SSRG International Journal of Recent Engineering Science
ISSN: 2349–7157 / https://doi.org/10.14445/23497157/IJRES-V10I4P103

Volume 10 Issue 4, 14-18, Jul-Aug 2023

Review Article

Modeling and Simulation of Reliability of Networked and Distributed Systems: A Case Data Reliability Model

Paul K.M. Masenya[1], Seleman Ismail[2]

[1]*Faculty of Information and Communication Technology Ruaha Catholic University, Iringa, Tanzania*
[2] *Faculty of Science, Technology and Environmental Studies, Open University of Tanzania, Dodoma, Tanzania.*

[1]*Corresponding Author : masenyapkm@gmail.com*

Received: 08 May 2023 Revised: 20 June 2023 Accepted: 07 July 2023 Published: 26 July 2023

Abstract *- Data are the raw facts, the material for obtaining information. Information is an increment in knowledge: it contributes to the general framework of concepts and facts we know. The information relies on the context and the recipient's general knowledge for its significance, i.e., output after data has been processed. Information systems (including networked and distributed systems) use data stored in computer databases to provide needed information. A database is an organized collection of interrelated data reflecting a major aspect of a firm's activities. Quality information (data) needs to possess several attributes such as Timely, Complete, Concise, Relevant, Precise, and Form (The level of detail, tabular versus graphic display, quantitative versus qualitative form, and relies upon either Internal or External situations). Failure to meet those attributes can result in a catastrophic impact. That is why we propose a formulation of the mathematical model to measure and predict the data reliabilities in networked and distributed systems. The proposed model discussed depends on general control, application control, and user control, whether internal or external, as well as techniques employed to ensure that data are accurate, complete, relevant, and valid.*

Keywords *- Data model, Data reliability model, Mathematical modeling of Data reliability.*

1. Introduction

The advancement of technology in computing has resulted in putting together a large number of networked computers, being geographically dispersed, forming a distributed system, where the size of a distributed system varies from a handful of devices to millions of computers, and the interconnection may be wired, wireless, or a combination of both. Today distributed systems are often highly dynamic, in the sense that computers can join and leave, with the topology and performance of the underlying network almost continuously changing.[1] The most evident rationale for the use of a distributed system is resource sharing and data sharing [2]. It has been reported that all information systems, regardless of their level of sophistication, are principally geared towards the output of one product, i.e., data. It is estimated that a rate of 0.5 to 30 percent of defective data exists for accuracy [3].

This paper discusses the issue of the reliability of systems, particularly the mathematical modeling of data reliability on networked and distributed systems. To accomplish this, we have outlined the literature review, systems and data reliabilities, existing general data models, formulation of data reliability models, modeling data reliability, a proposed mathematical data reliability model assumptions and concluded by pointing out the importance of the data reliability model as well as what to be done in next studies.

2. Literature Review

There are many sources of unreliable data, and each contributes its part to the total unreliable data problem in the system; these include Deliberate Errors [4], Null Problems [5], Flawed Data Entry Processes [6], and Mistakes [7].

A study by the Meta Group revealed that 41% of data warehouse projects fail, mainly due to insufficient data quality, leading to wrong decisions[8]. The quality of the input data strongly influences the quality of the results ("garbage in, garbage out" (GIGO) principle.[9,8]**.** Hence, a mathematical model is needed to study the reliability of data in Networked and distributed systems. This review paper proposes to formulate a mathematical model to study the data reliability of networked and distributed systems.

2.1. System Reliabilities

Reliability is the ability of an item (System) to perform a required function under given environmental and operational conditions and for a stated period (ISO8402). [10]. System Reliability is the probability that a system performs correctly

during a specific time duration. During this duration, it operates correctly, no repair is required or performed, and the system adequately follows the defined performance specifications[11]. Networked and Distributed systems contain multiple components connected. The effective reliability and availability of the system depend on the specifications of individual components, network configurations, and redundancy models. These components include hardware (Pc, network equipment, etc.), hardware-Software interaction (middleware), Software (operating system, programming languages, application software, etc.), people (users), data (information), and procedure (documents/knowledge)[12], This implies that every component should be reliable for the system to be reliable.

2.2. Data Reliability

The 1973 Webster's New Collegiate Dictionary defines data as "factual information (as measurements or statistics) used as a basis for reasoning, discussion, or calculation." The 1996 Webster's II New Riverside Dictionary Revised Edition defines data as "information, especially information organized for analysis[7].

Data reliability means that data is complete, valid, and accurate [13]. Data reliability is one of the main objectives of data integrity initiatives, which are also used to maintain data security, data quality, and regulatory compliance. With reliable data, It is an important component in such a way that institutional leaders rely on it for making decisions. It is the fuel that delivers trusted analytics and insights. It is one of the most important things to get right when it comes to improving the overall health of an organization's data [14].

Data reliability is classified as Generic data: (Data collected by an organization and published in handbooks. The collected data may be for specific component types (not brands) and maybe a combination of operating experience and manufacturer data applicable to a specific industry sector or specific usage conditions). Secondly is Brand data: (manufacturer specification for the particular component). Third is User-provided (experience) data: (Data collected by a specific user at a specific site). And lastly are expert judgments data (empirical judgment by experts through their opinion) [15].

3. Data Models

Data models represent an institution's data elements and the connections between them visually. Data models specify the data format in the context required for relevant business processes[16]. Models enable business and technical resources to collaboratively decide how data will be stored, accessed, shared, updated, and leveraged across an organization [18,20]. Data models play a key role in bringing together all segments of an enterprise, business analysts, management, and others to cooperatively design information systems (and the databases they rely on). Hence, these systems require properly defined and formatted data and models shine a clear light on what data is required and how it must be structured to support the desired business processes.

Most of the data models are concerned with data storage, and their development follows the same stages outlined in [20]. That is, its development starts with developing a conceptual model (domain model), which explores and provides an abstraction of the system. This gives way to the second stage of logical data models or specification models, which clarify the various logical entities (types or classes of data), the data attributes that define those entities, and the relationships between them. The third stage is the physical data modeling stage or computational model[20]. These models are usually directly translated into production database design, which supports further development of information systems.

Another model is a data-driven business model which frequently used in Online search engines, social media platforms, and targeted advertising services, often based on the large-scale collection, analysis, and monetization of personal data[19].

For Cloud Data, there is also a data reliability model which describes the reliability of the Cloud data throughout their life cycles, how they are stored with different redundancy levels, and stored on different storage devices with different failure rate patterns in different stages, respectively[24,25]. To have a reliable service, the architecture must include measurable reliability goals, Design principles, and Automated change management. Operational procedures that are easy to follow and let the user easily detect and mitigate failures[23].

4. Formulation of Data Reliability Model

This paper discusses developing a mathematical model of data reliability for networked and distributed systems. Data reliability assessment measures three different elements as Validity, Completeness, and Uniqueness [24]. Valid data refers to data that is correctly formatted and stored, and reliable data, on the other hand, refers to data that can be trusted, which is a basis for analysis and decision-making. Valid data is an important component of reliable data, but validity alone does not guarantee reliability[25].

As pointed out by[28,29], the main component of data reliability assessments is the consideration of internal controls by the institution, which is through policies and procedures intended to achieve the system's objectives over the information system concerned. Information system controls support the underlying mechanism and processes of the system where data are maintained. They consist of internal controls that depend on information systems processing (general, application, and user controls). Thus, data are to be complete, accurate, and Valid.

Application control is taken as a security practice that blocks or restricts unauthorized applications from executing in ways that put data at risk[28]. The control functions vary based on the institutional purpose of the specific application. The concern is to help ensure the privacy and security of data used by and transmitted between applications.

These controls include completeness and validity checks, authentication, authorization, identification, input controls, and forensic controls[29]. Failure of any of these control methods results into having unreliable data.

4.1. Modeling Data Reliability

Basing on the concept of Garbage in, Garbage out (GIGO) and condition put by [28] for data protection, Our model will consist of Certification/verification and validation techniques together with Authentication and authorization methods/schemes to control the accessibility of data, only allowed one should be able to access data, authentication, and authorization are two tightly coupled and interrelated concepts which are used to keep transactions secure and help in protecting confidential information[30]

Certification of data is a formal process of making certain that the data concerned are accurate, valid, complete, and relevant [31]. The certified data signify that the content is trustworthy. This implies there is a process involved to be certain that it is trustworthy and authoritative. The process content includes Inspection, Demonstration, Test, and Analysis, and there must be a 'certification scheme.' Based on such a scheme, data conformity is certified against a set of predefined criteria called requirements[31]; verification methods can also be applied instead; thus, data verification is the process of checking that the data a user has inputted is correct. This can be largely automated, provided someone has set up the rules on the system.

Validation can be defined as confirmation through the provision of objective evidence that the particular requirements for a specific intended use are fulfilled.[27, 35]. Thus data validation refers to the process of ensuring the accuracy and quality of data; on the other hand, it is the process of making sure that the data that has been transferred from one source to the other matches the original data[9]. These processes can be implemented by building several checks into a system or report to ensure the logical consistency of input and stored data (e.g., Code Check, Data Type Check, Format Check, Range Check, Uniqueness Check, Consistency Check,[32].

Authentication is a mechanism for allowing a user to use the system[33]. Different authentication techniques are used to create effective security and save the data from getting looted, which withstands all security breaches [34]. There are three main types of authentication mechanisms password entry, smart card, and biometric. Each authentication mechanism functions differently and has its strengths and weakness.

On the other hand, authorization is the task of defining access rights (privileges) to resources, which is part and parcel of information security and access control in particular[35]. Thus, "to authorize" is to specify an access policy. For example, system administrators are normally authorized to access computer/network systems resources.

In the next section, we highlight some assumptions with the argument that the proposed model will be data reliability is a function of validation rules/methods, certification methods, authorization, and authentication techniques employed for using data in the system.

4.2. Proposed Data Reliability Model

To model data, we have the following assumption regarding data used in networked and distributed systems:

1. We assume that there is a general control and application control, where policies and procedures that apply to all or a large segment of an entity's information systems and help ensure the proper operation of information systems, accessibility of any data authentication, and authorization are needed. There are methods for User controls, thus the controls that are performed by people interacting with information systems (user privileges, whose user is supposed to access/use what).
2. We assume there is a way to ensure that data are valid, completeness, accuracy, and confidentiality of transactions and data during application processing.
3. And we assume that each data error (caused by the unreliability of either technique or methods used) is unique and has an equal chance of causing a system failure.
4. Also, we assume that the time intervals between failures are random variables with an exponential distribution.

Therefore, our reliability data model will depend on authentication, authorization methods used to access data, Validation methods to ensure that data accessed are valid, and Certification methods used for inputting and outputting data are certified/verified to be correct, relevant, accurate, and complete.

Hence Reliability (R_d)

$$R_d(t) = f((R_{au}(t), R_{at}(t), R_c(t), R_v(t))$$

Where $R_{au}(t)$ is authorization method reliabilities, $R_{at}(t)$ is Authentication method reliabilities, $R_c(t)$ is Certification/verification technique reliabilities and $R_v(t)$ validation techniques reliabilities.

5. Conclusion

What makes the data reliability model so valuable is thus the model can first show you exactly where to fix data that you know is unreliable, secondly, reveal hidden issues with data you believed to be reliable, and lastly confirm in a quantifiable way that the data you are assessing is reliable and ready to work with.

Consequently, the Higher Learning Institution in Tanzania needs information and data governance programs to help manage the growing amount of electronic data and information. Hence, the necessity of the data reliability model.

This work is the first step towards the development or formulating of mathematical data reliability model sub-components; it is just an abstract model, next is develop a specification model, i.e., a more detailed one than conceptual, and this will include the developing algorithms (sub-model algorithms to be integrated into the main model) necessary for computing and computation model, i.e., an executable model, for simulation purposes. Then later, this sub-model (data reliability model) will be integrated with other User reliability sub-model [36] and System Unit (SU) sub-model to have A Networked and Distributed system model for higher learning Institutions in Tanzania in totality.

Acknowledgment

The authors wish to acknowledge the Open University of Tanzania (OUT) and Ruaha Catholic University (RUCU) for their support and for providing a conducive environment, friends, and other individuals for their support during the preparation and production of this paper.

References

[1] M. Van Steen, and A.S. Tanenbaum, CSE223B_textbook1_Distributed Systems, vol. 28, no. 11. 1985.

[2] Mona Shah, and Hiren Joshi, "Role of Distributed Systems in Data Mining," *International Journal of Computer Science and Information Technologies*, vol. 6, no. 3, pp. 3010–3013, 2015. [Publisher Link]

[3] T. Fletcher, and W.J. Howe, A Framework for Address DQ in Distributed Compu.pdf, 2000.

[4] Sarah Shi Hui Wong, and Stephen Wee Hun Lim, "The Derring Effect: Deliberate Errors Enhance Learning," *Journal of Experimental Psychology: General*, vol. 151, no. 1, pp. 25–40, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] Nathaniel Ayewah, and William Pugh, "Null Dereference Analysis in Practice," *ACM SIGPLAN/SIGSOFT Workshop on Program Analysis for Software Tools and Engineering*, pp. 65-72, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[6] J. Taylor, "Sources of Inaccurate Data in Data Entry Process" 2022.

[7] Brian Richmond, "Introduction to Data Analysis Handbook," *Academy for Educational Development*, 2006. [Google Scholar] [Publisher Link]

[8] Paulo Oliveira, Fatima Rodrigues, and Pedro Henriques, "A Formal Definition of Data Quality Problems," *Proceedings of the 2005 International Conference on Information Quality*, 2005. [Google Scholar] [Publisher Link]

[9] Siyavula. [Online]. Available: https://www.siyavula.com/read/cat/grade-11-cat/err

[10] Marvin Rausand, and Arnljot Hoyland, *System Reliability Theory Models, Statistical Methods, and Applications*, Wiley Interscience, 2004. [Google Scholar] [Publisher Link]

[11] Muhammad Raza, System Reliability & Availability Calculations, 2019, [Online]. Available: https://www.bmc.com/blogs/system-reliability-availability-calculations/

[12] M. Al-Kuwaiti, N. Kyriakopoulos, and S. Hussein, "A Comparative Analysis of Network Dependability, Fault-Tolerance, Reliability, Security, and Survivability," *IEEE Communications Surveys and Tutorials*, vol. 11, no. 2, pp. 106–124, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[13] T. Ravichandran, and Arun Rai, "Total Quality Management in Information Systems Development: Key Constructs and Relationships," *Journal of Management Information Systems*, vol. 16, no. 3. pp. 119–155, 1999. [CrossRef] [Google Scholar] [Publisher Link]

[14] Min Chen et al., "AIWAC: Affective Interaction through Wearable Computing and Cloud Technology," *IEEE Wireless Communication*, vol. 22, no. 1, pp. 20–27, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[15] M. Emerson, and J. Kofner, "Technical Product Standards and Regulations in the EU and EAEU – Comparisons and Scope for Convergence," *International Institute for Applied Systems Analysis*, 2018. [Google Scholar] [Publisher Link]

[16] Jinyong Wang et al., "An Optimized Method for Software Reliability Model Based on Nonhomogeneous Poisson Process," *Applied Mathematical Modelling*, vol. 40, no. 13–14, pp. 6324–6339, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[17] Antoniou, "What is a Data Model?," *European Journal for Philosophy of Science*, vol. 11, 2021. [CrossRef] [Publisher Link]

[18] International Atomic Energy Agency, Reliability Data Analysis - Use of Generic and/or Plant-Specific Data. [Online]. Available: http://www-ns.iaea.org/downloads/ni/training/specific_expert_knowledge/psa-level1/III3_1%20Reliability%20data%20analysis_use%20of%20generic%20plant%20specif.pdf

[19] Peter J. van de Waerdt, "Information Asymmetries: Recognizing the Limits of the GDPR on the Data-driven Market," *Computer Law and Security Review*, vol. 38, p. 105436, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[20] S. Tripathi, D. Muhr, M. Brunner, H. Jodlbauer, M. Dehmer, and F. Emmert-Streib, "Ensuring the Robustness and Reliability of Data-Driven Knowledge Discovery Models in Production and Manufacturing," Frontiers in Artificial Intelligence, vol. 4. 2021, doi: 10.3389/frai.2021.576892.

[21] Sourabh Kumar Jain et al., "Radiation Tolerant PLL for Onboard FPGAs," *SSRG International Journal of Electronics and Communication Engineering*, vol. 10, no. 4, pp. 51-62, 2023. [CrossRef] [Publisher Link]

[22] Ali Alzahrani et al., "Hybrid Approach for Improving the Performance of Data Reliability in Cloud Storage Management," *Sensors*, vol. 22, no. 16, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[23] Tuan Anh Nguyen et al., "Reliability and Availability Evaluation for Cloud Data Center Networks using Hierarchical Models," *IEEE Access*, vol. 7, pp. 9273–9313, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[24] R. L. Sarfin, "5 Characteristics of Data Quality" 2021.

[25] K. Thiruthanigesan, and N. Thiruchchelvan, "Data Verification and Validation Process in the Management System Development," *Middle-East Journal of Scientific Research*, vol. 25, no. 5, pp. 902–911, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[26] Christine Bellino, and Jefferson Wells, *Auditing Application Controls*, The Institute of Internal Auditors, 2007. [Google Scholar] [Publisher Link]

[27] GAO-20-283G, Assessing Data Reliability, Applied Research and Methods, 2019. [Online]. Available: https://www.gao.gov/assets/710/703275.pdf

[28] Data Insider, Data Protection 101, Digital Guardian. [Online]. Available: https://digitalguardian.com/blog/what-cyber-security

[29] Janis Bicevskis, Zane Bicevska, and Girts Karnitis, "Executable Data Quality Models," *Procedia Computer Science*, vol. 104, pp. 138–145, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[30] TechWeb, Understanding Authentication, Authorization, and Encryption. [Google Scholar] [Publisher Link]

[31] Eric Lachaud, "What GDPR tells about Certification," *Computer Law and Security Review*, vol. 38, p. 105457, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[32] Mark P.J. Van Der Loo, and Edwin De Jonge, "Data Validation," *Wiley StatsRef: Statistics Reference Online*, 2020. [CrossRef] [Publisher Link]

[33] Nilesh A. Lal, Salendra Prasad, and Mohammed Farik, "A Review of Authentication Methods," *International Journal of Scientific & Technology Research*, vol. 5, no. 11, pp. 246–249, 2016. [Google Scholar] [Publisher Link]

[34] D. Dinesh Kumar et al., "A Study on Different Types of Authentication Techniques in Data Security," *International Journal of Civil Engineering and Technology*, vol. 8, no. 12, pp. 194–201, 2017. [Google Scholar] [Publisher Link]

[35] Javad Shahreki, and Hiroshi Nakanishi, "Journal of Soft Computing and Decision Support Systems the Relationship between Task Technology Fit and Individual Performance : Case Study in Hotel Industry in Malaysia," *Journal of Soft Computing and Decision Support Systems*, vol. 3, no. 6, pp. 1–15, 2016. [Google Scholar] [Publisher Link]

[36] P.K.M. Masenya, "Modeling and Simulation of Reliability of Networked and Distributed Systems : A Case User Reliability Component," *International Journal of Academic Engineering Research*, vol. 6, no. 12, pp. 7–13, 2022.