

STOCHASTIC MODELS OF CASCADING FAILURES

ANDREW W. SWIFT,* *University of Nebraska at Omaha*

Abstract

What is a ‘cascading failure’? Whilst most people have an intuitive idea of what is meant by cascading, no definition exists within the framework of reliability theory. The aim of this paper is to explore the field of component dependence within a network structure, and to construct a definition of cascading failure.

Keywords: Reliability; cascading failure; parallel redundant system

2000 Mathematics Subject Classification: Primary 91B70

Secondary 90B25

1. Introduction

The notion of cascading and of cascading failures is not an abstract one, indeed it is very common in the real world, evident in such events as blackouts in the electric power grid and failures of electrical appliances. In these events the problem is initiated by the failure of one component of the system, for example, a resistor in the case of an electrical appliance or, in the case of the electric power grid, a transmission line or generator. The initial failure alters the structure function of the system, which leads to subsequent failures within a short period of time. The meaning of ‘within a short period of time’ is dependent on the type of system. In the case of an electrical appliance, ‘a short period of time’ is a fraction of a second, whereas in the case of the electric power grid, it could be minutes or even hours. Engineers refer to this short period of time as a ‘threshold’.

Whilst the idea of cascading failures is common in the real world, there does not exist an established mathematical definition, and, thus, there is no comprehensive mathematical model to represent such a phenomenon. A mathematical definition of cascading failures was suggested by Lindley and Singpurwalla [2], but the notion has not been fully explored. In this paper we aim to explore the notion of cascading failures using probability arguments. Using Lindley and Singpurwalla’s definition as a starting point, we construct a stochastic model representing a system operating under cascading failures. While our focus will primarily be on parallel redundant systems, we will also discuss other systems, such as ‘ k -out-of- n ’ systems.

1.1. Causality and causal failures

In order to define cascading and cascading failures, we first turn to the notion of causality. This topic has spawned a great deal of literature from philosophers, mathematicians, and engineers alike, without generating any kind of consensus as to what exactly causality means. However, one definition of causality, that by Suppes [5], provides us with a platform on which to construct a definition of cascading. (It is not our intention here to debate the philosophical

Received 28 February 2003; revision received 8 October 2008.

* Postal address: Department of Mathematics, University of Nebraska at Omaha, 6001 Dodge Street, Omaha NE 68182, USA. Email address: aswift@mail.unomaha.edu

and mathematical arguments surrounding the various approaches to causality; for a recent discussion of the various approaches, see Singpurwalla [4].)

Suppes [5, p. 12] stated that \mathcal{D} is a *prima facie probabilistic cause* of \mathcal{E} (denoted $\mathcal{D} \xrightarrow{P} \mathcal{E}$) if

- (a) \mathcal{D} occurs before \mathcal{E} (in time);
- (b) $\mathcal{P}(\mathcal{D}) > 0$; and
- (c) $\mathcal{P}(\mathcal{E} \mid \mathcal{D}) > \mathcal{P}(\mathcal{E})$.

This definition satisfies our intuitive notion of what it means for one event to ‘be the cause of’ another, that is, that \mathcal{D} is a (*prima facie probabilistic*) cause of \mathcal{E} if the probability of occurrence of event \mathcal{E} is increased by the preceding occurrence of event \mathcal{D} .

For our purposes, it will be more useful to have a definition of causality in terms of random variables. Fortunately, there is a direct relationship between events and random variables that makes this transition very simple. For example, suppose that we have two binary random variables X and Y . We can define the event \mathcal{D} to be the event that $X = 1$, and, similarly, we can define the event \mathcal{E} to be the event that $Y = 1$.

Freund’s [1] bivariate exponential distribution allows us to create a model of a parallel redundant system with two components. In this system, the lifetimes of the two components behave as if they are independent, until one of the components fails, after which the remaining component suffers increased stress. As a simple example, suppose that, given θ , the lifetimes of the individual components are independent, exponential distributions with parameter θ . Furthermore, suppose that when one of the two components fails, the increased stress on the surviving component is such that θ doubles to 2θ and its remaining lifetime is exponential with parameter 2θ . It can be shown by a simple calculation that the time to failure of the system has density at u of the form $4u\theta^2e^{-2\theta u}$ and that the mean time to failure is $1/\theta$. Thus, in this model, the failure of the first component *permanently* changes the parameter θ , the failure rate of the surviving component, to 2θ . This increase in the failure rate, shown in Figure 1, increases the probability of failure of the surviving component. Thus, if $X_1 = t_1$ denotes the event that the first component to fail fails at time t_1 and if $X_2 = t_2$, $t_1 < t_2$, denotes the event that the surviving component fails at time t_2 , then the event $(X_1 = t_1)$ is a *prima facie* cause of the event $(X_2 = t_2)$, and, thus, Freund’s model is also a description of causal failures.

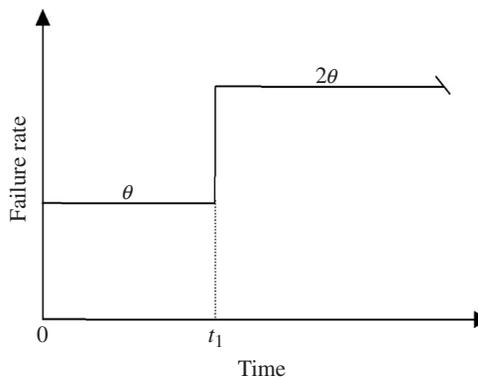


FIGURE 1: Failure rate of the second component to fail in Freund’s [1] model.

This model of causal failures provides the basis for the construction of the model for cascading and cascading failures, first introduced in [2] and expanded upon in this paper. It should be noted that many other models satisfy the definition of a causal model (for example, the bivariate exponential distribution of [3]); however, it is Freund’s model that best provides a link between causality and cascading.

1.2. Cascading and cascading failure models

To get an intuitive idea of what we mean by cascading failures, picture a row of dominos each standing on their end. Now, suppose that the first domino topples and collides with its neighbor, causing it to topple a short period of time after the first domino has toppled. However, now suppose that the second domino is sufficiently far from the first domino that it is not effected by its toppling; then, no matter how much time passes, the toppling of the initial domino will not result in the toppling of the second domino. Hence, the toppling of the first domino causes the toppling of the second domino, either almost immediately or not at all. This is the critical idea behind cascading.

How does this idea translate to a mathematical representation? Lindley and Singpurwalla [2] introduced a formulation of cascading by modifying the model for causal failures described in the previous subsection.

Suppose that the previous model is modified such that the failure rate of the surviving component changes at t_1 from θ to 2θ (as before), but at time $t_1 + \delta$ it reverts back to θ , as illustrated in Figure 2. The quantity $\delta > 0$ is called the critical time (or the threshold time), and the ensuing model is a description for a cascading failure. The choice for δ depends upon the scenario being modeled. Lindley and Singpurwalla considered only the case where δ is a deterministic quantity; however, in Section 2 we will consider the situation where δ is a random variable.

For the modified model described above, it can be shown that the time to failure of the parallel redundant system has density at u of the form

$$f_C(u) = \begin{cases} 4\theta^2 u e^{-2\theta u}, & u < \delta, \\ 4\theta^2 \delta e^{-2\theta u} + 2\theta e^{-\theta(u+\delta)} - 2\theta e^{-2\theta u}, & u \geq \delta. \end{cases}$$

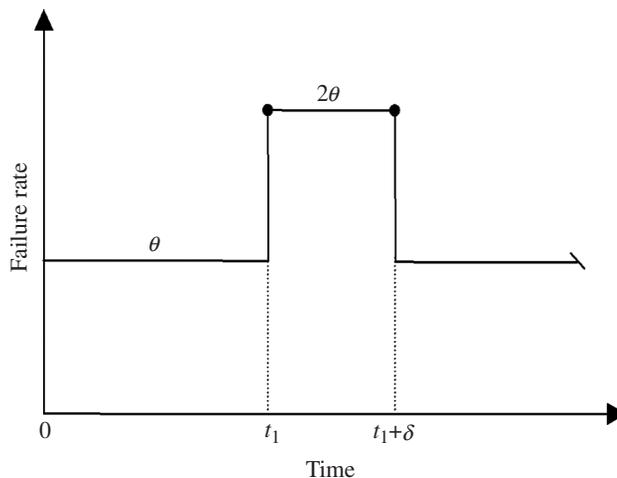


FIGURE 2: Failure rate of the second component to fail in a model for cascading failures.

Similarly, its survival function at u is of the form

$$S_C(u) = \begin{cases} 2\theta u e^{-2\theta u} + e^{-2\theta u}, & u < \delta, \\ 2\theta \delta e^{-2\theta u} + 2e^{-\theta(u+\delta)} - e^{-2\theta u}, & u \geq \delta, \end{cases}$$

and, thus, its failure rate at u is

$$h_C(u) = \begin{cases} 4\theta^2 \frac{u}{2\theta u + 1}, & u < \delta, \\ \frac{2\theta(2\theta \delta e^{-\theta u} + e^{-\theta \delta} - e^{-\theta u})}{2\theta \delta e^{-\theta u} + 2e^{-\theta \delta} - e^{-\theta u}}, & u \geq \delta. \end{cases}$$

The mean time to system failure is $1/\theta + (1/2\theta)e^{-2\theta\delta}$; this is larger than $1/\theta$, the mean time to failure under the causal model of Freund. This is to be expected. We also note that, as $\delta \uparrow \infty$, the mean time to failure converges to $1/\theta$, since the causal model can be thought of as the cascading model with an infinite value of δ . Furthermore, as $\delta \downarrow 0$, the mean time to failure of the system converges to $3/2\theta$, which is the mean time to failure of a parallel redundant system with two independent components, each with identical, exponentially distributed life lengths. Thus, for parallel redundant systems whose component life lengths have exponentially distributed life lengths, cascading failures result in a larger mean time to failure than causal failures, but a lower mean time to failure than that under independent failures.

2. Generalizing the two-component cascade model

The example in Section 1.2 gives the basic idea of what we are trying to achieve, however, it is a very specific example. Before we can look at networks of larger dimension, we need to generalize the previous two-component model. The most simple generalization is the jump in failure rate. Let us now suppose that instead of the failure rate of the second component jumping from θ to 2θ when the first component fails, suppose that it jumps to $c\theta$, where c is some constant and $c > 1$.

It is easy to show that the time to failure of the system with cascading failures has density at u of the form

$$f_C(u) = \begin{cases} \frac{2c\theta}{c-2}(e^{-2\theta u} - e^{-c\theta u}), & u < \delta, \\ 2\theta e^{-\theta(u-\delta+c\delta)} - 2\theta e^{-\theta(2u-2\delta+c\delta)} + \frac{2c\theta}{c-2}(e^{-2\theta u} - e^{-\theta(2u-2\delta+c\delta)}), & u \geq \delta. \end{cases}$$

Thus, the expected life length of the system can be calculated and is

$$\frac{2(2 - 3c + c^2)e^{-c\theta\delta} - 4 + c^2}{2c\theta(c - 2)},$$

which tends to $3/2\theta$ as $\delta \rightarrow 0$, as expected.

The generalization above is very basic and does not alter the basic idea of the model. However, now suppose that instead of regarding δ , the critical time, as some deterministic subjective value, consider δ to be the realization of a random variable Δ . Obviously, we must have $\Delta > 0$, and it also makes sense for Δ to be continuous. Thus, suppose that Δ has a probability density $\pi_\Delta(\delta)$ and that we are interested in $f_C(u)$, the probability density of the time to failure of the parallel system at u . Then

$$f_C(u) = \int_\delta^u \int_0^u f_C(u | t_1, \delta) f(t_1, \delta) dt_1 d\delta,$$

where t_1 is the time to failure of the first component. It is logical for the failure time of the first component and the value of δ to be independent; thus, the above equation becomes

$$f_C(u) = \int_{\delta} \left(\int_0^u f_C(u | t_1, \delta) f(t_1) dt_1 \right) \pi_{\Delta}(\delta) d\delta.$$

The choice still remains as to what distribution to assign to Δ . In the next section we choose Δ to have an exponential distribution, whereas in Section 2.2 we choose Δ to have a Pareto distribution. In both these cases we have taken $c = 2$.

Choosing Δ to have either an exponential or Pareto distribution was based on the desired properties of Δ . Clearly, as Δ is the critical time, an exponential distribution is a natural choice, since it is commonly used in many applications to describe the time until an event occurs. The choice of the Pareto distribution offers some differences over the exponential distribution, including a nonzero lower bound for the value of Δ (set to be 1 in Section 2.2) and ‘longer tails’. Other possible choices for the distribution of Δ are the generalized exponential, gamma, and Weibull. When modeling a specific application, it will be necessary to select an appropriate distribution for Δ .

2.1. Cascade model with an exponentially distributed critical time

Let us suppose that Δ is exponentially distributed with parameter λ . Thus, $\pi_{\Delta}(\delta) = \lambda e^{-\lambda\delta}$. We are interested in calculating $f_C(u)$, the probability density of the time to failure of the parallel system at u . From the previous subsection we know that

$$f_C(u) = \int_{\delta} \left(\int_0^u f_C(u | t_1, \delta) f(t_1) dt_1 \right) \pi_{\Delta}(\delta) d\delta,$$

and from Section 1.2 we know that

$$\int_0^u f_C(u | t_1, \delta) f(t_1) dt_1 = \begin{cases} 4\theta^2 u e^{-2\theta u}, & u < \delta, \\ 4\theta^2 \delta e^{-2\theta u} + 2\theta e^{-\theta(u+\delta)} - 2\theta e^{-2\theta u}, & u \geq \delta. \end{cases}$$

Thus, we have

$$\begin{aligned} f_C(u) &= \int_{\delta} \left(\int_0^u f_C(u | t_1, \delta) f(t_1) dt_1 \right) \pi_{\Delta}(\delta) d\delta \\ &= \int_0^u \left(\int_0^u f_C(u | t_1, \delta) f(t_1) dt_1 \right) \pi_{\Delta}(\delta) d\delta \\ &\quad + \int_u^{\infty} \left(\int_0^u f_C(u | t_1, \delta) f(t_1) dt_1 \right) \pi_{\Delta}(\delta) d\delta \\ &= \int_0^u (4\theta^2 \delta e^{-2\theta u} + 2\theta e^{-\theta(u+\delta)} - 2\theta e^{-2\theta u}) \lambda e^{-\lambda\delta} d\delta + \int_u^{\infty} 4\theta^2 u e^{-2\theta u} \lambda e^{-\lambda\delta} d\delta \\ &= \frac{2\theta e^{-2\theta u}}{\lambda(\lambda + \theta)} (2\theta^2 + \theta\lambda - \lambda^2 + \lambda^2 e^{\theta u} - e^{-\lambda u} (2\theta^2 + \theta\lambda)). \end{aligned}$$

The expected value of the time to failure of the system can be shown to be $(4\theta + 3\lambda)/2\theta(2\theta + \lambda)$. Note that, as $\lambda \downarrow 0$, the expected time to failure of the system converges to $1/\theta$ and, as $\lambda \uparrow \infty$, the expected time to failure of the system converges to $3/2\theta$. These results are consistent with the results in Section 1.2.

2.2. Cascade model with a Pareto distributed critical time

Suppose that Δ has a Pareto distribution with shape parameter a , that is, $\pi_{\Delta}(\delta) = a/\delta^{a+1}$ for $\delta \in (1, \infty)$. As in the previous section, we are interested in $f_C(u)$, the probability density of the time to failure of the parallel redundant system at u . In this case we have to separate the calculation into two ranges, $u < 1$ and $u \geq 1$. For $u < 1$,

$$\begin{aligned} f_C(u) &= \int_{\delta} \left(\int_0^u f_C(u | t_1, \delta) f(t_1) dt_1 \right) \pi_{\Delta}(\delta) d\delta \\ &= \int_1^{\infty} \left(\int_0^u f_C(u | t_1, \delta) f(t_1) dt_1 \right) \pi_{\Delta}(\delta) d\delta \\ &= \int_1^{\infty} 4\theta^2 u e^{-2\theta u} \frac{a}{\delta^{a+1}} d\delta \\ &= 4\theta^2 u e^{-2\theta u}, \end{aligned}$$

which is the expected result, since, as $u < 1 < \delta$, the failure rate never returns to its original level and we can treat this case as in the standard Freund model.

Now, for $u \geq 1$,

$$\begin{aligned} f_C(u) &= \int_{\delta} \left(\int_0^u f_C(u | t_1, \delta) f(t_1) dt_1 \right) \pi_{\Delta}(\delta) d\delta \\ &= \int_1^u \left(\int_0^u f_C(u | t_1, \delta) f(t_1) dt_1 \right) \pi_{\Delta}(\delta) d\delta \\ &\quad + \int_u^{\infty} \left(\int_0^u f_C(u | t_1, \delta) f(t_1) dt_1 \right) \pi_{\Delta}(\delta) d\delta \\ &= \int_1^u (4\theta^2 \delta e^{-2\theta u} + 2\theta e^{-\theta(u+\delta)} - 2\theta e^{-2\theta u}) \frac{a}{\delta^{a+1}} d\delta + \int_u^{\infty} 4\theta^2 u e^{-2\theta u} \frac{a}{\delta^{a+1}} d\delta \\ &= \frac{4\theta^2 e^{-2\theta u}}{a-1} (a - u^{-(a-1)}) + 2\theta e^{-2\theta u} (u^{-a} - 1) + 2a\theta \int_1^u e^{-\theta(u+x)} x^{-a-1} dx. \end{aligned}$$

Combining the two cases gives us the full probability density function:

$$f_C(u) = \begin{cases} 4\theta^2 u e^{-2\theta u}, & u < 1, \\ \frac{4\theta^2 e^{-2\theta u}}{a-1} (a - u^{-(a-1)}) + 2\theta e^{-2\theta u} (u^{-a} - 1) \\ \quad + 2a\theta \int_1^u e^{-\theta(u+x)} x^{-a-1} dx, & u \geq 1. \end{cases}$$

The remaining integral in the above expression has no closed-form analytical solution; hence, numerical integration methods are required to generate the result.

The use of the Pareto distribution for Δ rather than an exponential distribution results in a lowering of the expected time to failure of the system. Since the above expression for $f_C(u)$ is not in closed form, this can only be demonstrated numerically. For example, consider a system where $\theta = 0.1$. Now, if we use $\Delta \sim \text{Pareto}(2)$, we obtain an expected time to failure of the system of 13.52, whereas if we use $\Delta \sim \text{Exponential}(\frac{1}{2})$, we obtain an expected time to failure of 13.57. (Note that the comparison of Pareto(2) and Exponential($\frac{1}{2}$) was made as each has the same expected value (equal to 2).)

3. Extending the model to more than two components

Our initial starting point for defining a model for cascading failures was the bivariate exponential distribution of Freund [1]. However, there is no multivariate extension of the Freund bivariate model and, hence, this does not help us when it comes to defining a cascade model for networks of more than two components. From examination of the two-dimensional model and by intuitive thought about the nature of cascading, two possible extensions suggest themselves. In this section we will discuss these two extensions, namely, two-valued cascading and many-valued cascading, and compare their properties.

3.1. Two-valued cascading

In the two-dimensional cascade model first introduced in Section 1.2 and further discussed in Section 2, the failure rate of the second component takes one of only two possible values at any given time. That is, the failure rate of the second component is θ in the interval $[0, t_1]$, jumps up to $c\theta$ ($c > 1$) in the interval $(t_1, t_1 + \delta]$, and then returns to θ in the interval $(t_1 + \delta, \infty)$, where t_1 is the failure time of the first component. It is this feature that provides the motivation for two-valued cascading.

Assume that we have a three-component cascading system where each component has an initial failure rate of θ . Upon the failure of the first component, at time t_1 , the failure rate of the remaining two components jumps to $c\theta$. If neither of the remaining components fails in the interval $(t_1, t_1 + \delta]$ then their failure rates drop back to θ and the system from this point on acts exactly like the two-component cascade system described earlier (see Figure 3). If, however, the second component fails at time t_2 in the interval $(t_1, t_1 + \delta]$ then the failure rate of the third component no longer drops back to θ at time $t_1 + \delta$, instead it remains at the higher level of $c\theta$ until time $t_2 + \delta$ (see Figure 4). It is useful to note here that, unless otherwise stated, when we refer to the ‘first’, ‘second’, ‘third’, etc. component, we are not assigning labels to the components, we are instead referring to the order of failure. Hence, by definition, the ‘first’ component will always fail first, etc.

The failure rate of the third component at time t , $h_3(t)$, depends only on the failure time of the second component, t_2 , that is,

$$h_3(t) = \begin{cases} \theta, & t > t_2 + \delta, \\ c\theta, & t \leq t_2 + \delta. \end{cases}$$

Now, let $X_{(1)}$, $X_{(2)}$, and $X_{(3)}$ be random variables representing the failure times of the first, second, and third components, respectively. We are interested in finding $f_3(t)$, the probability density of the third component failing at some time t . This is fairly straightforward, using the law of the extended conversation:

$$f_3(t) = \int_0^t P(X_{(3)} = t \mid X_{(2)} = t_2) f_2(t_2) dt_2$$

$$= \begin{cases} c\theta e^{-c\theta t} \int_0^t \exp(c\theta t_2) f_2(t_2) dt_2, & t \leq \delta, \\ \theta e^{-\theta((c-1)\delta+t)} \int_0^{t-\delta} \exp(\theta t_2) f_2(t_2) dt_2 \\ \quad + c\theta e^{-c\theta t} \int_{t-\delta}^t \exp(c\theta t_2) f_2(t_2) dt_2, & t > \delta \end{cases}$$

(see Appendix A). Thus, in order to calculate $f_3(t)$, we need to know the form of $f_2(t_2)$.

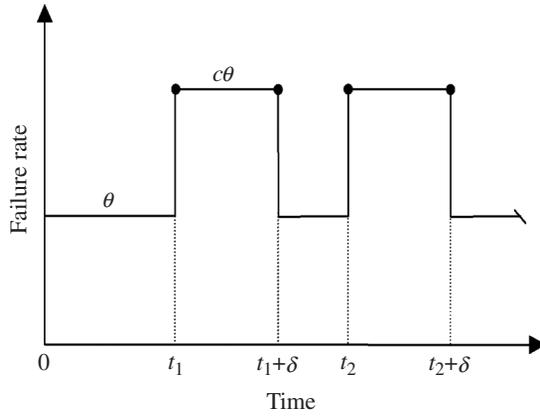


FIGURE 3: Failure rate of the third component to fail when the second component survives past time $t_1 + \delta$.

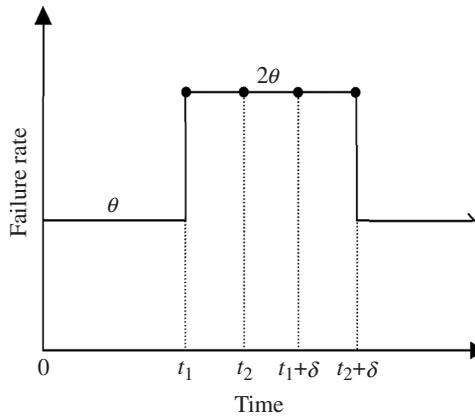


FIGURE 4: Failure rate of the third component to fail when the second component fails before time $t_1 + \delta$ and $c = 2$.

However, $f_2(t_2)$ can also be easily calculated as follows:

$$f_2(t_2) = \int_0^{t_2} P(X_{(2)} = t_2 \mid X_{(1)} = t_1) f_1(t_1) dt_1$$

$$= \begin{cases} 2c\theta \exp(-2c\theta t_2) \int_0^{t_2} \exp(2c\theta t_1) f_1(t_1) dt_1, & t_2 \leq \delta, \\ 2\theta \exp(-2\theta((c-1)\delta + t_2)) \int_0^{t_2-\delta} \exp(2\theta t_1) f_1(t_1) dt_1 \\ \quad + 2c\theta \exp(-2c\theta t_2) \int_{t_2-\delta}^{t_2} \exp(2c\theta t_1) f_1(t_1) dt_1, & t_2 > \delta \end{cases}$$

(see Appendix A). So, in order to calculate $f_2(t)$, we need to know $f_1(t)$. Note that $f_1(t)$ is the density of the time to failure of the first of three components, each with independent

Exponential(θ) life lengths; hence,

$$f_1(t_1) = 3\theta \exp(-3\theta t_1).$$

Thus, by substituting $f_1(t_1)$ into the formula for $f_2(t_2)$ we obtain

$$\begin{aligned}
 f_2(t_2) &= \begin{cases} 2c\theta \exp(-2c\theta t_2) \int_0^{t_2} \exp(2c\theta t_1) 3\theta \exp(-3\theta t_1) dt_1, & t_2 \leq \delta, \\ 2\theta \exp(-2\theta((c-1)\delta + t_2)) \int_0^{t_2-\delta} \exp(2\theta t_1) 3\theta \exp(-3\theta t_1) dt_1 \\ + 2c\theta \exp(-2c\theta t_2) \int_{t_2-\delta}^{t_2} \exp(2c\theta t_1) 3\theta \exp(-3\theta t_1) dt_1, & t_2 > \delta, \end{cases} \\
 &= \begin{cases} 6c\theta^2 \exp(-2c\theta t_2) \int_0^{t_2} \exp((2c\theta - 3\theta)t_1) dt_1, & t_2 \leq \delta, \\ 6\theta^2 \exp(-2\theta((c-1)\delta + t_2)) \int_0^{t_2-\delta} \exp(-\theta t_1) dt_1 \\ + 6c\theta^2 \exp(-2c\theta t_2) \int_{t_2-\delta}^{t_2} \exp((2c\theta - 3\theta)t_1) dt_1, & t_2 > \delta, \end{cases} \\
 &= \begin{cases} 6c\theta^2 \exp(-2c\theta t_2) \frac{\exp(t_2\theta(2c-3)) - 1}{\theta(2c-3)}, & t_2 \leq \delta, \\ 6\theta^2 \exp(-2\theta((c-1)\delta + t_2)) \frac{1 - \exp(-(t_2-\delta)\theta)}{\theta} \\ + 6c\theta^2 \exp(-2c\theta t_2) \frac{\exp(t_2\theta(2c-3)) - \exp((2c-3)(t_2-\delta)\theta)}{\theta(2c-3)}, & t_2 > \delta, \end{cases} \\
 &= \begin{cases} \frac{6c\theta}{2c-3} (\exp(-3\theta t_2) - \exp(-2c\theta t_2)), & t_2 \leq \delta, \\ 6\theta \exp(-2\theta((c-1)\delta + t_2)) (1 - \exp(-(t_2-\delta)\theta)) \\ + \frac{6c\theta}{2c-3} \exp(-2c\theta t_2) (\exp(t_2\theta(2c-3)) - \exp((2c-3)(t_2-\delta)\theta)), & t_2 > \delta. \end{cases}
 \end{aligned}$$

From the above, we can calculate the probability density of the failure time of the network, $f_3(t)$. Because the expression for $f_3(t)$ is long, it has been omitted. What we are mainly interested in is the survival function for the network, that is, $S_3(t) = P(X_{(3)} > t)$. We find that

$$S_3(t) = P(X_{(3)} > t) = \int_0^t P(X_{(3)} > t \mid X_{(2)} = t_2) f_2(t_2) dt_2,$$

where

$$P(X_{(3)} > t \mid X_{(2)} = t_2) = \begin{cases} \exp(-c\theta(t-t_2)), & t \leq t_2 + \delta, \\ \exp(-c\theta\delta - \theta(t-(t_2+\delta))), & t > t_2 + \delta. \end{cases}$$

Thus, it is straightforward to calculate the survival function for a three-component parallel redundant system under two-valued cascading. Again, the expressions are rather long and, hence, have been omitted.

The densities of the failure times are defined recursively, and because the system failure time is equal to the failure time of the final component we have a simple recursive definition for the probability density of the failure time of the system.

3.1.1. *A system with n components.* Now suppose that instead of three components we have n components, where $n \geq 2$. What is the probability density of the network failure time?

The network failure time is the same as the failure time of the last component. From above we know that we can define a recursive formula for the density of the failure time of any component. Let $f_i^n(t)$ be the probability density function of the i th failure in a system of n components. Then, by a simple extension of the three-component example above, we see that, for $i > 1$, we have

$$f_i^n(t_i) = \begin{cases} (n - i + 1)c\theta \exp(-(n - i + 1)c\theta t_i) \\ \quad \times \int_0^{t_i} \exp(-(n - i + 1)c\theta t_{i-1}) f_{i-1}^n(t_{i-1}) dt_{i-1}, & t_i \leq \delta, \\ (n - i + 1)\theta \exp(-(n - i + 1)\theta((c - 1)\delta + t_i)) \\ \quad \times \int_0^{t_i - \delta} \exp(-(n - i + 1)\theta t_{i-1}) f_{i-1}^n(t_{i-1}) dt_{i-1} \\ \quad + (n - i + 1)c\theta \exp(-(n - i + 1)c\theta t_i) \\ \quad \times \int_{t_i - \delta}^{t_i} \exp(-(n - i + 1)c\theta t_{i-1}) f_{i-1}^n(t_{i-1}) dt_{i-1}, & t_i > \delta, \end{cases}$$

and, for $i = 1$, we have

$$f_1^n(t_1) = n\theta \exp(-n\theta t_1).$$

We can thus obtain an expression for $f_n^n(t_n)$, the probability density function for the time to failure of the network.

3.1.2. *k-out-of-n networks.* Now suppose that we are not dealing with a parallel redundant system in which all the components must fail in order for the network to fail. Instead, suppose that we have a k -out-of- n network in which only k of the n components need to fail in order for the network to fail. Then, by the methodology detailed above, it is straightforward to compute the probability density of the network failure time. Indeed, this is simply $f_k^n(t)$.

3.2. Many-valued cascading

What is the difference between many-valued cascading and the two-valued cascading discussed in the previous subsection? Let us consider the three-component problem as we did previously. At the failure of the first component at time t_1 , the failure rate of the remaining two components jumps to $c\theta$ for the interval $(t_1, t_1 + \delta]$, just as in the two-valued cascading model. However, the difference between the two models becomes apparent if the second component fails at some time $t_2 < t_1 + \delta$. If this happens then the failure rate of the surviving component jumps up again to $2c\theta$ until time $t_1 + \delta$ when the failure rate drops back down to $c\theta$ until time $t_2 + \delta$, after which it returns to its original level of θ for all the remaining time. Whereas in the two-valued cascade model the failure rate at any given time is dependent only on whether or not there has been at least one failure in the previous δ units of time, in the many-valued cascade model the failure rate at a given time is dependent on the existence of failures in the previous δ units of time and on how many failures have occurred within that interval. The possible failure rate combinations for three components are shown in Figures 5 and 6.

The first thing to note, even from this simple illustration, is that the failure rate in the many-valued model is always equal to or greater than the equivalent failure rate in the two-valued

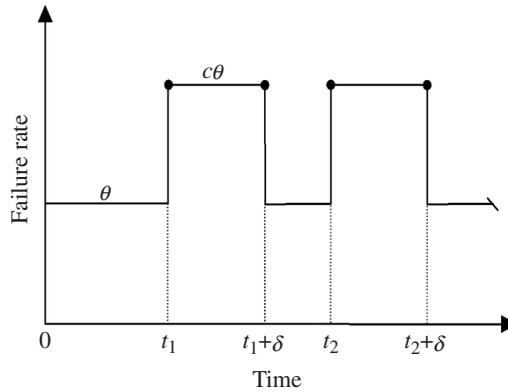


FIGURE 5: Failure rate of the third component to fail when the second component survives past time $t_1 + \delta$.

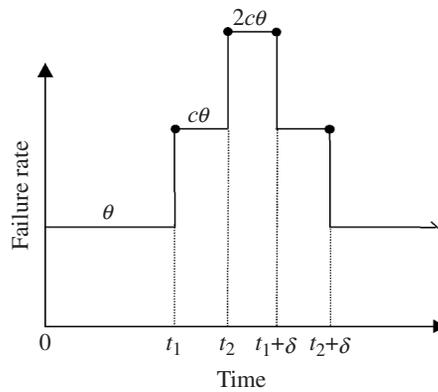


FIGURE 6: Failure rate of the third component to fail when the second component fails before time $t_1 + \delta$.

model. This model can easily be extended to more than three components. For the case of n components, the failure rate of the surviving components at any given time is equal to $(j + 1)c\theta$, where j is the number of components that have failed in the previous δ units of time.

The many-valued cascade model suffers from one major drawback. The computations to derive the probability density and survival functions for the time to failure of the system are very difficult. This complexity is a result of the dependence of the failure rate of surviving components on the number of components that have failed in the last δ units of time. Thus, a conditional probability argument like the one used in the two-valued cascading model does not work because we need to condition on the times of *all* previous failures. This in itself would not be too difficult except for the fact that every possible sequence of failures must be accounted for. Consider the three-component case described above. If the second component survives until after time $t_1 + \delta$, as in Figure 5, then there are two possible scenarios for the failure of the third component, either it can fail in the interval $(t_2, t_2 + \delta]$ or it can fail in the interval $(t_2 + \delta, \infty]$. Now, if the second component fails in the interval $(t_1, t_1 + \delta]$, as in Figure 6, then there are three possible scenarios for the failure of the third component. Either it can fail in the interval $(t_2, t_1 + \delta]$, in the interval $(t_1 + \delta, t_2 + \delta]$, or in the interval $(t_2 + \delta, \infty]$. Thus, for a three-component, many-valued cascade model there are five possible scenarios for the failure

TABLE 1: The number of scenarios for varying network sizes.

Number of components	Number of scenarios
2	2
3	5
4	14
5	42

of the third component that must all be considered when deriving the survival function of the network. Now, five scenarios is not a large number and can be handled with only a moderate amount of thought and computation. The problem occurs, however, because the number of scenarios increases rapidly as the number of components increases. Table 1 shows the number of scenarios for networks with two, three, four, and five components.

The numbers in the right-hand column of Table 1 are recognizable as the Catalan numbers, and, thus, for a system of n components, the number of scenarios would be $(2n)!/(n+1)n!$. The Catalan numbers increase at a rapid rate, for example, the 9th and 10th Catalan numbers are 4862 and 16 796; thus, it is virtually computationally impossible to obtain an analytical expression for the survival function of a network that contains more than a few components.

3.3. Uses of the many-valued cascade model

Although it is analytically impossible to derive the survival function of the network using the many-valued cascade model, there are situations where this type of model is appropriate for modeling a network. In addition, with slight modifications, it becomes very easy to analyze the network using this model.

Suppose that we have a network consisting of n components, each performing the same task, for example, engines on an aeroplane. Now, suppose that initially (at time 0) all components are operational and that the time to failure of each of the components are independent random variables with parameter θ . At some time, $t_{(1)} > 0$, the first component failure will be observed. At this point the failure rate of each remaining component will increase to $c\theta$. It is important to note that this increase in failure rate is not due to the fact that this is a network operating under the assumption of cascading failures. On the contrary this network is *not* operating under the cascade assumption. The increase in failure rate of the remaining components is due to the fact that there is now one less component performing the same task and, hence, each component is required to share more of the load.

Now, as soon as a component fails, work is started to repair it. Suppose that the repair takes exactly δ units of time, after which the component returns to operation and the failure rate returns to its initial value of θ . Of course, it is possible for another component to fail while the first component is under repair. That is, that $t_{(1)} < t_{(2)} < t_{(1)} + \delta$. If this is the case then the failure rate jumps again from $c\theta$ to $2c\theta$, and repair begins on that component also. Clearly, the model used to describe this situation is analogous to the many-valued cascade model discussed in the previous subsection. The main difference in this scenario is that the components are repaired; thus, there will be no time after which the network will cease to operate indefinitely. In this scenario we are interested in the percentage of time that the network is nonoperational, that is, the percentage of time when none of the components are operational.

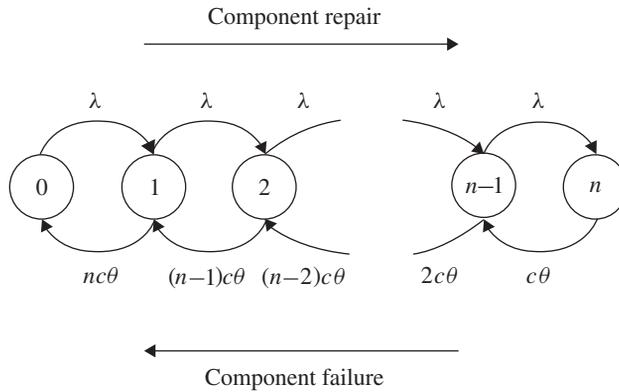


FIGURE 7: Transition rate diagram for a Markov process representing the number of working components.

Trying to analyze this situation has the same problem as analyzing the many-valued cascade model, namely, that there are far too many scenarios to consider. In this problem, because components can be repaired once they have failed, the analysis is even more complicated than that of the cascading model. However, suppose that instead of the repair time being a deterministic value δ it is a realization of an exponential random variable with parameter λ . Then we may set up a Markov process to represent the number of working components at any given time; see Figure 7.

From this we can use standard Markov process analysis to calculate steady-state probabilities and, hence, gain some insight into the amount of time that the network is nonoperational.

4. Summary and further investigation

In this paper we have introduced the idea of cascading failures of components in networks. We discussed in depth the motivation behind the problem and the philosophical arguments surrounding the field of analyzing networks of dependent components. We have suggested two possible models to representing cascading failures, and have discussed the problems with each. The main problem is that the analysis of cascading components on a continuous-time scale is very difficult, and requires a great deal of computational power even for the simplest of networks.

The choice of which of the two described cascading models to use will be motivated by the situation that is to be modeled. Both suggested models satisfy our philosophical definition of what we mean by cascading; however, the many-valued model will lead to system failure in a shorter amount of time than the two-valued model and, thus, is suited to modeling situations in which a component failure institutes a large increase in stress on the remaining working components.

We see this paper as an introductory step into the world of cascading, hopefully leading to more general results and the solutions to more complex problems. We hope that we have performed an important task by introducing the idea of continuous-time stochastic models to represent cascading failures.

Appendix A. Probability densities for two-valued cascading

We have

$$\begin{aligned}
 f_3(t) &= \int_0^t P(X_{(3)} = t \mid X_{(2)} = t_2) f_2(t_2) dt_2 \\
 &= \begin{cases} \int_0^t P(X_{(3)} = t \mid X_{(2)} = t_2) f_2(t_2) dt_2, & t \leq \delta, \\ \int_0^{t-\delta} P(X_{(3)} = t \mid X_{(2)} = t_2) f_2(t_2) dt_2 \\ \quad + \int_{t-\delta}^t P(X_{(3)} = t \mid X_{(2)} = t_2) f_2(t_2) dt_2, & t > \delta, \end{cases} \\
 &= \begin{cases} \int_0^t c\theta \exp(-c\theta(t - t_2)) f_2(t_2) dt_2, & t \leq \delta, \\ \int_0^{t-\delta} e^{-c\theta\delta} \theta \exp(-\theta(t - (t_2 + \delta))) f_2(t_2) dt_2 \\ \quad + \int_{t-\delta}^t c\theta \exp(-c\theta(t - t_2)) f_2(t_2) dt_2, & t > \delta, \end{cases} \\
 &= \begin{cases} c\theta e^{-c\theta t} \int_0^t \exp(c\theta t_2) f_2(t_2) dt_2, & t \leq \delta, \\ \theta e^{-\theta((c-1)\delta+t)} \int_0^{t-\delta} \exp(\theta t_2) f_2(t_2) dt_2 \\ \quad + c\theta e^{-c\theta t} \int_{t-\delta}^t \exp(c\theta t_2) f_2(t_2) dt_2, & t > \delta. \end{cases}
 \end{aligned}$$

Similarly,

$$\begin{aligned}
 f_2(t_2) &= \int_0^{t_2} P(X_{(2)} = t_2 \mid X_{(1)} = t_1) f_1(t_1) dt_1 \\
 &= \begin{cases} \int_0^{t_2} P(X_{(2)} = t_2 \mid X_{(1)} = t_1) f_1(t_1) dt_1, & t_2 \leq \delta, \\ \int_0^{t_2-\delta} P(X_{(2)} = t_2 \mid X_{(1)} = t_1) f_1(t_1) dt_1 \\ \quad + \int_{t_2-\delta}^{t_2} P(X_{(2)} = t_2 \mid X_{(1)} = t_1) f_1(t_1) dt_1, & t_2 > \delta, \end{cases} \\
 &= \begin{cases} \int_0^{t_2} 2c\theta \exp(-2c\theta(t_2 - t_1)) f_1(t_1) dt_1, & t_2 \leq \delta, \\ \int_0^{t_2-\delta} e^{-2c\theta\delta} 2\theta \exp(-2\theta(t_2 - (t_1 + \delta))) f_1(t_1) dt_1 \\ \quad + \int_{t_2-\delta}^{t_2} 2c\theta \exp(-2c\theta(t_2 - t_1)) f_1(t_1) dt_1, & t_2 > \delta, \end{cases}
 \end{aligned}$$

$$= \begin{cases} 2c\theta \exp(-2c\theta t_2) \int_0^{t_2} \exp(2c\theta t_1) f_1(t_1) dt_1, & t_2 \leq \delta, \\ 2\theta \exp(-2\theta((c-1)\delta + t_2)) \int_0^{t_2-\delta} \exp(2\theta t_1) f_1(t_1) dt_1 \\ \quad + 2c\theta \exp(-2c\theta t_2) \int_{t_2-\delta}^{t_2} \exp(2c\theta t_1) f_1(t_1) dt_1, & t_2 > \delta. \end{cases}$$

Acknowledgements

This research was supported by the Army Research Office under A MURI Grant DAAD 19-01-1-0502.

The author would like to thank the editors and anonymous referees for their time and effort in reviewing this paper.

References

- [1] FREUND, J. E. (1961). A bivariate extension of the exponential distribution. *J. Amer. Statist. Assoc.* **56**, 971–977.
- [2] LINDLEY, D. V. AND SINGPURWALLA, N. D. (2002). On exchangeable, causal and cascading failures. *Statist. Sci.* **17**, 209–219.
- [3] MARSHALL, A. W. AND OLKIN, I. (1967). A multivariate exponential distribution. *J. Amer. Statist. Assoc.* **62**, 30–44.
- [4] SINGPURWALLA, N. D. (2002). On causality and causal mechanisms. *ISI Rev.* **10**, 53–78.
- [5] SUPPES, P. (1970). *A Probabilistic Theory of Causality*. North-Holland, Amsterdam.