Calibrating a Dependent Failure Model for Computing Reliabilities on Telecommunication Networks

Omar Matus, Javiera Barrera, Eduardo Moreno

Faculty of Engineering and Sciences
Universidad Adolfo Ibáñez
Avda. Diagonal Las Torres 2700
Peñalolén, Santiago, CHILE

Gerardo Rubino

INRIA Rennes – Bretagne Atlantique
Campus de Beaulieu
35042 Rennes Cedex, FRANCE

Resumen

In this work, we propose a methodology to calibrate a dependent failure model to compute the reliability in a telecommunication network. We use the Marshall-Olkin (MO) copula model, which captures failures that arise simultaneously in groups of links. In practice, this model is difficult to calibrate because it requires the estimation of a number of parameters that is exponential in the number of links.

We formulate an optimization problem to calibrate a MO copula model to attain given marginal failure probabilities for all links and the correlations between them. Using a geographic failure model, we calibrate various MO copula models using our methodology, we simulate them, and we benchmark the reliabilities thus obtained.

Our experiments show that considering the simultaneous failures of small and connected subsets of links is the key to obtain a good approximation of reliability, confirming what is suggested by the telecommunication literature.

1. Introducción

Durante los últimos 40 años, el estudio del diseño y la evaluación de redes de telecomunicaciones se ha enfrentado de diversas maneras. Tomando en cuenta que existen medidas de funcionamiento que para una red, como velocidad, conectividad y seguridad entre otros, es importante notar que en este trabajo nos centraremos en el estudio de la confiabilidad de una red, entendida como la probabilidad de que un conjunto de componentes, que pueden tornarse no operativos a medida que pasa el tiempo, estén conectados en un tiempo determinado. Esta evaluación de la confiabilidad requiere un modelo de fallas, en conjunto con una metodología que permita calibrar dicho modelo. El estudio de la confiabilidad es un problema difícil de enfrentar, incluso cuando muchos modelos de estudio de fallas consideran el supuesto de independencia entre las fallas de los componentes el problema es #P-completo [PB83].

Si bien el supuesto de independencia es usado en varios modelos, la evidencia empírica, del estudio de diversos tipos de redes, demuestra que la correlación entre las fallas de los componentes existe e impacta significativamente la confiabilidad de redes en diferentes contextos [TLSS10], [HR01], [GHHK10], [GJN11].

Al enfrentarnos al problema de diseñar una red, debemos tener en consideración que el modelo de fallas seleccionado debe entregarnos, por un lado, rigurosidad en los resultados, al mismo tiempo que, flexibilidad en su calibración y uso, tal como señalan Barrera, Cancela y Moreno [BCM14].

El problema de elección de un modelo de fallas dependientes ha sido analizado durante los últimos años por diversos autores. En nuestro caso seleccionamos un modelo ampliamente estudiado y que ofrece una gran flexibilidad en su uso, el modelo de Marshall-Olkin [MO67]. A pesar de su versatilidad, este modelo ha enfrentado críticas a lo largo de los años, inducidas por la poca escalabilidad que posee su calibración debido a la gran cantidad de parámetros requeridos [Sin02].

2. Modelo de Marshall-Olkin

Sea $\mathscr{G}$ un grafo con un conjunto de nodos $\mathscr{N}$, conectados, no necesariamente todos con todos, por un conjunto de arcos denotados por $\mathscr{C} = \{1,2,\ldots,n\}$. Asumiremos que los nodos no pueden fallar. Para los arcos, tomemos los posibles subconjuntos de arcos, cada uno de los cuales puede ser afectado por un *shock* que vuelve no operativo a los componentes de ese subconjunto que estaban operativos al momento del *shock*. De esta forma, definimos un tiempo de vida de cada componente, que corresponde al mínimo tiempo hasta que un *shock* afecte a un subconjunto que posee el componente.

Si se considera que los *shocks* ocurren de acuerdo a procesos de Poisson, entonces el tiempo de vida de los componentes es exponencial. En caso de algún subconjunto no tenga un *shock* asociado, decimos que su tasa de falla es cero. De esta forma el estado de la red, operativa o no, es una función del tiempo y depende del estado de los arcos. Es importante notar que los arcos representan en este modelo los componentes de la red estudiada, mientras que los nodos son las conexiones entre dichos componentes.

Estudiamos la confiabilidad $R(t)$ de la red, seleccionando dos nodos, s y f, de forma tal que $R(t)$ representa la probabilidad de que en el tiempo t exista un camino entre estos nodos, formado por los arcos que aún permanecen operativos en ese instante. En nuestro caso, analizaremos la conectividad de los nodos en el momento $t = 1$, donde tasa de ocurrencia del *shock* que afecta al subconjunto de arcos i, estará dada por $\lambda_i = -\ln(1-p_i)$, así, con $R(1)$ se puede recuperar el modelo estático. Esto representa lo esencial del proceso de Elperin y Lomonosov [EGL91].

Dado un conjunto de probabilidades de fallas marginales de los arcos y las correlaciones entre estas fallas, podemos plantear un sistema de $n(n+1)/2$ ecuaciones y 2^n-1 variables, una ecuación para cada falla marginal de cada arco y una ecuación para cada correlación entre dos arcos, junto con una variable para cada subconjunto de arcos.

Debemos destacar que generalmente el sistema de ecuaciones posee infinitas soluciones, por lo cual es importante analizar el impacto que tiene en la estimación de la confiabilidad la elección de cada una de las soluciones posibles.

3. Escogiendo un conjunto apropiado de cópulas para MO

Buscamos determinar un conjunto de cópulas para satisfacer el sistema de ecuaciones, dicho problema no es fácil de resolver ya que el sistema tiene un numero exponencial de variables. Una solución para este problema puede obtenerse usando una técnica conocida como generación de columnas [DW60]. Esta técnica, permite incluir diferentes criterios de forma de priorizar las cópulas que se utilizarán. El modelo principal, denominado maestro en generación de columnas, para obtener las cópulas está dado en la ecuación (1).

$$\text{mín} \sum_{i,j\in\mathscr{C}} \left(t_{ij}^{+} + t_{ij}^{-}\right) \tag{1a}$$

$$\sum_{V\subseteq\mathscr{C}:\{i,j\}\in V} \lambda_V + t_{ij}^{+} - t_{ij}^{-} = \ln\left(\frac{\rho_{i,j}\sqrt{p_i}\sqrt{p_j}}{\sqrt{(1-p_i)(1-p_j)}}+1\right) \quad \forall i,j\in\mathscr{C}, i\neq j \tag{1b}$$

$$\sum_{V\subseteq\mathscr{C}:i\in V} \lambda_V + t_{ii}^{+} - t_{ii}^{-} = -\ln(1-p_i)\ , \ \ \forall i\in\mathscr{C} \tag{1c}$$

$$\lambda_V \geq 0\ , \ \ \forall V\subseteq\mathscr{C} \tag{1d}$$

$$t_{ij}^{+}, t_{ij}^{-} \geq 0\ , \ \ \forall i,j\in\mathscr{C}. \tag{1e}$$

En este modelo se busca encontrar un conjunto de cópulas que se acerquen lo más posible a satisfacer las ecuaciones del modelo de Marshall-Olkin, de esta forma tenemos que t_{ij}^{+} y t_{ij}^{-} representan la holgura que existe para satisfacer cada una de las ecuaciones, por sobre o bajo el valor del lado derecho de cada una, respectivamente. Mientras que los p_i y los $\rho_{i,j}$ representan respectivamente las probabilidades marginales de falla de cada componente y la correlación entre dos componentes.

Para verificar cuales cópula debe agregarse al problema, se puede elaborar un modelo *pricing* que permite determinar a través de conocer los costos reducidos de las cópulas candidatas a ingresar

$$\bar{C}_V = -\sum_{i\in V}\mu_i - \sum_{i,j\in V}\nu_{ij},$$

donde μ_i y ν_{ij} son las variables duales de las ecuaciones planteadas en (1c) y (1b), respectivamente. En este modelo se probaron diferentes criterios, tales como minimizar o maximizar la suma total de las tasas de ocurrencia de los *shocks* y minimizar la suma de las máximas distancias entre los nodos en cada subconjunto de arcos ponderados por su tasa.

4. Experimentos numéricos

Para probar las metodologías diseñadas, se escogió un sistema de fallas físicas, en donde una red de 36 arcos y 21 nodos es afectada por terremotos en posiciones y con intensidades aleatorias, tomando como base el modelo planteado por Agarwal et al. [AEG$^+$13]. La simulación de dicho modelo físico nos permite tener las probabilidades marginales de las fallas de cada arco, así como también las correlaciones existentes entre estos.

Con esta información aplicamos nuestro modelo de generación de columnas con diferentes criterios de elección, de forma tal de obtener los parámetros para calibrar MO. Una vez calibrado MO, realizamos simulaciones de forma tal de obtener la estimación de la confiabilidad de la red, para cada criterio escogido. Esta confiabilidad es comparada con la confiabilidad entregada por el modelo físico.

Es importante notar que en nuestro experimento conocemos el modelo físico detrás del funcionamiento de la red, por lo tanto el objetivo de estimar la confiabilidad por medio de MO es únicamente comparar la estimación de nuestro modelo con el resultado físico y analizar qué tanto nos acercamos. En la práctica muchas veces los modelos que subyacen a las redes son desconocidos o de difícil simulación o modelado. MO permite estimar la confiabilidad de las redes incluso sin conocer ese modelo.

Los resultados muestran que las confiabilidades obtenidas por medio del modelo de MO, calibrado utilizando nuestra metodología, son muy cercanos a las confiabilidades reales. En todos los casos, el error absoluto promedio en la estimación de la confiabilidad es de un 1 % mientras que los errores absolutos máximos son menores a 3 %. Dichos resultados están en la siguiente tabla, donde la notación es para cada caso Me para Media, M para máxima, AD para Diferencia Absoluta y DR para diferencia relativa. Y los modelos analizados son Independiente, Maximizar suma de Lambdas, Minimizar suma de Lambdas, Maximizar suma de Distancias en Lambdas.

Cuadro 1: Errores relativos y absolutos

Model	Me DA	Me DR	M DA	M DR
Indep.	0.041	0.048	0.156	0.196
Max L	0.006	0.007	0.028	0.036
Min L	0.002	0.003	0.016	0.019
Max D	0.003	0.003	0.026	0.033

5. Conclusiones

La dependencia entre las fallas de los componentes es significativamente importante en el estudio de la confiabilidad de redes e ignorarla puede traer consigo grandes divergencias en la estimación de la confiabilidad.

Por otro lado, mostramos que es posible calibrar MO teniendo las fallas marginales de los componentes y las correlaciones entre ellos; por medio de generación de columnas, se pueden determinar los parámetros necesarios para MO. Así también, considerando que las correlaciones son la probabilidad de falla simultánea de dos componentes, en caso de poseer la información de fallas simultáneas de más de dos componentes, el modelo es fácilmente adaptable a esta situación.

Finalmente corroboramos las conclusiones expuestas en la literatura, referentes a que los criterios de proximidad de los componentes y la cantidad de componentes involucrados en fallas simultaneas pueden ser utilizados para calibrar MO de forma tal de que estimemos de buena manera la confiabilidad de una red.

6. Trabajo Futuro

Como trabajo futuro pretendemos probar los métodos y modelos usados en más redes, de forma tal de analizar si los resultados obtenidos son aplicables a un numero mayor de ellas. Así también buscamos encontrar otro modelo de fallas distinto al de sismos, para ver una posible generalización de los resultados usando modelo de fallas de otra naturaleza.

Agradecimientos

Los autores agradecen el aporte financiero de los proyectos FONDECYT 1130681 y 1161064 INRIA-CIRIC y al Programa Iniciativa Científica Milenio NC130062.

Referencias

[AEG$^+$13] Pankaj K Agarwal, Alon Efrat, Shashidhara K Ganjugunte, David Hay, Swaminathan Sankararaman, and Gil Zussman. The resilience of wdm networks to probabilistic geographical failures. *IEEE/ACM Transactions on Networking (TON)*, 21(5):1525–1538, 2013.

[BCM14] Javiera Barrera, Hector Cancela, and Eduardo Moreno. Topological optimization of reliable networks under dependent failures. *Operations Research Letters*, 32(2):132–136, 2014.

[DW60] George B Dantzig and Philip Wolfe. Decomposition principle for linear programs. *Operations Research*, 8(1):101–111, 1960.

[EGL91] T Elperin, I Gertsbakh, and M Lomonosov. Estimation of network reliability using graph evolution models. *Reliability, IEEE Transactions on*, 40(5):572–581, 1991.

[GHHK10] Andres J. Gonzalez, Bjarne E. Helvik, Jon K. Hellan, and Pirkko Kuusela. Analysis of Dependencies between Failures in the UNINETT IP Backbone Network. In *2010 IEEE 16th Pacific Rim International Symposium on Dependable Computing*, pages 149–156. IEEE, 2010.

[GJN11] Phillipa Gill, Navendu Jain, and Nachiappan Nagappan. Understanding network failures in data centers. In *ACM SIGCOMM Computer Communication Review*, volume 41, page 350. ACM, 2011.

[HR01] Jane Nichols Hagstrom and Sheldon Ross. Component state dependence and error in reliability computation. Available in http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.20.8460, 2001.

[MO67] Albert W Marshall and Ingram Olkin. A multivariate exponential distribution. *Journal of the American Statistical Association*, 62(317):30–44, 1967.

[PB83] J Scott Provan and Michael O Ball. The complexity of counting cuts and of computing the probability that a graph is connected. *SIAM Journal on Computing*, 12(4):777–788, 1983.

[Sin02] Nozer D. Singpurwalla. Dependence in network reliability. *Proceedings of the 5th International Conference on Information Fusion, FUSION 2002*, 2:981–985, 2002.

[TLSS10] Daniel Turner, Kirill Levchenko, Alex C. Snoeren, and Stefan Savage. California fault lines: understanding the causes and impact of network failures. In *ACM SIGCOMM Computer Communication Review*, volume 40, pages 315–326. ACM, 2010.