Survival Analysis Approach to Reliability, Survivability and Prognostics and Health Management (PHM)

Zhanshan (Sam) Ma
sam@cs.uidaho.edu

Axel W. Krings
krings@cs.uidaho.edu

Computer Science Department,
University of Idaho
Moscow, ID 83844, USA

Abstract—Survival analysis, also known as failure time analysis or time-to-event analysis, is one of the most significant advancements of mathematical statistics in the last quarter of the 20th century. It has become the *de facto* standard in biomedical data analysis. Although reliability was conceived as a major application field by the mathematicians who pioneered survival analysis, survival analysis failed to establish itself as a major tool for reliability analysis. In this paper, we attempt to demonstrate, by reviewing and comparing the major mathematical models of both fields, that survival analysis and reliability theory essentially address the same mathematical problems. Therefore, survival analysis should become a major mathematical tool for reliability analysis and related fields such as Prognostics and Health Management (PHM). This paper is the first in a four part series in which we review state-of-the-art studies in *survival (univariate) analysis*, *competing risks analysis*, and *multivariate survival analysis*, with focusing on their applications to reliability and computer science. The present article discusses the *univariate* survival analysis (survival analysis hereafter).

INDEX TERMS: Survival Analysis, Reliability, Network Survivability, Prognostic and Health Management (PHM), Software Reliability.

TABLE OF CONTENTS[1][2]

1. INTRODUCTION

The modern mathematical theory of reliability was established more than half a century ago, predominantly based on probability theory (e.g., Bazovsky 1961). Indeed, the probability-based reliability theory has been considered a very successful field in applied mathematics. An example that shows the importance of reliability research in the mathematical literature is a survey conducted by Aven and Jensen (1999); they found that 1% of the literature (then indexed by Zentralblatt/ Mathematical Abstracts and Mathematical Reviews) are associated with the keyword *reliability*. In practice, the success of reliability engineering is obvious, embedded in every engineering endeavor performed by human being, ranging from the Apollo Moon-landing project, commercial airplanes, to consumer electronics. However, like many scientific theories, reliability theory is far from perfect. Computer networks and software, in particular, create serious challenges for traditional reliability theory. The challenges have been well recognized (e.g., Krings 2008, Munson 2003, Shooman 2002, Xie 1991). For example, in software engineering, the notion of failure time, which is the core of any traditional reliability model, often becomes less relevant, since most failures are latent until the software features which cause the failure are triggered. An even more obvious difference is that software does not *wear out*; nevertheless, the outdated-ness caused by the environment updates (such as Operating Systems) can cause massive failures or even render the software useless under new environments. Similarly, computer viruses or other security breaches render traditional reliability theory of limited value. In emerging technologies such as wireless sensors and ad hoc networks, spatial topology or organizational geometry may not be ignorable due to the limited signal range. All these challenges point to the same conclusion, there is an urgent need to develop new approaches for modeling reliability and survivability for computer software and networks.

The problems with software and computer networking are so serious that the paradigm of survivable network systems (SNS) was proposed as a new discipline in the late 1990s (e.g., Ellison et al. 1997, Krings and Ma 2006, Krings 2008). Survivability can conceptually be considered as a system's capability to endure catastrophic failures, such as a network system under malicious intrusions, but still preserve mission critical functionalities. To some extent, reliability is the foundation of security and survivability. A survivable system generally has to be reliable, and an insecure and/or unreliable system generally is not survivable. What makes survivable network system (SNS) or survivability so important is the fact that today's computer networks control critical national infrastructures.

An obvious point is that it is desirable to develop a theory of survivability that can be put into a unified framework with reliability. An intuitive idea could be to define both reliability and survivability in a single unified probability space with some mechanism to differentiate the malicious events over which the probability measures may be unknown. The difficulty lies in the reality that most events associated with survivability are generally unpredictable due to the nature of malicious intrusions. In addition, it appears that a probability definition for survivability alone, similar to that for reliability, may be equally infeasible, since mathematically the malicious intrusions correspond to the event points where *probability measures* may not exist. Indeed, despite its critical importance and significant efforts, there is not a well-accepted mathematical definition for survivability. One of the objectives of this article is to propose treating unpredictable events such as malicious intrusions as censoring in a survival analysis model; in this way, survivability can be assessed with a survival (survivor) function, which has the exact same definition as tradition reliability. We argue that this *unorthodox* use of censoring to model survivability with survival analysis should be feasible at least for some computer networks, such as wireless sensor networks, given the essential similarity between a *population of patients* in a clinical trial and a *population of sensor* nodes in a wireless sensor network.

1.1. Important Definitions in Reliability Theory

Let us recall some of the most essential definitions of reliability theory. Assume we are concerned with a device that fails at an unforeseen or unpredictable random time (or age), $T > 0$, with distribution function $F(t)$

$$F(t) = P(T \le t), t \in R^+ \tag{1}$$

and probability density function (pdf) $f(t)$.

The reliability $R(t)$ is defined as

$$R(t) = 1 - F(t) \tag{2}$$

This has the exactly same definition as the survival function in survival analysis.

The failure rate $\lambda(t)$ is defined as the ratio of pdf to reliability,

$$\lambda(t) = f(t) / R(t). \tag{3}$$

Failure rate $\lambda(t)$ measures inclination to failure at time t given $\lambda(t)\Delta t \approx P(T \le t + \Delta t \mid T > t)$ for all Δt. It is also called as *instantaneous failure rate*, *hazard rate*, *force of mortality*, *intensity rate*, or *conditional failure rate*. Obviously $\lambda(t)\Delta t$ is the conditional probability that a device surviving to time t will fail in the interval $(t, t + \Delta t]$.

The cumulative hazard rate (function), simply called hazard function, $H(t)$ is defined as,

$$H(t) = \int_0^t \lambda(s)ds = -\ln[R(t)] \tag{4}$$

This derives

$$R(t) = \exp[-H(t)] = \exp[-\int_0^t \lambda(s)ds]. \tag{5}$$

The reliability modeling is then mainly concerned with collect information about the state of the system being investigated, and one should be alerted that different information level might lead to different reliability models (Aven and Jensen 1991).

1.2. Important Modeling Strategies

The following is a brief description of important modeling approaches for reliability, mainly based on Aven and Jensen 1999).

1.2.1. Compounded Systems

Compounded Systems is what Aven and Jensen (1991) called complex systems. It is made of n components with positive random lifetimes $T_i, i = 1, 2, ..., n, n \in N$. Let $\Phi : \{0,1\}^n \to \{0,1\}$ be the *structure function* of the system, e.g., a parallel or series structure of n components. The possible states of a component can be *intact* and *failed*, which is indicated by the number *1* or *0*, respectively. Let $\Phi(t) = \Phi(X_t)$ denote the state of the system at time t, where $X(t) = (X_t(1), X_t(2), ..., X_t(n))$ and $X_t(i)$ denotes the indicator function for component i at time t, i.e., $X_t(i) = 1$, if $T_i > t$ and $X_t(i) = 0$ otherwise. The system lifetime is then given by,

$$T = \inf\{t \in R^+ : \Phi(t) = 0\}. \tag{6}$$

For example, in simple series or parallel systems, Φ may take the $\wedge$ (*min*) or $\vee$ (*max*).

Under the assumption of independent component failures, the random variables T_i are *i.i.d.* (independent and identically distributed). There are standard problems such as

(Aven and Jensen. 1999): (1) inferring the system lifetime distribution from the lifetime distributions of its components; (2) assessing the effects of component reliability on the whole system; (3) finding if certain properties of the component lifetime distribution, such as IFR (increasing failure rate), are lost by forming a *monotone* system. A *monotone system* is characterized by the property that its structure function $\Phi(t)$ is non-decreasing with each argument, which guarantees that system reliability never gets hurt by improving the reliability of a component (Aven and Jensen. 1999).

1.2.2. Damage Models.

Another reliability formulation can be based on the *damage* or state observation of the system at time t. Let $X(t)$ be the damage or state random variable at time t and let S be the level of damage that leads to failure. Then the system lifetime can be defined as

$$T = \inf\{t \in R^+ : X(t) \geq S\} \tag{7}$$

where S can be a constant or a random variable independent of the damage process. For example, a Wiener process with positive drift starting at *0* and the failure threshold of constant S may be used to describe the damage process, and the resulting system lifetime follows the inverse Gauss Distribution (Aven and Jensen. 1999).

1.2.3. Compound Point Process.

This category of model describes the so-called *shock* process where the *shocks* cause random amounts of damage to the system (Aven and Jensen. 1999). The consecutive occurrence times of shocks form an increasing sequence $0 < T_1 < T_2... < \infty$ of random variables. Each time point T_n is assigned a real-valued random *mark* V_n, which is the additional damage caused by the *n*-th shock. Let the marked point process be $(T,V) = (T_n, V_n)$, $n \in N$. The resulting *compound point process* (X) is defined as

$$X(t) = \sum_{n=1}^{\infty} I(T_n \leq t)V_n \tag{8}$$

where $I(T_n)$ is the *indicator* function for T_n. Note that $X(t)$ represents the accumulated damage up to time t. The simplest form for $X(t)$ is a compound Poisson process, in which the shock arrival time follows Poisson distribution and the shock amounts (V_n) are *i.i.d.* random variables (Aven and Jensen. 1999). The system lifetime T is the first hit time when the damage process $X(t)$ reaches level S. The level S may be constant or be a random variable. A random failure level S is adopted when the observed damage process does not carry deterministic information about the system failure state (Aven and Jensen. 1999).

The above process is elegantly described by the *martingale* theory as exemplified in Aven and Jensen 1999. A *martingale* is the mathematical model for a fair game with constant expectation function equal to zero across the time domain. It provides an effective and flexible mathematical tool for studying lifetime random processes, which is often properly described by a point process. A point process can be described by an increasing sequence of random variables, by a purely atomic random measure, or via corresponding counting process. The Poisson process is one of the simplest point processes with a semi-martingale representation. Markov chains generally have smooth semi-martingale (SSM) representation. A very desirable property of a stochastic process with SSM representation is that it can be decomposed into a *drift* or regression part and an *additive fluctuation* described by a martingale (Aven and Jensen 1999). The latter can have an expectation of zero across time domain. With respect to survival analysis, the rigorous mathematical theory of survival analysis can be derived from the *counting process* and its martingale models. This again demonstrate that both reliability and survival analysis address the same or very similar mathematical problems.

1.2.4. A General Failure Model.

Aven and Jensen (1999) summarized a general failure model based on modern stochastic and Martingale theory. Several terms need to be defined first.

Path set: A path set is a set of components whose functioning guarantees the functioning of the system.

Smooth semimartingale (SSM): A stochastic process $Z = Z(t)$ is termed SSM if it has a decomposition of the form

$$Z(t) = Z_0 + \int_0^t f_s d(s) + M_t \tag{9}$$

Filtration: Let (Ω, F, P) be the probability space, where Ω is the sample space, F is the $\sigma - $ algebra of measurable sets, and P is the probability measure. The information up to time t is expressed by the pre-t-history F_t, which consists of all events in F that can be distinguished up to and including time t. The filtration $\mathbf{F} = F(t), t \in R^+$, is the family of increasing pre-t-histories that follow the standard conditions of completeness and right continuity.

Let $Z(t) = I(T \leq t)$ denote the simple indicator process and T be the lifetime random variable defined on the basic probability space (Ω, F, P). Z is the counting process corresponding to the simple point process (T_n) with $T = T_1$ and $T_n = \infty$ for $n \geq 2$. The *paths* of this indicator process Z are constant, but one *jump* from *0* to *1* at *T*. Assume that this indicator process has a smooth **F**-semimartingale representation with an **F**-martingale $M \in M_0$ and a nonnegative stochastic process, $\lambda = \lambda(t)$:

$$I(T \le t) = \int_0^t I(T > s)\lambda(s)ds + M_t, t \in R^+. \tag{10}$$

The filtration **F** and the corresponding **F**-SSM representation of the indicator process then define a general lifetime model.

The process $\lambda = \lambda(t)$, $t \in R^+$ in the above SSM-representation is termed the **F**-*failure rate* (or **F**-*hazard rate process*).

The *compensator* is defined as,

$$H(t) = \int_0^t I(T > s)\lambda(s)ds \tag{11}$$

which is also called **F**-*hazard process*.

A key point in the above general model is that the failure rate $\lambda(t)$ is interpreted as the limit of the conditional expectation with respect to the pre-*t*-histories **F** $= F(t), t \in R^+$,

$$\lambda(t) = \lim_{h \to 0+} \{P(T \le t + h \mid F_t)/h\}. \tag{12}$$

Aven and Jensen (1999) demonstrated that this general failure model could conveniently derive all above models and resolve issues such as Simpson's paradox. They indicated that the main advantage of the semi-martingale representation is the random evolution of a stochastic process on different *information levels*. They warned that the monotonicity properties in general are not preserved when changing the observation or information level.

1.3. Does reliability exists independently of mathematical modeling?

Reliability, intuitively, should be the property of a physical system and exist independent of the mathematical theory adopted to measure it. Although it is beyond the scope of this paper to argue whether or not mathematical objects are discovered or created, the answer to the question seems elusive even if we choose to ignore the philosophical argument. Let us use some examples to make the points.

Assume that there is a well-accepted definition for reliability such as described above. However, depending on the information available to evaluate system reliability, the results can be different. *Simpson's paradox* is a revealing example. As Aven and Jensen (1999) demonstrated that Simpson's paradox may occur in parallel systems in some parameter ranges, but not in series systems. They revealed that this has to do with the fact that the failure time of a parallel system consisting of components with constant failure rates does not follow exponential distribution and the system failure rate is non-monotonic. In the case of series construction, however, the system lifetime still follows the exponential distribution. This problem is called "*change of information level*" in Aven and Jensen (1999). Unfortunately, monotonicity is not preserved when changing the observation or information level. Therefore, even with the well-defined probability theoretic definition, the measurement of reliability is strongly influenced by the *information* collected, and by the *assumptions* made about the components. This implies that the mathematically correct models supported with rational assumptions at particular levels may not be sufficient to determine the system level reliability (Aven and Jensen 1999).

1.4. Who studies reliability?

Is there a difference between reliability theory and reliability engineering? Researchers involved in reliability theory come largely from three groups. The first is pure mathematicians who are mostly interested in mathematical problems such as reliability on graphs, reliability polynomial, and connectivity of abstract electric networks (which leads to *random cluster model*, a field that integrates probability, graph theory and stochastic geometry) (Grimmett 2006). Reliability is an important topic in percolation theory, which started with the studies of disordered media (Grimmett 1999). Similarly, reliability has been studied in random graphs (e.g., Bollobás 2001, Graver and Sobel 2005). Overall, the reliability studied by pure mathematicians is often related to connectivity of graph models of networks. Although the reliability models of this type are rarely directly applicable in reliability engineering, their studies often provide deep insights to the second group of researchers, who are applied mathematicians and statisticians. They are responsible for much of the formal theory like what was described above. The third group refers to reliability engineers whose engineering designs require modeling of specific engineered systems, which demands both the deep understanding of the physical system involved as well as the reliability theory (Aven and Jensen (1999). As expected, there are no clear boundaries among the three groups. Two excellent monographs, among many others, show the cross-boundary examples: Aven and Jensen (1999) may belong to the first and second groups, while Meeker and Escobar (1998) may belong to the second and third groups.

The probabilistic nature and the mission criticality of applications, such as heart pacemaker, space shuttle, commercial airplane, and large scale civil architectures such as bridges and dams, make reliability engineering among the most challenging and demanding profession. Understandably, the high dependability demands the extreme caution when transferring theory into practice. That may explain apparently occasional disconnection between reliability theory and practice, since reliability engineering may be hesitant to adopt untested reliability theory due to the huge stakes involved. This paper is an attempt to bring survival analysis or failure time analysis, largely done by mathematicians (especially bio-mathematicians and bio-statisticians) and widely adopted as

4

the *de facto* standard in biomedicine fields to the attention of engineering reliability research.

In this introduction, we briefly introduced the essential models in reliability theory. The discussion of survival analysis actually starts in the next section. However, it is *necessary* to note two points. (1) The major definitions in subsection 1.1 in reliability theory [equations (1)–(5)] are mathematically the same as their counterparts in survival analysis. The only difference is that *reliability R(t)* is equivalent to *survivor function S(t)*. Therefore, we will not repeat these essential definitions, but they form the foundational definitions in survival analysis too. (2) The modeling strategy, especially the general failure model of Aven and Jensen (1999), represents the state-of-the-art of theoretical modeling in reliability theory, which is based on measure theoretic stochastic processes. This approach is very similar to the formal theory of survival analysis, which is based on the *counting stochastic process* and *Martingale central limit theorem*. We will have a brief introduction on this in section 2.6.

The remaining of this paper presents a review of the essential elements of survival analysis (Section 2) and advantages of survival analysis over tradition reliability theory. In Section 3, we briefly survey and analyze the status of survival analysis applications in IEEE related engineering and computer science fields. Specifically, we analyze why survival analysis has been frequently associated with artificial neuronal networks (ANN) in engineering and computer science applications and its relationships with ANN. The final section is a brief summary and perspective for its further applications in reliability engineering, survivability modeling, and computer science. It should be noted that in this paper we limit the discussion of survival analysis to *univariate* survival analysis. We postpone the discussion of competing risks analysis and multivariate survival analysis to Ma and Krings (2008a, b & c).

2. ESSENTIALS OF SURVIVAL ANALYSIS THEORY

2.1. A Brief History.

Survival analysis (SA), or failure time analysis, is a specialized field of mathematical statistics, developed to study a special type of random variable of positive values with censored observations, of which failure time or survival time events are the most common. The events are not necessarily associated with failure or survival at all, and perhaps *time-to-event* is a better term. Examples of time-to-event random variables are the lifetimes of organisms, failure time of machine components, survival times of cancer patients, occurrence of the next traffic accident, or durations of economic recessions. *Survival* or *failure* times are often used to represent time-to-event random variables. However, the scope and generality of time-to-event random variables are much broader than the "survival" or "failure"

times in their strict literal meaning. The generality and ubiquitousness of time-to-event random variables will become self-evident if we look into the nature of time-dependent processes. There are two major categories of observation data in studying time-dependent processes: one is the *time-series data*, which is well recognized, and the other is *time-to-event data*, which is perhaps less recognized but equally important and ubiquitous.

A particular challenge in analyzing survival data is information censoring, i.e., the observation of survival times is often incomplete. For example, in a medical clinical trial of a new drug, some individuals under observation may be dropped from the experiment for various reasons, ranging from complications from other diseases to death caused by accident. This kind of censoring may be treated as random censoring. In other cases, the observation may be terminated before the process is fully developed, which leads to observation truncation. One example is the information recorded in the black box of an airplane up to a crash. This is a typical case of *right censoring* or truncation where the observation is terminated at a fixed time. Besides random and right censoring, *left censoring* or truncation is also possible. For example, the time an enemy aircraft crosses a border may never be known exactly and thus the radar detection time may be left censored. Traditional statistics and reliability theory simply do not have mechanisms to extract the *partial* information from the *censored* observations. Either including or excluding the censored observations from analysis will induce statistical bias, which could lead to incorrect inferences. In survival analysis, unique mathematical models and methodologies have been developed to extract the *partial* information from the censored observations maximally, without inducing unwarranted bias, as long as censoring is not too severe. The ability to handle censoring in an effective manner is the most important and unique advantage of survival analysis. However, observation censoring or incomplete information are not necessary for the application of survival analysis.

Survival analysis was initially advanced at the UC Berkeley in the 1960s to provide a better analysis method for Life Table data. Chiang's (1960) papers expanded Kaplan & Meier's (1958) classic work on the survivor function. Aalen's (1975) established rigorous mathematical theory of survival analysis based on the *Martingale theory* and *Counting* stochastic process. It is interesting that Aalen's work probably was not fully appreciated until the late 1980s. Fleming & Harrington (1991) and Andersen, Borgan, Gill & Keiding's (1995) seem to be the only expansion (in the form of monographs) of the theoretic foundation set by Aalen (1975). As expected, besides the monographs, there are numerous theoretic research articles, which we choose to skip given our discussion is largely application-oriented.

On the statistical science frontline, the development of statistical procedures and models for survival analysis exploded in the 1970s and 1980s were mostly derived from

asymptotic methodology. By the late 1980s and early 1990s, survival analysis had established itself as the *de facto* standard statistical method in biomedical research. In medical schools, survival analysis became a dominant part of the standard biostatistics curriculum, and survival analysis is almost universally adopted for data analysis in major medical journals such as *JAMA* and the *New England Journal of Medicine*. Between the publishing of the first version of key monographs (such as Kalbfleisch & Prentice 1980 and Lawless 1982) to their second editions (Kalbfleisch & Prentice 2002 and Lawless (2003)), more than one dozen of monographs about survival analysis have been published. Besides the classics, such as the above-mentioned two, Cox and Oakes (1984), and Klein and Moeschberger (2003), two other monographs stand out: "*Analysis of Multivariate Survival Data*" by Hougaard (2000) and "*Bayesian Survival Analysis*" by Ibrahim, Chen & Sinha (2004). The titles of the later two monographs indicate their special contributions.

If asked to pick out a single *fundamental model of survival analysis*, it would be the Proportional Hazard Model originally proposed by D. R. Cox (1972). The Cox proportional model was initially developed as an empirical regression model. Furthermore, it was found that the *empirical* model captures some of the most fundamental properties of hazard functions and the basic framework of Cox's model still holds after numerous extensions (e.g., Therneau and Grambsch 2000).

Outside of biomedical fields, survival analysis has not been widely adopted, not even in ecology and other non-medical biology fields. Industry reliability analysis was actually envisioned as one of the most promising fields for survival analysis, besides the biomedical statistics. In 1991, NATO organized an Advanced Research Workshop on survival analysis and published proceedings from the workshop (Klein & Goel 1992), including a section on the application of survival analysis in industrial reliability. Meeker and Escobar (1998) published an excellent monograph, with strong emphasis on survival analysis approach to reliability modeling.

Two relevant international conferences are particularly worthy mentioning: the first is the biannual *International Conference on Mathematical Methods in Reliability* (MMR) and the latest conference proceedings was published in a volume edited by Wilson et al. (2005). The second is the *International Conference on Reliability and Survival Analysis* (ICRSA), last held in 2005. It appears that no conference proceedings were published other than the abstracts posted at http://www.isid.ac.in/~statmath/smuconferences/conference05/abstracts.html.

In our opinion, the potential of survival analysis in engineering reliability has not been fully explored yet. It appears that the application in computer science and IEEE related engineering is still limited to a handful of *ad hoc* applications. For example, in a recent online search of the

IEEE digital library, we found about 40 entries with the keyword *survival analysis*, however only about a half of which is the application of survival analysis in engineering reliability and the others are survival analysis application in biomedicine or other fields. In addition, among the 40 papers, about one-fourth is the application of artificial neural networks (ANN) to model fitting for survival analysis. Furthermore, with the keyword *competing risks analysis*, we obtained about 20 papers, but found no paper with the keywords *multivariate survival analysis* or *shared frailty*. The latter two fields are the areas that most of the new advances in survival analysis are made (which are discussed in Ma and Krings 2008a, b & c, and this article is limited to univariate survival analysis).

2.2. Censoring and Statistical Models

It is largely accurate to treat survival analysis as the statistics for lifetime data with censoring. Information or observation censoring is common in failure time data and reliability analysis. Censoring refers to the situations in which exact lifetimes are fully observed for only a portion of the individuals in a statistical population sample. Formally, an observation is *right censored* at C if the lifetime (T) is only known to be greater than or equal to C. Similarly, an observation is left censored if the lifetime is only known to be less than or equal to C (Lawless 1982, 2003). More precise definitions can be achieved by further distinguishing censoring as *type-I, type-II* and *random censoring*, each of which can be referred to as either left or right censoring. For simplicity, we only discuss their meaning for right censoring, which is the most common in failure time data.

Type II censoring only knows the r smallest (shortest lifetimes) in a random sample of n individuals $(1 \leq r \leq n)$. The observation data consists of the r smallest lifetimes $T_{(1)} \leq T_{(2)} ... T_{(r)}$ out of a random sample of n lifetimes $T_1, T_2,...,T_n$. If $T_1, T_2,...,T_n$ are *i.i.d.* and have a continuous distribution with pdf $f(t)$ and survivor function $S(t)$, then the joint pdf of right censoring $T_{(1)} \leq T_{(2)} ... T_{(r)}$ is:

$$\frac{n!}{(n-r)!} f(t_{(1)}) f(t_{(2)}) ... f(t_{(r)}) [S(t_{(r)})]^{n-r} \qquad (14)$$

With the pdf for censored sample individuals, statistical inference can be made based on it.

In *type-I censoring*, observation is cut off at some pre-specified time and an individual's lifetime will be known exactly only if it is less than the pre-specified value. One difference between the two types of censorings is that for *type-I* the number of fully observed individuals is a random variable, but for the *type-II* it is fixed. In a sample of n individuals, each individual (i) is associated with a lifetime T_i and a possible fixed censoring time C_i. The T_i are assumed to be *i.i.d.* with pdf $f(t)$ and survivor function $S(t)$. The exact lifetime T_i of an individual will be observed only

if $T_i \le C_i$. When all the C_i are equal, it is called singly Type-I censored. The data from type-I censoring consists of n pairs of random variables (t_i, δ_i), where $t_i = \min(T_i, C_i)$

$$\delta_i = 1 \text{ if } T_i \le C_i \ \delta_i = 0 \text{ if } T_i > C_i \qquad (15)$$

The joint pdf of (t_i, δ_i) is:

$$f(t_i)^{\delta_i} S(C_i)^{1-\delta_i} \qquad (16)$$

t_i is a *mixed random variable* with both discrete (if censored) and continuous parts.

If the pairs (t_i, δ_i) are independent, the likelihood function is:

$$L = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(C_i)^{1-\delta_i} \qquad (17)$$

An observation from the above likelihood function is: that each observed lifetime contributes a term *f(t)* and each censored time contributes a term *S(C_i)*. This observation has wide applicability (Lawless 2003, Kalbfleisch and Prentice 2002).

Random censoring occurs naturally. For example, in medical research, patients may be admitted into the study in somewhat random fashion contingent on their diagnostic times. If the study is terminated at some pre-specified time, then the censoring times, which start from the random entry, points until the termination, are random. If C_i is specified with different distributions, each with density mass at one fixed point, Type-I censoring can then be treated as a special case of random censoring. Often the likelihood function for type-I can be used to approximate that of random censoring.

If censorings and failure events are dependent, although it might be difficult to write down a model that represent the process, the above likelihood function structure still largely holds (Lawless 2003, Kalbfleisch and Prentice 2002). The above discussion of censoring is simplified greatly to highlight the principles of censoring and lifetime observations. The study of the failure-censoring pair random variables influences nearly every aspect of survival analysis.

2.3. Modeling Survivability with Random Censoring

From the discussion of censoring in the previous section, in particular the random censoring mechanism, we propose to use the *random censoring* mechanism to describe the unpredictable events such as malicious intrusions in the modeling of network survivability. A comparative analysis of random censoring and malicious intrusions to computer networks (or any survivable systems) will support our arguments. The most significant similarity is that both random censoring and malicious intrusions are unpredictable. In other words, we generally do not even know the probability that the event may happen. Specifically, the striking time of malicious act can be considered largely random but the event is only describable

after it occurs. Therefore, the striking times can be treated as the censoring events. Similarly, in a clinical trial of biomedicine, censoring should be unpredictable, otherwise the experiment is biased. Furthermore, by introducing different levels of censoring (e.g., percentage of censored individuals), one can simulate the effects of the strike *intensity* on the survivor function. If one defines survivability as a *threshold* of reliability breakdown (e.g., survivor function crosses some *threshold* value), then this kind of simulation can produce very important insights.

2.4. Survival Distribution Models—Parametric Models

As we have mentioned previously, the basic definitions of survival analysis, such as hazard rate (failure rate), cumulative hazard function, probability density function, distribution functions are exactly the same as those in reliability theory introduced in Subsection 1.1 [Equations (1)–(5)]. The only difference is the terminology, i.e. *survivor function* vs. *reliability*, but both have exactly the same mathematical definitions. However, beyond the similarity at the definition level the two fields diverged significantly. Through the remainder of this paper, the major concepts and models are drawn from the survival analysis literature.

The following three concepts exist in both fields; however, there are slight differences in their formulations: *mean residual lifetime*, *median lifetime*, and *p-th quantile* (also known as the *100p-th percentile*).

The mean residual life (*mrl*) is the expected remaining lifetime for an individual of age *t* and is defined as $mrl(t) = E(T - t \mid T > t)$. It can be proved that the following equation holds.

$$mrl(t) = \frac{\int_t^\infty (T-t)f(x)dx}{S(t)} = \frac{\int_t^\infty S(x)dx}{S(t)} \qquad (18)$$

The *mrl* is the area under the survival curve to the right of *t* divided by *S(t)*. In contrast, the *mean lifetime*, equivalent to the "*mean time to failure*" (MTTF) in reliability theory, is the total area under the survival curve

$$MTTF = E(T) = \int_0^\infty tf(t)dt = \int_0^\infty S(t)dt. \qquad (19)$$

The variance of *T* is related to the survival function by

$$Var(T) = 2\int_0^\infty tS(t)dt - \left[\int_0^\infty S(t)dt\right]^2 \qquad (20)$$

The *p*-th quantile of the distribution of *T* is the smallest t_p such that,

$$S(t_p) \le 1 - p, i.e., t_p = \inf\{t : S(t) \le 1 - p\}. \qquad (21)$$

If T is a continuous random variable, then the p-th quantile can be found by solving $S(t_p) = 1 - p$.

The median lifetime (*mlt*) is the 50-th percentile $t_{0.5}$ of the distribution of T and is obtained by solving the following equation: $S(t_p) = 1 - 0.5 = 0.5$.

Since failure time (or lifetime) is simply a non-negative random variable, the most natural approach is to study its probability distributions. Again, the probability distributions used in both reliability and survival analysis are often the same. The most commonly used distributions include: exponential, Weibull, extreme value distribution, gamma distribution, log-gamma, lognormal, generalized gamma, logistic, log-logistic, and inverse Gaussian. Distribution models discussed in this section are referred to as parametric models in the literature.

We will look at two examples, exponential distribution and Weibull distribution. The *exponential distribution* can serve as the baseline for more complex models, given its constant failure rate

$$h(t) = \lambda \quad t \geq 0, \lambda > 0 \tag{22}$$

with pdf

$$f(t) = \lambda e^{-\lambda t}. \tag{23}$$

The survivor function is

$$S(t) = e^{-\lambda t} \tag{24}$$

and the mean and variances are $\theta = 1/\lambda$ and θ^2, respectively.

When $\theta = \lambda = 1$, it is termed standard exponential distribution. In addition, exponential distribution is special cases of both the Weibull and gamma distributions.

The *Weibull distribution* is perhaps the most widely used lifetime distribution. Its hazard rate is:

$$h(t) = \lambda \beta (\lambda t)^{\beta - 1} \tag{25}$$

where $\lambda > 0$ and $\beta > 0$ are parameters. When $\beta = 1$, Weibull distribution becomes exponential distribution. Its pdf and survivor functions are:

$$f(t) = \lambda \beta (\lambda t)^{\beta - 1} \exp[-(\lambda t)^\beta] \quad t > 0 \tag{26}$$

$$S(t) = \exp[(-\lambda t)^\beta] \quad t > 0. \tag{27}$$

The hazard function of Weibull distribution is monotonic increasing if $\beta > 1$, decreasing if $\beta < 1$, and constant for $\beta = 1$, β is therefore termed shape parameter. The λ is called scale parameter.

Weibull distribution can be derived from uniform distribution U in the interval [0, 1] conveniently,

$$T = [(-\ln(U))^{1/\beta}]/\lambda. \tag{28}$$

This formula can be used for Monte Carlo simulation of Weibull distribution.

There is also a three-parameter version Weibull distribution by replacing the variable t with (t–u), where u is called location parameter.

The following two distributions are with bathtub-shaped hazards functions:

$$h(t) = \frac{\beta}{t + \gamma} + \delta t \tag{29}$$

$$h(t) = \frac{\beta}{\alpha} \left(\frac{t}{\alpha} \right)^{\beta - 1} \exp\left(\frac{t}{\alpha} \right)^\beta. \tag{30}$$

Klein and Moeschberger (2003) summarized a near exhaustive list of distributions used in survival analysis. The distribution models described above are for continuous failure time data. Any of the continuous failure time models can be used to generate a discrete model by introducing grouping on the time axis (Kalbfleisch and Prentice 1980, 2002, Klein and Moeschberger (2003).

The estimation and comparison of survival function $S(t)$, which is equivalent to the reliability $R(t)$ in reliability theory, with the above distribution models form a significant part of the early development of survival analysis. Estimating survival function or fitting lifetime data to survival distributions was the earliest endeavor in this field. Three approaches stand out: the *Kaplan-Meier estimator* for survivor function (Kaplan and Meier 1958), Chiang's (1960) life table approach, and the *Nelson-Aalen estimator* for cumulative hazard functions (Nelson 1969). The life table approach is particularly important in population demography and actuarial study. To compare the survival functions, several statistics have been devised or extended from classical statistics to censored data, such as the *Wilcoxon* test, *Savage log-rank* test, and the *Kruskal-Wallis* test (Kalbfleisch and Prentice 1980, 2002). In recent years, great progress has been made in improving the estimations and comparisons of survivor functions, such as the integration of Bayesian approach and Markov Chain Monte Carlo (MCMC) simulation, and kernel-smoothed estimation. One can refer to Klein and Moeschberger (2003) for comprehensive details.

2.5. Covariates Regression Models: Proportional Hazards Models and Accelerated Failure Time Models

The previous section focused on using probability distributions to model survival of a homogenous population. Translated into the reliability field, a homogenous population implies that devices experience exactly the same environmental covariates, and the covariates are not

correlated to failure times. For example, this would imply that operation environment parameters such as temperature have no effects on the failure rates of devices. The covariates regression approach introduced in this section eliminate this restriction. The basic problem can be formed as follows: given survival time $T > 0$, a vector $z = (z_1, z_2, ...z_s)$ of covariates that may influence T is observed. Furthermore, the vector z itself can also be time dependent and may include both quantitative and qualitative indicative variables. In this approach, Cox's (1972) proportional hazards model was initially treated as largely an empirical regression model, but later it was found that the framework of the model possesses exceeding flexibility to capture major hazards effects and failure mechanisms. The Cox model has been extended numerously and it forms the foundation for the covariates regression approach (Therneau and Grambsch 2000).

There are two ways to construct covariates regression models. The first is to extend survivor distributions such as the exponential and Weibull distribution, and this approach forms so-called mixture models. The second approach has no assumptions on the survival distribution, also called distribution-free approach, which is the approach adopted by Cox (1972, 1975). We briefly introduce the first approach by using exponential and Weibull distributions as examples. The approach can be extended to other distributions, including discrete distributions (Kalbfleisch and Prentice 1980, 2002).

2.5.1. Distribution-Dependent Regression Models: Parametric Regression Models

With exponential distribution, the hazard rate λ is constant. The hazard rate at time t for an individual with covariates z can be written as

$$\lambda(t;z) = \lambda(z), \qquad (31)$$

i.e, the hazard for a given vector z is constant; however, it varies with covariates z. The $\lambda(z)$ may be modeled in many ways. In the simplest case it assumes a linear relationship, $z\beta$, and thus

$$\lambda(t;z) = \lambda g(z\beta), \qquad (32)$$

where β is the vector of regression coefficients for vector z, and $z\beta$ is the inner product of the two vectors. Here g is a function form, for example,

$$g(x) = \exp(x). \qquad (33)$$

With $g(.)$ taking on an exponential form — the exponential distribution models under covariates (z) — the hazard function is

$$\lambda(t;z) = \lambda \exp[(z\beta)]. \qquad (34)$$

The conditional *pdf* of T, given z, is

$$f(t,z) = \lambda e^{z\beta} \exp(-\lambda t e^{z\beta}). \qquad (35)$$

The above hazard function implies that the *log* failure rate is a linear function of covariates z. Let $Y = \ln(T)$, $\alpha = -\ln(\lambda)$, and W be a random variable with *extreme value distribution*, the *log* of the exponential failure time (T) under covariates z can be expressed as

$$Y = \alpha - z\beta + W. \qquad (36)$$

This is a *log-linear model* with many nice properties.

Similar to the previous extension, Weibull distribution can be extended with covariates regression. The conditional hazard for Weibull distribution is

$$\lambda(t;z) = \lambda\beta(\lambda t)^{\beta-1} e^{z\gamma}. \qquad (37)$$

where γ represent the covariate regression vector to avoid the confusion with the parameter β of Weibull distribution. The conditional pdf for Weibull distribution is then:

$$f(t;z) = \lambda\beta(\lambda t)^{\beta-1} e^{z\gamma} \exp[-(\lambda t)^{\beta} e^{z\gamma}] \qquad (38)$$

The effects of covariates also act multiplicatively on the hazard in the Weibull distribution. In terms of the *log* failure time, the above conditional hazard function specifies the log-linear model. With $Y = \ln(T)$, $\alpha = -\ln(\lambda)$, $\sigma = 1/\beta$, and $\gamma^* = -\sigma\gamma$, we get the log-linear Weibull model

$$Y = \alpha + z\gamma^* + \sigma W \qquad (39)$$

where W follows the extreme value distribution. The above extensions of exponential and Weibull distributions demonstrate two important properties (Lawless 2003, Kalbfleisch and Prentice 2002). First, the effects of covariates act *multiplicatively* on the *hazard function*. This general relationship inspired the general proportional hazards model. Second, the log-linear models suggest that the covariates act *additively* on the *log failure time*. This inspired the development of the accelerated failure time model. Both models are described briefly below.

2.5.2. Distribution-Free Regression Models: Semi-parametric Regression Models.

The next two subsections introduce two of the most important survival analysis models. They do not assume any specific statistical distributions, and are therefore called distribution-free or semi-parametric approaches. Both models have been extended numerously. The Cox proportional hazard model is probably the most influential one in the entire survival analysis subject.

Cox's Proportional Hazards Model (PHM) — Let $\lambda(t;z)$ be the hazard function at time t for an individual with

covariates z. The proportional hazards model (PHM), first proposed by Cox (1972, 1975), is

$$\lambda(t;z) = \lambda_0(t)e^{z\beta}, \qquad (40)$$

where $\lambda_0(t)$ is an arbitrary unspecified base-line hazard function for continuous failure time T. In the PHM, the covariates act multiplicatively on the hazard function. By substituting $\lambda_0(t)$ with the corresponding hazard function for exponential or Weibull distributions, the previous distribution-dependent models for exponential or Weibull distribution can be derived as special cases.

The conditional pdf of T, given z, corresponding to the general hazard function $\lambda(t;z)$ is

$$f(t;z) = \lambda_0(t)e^{z\beta}\exp\left[-e^{z\beta}\int_0^t \lambda_0(u)du\right]. \qquad (41)$$

The conditional survival function for T under z is

$$S(t;z) = [S_0(t)]^{\exp(z\beta)}, \qquad (42)$$

where

$$S_0(t) = \exp\left[-\int_0^t \lambda_0(u)du\right]. \qquad (43)$$

The allowance of an arbitrary $\lambda_0(t)$ makes the model sufficiently flexible for many applications.

Extensions of Cox's Proportional Hazards Model (PHM)—There are numerous extensions to Cox's PHM (Therneau and Grambsch 2000). Among those extensions, two of the extensions: *strata* and *time-dependent covariates* are particularly important, but do not substantially complicate the parameters estimation (Kalbfleisch and Prentice 2002, Lawless 2003).

Suppose there is a factor that occurs on q levels and for which the proportionality assumption of PHM may be violated. The hazard function for an individual in the j-th stratum or level of this factor is

$$\lambda_j(t,z) = \lambda_{0j}(t)\exp(z\beta) \qquad (44)$$

for $j = 1,2,...q,$ where z is the vector of covariates for PHM. The baseline hazard functions, $\lambda_{01}(.),...,\lambda_{0q}(.)$ for the q strata are permitted to be arbitrary and are completely unrelated. The direct product of the group (stratum) likelihood can be utilized to estimate the common parameter vector, β. Once the parameter β is estimated, the survivor function for each stratum can be estimated separately.

The second generalization to the proportional hazards model is to let the variable z depend on time itself. For un-stratified PHM, the hazard function is

$$\lambda[t;z(t)] = \lambda_0(t)\exp[z(t)\beta] \qquad (45)$$

and for stratified PHM it is

$$\lambda_j[t;z(t)] = \lambda_{0j}(t)\exp[z(t)\beta] \quad j = 1,2,...,r. \qquad (46)$$

The procedure used to estimate the β is to maximize the so-called *partial likelihood functions* as described by Cox (1975) and Kalbfleisch & Prentice (1980, 2002).

Time-dependent covariates may be classified into two broad categories: external and internal (Kalbfleisch and Prentice 2002, Lawless 2003, Klein and Moeschberger. 2003). The failure mechanism does not directly involve external covariates. External covariates may be distinguished into (1) *fixed* (measured in advance and fixed during the observation), (2) *defined* (determined in advance but not fixed, for example, controlled stress levels), and (3) *ancillary* (output of a stochastic process external to an observed object). An internal covariate is *attached* to the individual, and its existence depends on the survival of the individual. It is the output of a stochastic process that is induced by the individual under study. The process is only observable if the individual survives and is uncensored.

The Accelerated Failure Time (AFT) model—The PHM represents the multiplicative influences of covariates (z) on the hazard. However, without specific $\lambda_0(t)$, this model does not tell how z affects failure time itself. The previous log-linear models for T answer this question.

Suppose that $Y = \ln(T)$ is related to the covariance z in a linear model,

$$Y = \alpha + z\beta + \sigma W, \qquad (47)$$

where α is a constant or intercept parameter, z is a vector of covariates, β is a vector of parameters to be estimated, σ is a scale parameter which is the reciprocal of the shape parameter, and W is a random variable with a specified distribution. Many distributions, including the Weibull, exponential, log-normal, and log-logistic, can be used to describe W.

Exponentiation of Y gives

$$T = \exp(\alpha + z\beta)T' \qquad (48)$$

where $T' = \exp(\sigma W) > 0$ has a hazard function $\lambda_0(t')$ that is independent of β. It follows that the hazard function for T can be written in terms of this baseline hazard $\lambda_0(.)$ as

$$\lambda(t;z) = \lambda_0[t\exp(\alpha - z\beta)]\exp(\alpha - z\beta) \qquad (49)$$

The survival function is

$$S(t;z) = \exp\left[-\int_0^{t\exp(\alpha - z\beta)} \lambda_0(u)du\right] \qquad (50)$$

and the pdf is the product of $\lambda(t;z)$ and $S(t;z)$.

This model specifies that covariates act multiplicatively on time (t), rather than on the hazard function. That is, we assume a baseline hazard function to exist and that the effects of the covariates are to alter the rate at which an individual proceeds along the time axis. In other words, the covariates z accelerates or decelerates the time to failure (Kalbfleisch and Prentice 2002, Lawless 2003).

It should be pointed out that the distribution-based regression models for exponential and Weibull distributions in the previous section are the special cases of both PHM and AFT. This correspondence is not necessarily true for models based on other distributions. Indeed, two-parameter Weibull distribution has the unique property that it is closed under both multiplication of failure time and multiplication of the hazard function by an arbitrary nonzero constant (Lawless 2003, Kalbfleisch & Prentice 2002, Klein & Moeschberger 2003).

2.6. Counting Process and Survival Analysis

In the previous sections, we introduced censoring and survival analysis models that can handle the censored information; however, we did not discuss how the censored information is processed. Accommodating and maximally utilizing the partial information from the censored observations is the most challenging and also the most rewarding task in survival analysis. This also establishes survival analysis as a unique field in mathematical statistics. Early statistical inferences for censored data in survival analysis were dependent on asymptotic likelihood theory (Severini 2000). Cox (1972, 1975) proposed *partial likelihood* as an extension to classical maximum likelihood estimation in the context of his proportional hazards model as a major contribution. Asymptotic likelihood has been and still is the dominant theory for developing survival analysis inference and hypothesis testing methods (Klein and Moeschberger 2003, Severini 2000). There are many monographs and textbooks of survival analysis containing sufficient details for applying survival analysis (Cox and Oakes 1984, Kalbfleisch and Prentice 1980, 2002, Lawless 1982, 2003, Klein and Moeschberger, 2003). A problem with traditional asymptotic likelihood theory is that the resulting procedures can become very complicated when handling more complex censoring mechanisms (Klein & Moeschberger 2003). A more elegant but requiring rigorous measure-theoretic probability theory is the approach with counting stochastic processes and the Martingale central limit theorems. Indeed, this approach was used by Aalen (1975) to set the rigorous mathematical foundation for survival analysis, and later further developed and summarized by Fleming and Harrington (1991), Andersen et al. (1993) and several research papers. In reliability theory, Aven and Jensen (1999) demonstrated such an approach by developing a general failure model, which we briefly introduced in Section 1.2. However, the counting process and Martingale approach require measure theoretic treatments of probability and stochastic processes, which is often not used in engineering or applied statistics. A detailed introduction of the topic is obviously beyond the scope of this paper, and we only present a brief sketch of the most important concepts involved. Readers are referred to the excellent monographs by Andersen et al. (1993), Fleming and Harrington (1991), Aven and Jensen (1999) for comprehensive details, and Kalbfleisch and Prentice (2002), Klein and Moeschberger (2003), Lawless (2003) for more application–oriented treatments. The following discussion on this topic is drawn from Klein and Moeschberger (2003).

A counting stochastic process $N(t)$, $t \geq 0$, possesses the properties that $N(0)$ is zero and $N(t) < \infty$ with probability one. The *sample paths* of $N(t)$ are right continuous and piecewise constant with jumps of size *1* (*step function*). In a right-censored sample, (we assume only right censoring in this section), the processes, $N_i(t) = I[T_i \leq t, \delta_i = 1]$, which keep the value *0* until individual *i* fails and then jump to *1*, are counting processes. The accumulation of $N_i(t)$ process, $N(t) = \sum_{i=1}^{n} N_i(t)$, is again a counting process, which counts the number of failures in the sample at or before time *t*.

The counting process keeps track of the information on the occurrences of events, for instance, the history information such as which individual was censored prior to time *t* and which individual died at or prior to time *t*, as well as the covariates information. This accumulated history information of the counting process at time *t* is termed *filtration* at time *t*, denoted by $\mathbf{F}_t$. For a given problem, $\mathbf{F}_t$ rests on the observer of the counting process. Thus, two observers with different recordings at different times will get different filtrations. This is what Aven and Jensen (1999) referred to as "*different information levels*", or the amount of actual available information about the state of a system may vary.

If the failure times X_i and censoring times C_i are independent, then the probability of an event occurs at time *t*, given the history just prior to *t*, (F_{t-}), can be expressed as:

$$P_r[t \leq T_i \leq t + dt, \delta_i = 1 \mid F_{t-}]$$
$$= P_r[t \leq X_i \leq t + dt, C_i > t + dt_i \mid X_i \geq t, C_i \geq t] = h(t)dt \; if \, T_i \geq t$$

$$P_r[t \leq T_i \leq t + dt, \delta_i = 1 \mid F_{t-}] = 0 \; if \, T <_i t \qquad (51)$$

Let $dN(t)$ be the change in the process $N(t)$ over a short time interval $[t, t + \Delta t)$. Ignoring the negligible chance of ties, $dN(t) = 1$ if a failure occurred and $dN(t) = 0$ otherwise.

Let $Y(t)$ denote the number of individuals with an observation time $T_i \geq t$. Then the conditional expectation of $dN(t)$ is:

$$E[dN(t) \mid F_{t-}] = E[number \; of \; observations \; with$$
$$t \leq X_i \leq t + dt, C_i > t + dt_i \mid F_{t-}] = Y(t)h(t)dt \qquad (52)$$

The process $\lambda(t) = Y(t)h(t)$ is called the *intensity process* of the counting process. It is a stochastic process that is determined by the information contained in the history process, F_t, via $Y(t)$. Note that $Y(t)$ is the process that records the number of individuals experiencing the risk at a given time t.

As it will become clear that $\lambda(t)$ is equivalent to the failure rate or hazard function in traditional reliability theory, but here it is treated as a stochastic process, the most general form one can assume for it. It is this generalization that encapsulates the power that counting stochastic process approach can offer to survival analysis.

Let the cumulative intensity process $H(t)$ be defined as:

$$H(t) = \int_0^t \lambda(s)ds, \; t \geq 0 \tag{53}$$

It has the property that

$$E[N(t)|F_{t-}] = E[H(t)|F_{t-}] = H(t). \tag{54}$$

This implies that filtration, F_{t-}, is known, the value of $Y(t)$ is fixed and $H(t)$ becomes deterministic. $H(t)$ is equivalent to the cumulative hazards in traditional reliability theory.

A stochastic process with the property that its *expectation* at time t, given history at time $s < t$, is equal to its value at s, is called a *martingale*. That is, $M(t)$ is a martingale if

$$E[M(t)|F_s] = M(s), \forall s < t. \tag{55}$$

It can be verified that the stochastic process

$$M(t) = N(t) - H(t) \tag{56}$$

is a martingale, and it is called the *counting process martingale*. The increments of the counting process martingale have an expected value of zero, given its filtration F_{t-}. That is,

$$E[dM(t)|F_{t-}] = 0. \tag{57}$$

The first part of the counting process martingale [Equation (56)] $N(t)$ is a non-decreasing *step* function. The second part $H(t)$ is a smooth process which is *predictable* in that its value at time t is fixed just prior to time t. It is known as the compensator of the counting process and is a random function. Therefore, the martingale can be considered as mean-zero noise and that is obtainable when one subtracts the smoothly varying compensator from the counting process.

Another key component in the counting process and martingale theory for survival analysis is the notion of the *predictable variation process* of $M(t)$, denoted by $\langle M \rangle(t)$. It is defined as the *compensator* of process $M^2(t)$.

The term *predictable variation process* comes from the property that, for a martingale $M(t)$, it can be verified that the conditional variance of the increments of $M(t)$ [i.e., $dM(t)$] equals the increments of $\langle M \rangle(t)$. That is,

$$Var[dM(t)|F_{t-}] = d\langle M \rangle(t). \tag{58}$$

To obtain $Var[dM(t)|F_{t-}]$, one needs the random variable $N(t)$, which is a zero-one random variable with probability $\lambda(t)$ of having a jump of size *1* at time t. The variance of $N(t)$ is $\lambda(t)[1 - \lambda(t)]$ since it follows binomial distribution. Ignoring the ties in the censored data, $\lambda^2(t)$ is close to zero and $Var[dM(t)|F_{t-}] = \lambda(t) = Y(t)h(t)$. This implies that the counting process $N(t)$, conditional on the filtration F_t, behaves like a Poisson process with rate $\lambda(t)$.

Why do we need to convert the previous very intuitive concepts in survival analysis into more abstract martingales? The key is that many of the statistics in survival analysis can be derived as the *stochastic integrals* of the basic martingales described above. The stochastic integral equations are mathematically well structured and some standard mathematical techniques for studying them can be adopted.

Here, let $K(t)$ be a *predictable process*. An example of a predictable process is the process $Y(t)$. Over the interval 0 to t, the stochastic integral of such a process, with respect to a martingale, is denoted by $\int_0^t K(u)dM(u)$. It turns out that such stochastic integrals themselves are martingales as a function of t, and their predictable variation process can be found from the predictable variation process of the original martingale by

$$\left\langle \int_0^t K(u)dM(u) \right\rangle = \int_0^t K^2(u)d\langle M \rangle(u). \tag{59}$$

The above discussion was drawn from Klein and Moeschberger (2003). They also provide examples of how the above process is applied. In the following, we briefly introduce one of their examples — the derivation of the Nelson-Aalen cumulative hazard estimator.

First, the model is formulated as:

$$dN(t) = Y(t)h(t)dt + dM(t) \tag{60}$$

If $Y(t)$ is non-zero, then

$$\frac{dN(t)}{Y(t)} = h(t)d(t) + \frac{dM(t)}{Y(t)} \tag{61}$$

Let $I(t)$ be the indicator of whether $Y(t)$ is positive and define 0/0 = 0, then integrating both sides of above (61), One get

$$\int_0^t \frac{I(u)}{Y(u)} dN(u) = \int_0^t I(u)h(u)d(u) + \int_0^t \frac{I(u)}{Y(u)} dM(u) \qquad (62)$$

The left side integral is the Nelson-Aalen estimator of $H(t)$.

$$\tilde{H}(t) = \int_0^t \frac{I(u)}{Y(u)} dN(u) \qquad (63)$$

The second integral on the right side,

$$W(t) = \int_0^t \frac{I(u)}{Y(u)} dM(u) \qquad (64)$$

is the stochastic integral of the predictable process $I(u)/Y(u)$ with respect to a martingale, and hence is also a martingale.

The first integral on the right side is a random quantity $H^*(t)$

$$H^*(t) = \int_0^t I(u)h(u)du \qquad (65)$$

For right-censored data it is equal to $H(t)$ in the data range, if the stochastic uncertainty in the $W(t)$ is negligible.

Therefore, the statistic $\tilde{H}(t)$ is a nonparametric estimator of the random quantity $H^*(t)$.

We would like to mention one more advantage of the new approach, that is, the martingale central limit theorem. The central limit theorem of martingales ensures certain convergence property and allows the derivations of confidence intervals for many statistics. In summary, most of the statistics developed with asymptotic likelihood theory in survival analysis can be derived as the stochastic integrals of some martingale. The large sample properties of the statistics can be found by using the predictable variation process and martingale central limit theorem (Klein and Moeschberger (2003).

2.7. Bayesian Survival Analysis

Like many other fields of statistics, survival analysis has also witnessed the rapid expansion of the Bayesian paradigm. The introduction of the full-scale Bayesian paradigm is relative recent and occurred in the last decade, however, the "invasion" has been thorough. Until the recent publication of a monograph by Ibrahim, Chen and Sinhaet (2005), Bayesian survival analysis has been either missing or occupy at most one chapter in most survival analysis monographs and textbooks. Ibrahim's et al. (2005) changed the landscape, with their comprehensive discussion of nearly every counterparts of frequentist survival analysis, from univariate to multivariate, from nonparametric, semi-

parametric to parametric models, from proportional to non-proportional hazards models, as well as the joint model of longitudinal and survival data. It should be pointed out that Bayesian survival analysis has been studied for quite a while and can be traced back at least to the 1970s.

A natural but fair question is what advantages the Bayesian approach can offer over the established frequentist survival analysis. Ibrahim et al. (2005) identified two key advantages. First, survival models are generally very difficult to fit, due to the complex likelihood functions to accommodate censoring. A Bayesian approach may help by using the MCMC techniques and there is available software, e.g., BUGS. Second, the Bayesian paradigm can incorporate prior information in a natural way by using historical information, e.g., from clinical trials. The following discussion in this subsection draws from Ibrahim's et al. (2005).

The Bayesian paradigm is based on specifying a probability model for the observed data, given a vector of unknown parameters θ. This leads to likelihood function $L(\theta|D)$. Unlike in traditional statistics, θ is treated as random and has a prior distribution denoted by $\pi(\theta)$. Inference concerning θ is then based on the posterior distribution, which can be computed by Bayes theorem

$$\pi(\theta|D) = \frac{L(\vartheta|D)\pi(\theta)}{\int_\Theta L(\vartheta|D)\pi(\theta)d\vartheta}, \qquad (66)$$

where Θ is the parameter space.

The term $\pi(\theta|D)$ is proportional to the likelihood $L(\theta|D)$, which is the information from observed data, multiplied by the prior, which is quantified by $\pi(\theta)$, i.e.,

$$\pi(\theta|D) \propto L(\theta|D)\pi(\theta). \qquad (67)$$

The denominator integral, $m(D)$, is the normalizing constant of $\pi(\theta|D)$ and often does not have an analytic closed form. Therefore, $\pi(\theta|D)$ often has to be computed numerically. The Gibbs sampler or other MCMC algorithms can be used to sample $\pi(\theta|D)$ without knowing the normalizing constant $m(D)$. There exist large amount of literature for solving the computational problems of $m(D)$ and $\pi(\theta|D)$.

Given that the general Bayesian algorithms for computing the posterior distributions should equally apply to Bayesian survival analysis, the specification or elicitation of informative *prior* needs much of the attention. In survival analysis with covariates such as Cox's proportional hazards model, the most popular choice of informative prior is the *normal* prior, and the most popular choice for non-informative is the *uniform*. Non-informative prior is easy to use but they cannot be used in all applications, such as model selection or model comparison. Moreover, non-informative prior does not harness the real prior information. Therefore, research for informative prior specification is crucial for Bayesian survival analysis.

Reliability estimation is influenced by the level of information available such as information on components or sub-systems. Bayesians approach is likely to provide such flexibility to accommodate various levels of information. Graves and Hamada (2005) introduced the *YADAS*, a statistical modeling environment that implements the Bayesian hierarchical modeling via MCMC computation. They showed the applications of *YADES* in reliability modeling and its flexibility in processing hierarchical information. Although this environment seems not designed with Bayesian survival analysis, similar package may be the direction if Bayesian survival analysis is applied to reliability modeling.

2.8. Spatial Survival Analysis

To the best of our knowledge, *spatial survival analysis* is an uncharted area, and there has been no spatial survival analysis reported with rigorous mathematical treatment. There are some applications of survival analysis to spatial data; however, they do not address the spatial process, which in our opinion should be the essential aspect of any spatial survival analysis. To develop formal spatial survival analysis, one has to define the spatial process first.

Recall, for survival analysis in the time domain, there is survival function

$$S(t) = \Pr(T > t), t \in R^+ , \qquad (68)$$

where T is the random variable and $S(t)$ is the cumulative probability that the lifetime will exceed time t. In spatial domain, what is the counterpart of t? One may wonder why do not we simply define the survival function in the spatial domain as

$$S(s) = \Pr(S > s), s \in R^d , R > 0, \qquad (69)$$

where s is some metric for d-dimensional space R^d, and the space is restricted to the positive region. S is the "*space to event*" measurement, e.g., the distance from some origin where we detect some point object. The problem is that the metric itself is an attribute of space, rather than space itself. Therefore, it appears to us that the basic entity for studying the space domain has to be *broader* than in the time domain. This is probably why spatial process seems to be a more appropriate entity for studying.

The following is a summary of descriptions of spatial processes and patterns, which intends to show the complexity of the issues involved. It is *not* an attempt to define the similar survival function in spatial domain because we certainly understand the huge complexity involved. There are several monographs discussing the spatial process and patterns (Schabenberger and Gotway 2005). The following discussion heavily draws from Cressie (1993) and Schabenberger and Gotway (2005).

It appears that the most widely adopted definition for spatial process is proposed in Cressie (1993), which defines a spatial process $Z(s)$ in d-dimensions as

$$\{ Z(s) : s \in D \subset R^d \} \qquad (70)$$

Here, $Z(s)$ denotes the attributes we observe, which are space dependent. When $d = 2$ the space (R^2) is a plane.

The problem is how to define randomness in this process? According to Schabenberger and Gotway (2005), $Z(s)$ can be thought of as located (indexed) by spatial coordinates $s = [s_1, s_2, ..., s_n]$, the counterpart of time series $Y(t)$, $t = t_1$, t_2, ...,t_n, indexed by time. The spatial process is often called *random field*.

To be more explicit, we denote $Z(s)$ as $Z(s, \omega)$ to emphasize that Z is the outcome of a random experiment ω. A particular realization of ω produces a surface $Z(s, \omega)$. Because the surface from which the samples are drawn is the result of the random experiment, $Z(s)$ is also called a random function.

One might ask what is a random experiment like in a spatial domain? Schabenberger and Gotway (2005) offered an imaginary example briefly described below. Imagine pouring sand from a bucket onto a desktop surface, and one is interested in measuring the depth of the poured sand at various locations, denoted as $Z(s)$. The sand distributes on the surface according to the laws of physics. With enough resources and patience, one can develop a deterministic model to predict exactly how the sand grains are distributed on the desktop surface. This is the same argument used to determine the head-tail coin flipping experiment, which is well accepted in statistical science. The probabilistic coin-flip model is more parsimonious than the deterministic model that rests on the exact (perfect) but hardly feasible representation of a coin's physics. Similarly, deterministically modeling the placement of sand grains is equally daunting. However, the issue here is not placement of sand as a random event, as Schabenberger and Gotway (2005) emphasized. The issue is that the sand was poured *only once*, regardless how many locations one measures the depth of the sand. With that setting, the challenge is how do we define and compute the expectation of the random function $Z(s)$? Would $E[Z(s)] = \mu(s)$ make sense?

Schabenberger and Gotway (2005) further raised the questions: (1) to what distribution is the expectation being computed? (2) if the random experiment of sand pouring can only be counted once, how can the expectation be the long-term average?

According to the definition of expectation in traditional statistics, one should repeat the process of sand pouring many times and consider the probability distributions of all surfaces generated from the repetitions to compute the expectation of $Z(s)$. This complication is much more serious

than what we may often realize. Especially, in practice, many spatial data is obtained from one time space sample only. There is not any independent replication in the sense of observing several independent realizations of the spatial process (Schabenberger and Gotway 2005).

How is the enormous complexity in spatial statistics currently dealt with? The most commonly used simplification, which has also been vigorously criticized, is the stationarity assumption. Opponents claim that stationarity often leads to erroneous inferences and conclusions. Proponents counter-argue that little progress can be made in the study of non-stationary process, without a good understanding of the stationary issues (Schabenberger and Gotway 2005).

The strict stationarity is a *random field* whose spatial distribution is invariant under translation of the coordination. In other words, the process repeats itself throughout its domain (Schabenberger and Gotway 2005). There is also a second-order (or weak) stationarity of a random field.

For random fields in the spatial domain, the model of Equation (70), i.e., $\{Z(s):s \in D \subset R^d\}$, is still too general to allow statistical inference. It can be decomposed into several sub-processes (Cressie 1993):

$$Z(s) = \mu(s) + W(s) + \eta(s) + \varepsilon(s), \quad s \in D, \qquad (71)$$

where $\mu(s) \equiv E[Z(.)]$ is a deterministic mean structure called large-scale variation. $W(s)$ is the zero-mean intrinsically stationary process, (with second order derivative), and it is called smooth *small-scale* variation. $\eta(s)$ is the zero-mean stationary process independent of $W(s)$ and is called *microscale* variation. $\varepsilon(s)$ is zero-mean *white noise*, also called measurement error. This decomposition is not unique, and is largely operational in nature (Cressie 1993). The main task of a spatial algorithm is to determine the allocations of the large, small, and microscale components. However, the form of the above equation is *fixed* (Cressie (1993), implying that it is not appropriate to sub-divide one or more of the items. Therefore, the key issue here is to obtain the deterministic $\mu(s)$, but in practice, especially with limited data, it is usually very difficult to get a unique $\mu(s)$. Alternative to the *spatial domain decomposition* approach, the frequency domain methods or *spectral analysis* used in time series analysis can also be used in spatial statistics. Again, one may refer to Schabenberger and Gotway (2005).

So, what are the implications of the general discussion on spatial process above to spatial survival analysis? One point is clear, Equation (69), $S(s) = \Pr(S > s), \quad s \in R^d$ is simply too naive to be meaningful. There seem, at least, four fundamental challenges when trying to develop survival analysis in space domain. (1) Space process is often multidimensional, while time process can always be treated as uni-dimensional in the sense that it can be represented as

$\{Z(t):t \in R^1\}$. The multidimensionality certainly introduces additional complications, but that is still not the only complication, perhaps not even the most significant. (2) One of the fundamental complications is the frequent lack of independent replication in the sense of observing several independent realizations of the spatial process, as pointed out by Cressie (1993). (3) The superposition of (1) and (2) brings up even more complexity, since coordinates (*s*) of each replication are a set of stochastic spatial process, rather than a set of random variables. (4) Even if the modeling of the spatial process is separable from the time process, it is doubtful how useful the resulting model will be. In time process modeling, if a population lives in a homogenous environment, the space can be "*condensed*" as a single point. However, the freezing of time seems to leave out too much information, at least for survival analysis. Since the traditional survival analysis is essentially a time process, therefore, it should be expanded to incorporate spatial coordinates into original survival function. For example, when integrating space and time, one gets a space-time process, such as

$$\{Z(s,t):s \in D \subset R^d, \quad t \in R^+\},$$

where *s* is the spatial index (coordinate) and *t* is time. We may define the spatial-temporal survival function as

$$S(s,t) = \Pr\{T > t, \ s \in D\}, \qquad (72)$$

where *D* is a subset of the *d*-dimensional Euclidean space. That is, the spatial-temporal survival function represents the cumulative probability that an individual will survive up to time *T*, within hyperspace *D*, which is a subset of the *d*-dimensional Euclidian space. One may define different scales for *D*, or even divide *D* into a number of unit hyperspaces of measurement *1* unit.

2.9. Survival Analysis and Artificial Neural Network

In the discussion on artificial neuron networks (ANN), Robert & Casella (2004) noted, "Baring the biological vocabulary and the idealistic connection with actual neurons, the theory of neuron networks covers: (1) modeling nonlinear relations between explanatory and dependent (explained) variables; (2) estimation of the parameters of these models based on a (training) sample." Although Robert & Casella (2004) did not mentioned survival analysis, their notion indeed strike us in that the two points are also the essence of Cox's (1972) Proportional Hazards model.

We argue that the dissimilarity might be superficial. One of the most obvious differences is that ANN usually avoids probabilistic modeling, however, the ANN models can be analyzed and estimated from a statistical point of view, as demonstrated by Neal (1999), Ripley (1994), (1996), Robert & Casella (2004). What is more interesting is that the most hailed feature of ANN, i.e., the "*identifiability*" of model structure, if one review carefully, is very similar to the work done in survival analysis for the structure of the

Cox proportional hazards model. The multilayer ANN model, also known as back-propagation ANN model, is again very similar to the stratified proportional hazards model.

There are at least two advantages of survival analysis over the ANN. (1) Survival analysis has a rigorous mathematical foundation. Counting stochastic processes and the Martingale central limit theory form the survival analysis models as stochastic integrals, which provide insight for analytic solutions. (2) In ANN, simulation is usually necessary (Robert & Casella (2004), which is not the case in survival analysis.

Our conjecture may explain a very interesting phenomenon. Several studies have tried to integrate ANN with survival analysis. As reviewed in the next section, few of the integrated survival analysis and ANN made significant difference in terms of model fittings, compared with the native survival analysis alone. The indifference shows that both approaches *do not* complement each other. If they are essentially different, the integrated approach should produce some results that are significantly different from the pure survival analysis alone, either positively or negatively. Again, we emphasize that our discussion is still a pure conjecture at this stage.

3. BRIEF CASE REVIEW OF SURVIVAL ANALYSIS APPLICATIONS

3.1. Applications Found in IEEE Digital Library.

In this section, we briefly review the papers found in the IEEE digital library with the keyword of "*survival analysis*" search. The online search was conducted in the July of 2007, and we found about 40 papers in total. There were a few biomedical studies among the 40 survival-analysis papers published in IEEE publications. These are largely standard biomedical applications and we do not discuss these papers, for obvious reasons.

Mazzuchi et al. (1989) seemed to be the first to *actively* advocate the use of Cox's (1972) proportional hazards model (PHM) in engineering reliability. They quoted Cox's (1972) original words "*industrial reliability studies and medical studies*" to show Cox's original motivation. Mazzuchi et al (1989) stated, "*while this model had a significant impact on the biomedical field, it has received little attention in the reliability literature.*" Nearly two decades after the introduction paper of Mazzuchi et al. (1989), it appears that little significant changes have occurred in computer science and IEEE related engineering fields with regard to the proportional hazards models and survival analysis as a whole.

Stillman et al. (1995) used survival analysis to analyze the data for component maintenance and replacement programs. Reineke et al. (1998) conducted a similar study for determining the optimal maintenance by simulating a series system of four components. Berzuini and Larizza (1996) integrated time-series modeling with survival analysis for medical monitoring. Kauffman and Wang (2002) analyzed the Internet firm survival data from IPO (initial public offer) to business shutdown events, with survival analysis models.

Among the 40 survival analysis papers, which we obtained from online search of the IEEE digital library, significant percentage is the integration of survival analysis with artificial neural networks (ANN). In many of these studies, the objective was to utilize ANN to modeling fitting or parameter estimation for survival analysis. The following is an incomplete list of the major ANN survival analysis integration papers found in IEEE digital library, Arsene et al. (2006), Bakker and Heskes (1999), Cawley et al. (2006), Eleuteri et al. (2003), Lisboa and Wong (2001), Mani et al. (1999). The parameter estimation in survival analysis is particular complex due to the requirements for processing censored observations. Therefore, approach such as ANN and Bayesian statistics may be helpful to deal with the complexity. Indeed, Bayesian survival analysis has been expanded significantly in recent years (Ibrahim, et al. 2005). We expect that evolutionary computing will be applied to survival analysis, in similar way to ANN and Bayesian approaches.

With regard to the application of ANN to survival analysis, we suggest three cautions: (1) The *integrated* approach should preserve the capability to process censoring; otherwise, survival analysis loses its most significant advantage. (2) Caution should be taken when the integrated approach changes model structure because most survival analysis models, such as Cox's proportional hazards models and accelerated failure time models, are already complex nonlinear models with built-in failure mechanisms. The model over-fitting may cause model *identifiability* problems, which could be very subtle and hard to resolve. (3) If the integrated approach does not produce significant improvement in terms of model fitting or other measurements, which seemed to be the case in majority of the ANN approach to survival analysis, then the extra complication should certainly be avoided. Even if there is improvement, one should still take caution with the *censor-handling* and model *identifiability* issues previously mentioned in (1) and (2).

3.2. Selected Papers Found in MMR-2004.

In the following, we briefly review a few survival analysis related studies presented in a recent International Conference on Mathematical Methods in Reliability, MMR 2004 (Wilson et al. 2005).

Pena and Slate (2005) addressed the dynamic reliability. Both reliability and survival times are more realistically described with dynamic models. Dynamic models generally refer to the models that incorporate the impact of actions or interventions as well as their accumulative history, which

can be monitored (Pena and Slate 2005). The complexity is obviously beyond simple regression models, since the dependence can play a crucial role. For example, in a load-sharing network system, failure of a node will increase the loads of other nodes and influences their failures.

Duchesne (2005) suggested incorporating usage accumulation information into the regression models in survival analysis. To simplify the model building, Duchesne (2005) assumes that the usage can be represented with a single time-dependent covariate. Besides reviewing the hazard-based regression models, which are common in survival analysis, Duchesne (2005) reviewed three classes of less commonly used regression models: models based on transfer functionals, models based on internal wear and the so-called collapsible models. The significance of these regression models is that they expand reliability modeling to two dimensions. One dimension is the calendar time and the other is the usage accumulation. Jin (2005) reviewed the recent development in statistical inference for accelerated failure time (AFT) model and the linear transformation models that include Cox proportional hazards model and proportional odds models as special cases. Two approaches, rank-based approach and least-squares approach were reviewed in Jin (2005). Osborn (2005) presented a case study of utilizing the remote diagnostic data from embedded sensors to predict system aging or degradation. This example should indicate the potential of survival analysis and competing risks analysis in the *prognostic and health management* (PHM) since the problem Osborn (2005) addressed is very similar to PHM. The uses of embedded sensors to monitor the health of complex systems, such as power plants, automobile, medical equipment, and aircraft engine, are common. The main uses for these sensor data are real time assessment of the system health and detection of the problems that need immediate attention. Of interest is the utilization of those remote diagnostic data, in combination with historical reliability data, for modeling the system aging or degradation. The biggest challenge with this task is the proper *transformation* of *wear time*. The wear is not only influenced by internal (temperature, oil pressures, etc) and external covariates (ambient temperature, air pressure, etc), but also different each time.

4. SUMMARY AND PERSPECTIVE.

4.1. Summary

Despite the common foundation with traditional reliability theory, such as the same probability definitions for *survival function* [$S(t)$] and *reliability* [$R(t)$], i.e., $S(t)=R(t)$, as well as the *hazards function* (the exact same term and definition are used in both fields), survival analysis has not achieved similar success in the field of reliability as in biomedicine. The applications of survival analysis seem still largely limited to the domain of biomedicine. Even in the sister fields of biomedicine such as biology and ecology, few applications have been conducted (Ma 1997, Ma and

Bechinski 2008). In the engineering fields, the Meeker and Escobar (1998) monograph, as well as the Klein and Goel (1992) still seem to be the most comprehensive treatments.

In Section 2, we reviewed the essential concepts and models of survival analysis. In Section 3, we briefly reviewed some application cases of survival analysis in engineering reliability. In computer science, survival analysis seems to be still largely unknown. We believe that the potential of survival analysis in computer science is much broader than network and/or software reliability alone. Before suggesting a few research topics, we note two additional points.

First, in this article, we exclusively focused on univariate survival analysis. There are two other related fields: one is *competing risks analysis* and the other is *multivariate survival analysis*. The relationship between multivariate survival analysis and survival analysis is similar to that between multivariate statistical analysis and mathematical statistics. The difference is that the extension from univariate to multivariate in survival analysis has been much more difficult than the development of multivariate analysis, because of (1) observation censoring, and (2) dependency, which is much more complex when multivariate normal distribution does not hold in survival data. On the other hand, the two essential differences indeed make multivariate survival analysis unique and extremely useful for analyzing and modeling time-to-event data. In particular, the advantages of multivariate survival analysis in addressing the dependency issue are hardly matched by any other statistical method. We discuss competing risks analysis and multivariate survival analysis in separate articles (Ma and Krings 2008a, b, & c).

The second point we wish to note is: if we are asked to point out the counterpart of survival analysis model in reliability theory, we would suggest the *shock* or *damage* model. The shock model has an even longer history than survival analysis and the simplest shock model is the *Poisson process* model that leads to exponential failure rate model. The basic assumption of the shock model is the notion that engineered systems endure some type of wear, fatigue or damage, which leads to the failure when the strengths exceed the tolerance limits of the system (Nakagawa 2006). There are extensive research papers on shock models in the probability theory literature, but relatively few monographs. The monograph by Nakagawa (2006) seems to be the latest. The shock model is often formulated as a *Renewal stochastic process* or *point process* with Martingale theory. The rigorous mathematical derivation is very similar to that of survival analysis from *counting stochastic processes,* which we briefly outlined in section 2.6.

It is beyond the scope of the paper to compare the shock model with survival analysis; however, we would like to make the two specific comments. (1) Shock models are essentially the stochastic process model to capture the failure mechanism based on the damage induced on the system by shocks, and the resulting statistical models are

often the traditional reliability distribution models such as exponential and Weibull failure distributions. Survival analysis does not depend on specific shock or damage. Instead, it models the failure with abstract time-to-event random variables. Less restrictive assumptions with survival analysis might be more useful for modeling software reliability where the notions of fatigue, wear and damage apparently do not apply. (2) Shock models do not deal with information censoring, the trademark of failure time data. (3) Shock models, perhaps due to mathematical complexity, have not been applied widely in engineering reliability yet. In contrast, the applications of survival analysis in biomedical fields are much extensive. While there have been about a dozen monographs on survival analysis available, few books on shock models have been published. We believe that survival analysis and shock models are complementary, and both are very needed for reliability analysis, with shock model more focused on failure mechanisms and the survival analysis on data analysis and modeling.

4.2. Perspective.

Besides traditional industrial and hardware reliability fields, we suggest that the following fields may benefit from survival analysis.

Software reliability—Survival analysis is not based on the assumptions of wear, fatigue, or damage, as in traditional reliability theory. This seems close to software reliability paradigm. The single biggest challenge in applying above discussed approaches to software systems is the requirement for a new metric that is able to replace the survival time variable in survival analysis. This "time" counterpart needs to be capable of characterizing the "vulnerability" of software components or the system, or the "*distance*" to the next failure event. In other words, this software metric should represent the metric-to-event, similar to time-to-event random variable in survival analysis. We suggest that the Kolmogorov complexity (Li and Vitanyi 1997) can be a promising candidate. Once the metric-to-event issue is resolved, survival analysis, both univariate and multivariate survival analysis can be applied in a relative straightforward manner. We suggest that the shared frailty models are most promising because we believe latent behavior can be captured with the shared frailty (Ma and Krings 2008b).

There have been a few applications of univariate survival analysis in software reliability modeling, including examples in Andersen et al.'s (1995) classical monograph. However, our opinion is that without the fundamental shift from time-to-event to new metric-to-event, the success will be very limited. In software engineering, there is an exception to our claim, which is the field of software test modeling, where time variable may be directly used in survival analysis modeling.

Modeling Survivability of Computer Networks—As described in Subsection 2.3, random censoring may be used to model network survivability. This modeling scheme is particularly suitable for wireless sensor network, because (1) of the population nature of wireless nodes and (2) the limited lifetime of the wireless nodes. Survival analysis has been advanced by the needs in biomedical and public health research where population is the basic unit of observation. As stated early, a sensor network is analogically similar to a biological population. Furthermore, both organisms and wireless nodes are of limited lifetimes. A very desirable advantage of survival analysis is that one can develop a unified mathematical model for both the reliability and survivability of a wireless sensor network (Krings and Ma 2006).

Prognostic and Health Management in Military Logistics—PHM involves extensive modeling analysis related to reliability, life predictions, failure analysis, burn-in elimination and testing, quality control modeling, etc. Survival analysis may provide very promising new tools. Survival analysis should be able to provide alternatives to the currently used mathematical tools such as ANN, genetic algorithms, and Fuzzy logic. The advantage of survival analysis over the other alternatives lies in its unique capability to handle information censoring. In PHM and other logistics management modeling, information censoring is a near universal phenomenon. Survival analysis should provide new insights and modeling solutions.

5. REFERENCES

Aalen, O. O. 1975. Statistical inference for a family of counting process. Ph.D. dissertation, University of California, Berkeley.

Andersen, P. K., O. Borgan, R. D. Gill, N. Keiding. 1993. Statistical Models based on Counting Process. Springer.

Arsene, C. T. C., et al. 2006. A Bayesian Neural Network for Competing Risks Models with Covariates. MEDSIP 2006. Proc. of IET 3rd International Conference On Advances in Medical, Signal and Info. 17-19 July 2006.

Aven, T. and U. Jensen. 1999. Stochastic Models in Reliability. Springer, Berlin.

Bazovsky, I. 1961. Reliability Theory and Practice. Prentice-Hall, Englewood Cliffs, New Jersey.

Bakker, B. and T. Heskes. 1999. A neural-Bayesian approach to survival analysis. IEE Artificial Neural Networks, Sept. 7-10, 1999. Conference Publication No. 470. pp832-837.

Berzuini, C. and C. Larizza. 1996. A unified approach for modeling longitudinal and failure time data, with application in medical monitoring. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 18(2):109-123.

Bollobás, B. 2001. Random Graphs. Cambridge University Press; 2nd edition. 500pp.

Cawley, G. C., B. L. C. Talbot, G. J. Janacek, and M. W. Peck. 2006. Sparse Bayesian Kernel Survival Analysis for Modeling the Growth Domain of Microbial Pathogens.

Chiang C. L. 1960. A stochastic study of life tables and its applications: I. Probability distribution of the biometric functions. Biometrics, 16:618-635.

Cox, D. R. 1972. Regression models and life tables. *J. R. Stat. Soc. Ser. B*. 34:184-220.

Cox, D. R. 1975. Partial likelihood. *Biometrika* 62:269-276.

Cox, D. R. & D. Oakes. 1984. *Analysis of Survival Data*. Chapman & Hall. London.

Cressie, N. A. 1993. *Statistics for Spatial Data*. John Wiley & Sons. 900pp.

Duchesne, T. 2005. Regression models for reliability given the usage accumulation history. In "Modern Statistical and Mathematical Methods in Reliability". Edited by A. Wilson, N. Limnios, S. Kelly-McNulty, Y. Armijo. pp.29-40. World Scientific, New Jersey.

Eleuteri, A., R. Tagliaferri, L. Milano, G. Sansone, D. D'Agostino, S. De Placido, M. Laurentiis. 2003. Survival analysis and neural networks. Proceedings of the International Joint Conference on Neural Networks, Vol. 4, 20-24 July 2003 Page(s):2631 - 2636.

Ellison, E., L. Linger, and M. Longstaff. 1997. Survivable Network Systems: An Emerging Discipline, Carnegie Mellon, SEI, Technical Report CMU/SEI-97-TR-013, 1997.

Fleming, T. R. & D. P. Harrington. 1991. Counting process and survival analysis. John Wiley & Sons. 429pp.

Graver, J. and M. Sobel 2005. You may rely on the Reliability Polynomial for much more than you might think. *Communications in Statistics: Theory and Methods*. 34(6):1411-1422

Graves, T. and M. Hamada. 2005. Bayesian methods for assessing system reliability: models and computation. In "Modern Statistical and Mathematical Methods in Reliability". Edited by A. Wilson, et al. pp.41-53.

Grimmett, G. 2006. *The Random-Cluster Model*. Springer.

Grimmett, G. 1999. *Percolation*. Springer.

Hougaard, P. 2000. Analysis of Multivariate Survival Data. Springer. 560pp.

Ibrahim, J. G., M. H. Chen and D. Sinha. 2005. Bayesian Survival Analysis. Springer. 481pp.

Jin Z. 2005. Non-proportional semi-parametric regression models for censored data. In "Modern Statistical and Mathematical Methods in Reliability". Edited by A. Wilson, N. Limnios, S. Kelly-McNulty, Y. Armijo. pp.279-292. World Scientific.

Kalbfleisch, J. D. & R. L. Prentice. 1980. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons. New York. 1980.

Kalbfleisch, J. D. & R. L. Prentice, 2002. The Statistical Analysis of Failure Time Data. Wiley-InterScience, 2nd ed. 462pp.

Lisboa, P. J. G. and H. Wong. 2001. Are neural networks best used to help logistic regression? Proceedings of International Joint Conference on Neural Networks, IJCNN '01. Volume 4, 15-19, July 2001. Page(s):2472 - 2477 vol.4

Kauffman, R. J. and B. Wang. 2002. Duration in the Digital Economy. Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03). 6-9 Jan 2003

Kaplan, E. L. & P. Meier. 1958. Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc*. 53:457-481.

Klein, J. P. and P. K. Goel. 1992. Survival Analysis: State of the Art. Kluwer Academic Publishes. 450pp.

Klein, J. P. and M. L. Moeschberger. 2003. Survival analysis techniques for censored and truncated data. Springer.

Krings, A. and Z. S. Ma. 2006. "Fault-Models in Wireless Communication: Towards Survivable Ad Hoc Networks", MILCOM 2006, Military Communications Conference, 23-25 October, 7 pages, 2006.

Krings, A. W. 2008. Survivable Systems. in "*Information Assurance: Dependability and Security in Networked Systems*." Yi Qian, James Joshi, David Tipper, and Prashant Krishnamurthy, Morgan Kaufmann Publishers. (in press)

Lawless, J. F. 1982. Statistical models and methods for lifetime data. John Wiley & Sons. 579pp.

Lawless, J. F. 2003. Statistical models and methods for lifetime data. John Wiley & Sons. 2nd ed. 630pp.

Li, M. and P. Vitanyi. 1997. Introduction to Kolmogorov Complexity and Its Applications. 2nd ed, Springer.

Ma, Z. S. 1997. Survival analysis and demography of Russian wheat aphid populations. Ph.D dissertation, 307pp, University of Idaho, Moscow, Idaho, USA.

Ma, Z. S., and E. J. Bechinski. 2008. Developmental and Phenological Modeling of Russian Wheat Aphid. *Annals of Entomological Soc. Am*. Vol. 101 (2) (In press).

Ma, Z. S. and A. W. Krings. 2008a. The Competing Risks Analysis Approach to Reliability Survivability, and Prognostics and Health Management. The 2008 IEEE-AIAA AeroSpace Conference. BigSky, Montana, March 1-8, 2008. (In Press, in the same volume)

Ma, Z. S. and A. W. Krings. 2008b. Multivariate Survival Analysis (I): The Shared Frailty Approaches to Reliability and Dependence Modeling. The 2008 IEEE-AIAA AeroSpace Conference. BigSky, Montana, March 1-8, 2008. (In Press, in the same volume).

Ma, Z. S. and A. W. Krings. 2008c. Multivariate Survival Analysis (II): Multi-State Models in Biomedicine and Engineering Reliability. 2008 IEEE International Conference on Biomedical Engineering and Informatics (BMEI 2008). May 27th-30th, 2008. (*Accepted*)

Mani, R., J. Drew, A. Betz, P. Datta. 1999. Statistics and Data Mining Techniques for Lifetime Value Modeling ACM Conf. on Knowledge Discovery and Data Mining.

Mazzuchi, T. A., R. Soyer., and R. V. Spring. 1989. The proportional hazards model in reliability. IEEE Proceedings of Annual Reliability and Maintainability Symposium. pp.252-256.

Meeker, W. Q. and L. A. Escobar. 1998. Statistical Methods for Reliability Data. Wiley-Interscience.

Munson, J. C. 2003. Software Engineering Measurement. Auerbach Publications.

Nelson, W. 1969. Hazard plotting for incomplete failure data. *J. Qual. Tech*. 1:27-52.

Nakagawa, T. 2006. Shock and Damage Models in Reliability Theory. Springer.

Osborn, B. 2005. Leveraging remote diagnostics data for predictive maintenance. In "Modern Statistical and Mathematical Methods in Reliability". Edited by A. Wilson, N. Limnios, S. Kelly-McNulty, Y. Armijo. pp. 353-363.

Pena, E. A. and E. H. Slate. 2005. Dynamic modeling in reliability and survival analysis. In "Modern Statistical and Mathematical Methods in Reliability". Edited by A. Wilson, N. Limnios, S. Kelly-McNulty, Y. Armijo. pp.55-71.

Reineke, D. M., E. A. Pohl, and W. P. Murdock. 1998. Survival analysis and maintenance policies for a series system, with highly censored data. 1998 Proceedings, Annual Reliability and Maintainability Symposium. pp. 182-188.

Schabenberger, O. and C. A. Gotway. 2005. Statistical Methods for Spatial Data Analysis. Chapman & Hall/CRC.

Severini, T. A. 2000. Likelihood methods in statistics. Oxford University Press.

Shooman, M. L. 2002. Reliability of Computer Systems and Networks: Fault Tolerance, Analysis and Design. John Wiley and Sons. 551pp.

Stillman, R. H. and M. S. Mackisack, B. Sharp, and C. Lee. 1995. Case studies in survival analysis of overhead line components. IEE Conference of the Reliability and Distribution Equipment. March 29-31, 1995. Conference Publication No. 406. pp210-215.

Therneau, T. and P. Grambsch. 2000. *Modeling Survival Data: Extending the Cox Model*. Springer.

Wilson, A. N. Limnios, S. Kelly-McNulty, Y. Armijo. 2005. Modern Statistical and Mathematical Methods in Reliability. World Scientific, New Jersey.

Xie, M. 1991. Software Reliability Modeling. World Scientific Press.

BIOGRAPHY

Zhanshan (Sam) Ma holds a Ph.D. in Entomology and is a Ph.D. candidate in Computer Science at the University of Idaho. He has published approximately 30 journal and 30 conference papers, mainly in the former field. Prior to his recent return to academia, he worked as senior network/software engineers in software industry. His current research interests include reliability and survivability of wireless sensor networks, fault tolerance, survival analysis, evolutionary game theory, evolutionary computation and bioinformatics.

Axel W. Krings is a professor of Computer Science at the University of Idaho. He received his Ph.D. (1993) and M.S. (1991) degrees in Computer Science from the University of Nebraska - Lincoln, and his M.S. (1982) in Electrical Engineering from the FH-Aachen, Germany. Dr. Krings has published extensively in the area of Computer & Network Survivability, Security, Fault-Tolerance and Real-time Scheduling. In 2004/2005 he was a visiting professor at the Institut d'Informatique et Mathématiques Appliquées de Grenoble, at the Institut National Polytechnique de Grenoble, France. His work has been funded by DoE/INL, DoT/NIATT, DoD/OST and NIST.