrow ComputeGrid(Rec, r, \epsilon)$
for every $u, v \in Grid$ **do**
// Compute the expected capacity between two points u and v: $g(u, v) \leftarrow \omega(u, v) \int_0^{max\{capacity\}} h(c, u, v)c\, dc$
The following steps are calculated over $Grid$:
$E[X_\alpha] \leftarrow \frac{1}{2} \iint\limits_{D} \iint\limits_{D} f(u)f(v)g(u, v)dudv$
$E[X_\beta] \leftarrow \iint\limits_{Rec-D} \iint\limits_{D} f(u)f(v)g(u, v)dudv$
$E[X_\gamma] \leftarrow \frac{1}{2} \iint\limits_{Rec-D} f(u)$**evaluateGamma**$(u, D, Grid)du$
return $E[X_\alpha] + E[X_\beta] + E[X_\gamma]$
Procedure evaluateGamma($u, D, Grid$)
Create two tangents t_1, t_2 to D going out from u.
Denote by K_u the set of points which is bounded by t_1, t_2, D and the boundary of the rectangle Rec (see Fig. 5).
return $\iint\limits_{K_u} f(v)g(u, v)dv$

Proof: By standard arguments on numerical integration the error in calculating the integral over any region is bounded by $M\Delta$ times the area of integration, where M is the bound on the variation rate for the product $f(\cdot)f(\cdot)g(\cdot, \cdot)$ over Rec. The integration here is performed over pairs of points that are bounded inside the area of Rec (denoted by $|Rec|$). Thus, the error is bounded by $M\Delta|Rec|^2$.

Additionally, the cumulative error value $|C - \tilde{C}|$ consists of the following:

Any point in a grid square is within a distance of $\Delta/\sqrt{2} < \Delta$ of the grid point (square center). The additional difference in the integral over α and β links is bounded by the integral over the area of inaccuracy around the circular cut (grid squares which are partially in the cut and partially outside). This is bounded by an annulus of radii $[r - \Delta/\sqrt{2}, r + \Delta/\sqrt{2}]$ around the center of the circular cut of area $2\sqrt{2}\pi r\Delta$. Thus, we obtain an error which is bounded by $2\sqrt{2}\pi r\Delta T|Rec|$ where $T = \max_{u,v\in Rec}\{f(u)f(v)g(u, v)\}$, which is the maximum value the function takes over all points in the rectangle.

The additional error in the calculation of γ-links is obtained in two terms, one term is determined in the procedure $evaluateGammma(u, D, Grid)$ where the area of inaccuracy is around K_u (grid squares which are partially in K_u and partially outside). This area is bounded by $\sqrt{2}\mathcal{L}\Delta$ where $\mathcal{L}$ is the diagonal length of Rec. This gives an error term which is bounded by $\sqrt{2}\mathcal{L}\Delta T|Rec|$.

The second error term is obtained using Lemma 3 which implies that for any point u in a grid square, the area of symmetric difference between K_u and K_v for the grid point (square center) v nearest to u is bounded by $a\sqrt{\Delta}$, where $a \leq const \cdot \frac{\mathcal{L}^2}{\sqrt{r}}$ (see Appendix). Thus, we obtain an error bounded by $a|Rec|T\sqrt{\Delta}$.

Taking into account the errors in this numerical integration from all terms above, one obtains that the leading term in the error when $\Delta \to 0$ is $|C - \tilde{C}| \leq const\mathcal{L}^2|Rec|T\sqrt{\frac{\Delta}{r}}$. Thus, the accuracy depends on $\sqrt{\Delta}$, as well as on the sides of the rectangle, the radius of the cut, the maximum values of the functions $f(\cdot)$, $g(\cdot, \cdot)$ and their variation bound over Rec. ∎

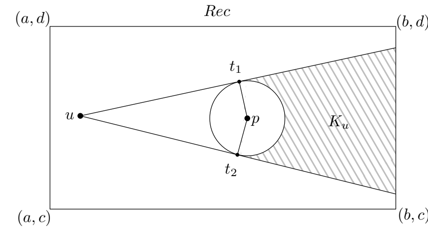

Fig. 5. The area K_u of the shadow by the circular cut from the point u.

Using Theorem 1, the function $ComputeGrid(Rec, r, \epsilon)$ can be implemented by selecting the value of Δ guaranteeing that the additive error will be at most ϵ.

Since the constant depends on the maximum value of the functions f and g, which may be undesirable in case these maxima are high, we have the following theorem, giving a combined additive and multiplicative accuracy with the constants independent of the maxima of f and g. Using the following Theorem 2, the function $ComputeGrid(Rec, r, \epsilon)$ can be modified to a new function $ComputeGrid(Rec, r, \epsilon, \varepsilon)$ which can be implemented by selecting the value of Δ guaranteeing that the additive error will be at most ϵ and the multiplicative error will be at most ε.

Theorem 2: For a grid of constant Δ, a point $p \in Rec$, and the result $\tilde{C}$ for $\mathrm{cut}(p, r)$ obtained by Algorithm 1, it holds that $(1 - \varepsilon)C - \epsilon \leq \tilde{C} \leq (1 + \varepsilon)C + \epsilon$, where C is the actual TEC value for $\mathrm{cut}(p, r)$, for $\epsilon = c_1 \cdot \sqrt{\Delta}$, $\varepsilon = c_2 \cdot \sqrt{\Delta}$, such that c_1 and c_2 depend only on Rec, r, and M the bound on the variation of $f(\cdot)f(\cdot)g(\cdot, \cdot)$, but are independent of the maximum values of f and g.

Proof: From the proof of Theorem 1, the standard error in the numerical integration over the grid depends only on the grid constant Δ, the radius of the cut r, and the bounded variation rate M of the integrated function.

Now denote the integrated function by $\tilde{f}(u)$, and write it in polar coordinates relative to the center of the cut. Let $R(\theta)$ denote the boundary of the integrated area and $R'(\theta)$ denote the boundary as approximated by the algorithm. We have for each θ, $|R(\theta) - R'(\theta)| \leq c\mathcal{L}\sqrt{\Delta}$ for some constant c. From the bounded variation it follows that $\iint d\theta r dr \tilde{f}(r, \theta) \geq \iint_0^{R(\theta)} d\theta r dr \max\{0, \tilde{f}(R(\theta), \theta) - M(R(\theta) - r)\}$. This integral evaluates to $\int d\theta \tilde{f}(R(\theta), \theta)R^2(\theta)/2 - O(MR^3(\theta))$, whereas the error term is at most

$$\mathcal{ERR} \leq c\mathcal{L}^2\sqrt{\Delta} \int d\theta \tilde{f}(R(\theta), \theta) + O(M\sqrt{\Delta}),$$

where the second summand depends only on $\mathcal{L}$ and M. Thus the relative error is bounded by

$$\frac{\mathcal{ERR}}{\iint d\theta r dr \tilde{f}(r, \theta) + aM} \leq b\sqrt{\Delta},$$

where a and b are constants independent of the maximum value of $\tilde{f}$. ■

The algorithm is based on running over pairs of points in the grid, and thus

Theorem 3: For a given additive accuracy parameter $\epsilon > 0$, the ϵ-factor[2] for the running time of Algorithm 1 is $O(\epsilon^{-8})$. Similarly, for a multiplicative accuracy parameter $\varepsilon > 0$ an additional term for the running time has ε-factor of $O(\varepsilon^{-8})$.

Proof: The number of grid points in the rectangle is $|Rec|/\Delta^2$. The algorithm performs integration over pairs of points in the grid. Thus the running time is at most proportional to the number of pairs of grid points, which is $O(|Rec|^2/\Delta^4)$. From Theorem 1, $\epsilon = O(\sqrt{\Delta})$ and thus we obtain a factor of $O(\epsilon^{-8})$. Similarly, for the combined ϵ-additive ε-multiplicative approximation, it can be derived from Theorem 2 that $\varepsilon = O(\sqrt{\Delta})$, giving the additional term. ■

V. Find Sensitive Locations Scheme

Using Algorithm 1 ($EDCC$) one can approximate the TEC for an attack at every point, and in particular, find an approximated worst case attack (one with the highest TEC value) or the expected damage by a random attack (where its location is probabilistically distributed).

To achieve this goal, we divide Rec into squares, forming a grid. Then, we execute $EDCC$ algorithm from the previous section for every grid point (squares center-points) such that it is a center-point of a cut of radius r. This leads to a "network sensitivity map", i.e., for every point we have an approximation of the damage by a possible attack in that point.

The approximated worst cut is given by taking the point with the highest TEC value among all the centers of grid squares. The actual worst case cut can be potentially located at any point within the grid squares whose centers' calculated TECs are approximated by the $EDCC$ algorithm. To guarantee an attack location with TEC of at least $C - \epsilon$ where C is the TEC of the actual worst cut and an additive accuracy parameter $\epsilon > 0$, we provide the following algorithm:

1) For a cut of radius r, and accuracy parameter $\epsilon > 0$, apply the function $computeGrid(Rec, r, \epsilon/2)$ to find $\Delta > 0$ such that the accuracy of Algorithm 1, given by Theorem 1 is $\epsilon/2$.
2) Form a grid of constant Δ (found in step 1) from Rec. For every grid point p, apply procedure $EDCC(N, \mathrm{cut}(p, r), \epsilon/2)$. The grid point with the highest calculated TEC is the center of the approximated worst cut.

In the following theorems we prove the correctness of the above algorithm and refer its running time.

Theorem 4: For an accuracy parameter $\epsilon > 0$, the attack with the highest TEC value $\tilde{C}$ found by the above algorithm satisfies $\tilde{C} \geq C - \epsilon$, where C is the TEC value for the actual worst cut.

[2] The running time depends also on the network parameters. Computational complexity analysis which depends on the network parameters can be derived from the theorems and the Appendix of the paper.

Proof: By Theorem 1 for every $\epsilon' > 0$ one can find a grid constant $\Delta > 0$ such that for any point $p \in Rec$ the TEC value of $\mathrm{cut}(p, r)$ obtained by algorithm $EDCC$ is within ϵ'-accuracy (additive) from the actual TEC value of $\mathrm{cut}(p, r)$.

For any cut located at a grid point, take a cut located at some other point within the grid square (within a distance $d < \Delta$ from the center of the square). The difference between the TEC for these two cases is exactly the same as in the symmetric case, where the functions f, g and the grid, are shifted a distance d in the other direction. Thus, using similar arguments as in the proof of Theorem 1, we obtain that the difference is at most ϵ'. Taking $\epsilon' = \epsilon/2$ completes the proof. ■

Theorem 5: For an additive accuracy parameter $\epsilon > 0$, the ϵ-factor on the running time of the above algorithm is $O(\epsilon^{-12})$.

Proof: The algorithm samples a circular cut of radius r at the center point p of each grid square. For each such cut, the algorithm executes $EDCC(N, \mathrm{cut}(p, r), \epsilon/2)$. The grid has $O(|Rec|/\Delta^2)$ points. Thus, from Theorem 3 the Δ-factor on the running time is of $O(\Delta^{-6})$. From Theorem 1, $\epsilon = O(\sqrt{\Delta})$ and thus we obtain a factor of $O(\epsilon^{-12})$. ■

Using the multiplicative approximation in the modified algorithm, similar results can be obtained for a combined ε-multiplicative and ϵ-additive approximation of the worst case cut (resulting an additional term for the running time with ε-factor of $O(\varepsilon^{-12})$).

Using the methods above, other valuable measures besides the worst case cuts can be obtained as well. For example, the expected impact caused by a 'random disaster' (where its location is randomly distributed), which can model failure resulting from a natural disaster such as a hurricane or collateral (non-targeted) damage in an EMP attack.

VI. Numerical Results

In order to test the algorithm presented, and estimate the expected impact of cuts on communication networks in the USA, we implemented the algorithm as a C program. Data for the population density of the USA [7], downsampled to a resolution of $\Delta = 27$km was taken as the intensity function $f(u)$. The function $\omega(u, v)$ was taken to be $1/\mathrm{dist}(u, v)$, where dist is the Euclidean distance between the points [3]. The capacity probability function h was taken to be constant, independent of the distance, reflecting an assumption of standard equipment.

Results are given in Fig. 6. As can be seen , the most harmful cut would be around NYC, as expected, where for the larger cut radius, a cut between the east and west coast, effectively disconnecting California from the north-east, would also be a worst case scenario. Lower, but still apparent maxima are observed in the Los-Angeles, Seattle and Chicago areas. Comparing the results to the results obtained for the fiber backbone in [13] it can be seen that some similarities and some dissimilarities exist. While the NYC maxima is apparent in all measures, California seems to be missing from the maxima in [13], probably reflecting the low density of fibers in that area in the map studied in [13]. As many fibers may exist that are not represented on the map in [13], it may be reasonable to assume that a cut around California would have a more significant effect

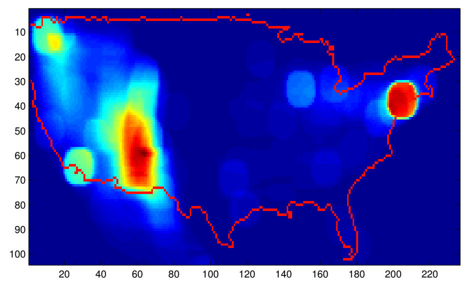

Fig. 6. Color map of the centers of circular cuts with radius $r = 8$ (approximately 216km) and $\omega(u, v) \sim 1/\mathrm{dist}(u, v)$. Red is most harmful.

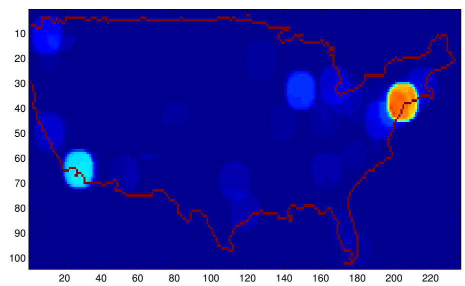

Fig. 7. Color map of the centers of circular cuts with radius $r = 8$ (approximately 216km) and $\omega(u, v) \sim 1/\mathrm{dist}^2(u, v)$. Red is most harmful.

than reflected there. On the other hand, several worst case cuts in Texas, and especially in Florida are apparent in [13] and are missing in the current simulation results. It seems that the effect of the narrow land bridge in Florida makes cuts there especially harmful, whereas our simulation assumes links are straight lines, which will make links to both east and west coast pass through the ocean, thus making cuts less harmful.

In order to examine a different assumption on the link distribution, we also simulated a distribution with $\omega(u, v) \sim 1/\mathrm{dist}^2(u, v)$. Results are given in Fig. 7. As can be seen, the results are pretty similar, with the exception of the maximum between California and the east coast, which is probably only expected for a length distribution allowing very long links. In fact, these results are closer to the results in [13].

As a full map of communication lines is not available, it is still unclear how good of an approximation the results here supply. However, the tool can be used in conjunction with more complicated modeling assumptions, including topographic features and economic considerations to give more exact results.

VII. Conclusions

Motivated by applications in the area of network robustness and survivability, we focused on the problem of geographically correlated network failures. While previous works in this emerging field focused mainly on deterministic networks, we studied the problem under non-deterministic network models.

We described polynomial time approximation algorithms for finding the damage caused by cuts at different points in our spatial random graph model and to approximate the location and damage of the worst case cuts. We proved the correctness of the scheme and mentioned the trade between running time and accuracy, for both the additive and multiplicative terms.

We applied the method to approximate the damage caused by cuts in different locations in the USA where the nodes were modeled probabilistically based on the population density map and the links based on distances. Some of the results agree with the exact results obtained before on a fiber backbone map of the USA and some do not. It is yet to be determined if taking into account more links would lead to results closer to our scheme's prediction or whether the results are fundamentally different due to an inexact link model. However, the tool can be used in conjunction with more complicated modeling assumptions, including topographic features and economic considerations to give more accurate results.

Thus, this scheme can be used as a tool for evaluating network vulnerabilities, and also as a tool for policy makers and engineers to design more robust networks by placing links along paths that avoid areas of high damage cuts.

Some future research directions include robust network design in the face of geographical failures.

Acknowledgments

We thank the BSF for support under grant 2010414. We thank David Hay for his valuable comments and discussion.

References

[1] P. K. Agarwal, H. Kaplan, and M. Sharir, "Union of random minkowski sums and network vulnerability analysis," in *Proceedings of the twenty-ninth annual symposium on Computational geometry*, ser. SoCG '13. New York, NY, USA: ACM, 2013, pp. 177–186.

[2] P. Agarwal, A. Efrat, S. Ganjugunte, D. Hay, S. Sankararaman, and G. Zussman, "The resilience of wdm networks to probabilistic geographical failures," *Networking, IEEE/ACM Transactions on*, vol. 21, no. 5, pp. 1525–1538, 2013.

[3] A. L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, October 1999.

[4] R. Bhandari, *Survivable networks: algorithms for diverse routing*. Kluwer, 1999.

[5] A. Fabrikant, A. Luthra, E. N. Maneva, C. H. Papadimitriou, and S. Shenker, "On a network creation game," in *PODC*, 2003, pp. 347–351.

[6] O. Gerstel and R. Ramaswami, "Optical layer survivability: a services perspective," *IEEE Commun.*, vol. 38, no. 3, pp. 104–113, Mar. 2000.

[7] Global Rural-Urban Mapping Project, Version 1, Network Data, 2000. [Online]. Available: http://sedac.ciesin.columbia.edu/data/set/grump-v1-population-density

[8] J. Manchester, D. Saha, and S. K. Tripathi (eds.), "Protection, restoration, and disaster recovery," *IEEE Network, Special issue*, vol. 18, no. 2, Mar.–Apr. 2004.

[9] E. Modiano and A. Narula-Tam, "Survivable lightpath routing: a new approach to the design of WDM-based networks," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 4, pp. 800–809, May 2002.

[10] A. Narula-Tam, E. Modiano, and A. Brzezinski, "Physical topology design for survivable routing of logical rings in WDM-based networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 8, pp. 1525–1538, Oct. 2004.

[11] S. Neumayer, A. Efrat, and E. Modiano, "Geographic max-flow and min-cut under a circular disk failure model," in *INFOCOM*, 2012, pp. 2736–2740.

[12] S. Neumayer and E. Modiano, "Network reliability under random circular cuts," in *GLOBECOM*, 2011, pp. 1–6.

[13] S. Neumayer, G. Zussman, R. Cohen, and E. Modiano, "Assessing the vulnerability of the fiber infrastructure to disasters," *IEEE/ACM Trans. Netw.*, vol. 19, no. 6, pp. 1610–1623, 2011.

[14] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach*. New York, NY, USA: Cambridge University Press, 2004.

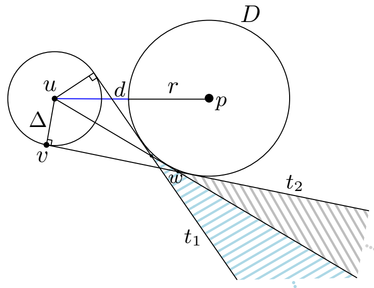

Fig. 8. t_1 and t_2 are tangents to a circle of radius Δ centered at u and to the circular cut. The colored areas depicts the extremum of difference for possible γ-links endpoints emanating from a point within the circle centered at u at one side of the cut.

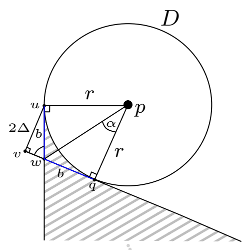

Fig. 9. The perpendicular tangents from a point u on the circular cut and from a point v at distance 2Δ from the circular cut. The area in grey depicts the extremum of difference for possible γ-links endpoints emanating from u and v at one side of the cut.

[15] Y. Shavitt and E. Shir, "Dimes: let the internet measure itself," *Computer Communication Review*, vol. 35, no. 5, pp. 71–74, 2005.

[16] J. P. G. Sterbenz, D. Hutchison, E. K. Çetinkaya, A. Jabbar, J. P. Rohrer, M. Schöller, and P. Smith, "Resilience and survivability in communication networks: Strategies, principles, and survey of disciplines," *Computer Networks: Special Issue on Resilient and Survivable Networks (COMNET)*, vol. 54, no. 8, pp. 1245–1265, June 2010.

[17] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications, 2nd Edition*, 2nd ed. Wiley, Jul. 1996.

[18] S.-H. Yook, H. Jeong, and A.-L. Barabási, "Modeling the Internet's large-scale topology," *Proceedings of the National Academy of Sciences*, vol. 99, no. 21, pp. 13 382–13 386, Oct. 2002.

[19] D. Zhang and J. P. G. Sterbenz, "Modelling critical node attacks in manets," in *Proceedings of IWSOS : Third International IFIP/IEEE Workshop on Self-Organizing Systems*, ser. Lecture Notes in Computer Science. Springer, 2013.

[20] D. Zhou and S. Subramaniam, "Survivability in optical networks," *IEEE Network*, vol. 14, no. 6, pp. 16–23, Nov.-Dec. 2000.

APPENDIX

Some technical results are needed for proving the main theorems. We first notice that, given a point u at a distance d from a circular cut, and a point v at a distance Δ from u, the difference between the area of endpoints of γ-links starting at point u to those starting at point v is bounded in the area between (a) the segment from u to the boundary of Rec tangent to the cut (b) the segment from v to the boundary of Rec tangent to the cut (c) the boundary of Rec. Plus the area bounded by (a), (b) and the boundary of the circular cut. Furthermore, the extrema of the area of difference are obtained in the cases where the segment from v to the boundary of Rec tangent to the cut is also tangent to the circle centered at u with radius Δ. See Fig. 8.

For simplicity, in the following lemmas we assume the tangents hit the same side of Rec. Other cases can be handled similarly, resulting an additional factor of 4. When bounding the error in the theorems of this paper, one should consider a bound, feasible for all cases, and add a factor of 2 (to bound also the area of possible γ-links endpoints obtained by the additional tangents emanating from u and v to the other side of the cut).

Lemma 2: The area bounded by (a) the tangent at a point u on the circumference of the circular cut (b) the tangent to the cut from any point v at distance 2Δ from u (c) the circumference of the cut and (d) the boundary of Rec (same side), is bounded by $c\sqrt{\Delta}$ for some constant c. See Fig. 9.

Proof: The extremum of the difference between the two tangents is when the tangent to the circle is also a tangent to the circle of radius 2Δ centered at u, i.e., when the segment of length 2Δ starting at point u is perpendicular to the tangent at its other endpoint, v. See Fig. 9.

Let q be the point of intersection of the second tangent and the circular cut, and let w be the point of intersection between the two tangents. Let p be the center of the circle. We have $\angle upw = \angle wpq = \alpha$ and angle $\angle uwv = 2\alpha$.

Let $b = d(u,w) = d(w,q)$ and $d(w,v) = a = \sqrt{b^2 - 4\Delta^2}$. We have $\tan\alpha = b/r$, where r is the radius of the circle, and $\tan 2\alpha = \frac{2\Delta}{\sqrt{b^2-4\Delta^2}}$. Using $\tan 2\alpha = 2\tan\alpha/(1-\tan^2\alpha)$ one obtains $b^2 = (\Delta^2 r^2 + \Delta r^3)/(r^2 - \Delta^2)$. Thus, $r\Delta \leq b^2 \leq 2r\Delta$. The area of the triangle bounded by both tangents and the boundary of Rec is bounded by

$$\frac{1}{2}\mathcal{L}^2 \sin 2\alpha \leq \mathcal{L}^2 \sin\alpha \leq \mathcal{L}^2 \tan\alpha \leq \mathcal{L}^2 \sqrt{\frac{2\Delta}{r}},$$

where $\mathcal{L}$ is the diagonal length of Rec.

The area bounded by the segments uw and wq and the circular arc uq is bounded by the area of triangle $\triangle uwq$, which, in turn is bounded by $b^2 \leq 2r\Delta \leq \mathcal{L}^2\sqrt{\frac{2\Delta}{r}}$. ■

Lemma 3: For a point u at a distance d from a circular cut $\mathrm{cut}(p,r)$ and a given $\Delta > 0$, the area between (a) the circumference of the cut (b) the tangent to the cut from u (c) the tangent to the cut from a point v located at a distance Δ from u and (d) the boundary of Rec (same side), is bounded by $c\sqrt{\Delta}$ for some constant c. See Fig. 8.

Proof: Let w be the point of intersection between the two tangents (from point u to the cut and from point v to the cut). We have $\sin(\angle uwv) = \Delta/(\sqrt{(r+d)^2 - r^2} + b)$, where b is the distance between the point of intersection of each of the tangents with the cut and the point w. we have $\sin(\angle uwv) = \Delta/(\sqrt{2rd + d^2} + b)$.

Now for $d \leq \Delta$ the point v is located within a distance of $d + \Delta \leq 2\Delta$ from the circumference of the cut, and the Lemma follows from Lemma 2. If $d > \Delta$ we have $\sin(\angle uwv) < \Delta/(\sqrt{2rd}) < \Delta/(\sqrt{2r\Delta})$. Thus, the area between the tangents within Rec is bounded by $\mathcal{L}^2\sqrt{\Delta/r}$ where $\mathcal{L}$ is the diagonal length of Rec. Similarly to Lemma 2 the area between the two tangents and the circle is also bounded by $b^2 \sin(\pi - \angle uwv) = b^2 \sin(\angle uwv) < \sqrt{r^3\Delta}$, as $b < r$ since both tangents intersect the same quadrant of the cut. ■