An Approach to Modelling Information System Availability by Using Bayesian Belief Network

Semir Ibrahimović
BBI Bank, School of Economics and Business Sarajevo, Bosnia and Herzegovina
Nijaz Bajgorić
School of Economics and Business Sarajevo, Bosnia and Herzegovina

Abstract

In today's era of the ubiquitous use of information technology (IT), it is expected that the information systems provide services to end-users on continuous basis, regardless of time and location. This is especially true in organizations where information systems (IS) support real-time critical operations, particularly, in the industries in which these systems must continuously operate 24x7x365. This paper presents a modified Bayesian Belief Network model for predicting IS availability. Based on a thorough review of all IS availability dimensions, we proposed a modified set of determinants. The model is parametrized using probability elicitation process with the participation of experts from the BiH financial sector. The results showed that most influential determinants of the IS availability are a timely and precise definition of the availability requirements, quality of IT operations, management and network.

Keywords: Information Systems, Business Continuity, Availability, Bayesian Belief Network, Monte-Carlo Simulation
JEL classification: M15

Introduction

Modern information systems (IS) are expected to be up and running,available and scalable, in providing services on always-on basis. This is particularly important for organizations and industries where information systems support real-time and mission-critical operations that need to be available "24x7x365". Examples of such organizations include process industries, telecommunications, healthcare, energy, banking, electronic commerce and avariety of cloud services.The primary objective of this research is to determine aset of factors that have the greatest influence on IS availability in BiH financial institutions, while the second objective is to compare the results of model parameterization with the results provided by Franke et al. (2012). The first section presents the literature review. Second sectiondeals withthe description of BBN which are used as a tool for the analysis of the factors influencing the IS availability. In the rest of the paper, the results areshortly presented and discussed, as well as conclusions and future research directions.

Modelling IS Availability – Literature Review

IEEE defined IS availability as "The degree to which a system or component is operational and accessible when required for use. Often expressed as a probability."(IEEE, 1990). ISO 27000 series of standards bound the availability to the concept of organizational assets. The asset is available if it is available and ready for use at the request of an authorized person. In the context of this standard, the property includes things such as information systems, facilities, networks and computers (ISO / IEC, 2012). Singh (2009) provided a quantitative definition: "Ps - system availability of the observed system S, is probability that a system is operational and ready to provide services. As the Psnumber should be close to 100% as possible, the usual way to view the availability of the system is in “number ofnines”. By using this approach, the 99.999% availability is called “five nines”. Availability can also be seen as a combination of three concepts: reliability, availability (accessibility) and timeliness (Somasundaram et al. 2009). The IS availability is referred to as part of the of the CIA (confidentiality, integrity, availability) triangle of information security (Martin et al., 2006), (Bodin et al., 2005), (Bodin et al., 2008), (Ioannidis et al., 2009), (Faisst et al.2008), (Kouns et al., 2010), (Chen et al., 2011), (Hole, 2010).Another term is used as well, “IS-resilience”, which implies that the system "must remain available and maintain an acceptable level of performance when faced with various types of errors affecting the normal operation" (Liu et al., 2010). Gaddum (2004) discussed the concept of resilience as IT,organizational and business issue, and introduced a model with six layers of resistance: strategy, organization, processes, data and applications, technologies and facilities. Shiesser (2010) observed availability as optimization process of the productions systems’readiness for accurately measuring, analysing and reducing the system downtime.Martin (2003) identified six major determinants of IS availability, namely: physical security, audit and evaluation of the system efficiency, security policy, system monitoring and control of operations, business continuity management and backup management. Franke et al. (2012) used 16 determinants that affect the availability of the system based on the "Index of availability", introduced by Marcus et al. (2003). Rauscher et al. (2006) proposed a model for the reliability of communications infrastructure and identified the following components that affect reliability: human factor, policies, hardware, software, network, load, environment and power. Different recommendations to raise the IS availability aregiven in: Liu et al., (2010), Raderius et al. (2009). Franke et al. (2009) suggested improvements of IS architecture, while Martin (2003) recommended improving security policies. Gay (2007) suggestedusing virtualization technology, while Calzolari (2006) recommended clustering and virtualization. Chen et al. (2011) suggested a strategy of diversification as a possible solution for reducing the IS downtime. Bell (2005) proposed using best practices in designing a data centreto improve the IS availability. Raderius et al. (2009) identified block diagrams reliability and Monte-Carlo simulations as the most frequentlyused reliability modeling tool. Markov chains were used as an IS-availability modeling tool by Goyal et al. (1987). Immonen et al. (2007) defined the framework for the comparison of methods for reliability and availability modelingfrom the software architecture perspective.

Methodology

In this article, Bayesian Belief Networks (BBN) are used as a tool for the analysis of the factors influencing the IS availability. Although Bayesian networks significantly reduce the number of parameters, which needs to be determined by specifying the joint probability distribution, the number of parameters in the model remains one of the major bottlenecks of this framework. One way to reduce this number is to assume a functional relationship that defines the interaction between all the parents of a node. The most widely accepted and applied solutions for this problem is the Noisy-OR model (Pearl, 1988). Noisy-OR model gives acausal interpretation to theinteraction between the parent node and child node. It assumes that all causes (parents) are independent of each other in terms of their ability to influence the variable effect (the child). Given these assumptions Noisy-OR model provides a logarithmic reduction in the number of parameters required for the construction of the CPT, which effectively makes the building of large models for real life problems feasible. In practical models, a situation where the absence of all causes modelled ensures the lack of impact almost never happens. To solve that weakness of Noisy-OR model, Henrion (1989) introduced the concept of a leakage or background probabilities that allows modeling the impact of a combination of factors that are not included in the model.

BBN have been widely applied in OpRisk, INFOSEC and availability modeling. Raderius et al. (2009) presented a case study where the availability of the information system was estimated using the "extended influence diagrams" combined with an architectural metamodel. Hinz et al. (2006) presented BBN model for assessing the risk of IT infrastructure whileWeber et al. (2001) used the influence diagrams for the economic analysis of the IS availability. Neil et al. (2008) suggested a methodology for developing BBN model representing the operational risk of IT infrastructure in the financial institutions. Wei et al. (2011) developed an integrated process, based on BBN, for the purposes of efficient IT services management. Sommestad et al. (2009) suggested a model based on the extended influence diagram, which enable the analysis of thecyber security of different architectural solutions. Simonsson et al. (2008) proposed a model for measuring IT governance efficiency based on BBN. Lande et al. (2010) modelled the critical information systems, using BBN. Zhang et al. (2009) presented an innovative model to improve the availability of the system based on the BBN in which the data for the CPT obtained from the system logs. Bonafede (2007) made areview of statistical methods that can be used to model business continuity and gave anexample of BBN use for that purpose. Different models based on BBN were created in the area of software reliability as well (Yu et al. 2009), (Dejaeger et al. 2012), (Gran 2002) and application development projects (Fenton et al. 2010), (Fineman 2010), (Radliński et al. 2007). Franke et al. (2012) presented the model in the area of IS availability based on Leaky Noisy-OR BBN. The model parameters were obtained based on the probability elicitation of 50 experts in the IS availability field. Their model, with modifications based on the theoretical dimension,has been applied in this research.

We propose a model that consists of 13 variables representing 13 domains affecting information systems availability. The probability elicitation used for determining the model parameters was done by interviewing 23 experts with experience in dealing with IT systems availability in the financial sector in BiH. The research focused on information systems in the BiH banking industry. During the elicitation, most experts agreed that the selection of variables in the model is adequate and that the model is comprehensive. Elicitation was conducted through structured interviews. In the first part of interviews, experts were trained and calibrated, while in the second part experts are asked to fill in the questionnaire. The questionnaire consisted of three sets of questions. Experts were first asked to estimate the impact of individual variables on system availability. In the second question, experts gave their assessment of the situation in the areas described by the variables in the financial sector in BiH. To answer the third question, they estimated the necessary investments to bring the field represented by variable to the level of best practices.

Overall system availability is defined as the average availability of each service weighted by a factor of importance of a service (for example different importance is given for payment card authorization service and service that calculates fixed assets depreciation). Equation below was used for availability calculation.

In this formula A- represents overall system availability, A_i- availability of service s_i and k_i coefficient significance of service s_i.

$$A = \frac{\sum_i A_i * k_i}{\sum_i k_i}$$

When calculating the availability of a particular service one should take into account the service operationg time, defined in the service level agreement, as well as the number of clients affected with the service interruption.

The availability of a particular service is calculated according to the following formula:

$$A_i = \frac{t_i}{t_i + ut_i * \frac{un_i}{n_i}}$$

In formula t_i is total time that service $s_i$was available under service level agreement, ut_i -total time for which the system was unavailable, n_i -the total number of the service users, un_i-the number of service users where experienced service interruption during time ut_i.Model is composed of 13 independent variables that affect the IS availability: the physical environment, availabilityrequirements management, operations management, change management, backup management , storage redundancy, avoiding errors in internal applications, avoiding errors in external services , network management, equipment and location of the DR datacentre, resistant client / server systems, monitoring of relevant components, human resources management.According to the Leaky Noisy-OR model, the following formula is used to calculate the probability of IS availability;

$$P(\mathrm{A}) = (1 - p_0) \prod_{i \in [1,\mathrm{n}]} (1 - p_i) = (1 - p_0) * \prod_{i \in [1,n]} (1 - \mathrm{k}V_i * (1 - B_i))$$

where n represents number of variables in the model, V_i percentage of the system's unavailability in case of application of the best practices, B_i - state of implementation of best practices in the areas of system, k transformation coefficient, p_0- leak representing probability that the system is unavailable in the case that for all domains included in the model, best practices are applied.

Results and Discussions

Research has shown that managing availability requirements makes the greatest impact on the availability (23.20%), followed by operations management (20.54%) and the location and equipment of the disaster recovery site (19.52%). The reduction of IS unavailability is least impacted by the physical environment (10.53%), followed by backup management (11.05%) and server platforms (11.81%). Research results showed that the state of implementation of best practices in the areas described by variables ranges from 4.60 to 6.85 on a scale from 1 to 10 depending on the area. The worst situation is in the fields of monitoring (4.6) and availability requirements management (4.94). The best state in the IS of financial institutions in BiH is in the primary infrastructure areas, backup management (6.85), network management (6:54), server infrastructure (6.39) and physical environment (6.05).According to the results of this research, the perception of experts is that the state of the essential IS infrastructure elements, including the server room, server and network infrastructure, data redundancy, backup management is much better than the process part, which includes change management, operations management, monitoring and requirements management.Assessment of current maturity level of backup management may explain why the experts estimated that implementing best practices in that area would have a small impact on reducing unavailability, as the situation in that field has been assessed as the best compared to all other areas that were part of the model. A similar explanation applies to the physical environment and server infrastructure. The above was the main reason to include assessed states of implementing best practices in the field as the prior probability for each parent node in BBN-based "Leaky Noisy-OR" model. The conditional probability table for the node that represents availability is filled based on a linear transformation of the elicited impact values. The model is set up assuming the initial system availability of 99% and a leak of 0.01%, which represents the unavailability of the system. Both of these parameters can be subsequently changed.

We compared the results with the studymade by Franke et al. (2012). To be able to compare the results, it was necessary to transform research results, since thedifferent method of calculating variables impact on the IS availability were used.Research findings and comparisonare shown in Table1. Column 1 represents theeffects of each variable on IS availability, where resulting percent is mean of experts opinion. The second column represents expert's opinions about the maturity level in the financial sector in BiH using scale 1-10. Third columns represent research results, whereasresulting impact was taken themode of experts' opinions modified with adjustment intervals. The fourth column represents Franke et al.'s results. The fourth, fifth and sixth columns represent ranking of data presented in first, third and fourth columns respectively.

Table 1
Elicitation results compared with Franke et al. (2012)

Variable	Percent median	Maturity level	Percent as mode	Percent Franke	PS	PP	PF
Operations Management	20.54	4.94	8.85	23	2	4	4
Network management	19.53	5.24	6.99	12.95	3	7	6
Equipment and location of the DR data centre	19.52	6.54	2.31	3.35	4	13	12
Avoiding errors in internal applications	17.58	5.28	6.44	16.86	5	10	5
Avoiding errors in external services	17.52	5.93	7.81	8.66	6	5	7
Change management	17.15	5.88	25.65	28.08	8	2	1
Human resources management	15.45	5.76	6.92	-	9	8	-
Backup management	11.05	5.56	7.65	5.27	12	6	10
Availability requirements management	23.2	5.44	29.64	25.25	1	1	3
Storage redundancy	17.2	4.60	5	5.34	7	11	9
Resistant client / server systems	11.81	6.39	3.94	3.65	11	12	11
Physical environment	10.54	6.85	10	8.22	13	3	8
Monitoring	12.33	6.05	6.71	26.14	10	9	2

Source: Authors

Comparing data, we can notice that according to Franke et al.'s research, there are more IS availability determinants with animpact greater than 20%. In our study, only two variables have an impact over 20% while the other variables impact IS availability with the 10% and less. Among the four most influentialareas, both studies identified the same three areas: "Availability requirements management", "Change Management" and " Operations management". Also, both studies have shown that least influential are variables“resistant client/server systems” and “DR equipment and location”. The biggest differences are in the areas of “monitoring of relevant components” and “avoiding errors in internal applications”. In Franke et al research, they have a greater impact (2nd and 5th position) compared withourresearch (9th and 10th). Contrary “physical environment” has significant impact according our results (3th place) in comparison with Franke et al.’sresearch (8th place).

One of the biggest disadvantages of the proposed model is the deterministic determination of parameters. In other words, each parameter in the network is set on the basis of the weighted mean values obtained in the elicitation process, not reflecting the diversity of expert’s opinion. For this reason, the same mathematical model was implemented using Microsoft Excel and Oracle Crystal Ball software. The base values of the input variables were set in the same way as the parameters of BBN. However, each input variable is represented not only the mean value but also using the entire distribution obtained in the process of elicitation. In this way, we get the stochastic equivalent of the model based on BBN. This model we used to run Monte-Carlo simulations. The first simulation was run without optimization, just applying the distributions obtained by elicitation. Each simulation had a total of 10.000 trials. The results show the stochastic nature of availability prediction. If there are 13 variables,which can affect the availability and which are not at the best practices level, it is not possible to precisely determine the time and the effect that this weakness maycause. Thus it is not possible accurately to predict the IS availability percent, rather it is possible to predict that availabilitywillbe inside predicted range with particular certaintylevel. According to the results of the simulation, we got IS availability ranges from 98.33% to 99.76% with 90% confidence for the case in which best practice are not applied. Mean and median values were 98.93% and 98.97% respectively, which was close to the initial assumption of 99%.

Conclusions

An attempt is made to adapt the model created by Franke et al. (2012) in two aspects. We changed input variables of the model, and incorporated information on the previous states of the variables, which improved the predictability of the model. Also, we conducted a field research, providing probability elicitation on the entire population of InfoSec, IS audit and IS management experts from the BiH banking sector. We performed a comparative analysis of the research results, with results that Franke and Johnson (2012) obtained in a similar study conducted in Western Europe and USA. The basic assumption built into this model is the independence of the variables that enabled application of Leaky Noisy-OR approach.Anotherlimitation of presented model is the binary representation of variables. Investment in a domain does not always results in bringing this domain to the level of best practice, rather than improve the situation in the domain, thus reducing the impact to the IS unavailability. The empirical research was done on information systems of banks in Bosnia and Herzegovina. It would be interesting to conduct a similar survey in other industries as well. Further researches should lead to amodel that overcomes this limitation by using continuous variables instead of using binary and Noisy-MAX node instead of Noisy-OR.

References

1. Bell, M.A. (2005), "Use Best Practices to Design Data Center Facilities", Gartner Publication.
2. Bodin, L.D., Gordon, L.A., Loeb, M.P. (2005) "Evaluating Information Security Investments Using the Analytic Hierarchy Process", Comm. of the ACM, Vol. 48 No. 2, pp. 79–83.
3. Bodin, L.D., Gordon, L.A., Loeb, M.P. (2008) "Information security and risk management", Comm. of the ACM, Vol. 51 No. 4.
4. Bonafede, E., Cerchiello, P. (2007), "Statistical models for business continuity management", Journal of Operational Risk, Vol. 2 No. 4, pp. 79–96.
5. Calzolari, F. (2006), "High availability using virtualization", Universita di Pisa.
6. Chen, P., Kataria, G., Krishnan, R. (2011), "Correlated failures, diversification, and information security risk management", MIS Quarterly, Vol. 35 No. 2, pp. 397–422.
7. Dejaeger, K., Verbraken, T. (2012), "Towards comprehensible software fault prediction models using Bayesian network classifiers", Transactions on Software, Vol. 1 No. 1, pp. 1–22.
8. Faisst, U., Prokein, O. (2008), "Management of Security Risks - A Controlling Model for Banking Companies", in D. Seese, C. Weinhardt, & F. Schlottmann (Eds.), Handbook on Information Technology in Finance, Vol. 4801.
9. Fenton, N. (2012), "Risk Assessment and Decision Analysis with Bayesian Networks", CRC Press.
10. Fenton, N., Hearty, P., Neil, M., Radliński, Ł. (2010), "Software project and quality modelling using Bayesian networks", Artificial Intelligence Applications for Improved Software, pp. 1–25.
11. Fineman, M. (2010), "Improved Risk Analysis for Large Projects: Bayesian Networks Approach", Queen Mary, University of London.
12. Franke, U., Flores, W.R., Johnson, P. (2009), "Enterprise architecture dependency analysis using fault trees and bayesian networks", in SpringSim '09 Proceedings of the 2009 Spring Simulation Multiconference.
13. Franke, U., Johnson, P. (2012), "Availability of enterprise IT systems – an expert-based Bayesian model", Software Quality Journal, Vol. 20 No. 2, pp. 369–394.
14. Gaddum, R. (2004), "Business resilience–the next step forward for business continuity", available at: http://www.continuitycentral.com/feature083.htm (accessed April 27th 2014).
15. Gay, S. (2007), "An examination of virtualization's role in contingency planning", in InfoSecCD '07 Proceedings of the 4th annual conference on Information security curriculum development, available at: http://dl.acm.org/citation.cfm?id=1409927 (accessed April 27th 2015).
16. Goyal, A., Lavenberg, S.S. (1987), "Modelling and analysis of computer system availability", IBM Journal of Research and Development, Vol. 31 No. 6, pp. 651–664.
17. Gran, B. A. (2002), "Use of Bayesian Belief Networks when combining disparate sources of information in the safety assessment of software-based systems", International Journal of Systems Science, Vol. 33 No. 6, pp. 529–542.
18. Henrion, M. (1987), "Practical issues in constructing a Bayes' belief network", in: Proceedings of the Third Conference on Uncertainty in Artificial Intelligence. AUAI Press.
19. Hinz, D.J., Malinowski, J. (2006), "Assessing the Risks of IT Infrastructure – A Personal Network Perspective", in HICSS '06", Proceedings of the 39th Annual Hawaii International Conference on System Sciences, Vol. 00, pp. 1–8.
20. Hole, K. (2010), "Toward risk assessment of large-impact and rare events", IEEE Security & Privacy, pp. 21–27.
21. IEEE (1990), "IEEE Standard Glossary of Software Engineering Terminology", IEEE.
22. Immonen, A., Niemelä, E. (2007), "Survey of reliability and availability prediction methods from the viewpoint of software architecture", Software & Systems Modelling, Vol. 7 No. 1, pp. 49–65.
23. ISO/IEC (2005), "ISO/IEC 27001:2005 Information technology -- Security techniques -- Information security management systems – Requirements", ISO/IEC.

24. Ioannidis, C., Pym, D. (2009), "Investments and trade-offs in the economics of information security", In Financial Cryptography and Data Security, Vol. 5628, pp. 148–166, Springer Berlin Heidelberg.
25. Lande, S., Zuo, Y., Pimple, M. (2010), "A Survivability Decision Model for Critical Information Systems Based on Bayesian Network", in 5th Annual Symposium on Information Assurance.
26. Kouns, J., Minoli, D. (2010), "Information technology risk management in enterprise environments", Wiley
27. Liu, D., Deters, R., Zhang, W.J. (2010), "Architectural design for resilience", Enterprise Information Systems, Vol. 4 No. 2, pp. 137–152.
28. Marcus, E., Stern, H. (2003), "Blueprints for high availability (Second)", Wiley.
29. Martin, A.P. (2003), "Key determinants of information availability: A multiple case study", University of Nebraska.
30. Martin, A.P., Khazanci, D. (2006), "Information Availability and Security Policy", in Proceedings of the Twelfth Americas Conference on Information Systems.
31. Oniśko, A., Druzdzel, M.J., Wasyluk, H. (2001), "Learning Bayesian network parameters from small data sets: Application of Noisy-OR gates", International Journal of Approximate Reasoning, Vol. 27 No.2, pp. 165–182.
32. Pearl, J. (1988), "Probabilistic Reasoning in Intelligent Systems", Morgan-Kaufmann.
33. Raderius, J., Narman, P., Ekstedt, M. (2009), "Assessing System Availability Using an Enterprise Architecture Analysis Approach", In Service-Oriented Computing–ICSOC Workshops, pp. 351–362, Springer.
34. Radliński, Ł., Fenton, N., Neil, M. (2007), "Improved decision-making for software managers using Bayesian networks", in Software Engineering and Applications, pp. 1–13.
35. Schiesser, R. (2010), "IT systems management", Pearson Education, Inc.
36. Simonsson, M., Robert, L., Johnson, P. (2008), "A Bayesian network for IT governance performance prediction", ICEC '08 Proceedings.
37. Somasundaram, S., et al. (2009), "NSAID-induced gut inflammation and vasoconstriction: Causes and potential reversal with beta-CGRP - A hypothesis", Bioscience Hypotheses.
38. Weber, P., Suhner, M. (2001), "System architecture design based on a Bayesian Networks method", in 10th International Symposium on Applied Stochastic Models and Data Analysis. Compiègne, France.
39. Wei, W., Wang, H., Yang, B., Liu, L. (2011), "A Bayesian Network Based Knowledge Engineering Framework for IT Service Management", IEEE Transactions on Services Computing, Vol. 99, pp. 1–14.
40. Zhang, R., Cope, E., Heusler, L. (2009), "A Bayesian Network Approach to Modeling IT Service Availability using System Logs", In SOSP Workshop on the Analysis of System Logs.

About the authors

Semir Ibrahimovic has been working as Director of Information Technology Division at Bosna Bank International and senior TA in Department of Management and IT at the School of Economics and Business in Sarajevo. He was advisor to general manager in Vakufska bank Sarajevo and CTO for system integrator Alden Data. He was leading a number of IS implementation projects, including few core banking IS-projects and BCM projects. He worked as a consultant on IT projects for domestic and foreign institutions. Author can be contacted at **semir.ibrahimovic@efsa.unsa.ba**

Nijaz Bajgoric is a Professor of Business Computing and Information Technology Management in the School of Economics and Business, University of Sarajevo, BiH. He has published papers in: International Journal of Business Continuity and Risk Management, International Journal of Enterprise Information Systems, Cybernetics, Information Management and Computer Security, Information Systems Management, Industrial Management and Data Systems, International Journal of Production Research, European Journal of Operational Research, International

Journal of Agile Management Systems, Journal of Concurrent Engineering, International Journal of Agile Manufacturing and authored/co-authored chapters in the edited books published by: Wiley, IGI-Global, Elsevier Science, Taylor & Francis; Kluwer Academic Publishers, IOS Press, CRC Press, and Auerbach Publications.Author can be contacted at **nijaz.bajgoric@efsa.unsa.ba**