Chapter 13
Model–Based Methodology and Framework for Assessing Service and Business Process Availability

Nikola Milanovic
Model Labs - Berlin, Germany

Bratislav Milic
Humboldt University, Germany

ABSTRACT
In the world where on-demand and trustworthy service delivery is one of the main preconditions for successful business service and business process, availability is of the paramount importance and cannot be compromised. This chapter presents a framework for modeling business process availability that takes into account services, the underlying ICT-infrastructure and people. Based on a fault model, the chapter develops the methodology to map dependencies between ICT-components, services and business processes. The mapping enables automatically derived appropriate availability models and analytically assessed steady-state, interval and user perceived availability at all levels, up to the level of the business process. The chapter demonstrates the applicability of the mapping using two case studies. Finally, it defines a roadmap towards model-based service management and position this work within its context.

INTRODUCTION
Service Oriented Architecture (SOA) is the paradigm that pretends to play a dominant role in the shaping of the Information Technology (IT) landscape in the coming decades. It represents a logical and evolutionary step initiated by development in the areas of distributed computing, business process modeling, and the increased ubiquity of networking technologies. The main goal of SOA
is to introduce standard methodologies, architectures, tools, languages and protocols for development and integration of distributed applications based on the loosely coupled, independent and autonomous software artifacts, thus supporting the large-scale composability, reusability and agility.

Trustworthy service delivery in SOA is at present one of the main preconditions for the successful and sustainable business operations. Therefore, service and business process availability cannot be compromised. Even today, services are simply expected to be delivered reliably and on demand, and this requirement will be even more important in the near future. The unreliable and incorrect services (Amazon EC-2 Support Team, 2007) can corrupt business processes causing an impact such as lost opportunity or money. Common understanding of the service and business process availability properties is rather sketchy, limited and mostly empirical.

Several methodologies can be used to assess service and business process availability: quantitative, qualitative and analytical. Quantitative assessment is based on the real-time measurement and monitoring. Whereas it has proven itself in several areas (e.g., hardware benchmarks and testing), it is difficult to apply to services because of the lack of adequate metrics and instrumentation. Qualitative availability assessment is performed informally (e.g., through the questionnaires and interviews) and assigns an availability class to the system (service). The qualitative results are easy to misinterpret, difficult to compare and depend heavily on the consultant performing the analysis. The analytical methods are used to attempt to model services and their behavior and calculate or simulate their availability. Up to now, however, classical analytical methods have been applied to determine service availability with mixed success and relative low industry penetration due to the scalability, complexity and evolution problems.

DEFINITIONS

In this section we present the basic (and informal, that is, non-mathematical) definitions of reliability and availability, as well as definitions of services and business processes, as all these terms are used very colloquially today, and also frequently outside of the strict computer science vocabulary and context. Such definitions will suffice to comprehend the overall context of the proposed framework, for a deeper mathematical treatment of the subject many of the references in this section provide enough detailed information.

The events that lead to the system malfunction have intrinsic probabilistic nature. Therefore, the lifetime or time to failure of a system can usually be represented by a random variable. A phenomenon is considered random if its future behavior is not exactly predictable. An example is tossing a pair of dice or measuring the time between particle emissions by a radioactive sample. A function that associates a number with every possible outcome of an event is called a random variable. Let the random variable \( X \) represent the lifetime or time to failure of a system. The continuous random variable \( X \) can be characterized by the cumulative distribution function \( F(t) \), the probability density function \( f(t) \) and the hazard rate function \( h(t) \), also known as the instantaneous failure rate. The CDF represents the probability that the system will fail before a given time, the PDF describes the rate of change of the CDF, and the hazard rate function represents the conditional probability that a system that has survived until time \( t \) will fail in the given subsequent interval.

Time to failure or lifetime of a system may follow different distributions, such as exponential, Weibull, geometric, Erlang etc. These distributions are parameterized, e.g., with parameters such as failure or repair rate.