Preprint December 2003

A Loading-Dependent Model of Probabilistic Cascading Failure

Ian Dobson Benjamin A. Carreras David E. Newman

Abstract: Large blackouts of electric power transmission systems are typically caused by cascading failure of loaded system components. We propose an analytically tractable model of loading dependent cascading failure that captures some of the salient features of large blackouts. This leads to a new application and derivation of the quasibinomial distribution and its generalization to a saturating form with an extended parameter range. Suitably loading the model of cascading failure yields a power law region in the distribution of the number of failures similar to the distribution of blackout sizes observed in blackout data and simulations. The cascading failure model has many identical components randomly loaded. An initial disturbance adds load to each component and causes some components to fail by exceeding their loading limit. Failure of a component causes a fixed load increase for other components. As components fail, the system becomes more loaded and cascading failure of further components becomes likely. The probability distribution of the number of failed components is a saturating quasibinomial distribution. The saturation extends the parameter range of the quasibinomial distribution and the saturated distribution can represent highly stressed systems with a high probability of all components failing. Explicit formulas for the saturating quasibinomial distribution are derived using a recursion and via the quasimultinomial distribution of the number of failures in each stage of the cascade. The application of the saturating quasibinomial distribution is illustrated by increasing average initial component load. At low load, the probability distribution of the number of failed components has an exponential tail. At a critical load the distribution has a power law region that indicates a substantial risk of a large cascading failure.
Keywords: Blackout; Electric power transmission system; Infrastructure; Power law; Quasibinomial distribution.

Ian Dobson is Professor, Electrical and Computer Engineering Department, University of Wisconsin, Madison WI 53706, (email: dobson@engr.wisc.edu); Benjamin A. Carreras is Corporate Fellow, Oak Ridge National Laboratory, Oak Ridge TN 37831 (email: carrerasba@ornl.gov); and David E. Newman is Associate Professor, Physics Department, University of Alaska, Fairbanks AK 99775 (email: ffden@uaf.edu). The work was coordinated by the Consortium for Electric Reliability Technology Solutions and funded in part by the Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Power Technologies, Transmission Reliability Program of the U.S. Department of Energy under contract 9908935 and Interagency Agreement DE-A1099EE35075 with the National Science Foundation. The work was funded in part by NSF grants ECS-0214369 and ECS-0216053. Part of this research has been carried out at Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract number DE-AC05-00OR22725.

1. INTRODUCTION

Cascading failure is the usual mechanism for large blackouts of electric power transmission systems. For example, long, intricate cascades of events caused the Western North American blackouts in July and August 1996 (NERC 2002). The August 1996 blackout disconnected 30,390 MW of power to 7.5 million customers (Kosterev, Taylor, and Mittelstadt 1999). An even more spectacular example is the August 2003 blackout of a portion of Northeastern North America that disconnected 61,800 MW of power to an area spanning eight states and two provinces and containing 50 million people (U.S.-Canada Power System Outage Task Force 2003). The vital importance of the electrical infrastructure to society motivates the construction and study of models that capture salient features of cascading failure.

Bulk electrical power transmission systems are complex networks of large numbers of components that interact in diverse ways. For example, most of America and Canada east of the Rocky mountains is supplied by a single network running at a shared supply frequency. This network includes thousands of generators, tens of thousands of transmission lines and network nodes and about one hundred control centers that monitor and control the network flows. The flow of power and some dynamical effects propagate on a continental scale. All the electrical components have limits on their currents and voltages. If these limits are exceeded, automatic protection devices or the system operators disconnect the component from the system. We regard the disconnected component as failed because it is not available to transmit power (in practice it will be reconnected later). Components can also fail in the sense of misoperation or damage due to aging, fire, weather, poor maintenance or incorrect design or operating settings. In any case, the failure causes a transient and causes the power flow in the component to be redistributed to other components according to circuit laws, and subsequently redistributed according to automatic and manual control actions. The transients and readjustments of the system can be local in effect or can involve components far away, so that a component disconnection or failure can effectively increase the loading of many other components throughout the network. In particular, the propagation of failures is not limited to adjacent network components. The interactions involved are diverse and include deviations in power flows, frequency, and voltage as well as operation or misoperation of protection devices, controls, operator procedures and monitoring and alarm systems. However, all the interactions between component failures tend to be stronger when components are highly loaded. For example, if a more highly loaded transmission line fails, it produces a larger transient, there is a larger amount of power to redistribute to other components, and failures in nearby protection devices are more likely. Moreover, if the overall system is more highly loaded, components have smaller margins so they can tolerate smaller increases in load before failure, the system nonlinearities and dynamical couplings increase, and the system operators have fewer options and more stress.

A typical large blackout has an initial disturbance or trigger events followed by a sequence of cascading events. Each event further weakens and stresses the system and makes subsequent events more likely. Examples of an initial disturbance are short circuits of transmission lines through untrimmed trees, protection device misoperation, and bad weather. The blackout events and interactions are often rare, unusual, or unanticipated because the likely and anticipated failures are already routinely accounted for in power system design and operation. The complexity is such that it can take months after a large blackout to sift through the records, establish the events occurring and reproduce with computer simulations and hindsight a causal sequence of events.

The historically high reliability of North American power transmission systems is largely due to estimating the transmission system capability and designing and operating the system with margins with respect to a chosen subset of likely and serious contingencies. The analysis is usually either deterministic analysis of estimated worst cases or Monte Carlo simulation of moderately detailed probabilistic models that capture steady state interactions (Billington and Allan 1996). Combinations of likely contingencies and some dependencies between events such as common mode or common cause are sometimes considered. The analyses address the first few likely failures rather than the propagation of many rare or unanticipated failures in a cascade leading to a large blackout.

In this paper we seek to describe some of the salient features of cascading failure in blackouts with an analytically tractable probabilistic model. The features that we abstract from the formidable complexities of large blackouts are the large but finite number of components, components that fail when their load exceeds a threshold, and the additional loading of components by an initial disturbance and the failure of other components. The initial overall system stress is represented by upper and lower bounds on a range of initial component loadings. The model neglects the length of times between events and the diversity of power system components and interactions. Of course, an analytically tractable model is necessarily much too simple to represent with realism all the aspects of cascading failure in blackouts; the objective is rather to help understand some global systems effects that arise in more detailed models and in blackouts. Initial versions of the model appear in Dobson, Chen, Thorp, Carreras, and Newman (2002) and Dobson, Carreras, and Newman (2003). While our main motivation is large blackouts, the model is sufficiently simple and general that it could be applied to cascading failure of other large interconnected infrastructures.

We briefly review some other approaches to cascading failure in power system blackouts. Carreras, Lynch, Dobson, and Newman (2002) represent cascading transmission line overloads and outages in a power system model using the DC load flow approximation and standard linear programming optimization of the generation dispatch. The model shows critical point behavior as load is increased and can show power tails similar to those observed in blackout data. Chen and Thorp (2002) model power system blackouts using the DC load flow approximation and standard linear programming optimization of the generation dispatch and represent in detail hidden failures of the protection system. The expected blackout size is obtained using importance sampling and it shows some indications of a critical point as loading is increased. Rios, Kirschen, Jawayeera, Nedic, and Allan (2002) evaluate expected blackout cost using Monte Carlo simulation of a power system model that represents the effects of cascading line overloads, hidden failures of the protection system, power system dynamic instabilities, and the operator responses to these phenomena. Ni, McCalley, Vittal, and Tayyib (2003) evaluate expected contingency severities based on real time predictions of the power system state to quantify the risk of operational conditions. The computations account for current and voltage limits, cascading line overloads, and voltage instability. Roy, Asavathiratham, Lesieutre, and Verghese (2001) construct randomly generated tree networks that abstractly represent influences between idealized components. Components can be failed or operational according to a Markov model that represent both internal component failure and repair processes and influences between components that cause failure propagation. The effects of the network degree and the inter-component influences on the failure size and duration are studied. Pepyne, Panayiotou, Cassandras, and Ho (2001) also use a Markov model for discrete state power system nodal components, but propagate failures along the transmission lines of a power systems network with a fixed probability. Numerical simulations study the effect of the propagation probability and maintenance policies that reduce the probability of hidden failures. The challenging problem of determining cascading failure due to dynamic transients in hybrid nonlinear differential equation models is addressed by DeMarco (2001) using Lyapunov methods applied to a smoothed model and by Parrilo et al. (1999) using Karhunen-Loeve and

Galerkin model reduction. Watts (2002) describes a general model of cascading failure in which failures propagate through the edges of a random network. Network nodes have a random threshold and fail when this threshold is exceeded by a sufficient fraction of failed nodes one edge away. Phase transitions causing large cascades can occur when the network becomes critically connected by having sufficient average degree or when a highly connected network has sufficiently low average degree so that the effect of a single failure is not swamped by a high connectivity to unfailed nodes. Lindley and Singpurwalla (2002) describe some foundations for causal and cascading failure in infrastructures and model cascading failure as an increase in a component failure rate within a time interval after another component fails.

The paper is organized as follows. Section 2 describes the model and its normalization. Section 3 derives the saturating quasibinomial distribution of the number of failures and shows how the saturation generalizes the quasibinomial distribution and extends its parameter range. Section 4 illustrates the use of the model in studying the effect of system loading on blackout risk.

2. DESCRIPTION OF MODEL

The model has n identical components with random initial loads. For each component the minimum initial load is $L^{\min}$ and the maximum initial load is $L^{\max}$. For j = 1, 2, ..., n, component j has initial load L_j that is a random variable uniformly distributed in $[L^{\min}, L^{\max}]$. $L_1, L_2, \cdots, L_n$ are independent.

Components fail when their load exceeds L^{fail}. When a component fails, a fixed and positive amount of load P is transferred to each of the components.

To start the cascade, an initial disturbance loads each component by an additional amount D. Some components may then fail depending on their initial loads L_j and the failure of each of these components will distribute an additional load P that can cause further failures in a cascade. The components become progressively more loaded as the cascade proceeds.

In particular, the model produces failures in stages $i = 0, 1, 2, ...$ according to the following algorithm, where M_i is the number of failures in stage i.

CASCADE Algorithm

0. All n components are initially unfailed and have initial loads $L_1, L_2, \cdots, L_n$ that are independent random variables uniformly distributed in $[L^{\min}, L^{\max}]$.

1. Add the initial disturbance D to the load of each component. Initialize the stage counter i to zero.

2. Test each unfailed component for failure: For $j = 1, ..., n$, if component j is unfailed and its load $> L^{\text{fail}}$ then component j fails. Suppose that M_i components fail in this step.

3. Increment the component loads according to the number of failures M_i: Add M_iP to the load of each component.

4. Increment i and go to step 2

The CASCADE algorithm has the property that if there are no failures in stage j so that $M_j = 0$, then $0 = M_j = M_{j+1} = ...$ so that there are no subsequent failures (in step two, M_j can be zero either because all the components have already failed, or because the loads of the unfailed components are less than L^{fail}). Since there are n components, it follows that $M_n = 0$ and that the outcome with the maximum number of stages with nonzero failures is $1 = M_0 = M_1 = ... = M_{n-1}$. We are most interested in the total number of failures $S = M_0 + M_1 + ... + M_{n-1}$.

When interpreting the model in an application, the load increment P need not correspond only to transfer of a physical load such as the power flow through a component. Many ways by which a component failure makes the failure of other components more likely can be thought of as increasing an abstract "load" on the other components until failure occurs when a threshold is reached.

It is useful to normalize the loads and model parameters so that the initial loads lie in $[0, 1]$ and $L^{\text{fail}} = 1$ while preserving the sequence of component failures and $M_0, M_1, ...$

First note that the sequence of component failures and $M_0, M_1, ...$ are unchanged by adding the same constant to the initial disturbance D and the failure load L^{fail}. In particular, choosing the constant to be $L^{\max} - L^{\text{fail}}$, the initial disturbance D is modified to $D + (L^{\max} - L^{\text{fail}})$ and the failure load L^{fail} is modified to $L^{\text{fail}} + (L^{\max} - L^{\text{fail}}) = L^{\max}$.

Then all the loads are shifted and scaled to yield normalized parameters. The normalized initial load on component j is $l_j = (L_j - L^{\min})/(L^{\max} - L^{\min})$ so that l_j is a random variable uniformly distributed on $[0, 1]$. The normalized minimum initial load is zero, and the normalized maximum initial load and the normalized failure load are both one. The normalized modified initial disturbance and the normalized load increase when a component fails are

$$d = \frac{D + L^{\max} - L^{\text{fail}}}{L^{\max} - L^{\min}}, \qquad p = \frac{P}{L^{\max} - L^{\min}}. \qquad (1)$$

An alternative way to describe the model follows. It is convenient to use the normalized parameters (1). Let $N(t)$ be the number of components with loads in $(1 - t, 1]$. If the n initial component loadings are regarded as n points in $[0, 1] \subset \mathbb{R}$, then $N(t)$ is the number of points greater than $1 - t$. Then $0 \leq N(t) \leq n$, the sample paths of N are nondecreasing, and $N(t)=0$ for $t\le 0$ and $N(t)=n$ for $t\ge 1$.

Let the number of components failed at or before stage j be $S_j=M_0+M_1+...+M_j$. Then, assuming $S_{-1}=0$, the CASCADE algorithm generates $S_0,S_1,...$ according to

$$S_j=N(d+S_{j-1}p),\quad j=0,1,...\ . \tag{2}$$

Then $0\le S_j\le n$, S_j is nondecreasing and $S_k=S_{k+1}$ implies that $S_j=S_{j+1}$ for $j\ge k$. The minimum such k is the maximum stage number in which failures occur and $S_{-1}<S_0<S_1<...<S_k=S_{k+1}=...$ and the total number of failures $S=S_k$. That is,

$$N(d+Sp)\,=\,S, \tag{3}$$

$$N(d+S_jp)\,>\,S_j,\quad -1\le j<k. \tag{4}$$

Moreover, for $j<k$ and $r=0,1,...,M_{j+1}-1$,

$$N(d{+}(S_j{+}r)p)\ge N(d{+}S_jp)=S_{j+1}=S_j{+}M_{j+1}>S_j{+}r. \tag{5}$$

Therefore $N(d+sp)>s$ for $s=0,1,...,S-1$, and this inequality and (3) allow the total number of failures to be characterized as

$$S=\min\{s\mid N(d+sp)=s,\quad s\in\{0,1,2,...\}\}. \tag{6}$$

If, at stage j, $d+S_jp>1$, we say that the model saturates. Saturation implies $S_{j+1}=n$. Saturation never occurs if d and p are small enough that $d+np<1$.

The model can also be formulated as a queue with a single server. Exactly n customers arrive during a given hour independently and uniformly. The server is available to serve these customers at time d after the start of the hour because of completing some other task. The customer service time is p. Then S is the number of customers that arrive during the first busy period. The queue saturates when the first busy period runs past the end of the hour. Charalambides (1990) and Takács (1967) analyze this queue in the nonsaturating case as described in section 3.3.

3. DISTRIBUTION OF NUMBER OF FAILURES S

The main result is that the distribution of the total number of component failures is

$$P[S=r]=\begin{cases}\binom{n}{r}\phi(d)(d+rp)^{r-1}(\phi(1-d-rp))^{n-r}, & \\ & r=0,1,...,n-1,\\ 1-\displaystyle\sum_{s=0}^{n-1}f(s,d,p,n), & r=n,\end{cases} \tag{7}$$

where $p\ge 0$ and the saturation function is

$$\phi(x)=\begin{cases}0 & ;x<0,\\ x & ;0\le x\le 1,\\ 1 & ;x>1.\end{cases} \tag{8}$$

It is convenient to assume that $0^0\equiv 1$ and $0/0\equiv 1$ when these expressions arise in any formula in this paper.

If $d\ge 0$ and $d+np\le 1$, then there is no saturation ($\phi(x)=x$) and (7) reduces to the quasibinomial distribution

$$P[S=r]=\binom{n}{r}d(d+rp)^{r-1}(1-d-rp)^{n-r}. \tag{9}$$

The quasibinomial distribution was introduced by Consul (1974) to model an urn problem in which a player makes strategic decisions. Burtin (1980) derives the distribution of the number of initially uninfected nodes that become infected in an inverse epidemic process in a random mapping. This distribution is quasibinomial with d the fraction of initially infected nodes and p the uniform random mapping probability. Islam, O'Shaughnessy, and Smith (1996) interpret d and p as primary and secondary infection probabilities and apply the quasibinomial distribution to data on the final size of influenza epidemics. Jaworski (1998) generalizes the derivation to a random mapping with a general fixed point probability.

The cascading failure model gives a new application and interpretation of the quasibinomial distribution. Moreover, the saturation in (7) extends the range of parameters of the quasibinomial distribution to allow $d+np>1$. Section 4 shows that this extended parameter range can describe regimes with a high probability of all components failing.

The next two subsections derive (7) from the CASCADE algorithm in two ways by means of a recursion and by means of the quasimultinomial joint distribution of $M_0,M_1,...,M_{n-1}$.

3.1 Recursion

It is convenient to show the dependence of the distribution of number of failures on the normalized parameters by writing $f(r,d,p,n)=P[S=r]$.

In the case of $n=0$ components,

$$f(0,d,p,0)\,=\,1. \tag{10}$$

According to the CASCADE algorithm, when the initial disturbance $d\le 0$, no components fail, and when $d\ge 1$, all n components fail. Then

$$f(r,d,p,n)=\left\{\begin{array}{cl}1-\phi(d) & ;r=0\\ 0 & ;0<r<n\\ \phi(d) & ;r=n\end{array}\right\},\quad \begin{array}{c}(d\le 0\text{ or }d\ge 1)\\ \text{and }n>0.\end{array} \tag{11}$$

We assume $n>0$ and $0<d<1$ for the rest of the subsection.

The initial disturbance d causes stage zero failure of the components that have initial load l in $(1-d,1]$. Therefore the probability of any component failing in stage zero is d and

$$P[M_0 = k] = \binom{n}{k} d^k (1-d)^{n-k}. \tag{12}$$

Suppose that $M_0 = k$ and consider the $n-k$ components that did not fail in stage zero. Since none of the $n-k$ components failed in stage zero, their initial loads l must lie in $[0, 1-d]$ and the distribution of their initial loads conditioned on not failing in stage zero is uniform in $[0, 1-d]$. In stage one, each of the $n-k$ components has had a load increase d from the initial disturbance and an additional load increase kp from the stage zero failure of k components. Therefore the equivalent total initial disturbance for each of the $n-k$ components is $D = kp + d$.

To summarize, assuming $M_0 = k$, the failure of the $n-k$ components in stage one is governed by the model with initial disturbance $D = kp+d$, load transfer $P = p$, $L^{\min} = 0$, $L^{\max} = 1-d$, $L^{\text{fail}} = 1$, and $n-k$ components. Normalizing the parameters using (1) yields that the failure of the $n-k$ components is governed by the model with normalized initial disturbance $kp/(1-d)$ and normalized load transfer $p/(1-d)$. That is,

$$P[S = r | M_0 = k] = f(r-k, \frac{kp}{1-d}, \frac{p}{1-d}, n-k). \tag{13}$$

Combining (12) and (13) yields the recursion

$$\begin{aligned} f(r,d,p,n) &= \sum_{k=0}^{r} P[S = r|M_0 = k]\, P[M_0 = k] \\ &= \sum_{k=0}^{r} \binom{n}{k} d^k (1-d)^{n-k} f(r-k, \frac{kp}{1-d}, \frac{p}{1-d}, n-k) \\ &\qquad ;\ 0 \le r \le n, \quad 0 < d < 1, \quad n > 0. \end{aligned} \tag{14}$$

Equations (10), (11) and (14) define $f(r,d,p,n) = P[S = r]$ for all $n \ge 0$ and $p \ge 0$. Equations (10) and (11) agree with (7). Moreover it is routine to prove in the Appendix that (7) satisfies recursion (14). Therefore (7) is the distribution of S in the CASCADE algorithm. Thus the recursion offers a simple way to derive the saturating quasibinomial distribution that avoids complicated algebra or combinatorics. It is also straightforward to use (10) and (14) to confirm by induction on n that (7) is a probability distribution.

3.2 A Quasimultinomial Distribution

This subsection shows that the joint distribution of $M_0, M_1, ..., M_{n-1}$ is quasimultinomial and hence derives (7). It is convenient throughout to assume $d \ge 0$, restrict m_0, m_1, ... to nonnegative integers, and to write $s_i = m_0 + m_1 + ... + m_i$ for $i = 0, 1, ...$ and $s_{-1} = 0$.

Let $\alpha_0 = \phi(d)$, $\beta_0 = 1$, and, for $i = 1, 2, ...$,

$$\begin{aligned} \alpha_i &= \phi\Big(\frac{m_{i-1}p}{1-d-s_{i-2}p}\Big), \\ \beta_i &= \phi(1-d-s_{i-2}p). \end{aligned}$$

The identity

$$\beta_i(1-\alpha_i) = \beta_{i+1} \quad , i = 0, 1, 2, ... \tag{15}$$

can be verified using $1 - \phi(x) = \phi(1-x)$ and $d \ge 0$ and considering all the cases.

In step two of stage zero in the CASCADE algorithm, the probability that the load increment of d causes one of the components to fail is $\alpha_0 = \phi(d)$ and the probability of m_0 failures in the n components is

$$P[M_0 = m_0] = \binom{n}{m_0} \alpha_0^{m_0} (1-\alpha_0)^{n-m_0}. \tag{16}$$

Consider the end of step two of stage $i \ge 1$ in the CASCADE algorithm. The failures that have occurred are $M_0 = m_0, M_1 = m_1, ..., M_i = m_i$ and there are $n - s_i$ unfailed components, but the component loads have not yet been incremented by $m_i p$ in the following step three.

Suppose that $d + s_{i-1}p < 1$. Then, conditioned on the $n - s_i$ components not yet having failed, the loads of the $n - s_i$ unfailed components are uniformly distributed in $[d + s_{i-1}p, 1]$. In the following step three, the probability that the load increment of $m_i p$ causes one of the unfailed components to fail is α_{i+1} and the probability of m_{i+1} failures in the $n - s_i$ unfailed components is

$$\begin{aligned} P[M_{i+1} = m_{i+1} | M_i = m_i, ..., M_0 = m_0] &= \binom{n-s_i}{m_{i+1}} \\ \alpha_{i+1}^{m_{i+1}} (1-\alpha_{i+1})^{n-s_{i+1}}, &\quad m_{i+1} = 0, 1, ..., n - s_i. \end{aligned} \tag{17}$$

Suppose that $d + s_{i-1}p \ge 1$. Then all the components must have failed on a previous step and $P[M_{i+1} = m_{i+1} | M_i = m_i, ..., M_0 = m_0] = 1$ for $m_{i+1} = 0$ and is zero otherwise. In this case, $\alpha_{i+1} = 0$ and (17) is verified.

We claim that for $s_i \le n$,

$$\begin{aligned} P[M_i = m_i, ..., M_0 = m_0] &= \frac{n!}{m_0!\, m_1! \dots m_i!\, (n-s_i)!} \\ &\quad (\alpha_0\beta_0)^{m_0} (\alpha_1\beta_1)^{m_1} \dots (\alpha_i\beta_i)^{m_i} \beta_{i+1}^{n-s_i}. \end{aligned} \tag{18}$$

Formula (18) is proved by induction on i. For $i = 0$, (18) reduces to (16). The inductive step is verified by multiplying (17) and (18) and using (15) to obtain $P[M_{i+1} = m_{i+1}, ..., M_0 = m_0]$ in the form of (18).

An expression equivalent to (18) obtained using (15) is

$$\begin{aligned} P[M_i = m_i, ..., M_0 = m_0] &= \frac{n!}{m_0!\, m_1! \dots m_i!\, (n-s_i)!} \\ &\quad (\beta_0 - \beta_1)^{m_0} (\beta_1 - \beta_2)^{m_1} \dots (\beta_i - \beta_{i+1})^{m_i} \beta_{i+1}^{n-s_i}. \end{aligned} \tag{19}$$

The CASCADE algorithm has the property that if there are no failures in stage j so that $M_j = 0$, then $0 = M_j = M_{j+1} = ...$ and there are no subsequent failures. This property is verified by (19) because $m_j = 0$ implies $\beta_{j+1} = \beta_{j+2}$ so that the factor $(\beta_{j+1} - \beta_{j+2})^{m_{j+1}} = 0^{m_{j+1}}$, which vanishes unless $m_{j+1} = 0$. Iterating this argument gives $0 = M_j = M_{j+1} = ...$ Since the maximum number of failures is n, the longest sequence of failures has n stages with $M_0 = M_1 = ... = M_{n-1} = 1$. It follows that $0 = M_n = M_{n+1} = ...$ and that the nontrivial part of the joint distribution is determined by M_0, M_1, ..., M_{n-1}. It also follows that $M_{n-1} = 0$ if there are less than n stages with failures.

Formula (19) can now be rewritten for $i = n-1$. Let I be the largest integer not exceeding n such that $1-d-s_{I-2}p > 0$. Then (19) becomes, for $s_{n-1} \le n$,

$$\begin{aligned} &P[M_{n-1} = m_{n-1}, ..., M_0 = m_0] \\ &\quad = \frac{n!}{m_0!\, m_1! \dots m_{n-1}!\, (n - s_{n-1})!} \\ &\qquad (\phi(d))^{m_0} (m_0 p)^{m_1} (m_1 p)^{m_2} ... (m_{I-2} p)^{m_{I-1}} \\ &\qquad (\phi(1 - d - s_{I-2} p))^{n - s_{I-1}} A(\mathbf{m}, I), \end{aligned} \tag{20}$$

where $A(\mathbf{m}, n) = 1$ and $A(\mathbf{m}, I) = 0^{m_{I+1}} \dots 0^{m_{n-1}} 0^{n-s_{n-1}}$ for $I < n$. It follows from the definition of $A(\mathbf{m}, I)$ that (20) vanishes for $I < n$ unless $0 = M_{I+1} = ... = M_{n-1}$ and $S = M_0 + ... + M_I = n$. (Although (20) was derived assuming $d \ge 0$, it also holds for $d < 0$. In particular, for $d < 0$, (20) implies $P[M_{n-1} = 0, ..., M_0 = 0] = 1$.)

Distribution (20) generalizes the quasibinomial distribution and is a form of quasimultinomial distribution. It is a different generalization of the quasibinomial distribution than the quasitrinomial distribution considered by Berg and Mutafchiev (1990) to describe numbers of nodes in central components of random mappings.

Suppose that $S = M_0 + ... + M_{n-1} = r < n$. Then $M_{n-1} = 0$ and $M_0 + ... + M_{n-2} = r - M_{n-1} = r$ and (20) vanishes unless $I = n$. Summing (20) over nonnegative integers $m_0, ..., m_{n-1}$ that sum to r yields

$$\begin{aligned} P[S = r] &= \sum_{s_{n-1} = r} \frac{n!}{m_0!\, m_1! \dots m_{n-1}!\, (n - r)!} \\ &\qquad (\phi(d))^{m_0} (m_0 p)^{m_1} \dots (m_{n-2} p)^{m_{n-1}} (\phi(1 - d - rp))^{n-r} \\ &= \binom{n}{r} (\phi(1 - d - rp))^{n-r} p^r \\ &\qquad \sum_{s_{n-1} = r} \frac{r!}{m_0!\, m_1! \dots m_{n-1}!} \left(\frac{\phi(d)}{p}\right)^{m_0} m_0^{m_1} \dots m_{n-2}^{m_{n-1}}, \end{aligned}$$

which reduces to (7) using a lemma by Katz (1955). (The context of Katz's lemma assumes $\phi(d)/p$ is a positive integer, but the generalization is immediate.)

3.3 Applying a Generalized Ballot Theorem

Charalambides (1990) explains how the quasibinomial distribution appears as a consequence of generalized ballot theorems in the theory of fluctuations of stochastic processes (Takács 1967). We summarize this approach and comment that it derives only the nonsaturating cases of (7).

We assume $0 < d < 1$. Consider p multiplied by the number of components $N(t)$ with loads in $(1-t, 1]$. For $0 \le t \le 1$, $pN(t)$ is a stochastic process with interchangeable increments whose sample functions are nondecreasing step functions with $pN(0) = 0$. According to (6), the first passage time of $t - pN(t)$ through d is $\min\{t \,|\, pN(t) = t - d\} = \min\{d + sp \,|\, N(d + sp) = s\} = d + Sp$. Then, according to Takács (1967, sec. 17, thm. 4),

$$P[d + Sp \le t] = \sum_{d \le y \le t} \frac{d}{y} P[pN(y) = y - d] \tag{21}$$

for $0 < d \le t \le 1$. That is,

$$\sum_{k=0}^{\lfloor (t-d)/p \rfloor} P[S = k] = \sum_{k=0}^{\lfloor (t-d)/p \rfloor} \frac{d}{d + kp} P[N(d + kp) = k]. \tag{22}$$

Setting $t = d + rp$ in (22) for $r = 0, 1, ..., \min\{n, (1 - d)/p\}$, differencing the resulting equations and using the binomial distribution of $N(t)$ for $0 \le t \le 1$ yields the nonsaturating cases of (7). However, the approach does not extend to the saturating cases because $pN(t)$ does not have interchangeable increments when $t > 1$.

3.4 Approximate Power Tail Exponent at a Critical Case

We describe standard approximations of the quasibinomial distribution that yield a power tail exponent at the critical case. For parameters satisfying $np + d \le 1$ (no saturation), the distribution of S is quasibinomial and can be approximated by letting $n \to \infty$, $p \to 0$ and $d \to 0$ in such a way that $\lambda = np$ and $\theta = nd$ are fixed to give the generalized (or Lagrangian) Poisson distribution (Consul 1988, 1989; Consul and Shoukri 1988)

$$P[S = r] \approx \theta (r\lambda + \theta)^{r-1} \frac{\exp(-r\lambda - \theta)}{r!}, \tag{23}$$

which is the distribution of the number of offspring in a Galton-Watson-Bienaymé branching process with the first generation produced by a Poisson distribution with parameter θ and subsequent generations produced by a Poisson distribution with parameter λ. The critical case for the branching process is $np = \lambda = 1$ and Otter (1949) proved that at criticality the distribution of the number of offspring has a power tail with exponent -1.5. Further implications for cascading failure of the branching process approximation are considered in Dobson, Carreras, and Newman (2004).

4. EFFECT OF LOADING

How much can an electric power transmission system be loaded before there is undue risk of cascading failure? This section discusses qualitative effects of loading on the distribution of blackout size and then applies the model to describe the effect of loading and illustrate its use.

4.1 Distribution of Blackout Size at Extremes of Loading

Consider cascading failure in a power transmission system in the impractically extreme cases of very low and very high loading. At very low loading near zero, any failures that occur have minimal impact on other components and these other components have large operating margins. Multiple failures are possible, but they are approximately independent so that the probability of multiple failures is approximately the product of the probabilities of each of the failures. Since the blackout size is roughly proportional to the number of failures, the probability distribution of blackout size will have an exponential tail. The probability distribution of blackout size is different if the power system were to be operated recklessly at a very high loading in which every component was close to its loading limit. Then any initial disturbance would necessarily cause a cascade of failures leading to total or near total blackout. It is clear that the probability distribution of blackout size must somehow change continuously from the exponential tail form to the certain total blackout form as loading increases from a very low to a very high loading. We are interested in the nature of the transition between these two extremes.

4.2 Effect of Loading in Model

This subsection describes one way to represent a load increase in the model and how this leads to a parameterization of the normalized model. Then the effect of the load increase on the distribution of the number of components failed is described.

For purposes of illustration the system has $n = 1000$ components. Suppose that the system is operated so that the initial component loadings vary from $L^{\min}$ to $L^{\max} = L^{\text{fail}} = 1$. Then the average initial component loading $L = (L^{\min}+1)/2$ may be increased by increasing $L^{\min}$. The initial disturbance $D = 0.0004$ is assumed to be the same as the load transfer amount $P = 0.0004$. These modeling choices for component load lead via the normalization (1) to the parameterization $p = d = 0.0004/(2 - 2L)$, $0.5 \leq L < 1$. The increase in the normalized power transfer p with increased L may be thought of as strengthening the component interactions that cause cascading failure.

For average initial load $L = 0.5$, $d = p = 0.0004$ and the probability distribution of the number S of components failed is close to the binomial distribution obtained for $d = 0.0004$ and $p = 0$ and the tail is exponential. The distribution as L increases from 0.5 is shown in Fig. 1. The distribution for the nonsaturating case $L = 0.6$ has a tail slightly heavier than binomial but still approximately exponential. The tail becomes heavier as L increases and the distribution for the critical case $L = 0.8$, $np = 1$ has an approximate power law region over a range of S. The power law region has an exponent of approximately -1.4 and this compares to the exponent of -1.5 obtained by the analytic approximation in subsection 3.4. The distribution for the saturated case $L = 0.9$ has a slightly heavier than exponential tail for small r, zero probability of intermediate r, and a probability of 0.80 of all 1000 components failing. If an intermediate number of components fail in a saturated case, then the cascade always proceeds to all 1000 components failing.

The increase in the mean number of failures ES as the average initial component loading L is increased is shown in Fig. 2. The sharp change in gradient at the critical loading $L = 0.8$ corresponds to the saturation of (7) and the consequent increasing probability of all components failing. Indeed, at $L = 0.8$, the change in gradient in Fig. 2 together with the power law region in the distribution of S in Fig. 1 suggest a type two phase transition in the system. If we interpret the number of components failed as corresponding to blackout size, the power law region is consistent with North American blackout data and blackout simulation results (Chen, Thorp, and Parashar 2001;Dobson et al. 2002; Carreras et al. 2002). In particular, North American blackout data suggest an empirical distribution of blackout size with a power tail with exponent between -1 and -2 (Carreras, Newman, Dobson, and Poole 2001; Chen et al. 2001). This power tail indicates a significant risk of large blackouts that is not present when the distribution of blackout sizes has an exponential tail (Carreras, Lynch, Newman, and Dobson 2003).

The model results show how system loading can influence the risk of cascading failure. At low loading there is an approximately exponential tail in the distribution of number of components failed and a low risk of large cascading failure. There is a critical loading at which there is a power law region in the distribution of number of components failed and a sharp increase in the gradient of the mean number of components failed. As loading is increased past the critical loading, the distribution of number of components failed saturates, there is an increasingly significant probability of all components failing, and a significant risk of large cascading failure.

APPENDIX: SATURATING QUASIBINOMIAL FORMULA SATISFIES RECURSION

We prove that the saturating quasibinomial formula (7) satisfies recursion (14) for $0 < d < 1$ and $n > 0$.

In the case $d + rp < 1$ and $r < n$, since

$$d + rp < 1 \iff \frac{kp}{1-d} + (r-k)\frac{p}{1-d} < 1, \qquad (24)$$

none of the instances of f in the right hand side of (14) saturate so that the right hand side of (14) becomes

$$\begin{aligned}
&\sum_{k=0}^{r} \binom{n}{k} d^k (1-d)^{n-k} \\
&\quad \binom{n-k}{r-k} \frac{kp}{1-d} \left(\frac{rp}{1-d}\right)^{r-k-1} \left(1 - \frac{rp}{1-d}\right)^{n-r} \\
&= \binom{n}{r} \sum_{k=0}^{r} \binom{r}{k} \frac{k}{r} d^k (rp)^{r-k} (1-d-rp)^{n-r} \\
&= \binom{n}{r} d(d+rp)^{r-1}(1-d-rp)^{n-r}.
\end{aligned}$$

In the case $d + rp \geq 1$ and $r < n$, (24) and $r - k < n - k$ imply that all the instances of f in the right hand side of (14) vanish.

In the case $r = n$, substituting the expression from (7) for $f(n-k, (kp)/(1-d), p/(1-d), n-k)$ into the right hand side of (14) leads to

$$\begin{aligned}
1 - \sum_{t=0}^{n-1} \sum_{k=0}^{t} \binom{n}{k} d^k (1-d)^{n-k} f(t-k, \frac{kp}{1-d}, \frac{p}{1-d}, n-k) \\
= 1 - \sum_{s=0}^{n-1} f(s, d, p, n),
\end{aligned}$$

where the last step uses the result established above that (7) satisfies (14) for $r < n$.

REFERENCES

Berg, S., and Mutafchiev, L. (1990), "Random Mappings With an Attracting Center: Lagrangian Distributions and a Regression Function," *Journal of Applied Probability*, 27, 622-636.

Billington, R., and Allan, R. N. (1996), *Reliability Evaluation of Power Systems*, (2nd ed.), Plenum Press, NY.

Burtin, Y. D. (1980), "On a Simple Formula for Random Mappings and its Applications," *Journal of Applied Probability*, 17, 403-414.

Carreras, B. A., Newman, D. E., Dobson, I., and Poole, A. B. (2001), "Evidence for Self-organized Criticality in Electric Power System Blackouts," *34th Hawaii International Conference on System Sciences*, Maui, Hawaii.

Carreras, B. A., Lynch, V. E., Dobson, I., and Newman, D. E. (2002), "Critical Points and Transitions in an Electric Power Transmission Model for Cascading Failure Blackouts," *Chaos*, 12, 4, 985-994.

Carreras, B. A., Lynch, V. E., Newman, D. E., and Dobson, I. (2003), "Blackout Mitigation Assessment in Power Transmission Systems," *36th Hawaii International Conference on System Sciences*, Hawaii.

Chen, J., Thorp, J. S., and Parashar, M. (2001), "Analysis of Electric Power Disturbance Data," *34th Hawaii International Conference on System Sciences*, Maui, Hawaii.

Chen, J., and Thorp, J. S. (2002), "A Reliability Study of Transmission System Protection via a Hidden Failure DC Load Flow Model," *IEE Fifth International Conference on Power System Management and Control*, 384-389.

Charalambides, Ch. A. (1990), "Abel Series Distributions With Applications to Fluctuations of Sample Functions of Stochastic Functions," *Communications in Statistics - Theory and Methods*, 19, 1, 317-335.

Consul, P. C. (1974), "A Simple Urn Model Dependent Upon Predetermined Strategy," *Sankhyā: The Indian Journal of Statistics, Series B*, 36, 4, 391-399.

—— (1988), "On Some Models Leading to a Generalized Poisson Distribution," *Communications in Statistics - Theory and methods*, 17, 2, 423-442.

—— (1989), *Generalized Poisson Distributions*, Dekker, NY.

Consul, P. C., and Shoukri, M. M. (1988), "Some Chance Mechanisms Leading to a Generalized Poisson Probability Model," *American Journal of Mathematical and Management Sciences*, 8, 1&2, 181-202.

DeMarco, C. L. (2001), "A Phase Transition Model for Cascading Network Failure," *IEEE Control Systems Magazine*, December, 40-51.

Dobson, I., Chen, J., Thorp, J. S., Carreras, B. A., and Newman, D. E. (2002), "Examining Criticality of Blackouts in Power System Models With Cascading Events," *35th Hawaii International Conference on System Sciences*, Hawaii.

Dobson, I., Carreras, B. A., and Newman, D. E. (2003), "A Probabilistic Loading-Dependent Model of Cascading Failure and Possible Implications for Blackouts," *36th Hawaii International Conference on System Sciences*, Hawaii.

—— (2004), "A Branching Process Approximation to Cascading Load-Dependent System Failure," *37th Hawaii International Conference on System Sciences*, Hawaii.

Islam, M. N., O'Shaughnessy, C. D., and Smith, B. (1996), "A Random Graph Model for the Final-Size Distribution of Household Infections," *Statistics in Medicine*, 15, 837-843.

Jaworski, J. (1998), "Predecessors in a Random Mapping," *Random Structures and Algorithms*, 14, 501-519.

Katz, L. (1955), "Probability of Indecomposability of a Random Mapping Function," *The Annals of Mathematical Statistics*, 26, 512-517.

Kosterev, D. N., Taylor, C. W., and Mittelstadt, W. A. (1999), "Model Validation for the August 10, 1996 WSCC System Outage," *IEEE Transactions on Power Systems*, 13, 3, 967-979.

Lindley, D. V., and Singpurwalla, N. D. (2002) "On Exchangeable, Causal and Cascading Failures," *Statistical Science*, 17, 2, 209-219.

NERC (2002), *1996 System Disturbances*, North American Electric Reliability Council, Princeton Forrestal Village, 116-390 Village Boulevard, Princeton, New Jersey 08540-5731.

Ni, M., McCalley, J. D., Vittal, V., Tayyib, T. (2003) "Online Risk-Based Security Assessment," *IEEE Transactions on Power Systems*, 18, 1, 258-265.

Otter, R. (1949), "The Multiplicative Process," *The Annals of Mathematical Statistics*, 20, 206-224.

Parrilo, P. A., Lall, S., Paganini, F., Verghese, G. C., Lesieutre, B. C., and Marsden, J. E. (1999), "Model Reduction for Analysis of Cascading Failures in Power Systems," *Proceedings of the American Control Conference*, 6, 4208-4212.

Pepyne, D. L., Panayiotou, C. G., Cassandras, C. G., and Ho, Y.-C. (2001), "Vulnerability Assessment and Allocation of Protection Resources in Power Systems," *Proceedings of the American Control Conference*, 6, 4705-4710.

Rios, M. A., Kirschen, D. S., Jawayeera, D., Nedic, D. P., and Allan, R. N. (2002), "Value of Security: Modeling Time-Dependent Phenomena and Weather Conditions," *IEEE Transactions on Power Systems*, 17, 3, 543-548.

Roy, S., Asavathiratham, C., Lesieutre, B. C., and Verghese, G. C. (2001), "Network Models: Growth, Dynamics, and Failure," *34th Hawaii International Conference on System Sciences*, 728-737.

Takács, L. (1967), *Combinatorial Methods in the Theory of Stochastic Processes*, Wiley, NY.

U.S.-Canada Power System Outage Task Force (2003), *Interim Report: Causes of the August 14th Blackout in the United States and Canada*, United States Department of Energy and National Resources Canada, November.

Watts, D. J. (2002), "A Simple Model of Global Cascades on Random Networks," *Proceedings of the National Academy of Sciences USA*, 99, 9, 5766-5771.

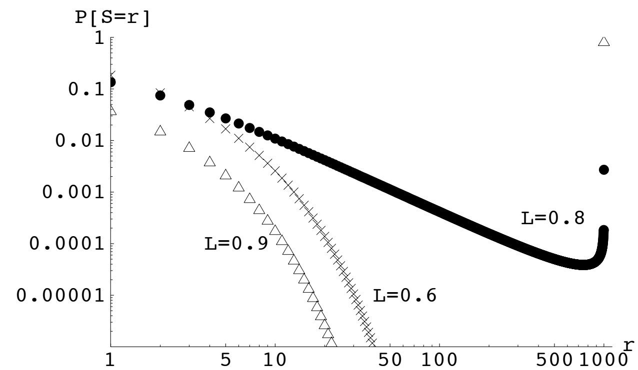

Figure 1: Log-log plot of distribution of number of components failed S for three values of average initial load L. Note the power law region for the critical loading $L = 0.8$. $L = 0.9$ has an isolated point at $(1000, 0.80)$ indicating probability 0.80 of all 1000 components failed. Probability of no failures is 0.61 for $L = 0.6$, 0.37 for $L = 0.8$, and 0.14 for $L = 0.9$.

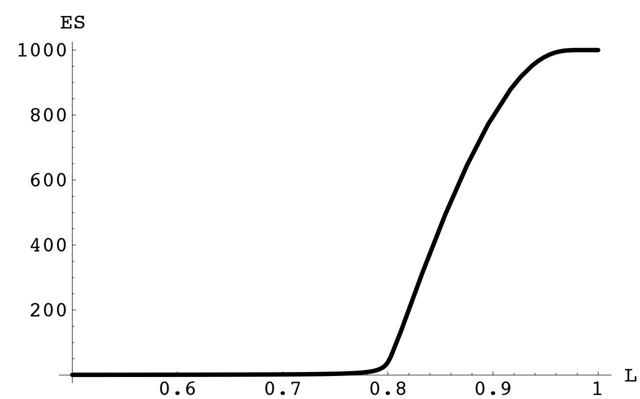

Figure 2: Mean number of components failed ES as a function of average initial component loading L. Note the change in gradient at the critical loading $L = 0.8$. There are $n = 1000$ components and ES becomes 1000 at the highest loadings.