1

On Storage and Renewables: A Theory of Sizing and Uncertainty

Rahul Singh, Student Member, IEEE, Le Xie, Member, IEEE, and P. R. Kumar, Fellow, IEEE

Abstract—This paper studies the fundamental problem of sizing energy storage given an uncertainty level of variable resources in a microgrid. A queuing theoretical model is introduced, which provides unique insights on the coupling between energy storage size and uncertainty level of the net load. The proposed model lends itself to 3 levels of details: a random walk model with single uncertainty from one net load, a reflected Brownian motion model with many more uncertain resources, and a model with a collection of Markov-type power producers and consumers. It is shown that the fundamental requirement of energy storage sizing can be approximately derived from the aforementioned 3 queuing theory models. Numerical examples suggest that this approach can be applied to microgrid planning and operation in assessing the optimal size of energy storage , as well as the potential curtailment of renewable energy.

I. INTRODUCTION

This paper is motivated by the increasing penetration of variable resources around the world. A fundamental question arises with more and more variable resources deployed: What is the amount of energy storage needed for high penetration of renewable power? We propose a theoretical framework that provides rigorous yet simple tools to address this question.

A historical comparison is in order. In the early 1900s, when telephone exchanges were being built across the country, there were several questions related to how large the exchanges had to be in terms of the number of lines, in order to ensure that the percentage of blocked calls was below a specified value, while carrying a specified volume of calls with certain holding times. In response to this challenge, queuing theory was invented by Erlang [1], [2].

A similar problem arises today in the context of the integration of clean energy and energy storage resources. We use the word "storage" in a rather broad context. It includes large batteries, large buildings with controllable thermal energy, as well as a large number of coordinated electric vehicles [3]–[6].

Storage can be used to mitigate the unreliability of such stochastic time variation of energy sources, but, depending on the demand, there will necessarily be a need for balancing power, likely from conventional generation (typically fossil fuel-based). In a future grid with objective to lower the carbon emission, it is important to characterize not only the mean power drawn from the fossil fuel, but also its variation, e.g., peak-to-mean ratio.

Thus, one would like to determine how the magnitude of the storage interacts with the stochastic temporal unreliability of renewables and their spectral content, in terms of determining what nature of demands can be supported, and what the resulting peak as well as mean need is for augmenting energy sources. This is the goal of the paper.

To elucidate how one may address these interrelated issues, we pose the problem in a simple yet fundamental mathematical model, and show what analytical tools can be brought to bear to address the nexus of these issues. The scope of this paper is limited. We attempt to only show how certain theoretical techniques that may be brought to bear to address these questions. Future extensions will, one hopes, obtain more useful results employing realistic models of the phenomena involved.

The rest of the paper is organized as follows. We provide a description of the system model in Section II, which is followed by a derivation of the system performance in Section III for the simple case when the energy delivery in a micro-grid follows a "random-walk" model. Section IV provides an exposition of the Brownian motion model, a model which is justified when one is dealing with a large number of power producers and consumers in the electricity market. Section V considers the case when we are dealing with more complex dynamics occurring at the level of producers and consumers. Illustrative examples are provided in Section VI. Section VI presents the concluding remarks.

II. SYSTEM MODEL

We consider the abstract model, shown in Figure 1, of a micro-grid which supplies its customers from a portfolio of renewable power, fossil fuel generation, and a storage device. The storage unit has an energy capacity of B. This is similar to a scenario of a micro-grid operator trying to schedule all the resources to a community of customers in the isolated operating mode.

Once the storage energy hits the level 0, the conventional generators are utilized to meet the power demand that exceeds renewable power supply since the operator has no more stored renewable energy to supply. Thus, at all times, the aggregate excess demand of the consumers

This material is based upon work partially supported by NSF under Contract Nos. CCF-1331863, ECCS-1150944, CPS-1239116 and Science & Technology Center Grant CCF-0939370.

R. Singh, L. Xie, and P. R. Kumar are with Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843. Email: rsingh1@tamu.edu, lxie@ece.tamu.edu, prk@tamu.edu.

is fulfilled either from renewable and storage, or from the conventional generation.

At the other extreme, whenever the storage reaches its maximum value of B, the supplier is forced to curtail any excess renewable energy being produced ("overflow of renewable energy"), thus leading to spill of some renewable energy.

With the above set-up in place, we will be interested in answering the following questions: What is the average amount of fossil power consumed? How does it change with the size of the storage? Does the amount of fossil power required also depend on the statistics of the renewable energy source ([7], [8]? How does it depend on the "spectral content" of the time variation of the renewable energy source (since high frequency variations are better buffered by the source rather than low frequency variations)? What parameters of the renewal energy supply process quantify this dependence? Similar questions are of interest concerning the quantity of renewable curtailment due to "overflows." As a finer measure of performance, one might also be interested in second-order moments of the amount of fossil energy utilized, since that is related to issues such as peak-to-average generation ratio, and the amount of renewable energy wasted.

Fundamentally it is the difference between two stochastic processes (renewable power and load) that is relevant to answer the above questions 1 Our goal in this paper is to show how these two stochastic processes interact with the storage capacity B to determine the answers to the above questions. To illustrate the methodology, we carry forward this analysis to obtain somewhat explicit back-of-the-envelope type answers for certain simple stochastic process models. Doubtless, it is important in practice to determine answers for models of higher fidelity, and this can be and we hope will be, pursued along the lines indicated in this paper, though at the expense of greater mathematical complexity.

In the next three sections, we provide illustratory answers to three models of stochastic time variation.

III. RANDOM WALK MODEL

We begin with the simplest stochastic process, a random walk, to model the renewable sources and loads. Let us denote the energy leve in the storage at time t by V(t), where the time parameter t assumes the discrete values $0,1,\ldots$. For simplicity, we suppose that the energy-levels of the storage are discretized, so that $V(t) \in \{0,1,2,\ldots,B\}$. An important quantity, in fact the only relevant quantity, is the net-put, which is the difference between the renewable power supply at a certain time minus the demand at that time. If the "net-put" to the storage at time t is X(t), then the value of storage at time t is given by

$$V(t+1) = (V(t) + X(t))^{+} \wedge B,$$

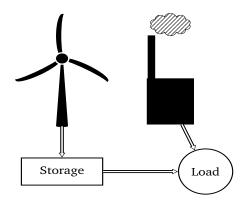


Fig. 1: Renewable and Fossil fuel energy consolidated into a microgrid.

where $a \wedge b := Min(a, b)$, and $a^+ := Max(a, 0)$. (Since we are in discrete time, the "energy" net-put in one time unit can also be called "power").

For simplicity, we begin by supposing that $X(0), X(1), \ldots$ are independent and identically distributed random variables, with a known distribution. Under these assumptions, V(t) is a Markov process evolving on a finite state space $\{0,1,\ldots,B\}$, and, under a mild irreducibility condition that we assume, has a unique stationary distribution. Let us denote by $V(\infty)$ the random variable having the stationary distribution of the process V(t).

Next, we define two stochastic processes that are relevant to the system performance. Let L(t) be the *loss* in the renewable energy at time t, i.e.,

$$L(t) = (V(t-1) + X(t-1))^{+} - V(t).$$

As its name implies, this is the renewable energy that is indeed curtailed because the storage is full. Also, denote by F(t) the fossil energy expended to meet the demands,

$$F(t):=V(t)-(V(t-1)+X(t-1))\wedge B.$$

Clearly,

$$V(t) = V(t-1) + X(t-1) + F(t) - L(t).$$
 (1)

The average energy wasted due to overflows is then simply,

$$\bar{L} := \mathrm{E}\left(V(\infty) + X\right)^{+} - \left[\left(V(\infty) + X\right)^{+} \wedge B\right]$$

where X is a random variable having the same distribution as $X(0), X(1), \ldots$ Using (1), under a mild aperiodicity assumption, in steady state the expected value of V(t) is the same as that of V(t+1), and so $\bar{L} = \bar{F} + \bar{X}$.

Similarly, the second moments of the steady state loss and fossil energy can be calculated once the stationary

¹In this paper we assume there is no line constraint.

distribution of the process V(t) is known. The following result from [9] is useful:

Theorem 1: Let $S(n):=\sum_{i=0}^n X(i)$, and $\tau[u,v):=\inf\{n\geq 0: S(n)\notin [u,v)\}$. Then

$$\mathbb{P}\left(V(\infty)\geq x\right)=\mathbb{P}\left(S\left(\tau\left(x-B,x\right)\right)\geq x\right).$$

To illustrate how we can determine the quantities of significant interest, let us consider a "simple random walk", where X assumes value 1 with a probability $p<\frac{1}{2}$ and -1 with probability $q=1-p>\frac{1}{2}$; a similar analysis can be carried out for more general models. The steady state distribution is

$$\mathbb{P}(V(\infty) = x) = \frac{1 - \rho}{1 - \rho^{B+1}} \rho^x$$
, for $x = 0, 1, \dots, B$,

where $\rho := \frac{p}{a}$. Thus the renewable-energy lost is given by,

$$\bar{L} = \frac{1 - \rho}{1 - \rho^{B+1}} \rho^B p \tag{2}$$

and its standard deviation is given by, $\sigma_L = \frac{1-\rho}{1-\rho^B+1}\rho^B p$. The formulas for loss and variability provide valuable

The formulas for loss and variability provide valuable insight. It follows from (2) that for a fixed value of ρ , the wastage suffered by the energy supplier decreases roughly exponentially as the size of the storage is increased. Since the storage capacity B comes at a cost, either fixed or operating, such an insight as that provided by (2) might be useful for windfarm operators faced with the issue of choosing the right location to place windfarms and the right size of storage. The location of the windfarm fixes the quantity ρ , and the storage-size corresponds to B. The ability to do such back-of-the-envelope calculations is potentially important for providing an understanding of sizing for storage.

IV. REFLECTED BROWNIAN MOTION MODEL

We next consider an alternative continuous-time model, where the "cumulative net-put process" (i.e. the net-put process of the previous section integrated over time) is a Brownian motion. One justification for such a model is as a limit of a sequence of markets operating with many renewable power producers, in which the n-th market has a number n of energy suppliers and consumers such that the mismatch between the total demand of the consumers and the energy supplied in a time-epoch is (after an appropriate scaling) of the order \sqrt{n} with a high probability. As the number of individual renewable resources, $n \to \infty$, it can be shown that the energy traded in the market converges weakly to the Brownian motion ([10]). The advantage of using such a Brownian approximation is that it allows us to stochastic calculus to obtain simple and elegant expressions for the relevant quantities. Such an approach has been successfully used in stochastic flow systems, and the analysis here follows the approach of modeling queueing systems by reflected Brownian motion [11]–[13].

As in the previous section, there is a single storage unit, which is "fed" by a Brownian motion. The system evolves

in continuous time, and the *cumulative net-put process* X(t) obtained by subtracting the total energy demand of the consumers till time t from the net renewable energy produced uptil time t, is a Brownian motion with a drift μ , and a variance σ .

Due to the storage capacity of B, the process denoting the storage level at time t, $V(t) \in [0, B]$, is a Brownian motion constrained between the barriers at levels 0 and B. We assume that the system starts at time t=0 with an initial storage level V(0)=x.

If L(t) and F(t) are the cumulative renewable energy wasted due to overflows, and the amount of fossil fuel energy used till time t, respectively, then,

- 1) $V(t) = X(t) + F(t) L(t), V(t) \in [0, B]$ for all $t \ge 0$.
- 2) F can increase only when V is 0, and L can increase only when V = B, i.e. the fossil fuel sources are turned on only to meet the excess demand when the energy level in the storage hits the level 0. Likewise, excess renewable energy is wasted only when the energy level in the storage is B.

Thus the average loss and the average fossil energy used are ([13]),

$$\bar{L} = \lim_{t \to \infty} \frac{L(t)}{t}, \bar{F} = \lim_{t \to \infty} \frac{F(t)}{t}.$$

The associated variances are given by

$$\sigma_L^2 = \lim_{t \to \infty} \frac{Var(L(t))}{t}, \sigma_F^2 = \lim_{t \to \infty} \frac{Var(F(t))}{t}.$$

To obtain the above quantities of interest, we decompose the paths of the process V(t) into i.i.d. cycles and use the resulting regenerative structure. Define the following,

$$T_0 := \inf\{t \ge 0 : V(t) = 0\}.$$

For n = 0, 1, 2, ..., and t > 0 define,

$$\begin{split} V_{n+1}^{\star}(t) &:= V(T_n+t), L_{n+1}^{\star}(t) := L(T_n+t) - L(T_n), \\ U_{n+1}^{\star}(t) &:= U(T_n+t) - U(T_n), \text{ where} \\ T_{n+1} &:= \text{ smallest } t > T_n \text{ such that} \\ V(t) &= 0 \text{ and } Z(s) = b \text{ for some } s \in (T_n,t) \,. \end{split}$$

We see that the regeneration times are $T_1, T_2, ...$, and letting $\tau = T_1 - T_0$, we have,

$$\bar{L} = \frac{\mathbb{E}_0 \left[L(\tau) \right]}{\mathbb{E}_0(\tau)}, \quad \bar{F} = \frac{\mathbb{E}_0 \left[F(\tau) \right]}{\mathbb{E}_0 \left(\tau \right)}.$$

Let us denote by $\pi(\cdot)$ the stationary distribution of V(t) (existence of which can be shown using renewal theory). Also assume that V(0)=0. Then if f is any real-valued twice continuously differentiable function, we have,

$$f(V(t)) = f(V(0)) + \sigma \int_0^t f'(V) dX$$
$$+ \int_0^t \Gamma f(V) ds + f'(0)F(t) - f'(B)L(t).$$

Let $t = \tau$ in the above, and note that $f(V(\tau)) = f(V(0)) =$ f(0). Taking the expectations of both sides, noting that $\mathbb{E}\int\limits_{0}^{s}f'(V)\mathrm{d}X=0$, and after performing some algebraic manipulations, we have,

$$\int_0^B \Gamma f(z) \pi(\mathrm{d}z) + f'(0)\bar{F} - f'(B)\bar{L} = 0.$$
 (3)

By proper choice of the functions f in equation (3), such as f(v) = v or $f(v) = v^2$, one can calculate relevant

quantities of interest, which leads us to, Theorem 2: If $\mu=0$, then $\bar{L}=\bar{F}=\frac{\sigma^2}{2B}$ and π is the uniform distribution on [0, B].

Otherwise, for $\mu < 0$, where the renewable energy is not sufficient to fully meet the loads, let $\theta = \frac{2\mu}{2}$. Then,

$$\bar{L} = \frac{\mu}{1 - \exp^{-\theta B}}, \bar{F} = \frac{\mu}{\exp^{\theta B} - 1}.$$
 (4)

$$SCV(L) = \begin{cases} \frac{2B}{3} & \text{if } \mu = 0, \\ \frac{2(1 - \exp(2\theta B)) + 4\theta B \exp(\theta B)}{-\theta (1 - \exp(\theta B))^2}, \end{cases}$$

where SCV(L) is the squared coefficient of variation, i.e. standard deviation divided by mean, of loss L.

The relation (4) is the limit of the relation (2), justifying the random-walk model as discussed in Section III as a possibly reasonable assumption for a market having a large number of players. Moreover, as earlier, we note that if $\mu < 0$ (i.e., on an average the renewable supply is lesser than the demand), then the renewable energy wastage decreases exponentially with the size of the storage B.

V. CORRELATED UNCERTAINTY BETWEEN LOADS AND RENEWABLES

The assumptions in the previous sections assume that the energy source and consumers behave in an independent and identically distributed manner over time. This is not exactly the practical case. Inspired by the stochastic fluid analysis, in this section we extend the analysis to allow for more detailed models of individual generators and loads ([14]–[18]). Suppose there are m sources and n consumers which are "coupled" through the storage. Thus the energy produced by the sources is used to fill the storage, and the consumers withdraw from it. This approach is potentially applicable for commercial and industrial loads.

To illustrate the approach, suppose that each source can be in one of two states, active or passive, and the time taken to transition from one state to the other is exponentially distributed. When a source is active, it produces energy at a constant rate c_1 , while no energy is produced when it is passive. The rates of transition from active to passive and vice versa are f_1 , r_1 respectively. Similarly each consumer transitions from active to passive at rate f_2 , and vice versa at a rate r_2 , and consumes energy at a constant rate c_2 units while in the active state. Such a system can be described by a Markov process with state (V,i,j), where V(t) is the storage level at time t, and market state (i(t), j(t)) where i and j are the numbers of active sources and consumers respectively, with $0 \le i \le n_1$ and $0 \le j \le n_2$; The processes describing the number of active sources/consumers are birth-death processes.

 $p_1(t;i) := P(i \text{ sources are active at time } t),$ $p_2(t;j) := P(j \text{ consumers are active at time } t)$, and $\mathbf{p}_i(t) := (p_2(t;0), p_2(t;1), \dots, p_2(t;n_i)),$

we have.

$$\frac{d}{dt}\mathbf{p}_i(t) = \mathbf{p}_i(t)\mathbf{M}_i,$$

where, M_i is the corresponding transition rate matrix. Let $p(t; i, j) = p_1(t; i)p_2(t; j),$

$$\mathbf{p}(t) := (\mathbf{p}(t; 0, 0), \mathbf{p}(t; 0, 1), \cdots, \mathbf{p}(t; n_1, n_2)),$$

(lexicographic ordering)

P(t, x; i, j) :=

 $P(\text{ storage level } \leq x, \text{ market state } = (i, j) \text{ at time } t),$

and $\mathbf{P}(t,x)$ the lexicographic arrangement $\{P(t, x; i, j)\}$. Then,

$$\frac{d}{dt}\mathbf{p}(t) = \mathbf{p}(t)\mathbf{M}, \text{ where } \mathbf{M} := \mathbf{M}_1 \otimes \mathbf{I}(n) + \mathbf{I}(m) \otimes \mathbf{M}_2,$$

and $\mathbf{I}(k)$ is the k+1 dimensional identity matrix and \otimes is the Kronecker product, and

$$\frac{\partial}{\partial t} \mathbf{P} + \frac{\partial}{\partial x} \mathbf{P} \mathbf{D} = \mathbf{P} \mathbf{M}, t \ge 0, 0 < x < B, \tag{5}$$

where the "drift matrix" D is given by, $\mathbf{D} := c_1 \mathbf{E}(n_1) \otimes$ $I(n_2) - c_2 I(n_1) \otimes E(n_2)$, with $E(n_2) = diag(0, 1, \dots, n_2)$. Letting π be the continuous steady state solution of the equation (5), we get,

$$\frac{d}{dx}\boldsymbol{\pi}(x)\mathbf{D} = \boldsymbol{\pi}(x)\mathbf{M}, \quad 0 \le x \le B.$$
 (6)

That is to say for any $x \in (0, B)$, P(storage content \leq x and market state is(i,j) = $\pi(x;i,j)$. Spectral expansion gives us, $\pi(x) = \sum_{l} a_{l} \exp(z_{l}x) \phi(l)$, where $\{z_{l}, \phi(l)\}$ are solutions of the eigenvalue problem, $z\phi D = \phi M$, with a_l to be determined by the boundary conditions.

A. Performance Analysis

Let $w_1(i), w_2(j), 0 \leq i \leq n_1, 0 \leq j \leq n_2$ be the stationary probabilities that i sources and j consumers are active, and let \mathbf{w}_k for k = 1, 2 be the vectors comprising these probabilities. $w_k(i)$'s can be easily solved to obtain the stationary distribution of the market process: w:= $\mathbf{w}_1 \otimes \mathbf{w}_2$. The total average energy produced is then simply $c_1 \sum_{i=0}^m iw_1(i)$, while the average demand is $c_2 \sum_{j=0}^n jw_2(j) = \frac{nc_2r_2}{f_2+r_2}$. Since the storage levels will hit the boundaries

at 0 and B, it is clear that the energy utilized

will be lesser than both of the above quantities. Thus the renewable energy lost due to the limitation on the size of the storage is simply the sum $\sum P(\text{storage is full and market state is }(i,j)) (c_1i-c_2j)$ over all the states in which the loss-rate (c_1i-c_2j) is positive.

VI. NUMERICAL SIMULATIONS

We used the Structured Markov Chain (SMC) solver Matlab Toolbox, which can be found at [19] for our numerical analysis. The wind power is assumed to be lower cost than the conventional generation. The energy storage round trip efficiency is assumed to be perfect. The wind and load data is scaled down from a representative profile in the Electric Reliability Council of Texas. Figure 2 shows the loss as a function of energy storage size when the energy produced is roughly equal to the demand, with the average load being 26 MWh. The variations of the loads are shown in the three different colors in Figure 2. We see that when the demand rate is roughly the same as renewable-generation rate, the energy curtailed indeed decreases exponentially in the size of the storage. This is in further agreement with the expression derived in (4).

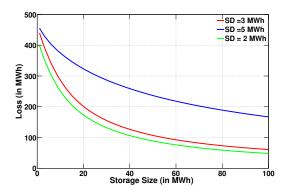


Fig. 2: Renewable energy curtailment as a function of storage size and load variability.

VII. CONCLUSIONS

This paper addresses the problem of energy storage sizing in a microgrid setting with high penetration of intermittent resources such as wind and solar. By considering the energy storage as a service provider, we propose a queueing theoretical approach to studying the fundamental coupling between energy storage sizing and uncertainty levels from the net load. Three models with different levels of details provide a suite of tools for microgrid operators to determine an optimal size of energy storage given a level of renewable and load uncertainties. Numerical examples based on realistic wind and load data suggest that the proposed approach could be a new avenue of research for optimal sizing of energy storage in renewable-rich power systems.

While this is a first step toward systematically understanding the fundamental role of uncertainty on sizing of energy storage. There are many fruitful directions for future research. One possible direction is to analyze the impact of line constraints on the optimal location of energy storage. Also it would be worthwhile to assess the effectiveness of this framework in a larger-scale realistic system.

REFERENCES

- [1] A. Erlang, "Probability and telephone calls," *Nyt Tidsskrift Mat. Ser. B*, vol. 20, pp. 33–39, 1909.
- [2] A. J. E. Brockmeyer, HL Halstrm and A. K. Erlang, "The life and works of AK Erlang." 1948.
- [3] D. Lindley, "Smart grids: The energy storage problem," *Nature*, pp. 18–20, 2010.
- [4] A. Thatte and L. Xie, "Towards a unified operational value index of energy storage in smart grid environment," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1418–1426, Sept 2012.
- [5] C. Guille and G. Gross, "A conceptual framework for the vehicle-to-grid (v2g) implementation," *Energy Policy*, vol. 37, no. 11, pp. 4379 4390, 2009.
- [6] F. Rahimi and A. Ipakchi, "Demand response as a market resource under the smart grid paradigm," *IEEE Transactions on Smart Grid*, vol. 1, no. 1, pp. 82–88, June 2010.
- [7] J. Apt, "The spectrum of power from wind turbines," *Journal of Power Sources*, vol. 169, no. 2, pp. 369–374, 2007.
- [8] A. E. Curtright and J. Apt, "The character of power output from utility-scale photovoltaic systems," *Progress in Photovoltaics: Research and Applications*, vol. 16, no. 3, pp. 241–247, 2008.
- [9] S. Asmussen, Applied Probability and Queues. Wiley, 1987.
- [10] P. Billingsley., Convergence of probability measures, ser. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., 1999.
- [11] Berger, Arthur W. and Whitt, Ward, "The Brownian approximation for rate-control throttles and the G/G/1/C queue," *Discrete Event Dynamic Systems*, vol. 2, no. 1, pp. 7–60, 1992.
- [12] Williams, R. J., "Asymptotic variance parameters for the boundary local times of reflected brownian motion on a compact interval," *Journal of Applied Probability*, vol. 29, no. 4, pp. pp. 996–1002, 1992.
- [13] Harrison, J. Michael, Brownian motion and stochastic flow systems, ser. Wiley series in probability and mathematical statistics. New York: Wiley, 1985.
- [14] Anick, D. and Mitra, D. and Sondhi, M. M., "Stochastic theory of a data-handling system with multiple sources," *Bell System Technical Journal*, vol. 61, no. 8, pp. 1871–1894, 1982.
- [15] Mitra, Debasis, "Stochastic theory of a fluid model of producers and consumers coupled by a buffer," *Advances in Applied Probability*, vol. 20, no. 3, pp. pp. 646–676, 1988.
- [16] Isi Mitrani and Ram Chakka, "Spectral expansion solution for a class of markov models: application and comparison with the matrix-geometric method," *Performance Evaluation*, vol. 23, no. 3, pp. 241 260, 1995.
- [17] L. C. G. Rogers, "Fluid models in queueing theory and wiener-hopf factorization of markov chains," *The Annals of Applied Probability*, vol. 4, no. 2, pp. 390–413, 1994.
- [18] Marcel F. Neuts, Matrix-geometric solutions in stochastic models an algorithmic approach. Dover Publications, 1994.
- [19] Benny Van Houdt, http://win.ua.ac.be/~vanhoudt/.