PEDECIBA Informática
Instituto de Computación – Facultad de Ingeniería
Universidad de la República
Montevideo, Uruguay

Reporte Técnico RT 12-09

The complexity of K-reliability in network constrained to diameter two

Eduardo Canale Héctor Cancela

Franco Robledo Pablo Sartor

2012

Modelado estocástico múltiple etapa de adquisición de combustible para la generación de electricidad bajo demanda incierta

Testuri, Carlos; Zimberg, Bernardo; Ferrari, Germán

ISSN 0797-6410
Reporte Técnico RT 12-09
PEDECIBA
Instituto de Computación – Facultad de Ingeniería
Universidad de la República
Montevideo, Uruguay, 2012

The complexity of the K-reliability in networks constrained to diameter two

Eduardo Canale, Héctor Cancela, Franco Robledo, Pablo Sartor

Instituto de Computación, Facultad de Ingeniería, Universidad de la República
Julio Herrera y Reissig 565 - Código Postal 11.300 - Montevideo, Uruguay
Tel. (+598) 2711 4244 - Fax (+598) 2711 0469

Email: {canale,cancela,frobledo,psartor}@fing.edu.uy

Abstract

Consider a communication network composed by sites that never fail and links between them that fail independently from one another. In any instant, every link (xy) is operational or failed according to certain known probabilities $p(xy)$ and $1 - p(xy)$. Let d be any positive integer. Computing the probability that a fixed subset K of sites remains connected by paths whose lengths do not exceed d (considering only non-failing links) is known as the d-DCKR (d-diameter constrained K-reliability) problem. Its general case is known to belong to the NP-hard complexity class; there are a number of particular cases whose complexity remains undetermined. In this paper we show that the computational complexity of the d-DCKR is linear in the number of sites of the network when $d = 2$ and $|K|$ is fixed (i.e. when $|K|$ is not a free input for the complexity analysis, but a fixed parameter of the problem).

Keywords: Network reliability, survivability, fault tolerance, diameter constraints, combinatorial problems, computational complexity.

1. Introduction

Consider a communication network with a set of sites and a set of links between them. Suppose that the sites are perfect but the links can fail independently from one another. Suppose also that at any given instant t, every link xy is operational or failed with probabilities denoted by $p(xy)$ and $1 - p(xy)$ respectively. Therefore, there is an "operational subnetwork" composed by all the sites and only those links that are operational. Computing the network reliability, i.e. the probability that a given subset K of "distinguished" sites is connected on the operational network yielded at a certain moment is known as the K-reliability problem and has been widely studied [1]. When additionaly requiring that the operational network be d-K-connected (i.e. that the distance between any pair of sites of K be bounded by a positive integer d) the problem is known as d-diameter-constrained K-reliability (d-DCKR). In this case the reliability is denoted by $R_K(G, d)$. First introduced in [2], this problem has recently gained relevance because it can model situations where limits exist on the acceptable delay times to propagate traffic (like in voice applications over IP networks) or in the amount of hops that packets can undergo (peer-to-peer networks). The general version is known to belong to the NP-hard complexity class [3]. In this paper we prove that *the computational complexity of the 2-DCKR is linear in the number of sites, whenever $|K|$ is a fixed parameter of the problem (and regardless its value).*

2. Definitions and Notation

Let us model the network by a simple, undirected and complete graph $G = (V, E)$, with $n = |V|$ and $E = \{\{x, y\} : x \in V \wedge y \in V \wedge x \neq y\}$. The nodes V correspond to the sites of the network. Each edge $xy \in E$ has a label $p(xy) \in [0, 1]$. If there is a link connecting the sites to which x and y correspond then $p(xy)$ is the probability that it is operational; otherwise we define[1] $p(xy) = 0$. We denote the probability of any event z as $\Pr(z)$. Additionally:

- given $K \subseteq E$, a *2-path* is any $E' \subseteq E$ such that the partial graph (V, E') is 2-K-connected;
- we denote by n' the number of nodes not in K, that is, $n' = n - |K|$;
- we denote by $O^2_K(E)$ the set of 2-paths determined by E and K (following the notation of [4]);
- we denote by $X^{\{m\}}$ the set of all subsets of a certain set X that have m different elements, that is, $X^{\{m\}} = \{Y \subseteq X : |Y| = m\}$;
- we denote by $X^{(m)}$ the set of all m-tuples built with different elements of a certain set X, that is, $X^{(m)} = \{Y \in X^m : i \neq j \rightarrow Y_i \neq Y_j\}$;
- we denote by $\otimes$ the binary operator defined as:
  $(a_1, a_2, \ldots, a_n) \otimes (b_1, b_2, \ldots, b_n) = \{(a_1, b_1), (a_2, b_2), \ldots, (a_n, b_n)\}$;
- we call *requirement* to any set of two nodes $\{x, y\} \subseteq K^{\{2\}}$ and denote it by xy. We say that a requirement is *satisfied* if there is a path of length below three, formed only by operating edges, that connects the two nodes that define the requirement;
- we denote by $\mathcal{P}(A)$ the powerset of a certain set A.

3. Demonstration

3.1. Demonstration Plan

We start by finding an analytical expression for computing $R_K(G, 2)$. To do so, we partition $O^2_K(E)$ in disjoint components whose probabilities are then computed and totaled. We build a function $f : O^2_K(E) \rightarrow A$ (with a certain discrete codomain A conveniently chosen) and then compute the probability of the domain by totaling the probabilities of all preimages of A:

$$R_K(G, 2) = \Pr(O^2_K(E)) = \sum_{a \in A} \Pr(f^{-1}(a)) \tag{1}$$

The set A is defined as

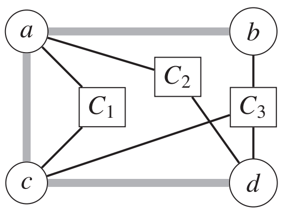
$$A = \bigcup_{\ell=0}^{n'} \Big((V \setminus K)^{(\ell)} \otimes \mathcal{P}(K)^{(\ell)} \Big).$$

[1] In other words we see the network as a complete graph where non-existent links correspond to edges with probability of operation equal to zero.

Each element of A is a set of pairs (t, C), where t is a node of $V \setminus K$ and C is a subset of K. We see each such element of A as a collection of sets of edges between nodes of $V \setminus K$ and K that belong to a 2-path (besides edges between nodes of K). Observe that any edge linking two nodes of $V \setminus K$ is irrelevant for the 2-K-connectivity, since they can not be part of a path of lenght one or two connecting two nodes of K. The function f is defined in Eq. (3). We first show that A can be built such that totaling the probabilities $\Pr(f^{-1}(a))$ involves a number of elementary operations that is polynomial in n; finally, we show that it is indeed linear in n.

3.2. *Partitioning $O^2_K(E)$*

We assume that there is a certain strict ordering within V. We say that a family $\mathcal{C} \subseteq \mathcal{P}(K)$ *covers (or is a cover of) K for $F \subseteq K^{\{2\}}$*, and denote it by $\mathcal{C} \sqsupset_F K$, if and only if for every requeriment xy at least one of the following applies: (i) $xy \in F$; (ii) $\exists z \in K : \{xz, zy\} \subseteq F$; (iii) $\exists C \in \mathcal{C} : \{x, y\} \subseteq C$.



$$K = \{a, b, c, d\}$$
$$F = \{ab, ac, cd\}$$
$$\mathcal{C} = \{\{a, c\}, \{a, d\}, \{b, c, d\}\}$$

Figure 1: Example of cover set

Figure 1 illustrates the concept of covers. The thick lines represent the elements of F. The nodes connected by thin lines to the square nodes represent each element of $\mathcal{C} = \{C_1, C_2, C_3\}$. Condition (i) applies for the pairs ab, ac and cd. Condition (ii) applies for bc and ad. Condition (iii) applies for ac, ad, bc, bd and cd. Thus all pairs satisfy at least one of (i), (ii), (iii) and then $\mathcal{C} \sqsupset_F K$.

Observe that if $E' \in O^2_K(E)$ then the family $\mathcal{C}_{E'}$ defined as the set

$$\mathcal{C}_{E'} = \{C_{E'}(t) : t \in V \setminus K\} \quad \text{being } C_{E'}(t) = \{x \in K : xt \in E'\}, \tag{2}$$

covers K for $F = E' \cap K^{\{2\}}$, and we say that E' *generates* $\mathcal{C}_{E'}$. Conversely, given a cover $\mathcal{C}$ of K for F, if $|\mathcal{C}| \leq n - |K|$ then there is some E' that generates $\mathcal{C}$; for example $E' = \{xy \in E : xy \in F \vee (\exists C \in \mathcal{C} : \{x, y\} \subseteq C)\}$ is a 2-path. The preceding definitons relate with the operational states of the network in the following way. Assume that the elements of F correspond exactly to the edges linking nodes of K that operate at a certain moment. Assume also that each element $C \in \mathcal{C}$ represents a node in $V \setminus K$ whose neighbors, ignoring failing edges, are exactly the elements of C (e.g. C_1, C_2, C_3 in Figure 1). Then, $\mathcal{C} \sqsupset_F K$ implies that every requirement is satisfied at that moment.

The powerset $\mathcal{P}(K)$ has $2^{|K|}$ elements (i.e. there are $2^{|K|}$ different subsets of K). Its powerset $\mathcal{P}(\mathcal{P}(K))$ has $2^{(2^{|K|})}$ elements that are sets of subsets of K. So, for the number of families $\mathcal{C}$ for which $\mathcal{C} \sqsupset_F K$, there is an upper bound $2^{(2^{|K|})}$ that depends only on $|K|$. Now, observe that when considering a family $\mathcal{C}_{E'}$, if there are two nodes $t \neq t' \notin K$ with $C_{E'}(t) = C_{E'}(t')$, one of t and t' can be removed from the graph G, yet obtaining the same family $\mathcal{C}_{E'}$. For example, in Figure 1, the addition of a node C_4 with exactly a and c as neighbors in K, would make no difference when producing $\mathcal{C}_{E'}$, due to the existence of a node C_1 with exactly the same neighbours in K.

In general, given a path E' and one element C of the cover $C_{E'}$ there are one or more $t \in V \setminus K$ such that $C_{E'}(t) = C$. Let us define $t_{E'}(C)$ as the minimum of them according to the ordering of V, that is

$$t_{E'}(C) = \min\{t \in V \setminus K : C_{E'}(t) = C\}.$$

Then, for every 2-path E', there is exactly one set $\{t_1, t_2, \ldots, t_\ell\}$ and one cover $C_{E'} = \{C_1, C_2, \ldots, C_\ell\}$ for $F = E' \cap K^{\{2\}}$, such that $t_i = t_{E'}(C_i)\ \forall i = 1, \ldots, \ell$. Now we can define our function f as follows:

$$f(E') = \{(t_1, C_1), (t_2, C_2), \ldots, (t_\ell, C_\ell)\}. \tag{3}$$

3.3. Computing the probabilities

Given $x \in V \setminus K$ and $C \subseteq K$, let us denote as $P(x, C)$ the event where the set of neighbors of node x that belong to K and are connected by working links is exactly C. Its probability, denoted as $p(x, C)$, is:

$$p(x, C) = \prod_{y \in C} p(xy) \prod_{y \in K \setminus C} (1 - p(xy)).$$

Now, given $C_1, \ldots, C_\ell \subseteq K$ and $t_1, \ldots, t_\ell \in V \setminus K$ such that $t_i < t_{i+1} \forall i = 1, \ldots, l-1$, let us define the event $P(t_1, \ldots, t_\ell, C_1, \ldots, C_\ell)$ as the event where:

- the nodes $t_1, \ldots, t_\ell$ are connected exactly to $C_1, \ldots, C_\ell$ in K respectively; and
- every node $t \in V \setminus K$ has no neighbors in K or is connected exactly in K to some C_i with $t_i < t$.

This is the event where all the following statements hold true (we denote its probability as $p(t_1, \ldots, t_\ell, C_1, \ldots, C_\ell)$):

- the nodes $t_1, \ldots, t_\ell$ are connected exactly to $C_1, \ldots, C_\ell$ in K respectively;
- the nodes t with $t < t_1$ have no neighbors in K;
- the nodes t with $t_1 < t < t_2$ have no neighbors in K or have exactly as neighbors the nodes C_1;
- the nodes t with $t_2 < t < t_3$ have no neighbors in K or have exactly as neighbors one of C_1 and $C_2 \ldots$;
- the nodes t with $t_\ell < t$ have no neighbors in K or have exactly as neighbors one of $C_1, \ldots, C_\ell$.

So whe have that

$$\begin{aligned} p(t_1, \ldots, t_\ell, C_1, \ldots, C_\ell) = &\left[\prod_{t<t_1} p(t, \emptyset)\right] p(t_1, C_1) \left[\prod_{t_1<t<t_2} [p(t, \emptyset) + p(t, C_1)]\right] p(t_2, C_2) \\ &\left[\prod_{t_2<t<t_3} [p(t, \emptyset) + p(t, C_1) + p(t, C_2)]\right] p(t_3, C_3) \ldots \\ &\left[\prod_{t_{\ell-1}<t<t_\ell} [p(t, \emptyset) + p(t, C_1) + \cdots + p(t, C_{\ell-1})]\right] p(t_\ell, C_\ell) \\ &\left[\prod_{t_\ell<t} [p(t, \emptyset) + p(t, C_1) + \cdots + p(t, C_\ell)]\right] \end{aligned}$$

To build a compact expression, let us define t_0 and $t_{\ell+1}$ as "virtual nodes" such that $t_0 < t_1$ and $t_\ell < t_{\ell+1}$; and $C_0 = \emptyset$. Then we have that:

$$p(t_1, \ldots, t_\ell, C_1, \ldots, C_\ell) = \left[\prod_{i=1}^{\ell} p(t_i, C_i)\right]\left[\prod_{i=0}^{\ell} \prod_{t_i<t<t_{i+1}} \sum_{j=0}^{i} p(t, C_j)\right]. \tag{4}$$

Note that the terms in the addition $p(t, \emptyset) + p(t, C_1) + \cdots + p(t, C_j)$ correspond to the probabilities of events that are pairwise disjoint due to the "is exactly" statement in the definition of $P(x, C)$. Therefore, this addition results in the probability of the union event. It follows that the probability that a 2-path E' generates the cover $\{C_1, C_2, \ldots, C_\ell\}$ and the ℓ-tuple $(t_1, t_2, \ldots, t_\ell)$ with $t_{E'}(C_i) = t_i \forall i \in 1, \ldots, \ell$ is:

$$p(\{E' : \mathcal{C}_{E'} = \{C_1, \ldots, C_\ell\} : \ t_{E'}(C_i) = t_i\}) = p(t_1, \ldots, t_\ell, C_1, \ldots, C_\ell). \tag{5}$$

Finally, from Eq. (1) and 5 we have that:

$$R_K(G, 2) = \sum_{F \subset K^{[2]}} \left[\prod_{f\in F} p(f)\right]\left[\prod_{f \in K^{[2]}\setminus F}(1 - p(f))\right] \sum_{\substack{(C_1,\ldots,C_\ell)\in\mathcal{P}(K)^{(\ell)}: \\ \{C_1,\ldots,C_\ell\}\sqsupset_F K}} \ \sum_{\substack{(t_1,\ldots,t_\ell)\in(V\setminus K)^{(\ell)}: \\ t_i<t_{i+1}, i=1,\ldots,l-1}} p(t_1, \ldots, t_\ell, C_1, \ldots, C_\ell). \tag{6}$$

3.4. Computational complexity

The first summation in Eq. (6) has $2^{\binom{|K|}{2}}$ terms. The second summation has no more than $2^{(2^{|K|})}(\ell!)$ terms with $\ell \leq 2^{(2^{|K|})}$. The third summation has $\binom{n'}{\ell}$ terms. The product operands involve $\binom{|K|}{2}$ products. Computing $p(t_1, \cdots, t_\ell, C_1, \cdots, C_\ell)$ involves n products and a number of additions bounded by $(\ell+1)n$ since, denoting as τ_i the position of t'_i within the ordering of $V \setminus K$, the number of additions is equal to $2(\tau_2 - \tau_1 - 1) + 3(\tau_3 - \tau_2 - 1) + \cdots + \ell(\tau_\ell - \tau_{\ell-1} - 1) + (\ell + 1)(n - \tau_\ell) = -2 - 3 - \cdots - \ell - 2\tau_1 - \tau_2 - \tau_3 - \cdots - \tau_\ell + (\ell + 1)n < (\ell + 1)n$. Hence the number of elemental operations (additions and products) needed to compute $R_K(G, 2)$ has order

$$2^{\binom{|K|}{2}}\binom{|K|}{2}2^{(2^{|K|})}2^{(2^{|K|})}!(n')^{2^{(2^{|K|})}}n(2^{(2^{|K|})} + 2) \tag{7}$$

which is a polynomial in n of degree $2^{(2^{|K|})}+1$, thus proving that the complexity is polynomial. It is easy to see that an enumeration of the sets $\mathcal{P}(K^{(2)})$, $\{C \sqsupset_F K\}$ and $(V \setminus K)^{(\ell)}$ can be done in a number of steps linear in their cardinality; then the computational complexity order of the three summations is the number of terms involved. Now let us see that the complexity is, indeed, linear in n. To do so, it is enough to show that the following summation can be computed in a time linear in n,

$$\sum_{\substack{(t_1,\ldots,t_\ell)\in(V\setminus K)^{(\ell)}: \\ t_i<t_{i+1}, i=1,\ldots,l-1}} p(t_1, \ldots, t_\ell, C_1, \ldots, C_\ell), = \sum_{\substack{(t_1,\ldots,t_\ell)\in(V\setminus K)^{(\ell)}: \\ t_i<t_{i+1}, i=1,\ldots,l-1}} \prod_{i=1}^{\ell} p(t_i, C_i) \prod_{i=0}^{\ell} \prod_{t'_i<t<t'_{i+1}} \sum_{j=0}^{i} p(t, C_j), \tag{8}$$

since the remaining summations and product operators in Eq. (6) multiply the order by factors that only depend on $|K|$, thus being constant with regard to n. For simplicity of notation let us assume that $V \setminus K = \{1, \cdots, n'\}$. Following Eq. (4), the summation of Eq. (8) coincides with the summation of all products that have the form

$$p(1, C_{a_1})p(2, C_{a_2}) \dots p(n', C_{a_{n'}})$$

where $a_t \in \{0, 1, \dots, \ell\}$ (recall that $C_0 = \emptyset$) and there are ℓ integers t_i with $0 < t_1 < \cdots < t_\ell \leq n'$ such that:

- $a_{t_i} = i \quad \forall i \in \{1, \dots, n'\}$,
- if $t < t_1$ then $a_t = 0$,
- if $t_1 < t < t_2$, then $a_t \in \{0, 1\}$,
- if $t_2 < t < t_3$, then $a_t \in \{0, 1, 2\}, \dots,$
- if $t_{\ell-1} < t < t_\ell$, then $a_t \in \{0, 1, \dots, \ell - 1\}$,
- if $t > t_\ell$, then a_t is any integer between 0 and ℓ.

These products can be associated to directed paths in the directed graph defined by:

$$\begin{aligned}
\vec{G} &= (\{1, \dots, n'\} \times \{0, \dots, \ell\}^2, \vec{E}) \\
\vec{E} &= \{((t, a, b), (t + 1, a', b')) : t \leq n' - \ell, t \leq b + 1, 0 \leq a' \leq b + 1, b' = \max(a', b)\} \cup \\
&\quad \{((t, a, b), (t + 1, a', b')) : n' - \ell \leq t \leq n' - \ell + b, \ell - (n' - t) < a' \leq b + 1, b' = \max(a', b)\}
\end{aligned}$$

that go from vertex $(1, 0, 0)$ to (n', ℓ, ℓ). Each vertex (t, a, b) is associated to the probability $p(t, C_a)$. Figure 2 shows, as an example, the progression of pairs (t, b) visited when $n' = 5$ and $\ell = 2$ (it can be seen as a projection of the graph in the plane t, b). The variable b cumulates the number of nodes of $t_1, \dots, t_n$ already visited when t moves from $t = 1$ to $t = n'$; while a represents the possible sets involved in the events $p(t, C_a)$. Computing the summation of Eq. (8) can be done by dynamic programming, proceeding from the vertices with the form (n', a, b) downwards to reach $(1, 0, 0)$. In each step, a value $s(t, a, b)$ is assigned to the vertex (t, a, b) satisfying

$$s(t, a, b) = \begin{cases} 1 & t = n', \\ p(t, C_{a'}) \sum_{(t,a,b) \rightsquigarrow (t+1,a',b')} s(t + 1, a', b') & t < n'. \end{cases}$$

Hence the number of operations for computing the summation of Eq. (8) will not exceed the number of edges of the graph $\vec{G}$, which is bounded by $(n' + 1)\ell^2$ times the maximum possible degree ℓ^2, that is $(n' + 1)\ell^4$, linear in n, which completes the proof.

4. Conclusions and Ongoing Work

This paper introduces a contribution regarding the complexity analysis of the d-DCKR problem; it is proved that computing the 2-DCKR is of polynomial complexity in n, the number of nodes of the network, provided that n is the only input of the problem (being $|K| \leq n$ any fixed parameter). Moreover, it is proved that the complexity order is indeed linear in n, according to the explicit formula given for computing the 2-DCKR. Recall that it was already known that for $d > 2$ the problem is NP-hard.

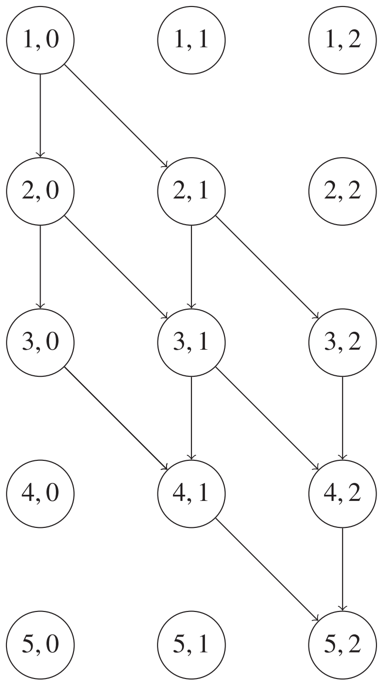


Figure 2: (t,b)-projection of the auxiliary graph for computing $p(t_1, \ldots, t_5, C_0, \ldots, C_2)$

The case where $|K|$ is also considered as an input for the complexity analysis seems to be also an NP-hard problem; although this has not yet been demonstrated. Observe that the problem P_1 of computing the 2-DCKR with $K = V$ and all edge reliabilities equal to 0.5 is equivalent to the problem P_2 of counting all the partial graphs $(V, E' \subseteq E)$ of G with diameter two. This is due to the fact that under such hypothesis every partial graph has the same probability of occurrence (2^{-n}) and each partial graph is $2 - V-$connected if and only if it has diameter two; it follows that $R_V(G, 2)$ equals the number of such partial graphs divided by 2^n. Then, proving that the problem $P2$ is NP-hard would suffice to prove that the 2-DCKR with $K = V$ is NP-hard too; since P_2 is a special case of the latter (where all edges are assigned reliabilities equal to 0.5). Furthermore, it would follow that the 2-DCKR with K as a free input for the complexity would be NP-hard, since so it is its particular case 2-DCKR with $K = V$. Currently our efforts head toward proving that P_1 is an NP-hard problem.

References

[1] C. J. Colbourn, The Combinatorics of Network Reliability, Oxford University Press, Inc., New York, NY, USA, 1987.

[2] L. Petingi, J. Rodriguez, Reliability of networks with delay constraints, in: Congressus Numerantium (2001), volume 152, pp. 117–123.

[3] H. Cancela, L. Petingi, Reliability of communication networks with delay constraints: computational complexity and complete topologies, International Journal of Mathematics and Mathematical Sciences 2004 (2004) 1551–1562.

[4] L. Petingi, A diameter-constrained network reliability model to determine the probability that a communication network meets delay constraints, WTOC 7 (2008) 574–583.