Modeling System Reliability For Digital Preservation: Model Modification and Four-Copy Model Study

Yan Han, Chi Pak Chan

The University of Arizona Libraries
1510 E University Blvd, Tucson, AZ, USA, 85721
{yhan, cpchan}@email.arizona.edu

Abstract

Research has been studied to evaluate the reliability of storage media and the reliability of a computer backup system. In this paper, we use the Continuous Time Markov Chain to model and analyze the reliability of a computer backup system. We propose a modified model from that of the Constantopoulos, Doerr and Petraki [1]. We analyze the difference, show computational results, and propose new input parameters (e.g. time to repair) for the model from our experience. Further we developed a four-copy data model to test if it fulfills the sample reliability rate set by the RLG-NARA. The modeling process can be applied to construct models for computer preservation systems using different storage media. The reliability of constructed models can be calculated so that preservation institutions can have quantitative data to decide their preservation strategies.

1. Introduction

Traditional preservation techniques have focused on longevity of the media since the only requirement has usually been human readability. With a growing number of born-digital data and digitized materials there is an urgent need for research on digital preservation. Unlike traditional preservation strategies, digital preservation fundamentally changes the nature and process of preservation while considering issues related to media, storage, access, representation, and authentication. Digital preservation is more complex, not only because of the information encoded in various IT standards or protocols, but also related to its context: metadata management and higher level of policy issues.

Digital preservation has two related components: physical and logical preservation. Physical preservation for digital assets is similar to preserving analog materials and ensuring bit-streams to be readable from storage media. Logical preservation is more complex because it requires technology and processes to ensure that bit-streams are renderable and accessible for computers and humans. This paper discusses using Continuous Time Markov Chain to measure capacity of physical preservation, including modeling, analyses, and comparisons between the CDP's model [1] and our modified model. We suggest new input parameters such as time to repair for the model and construct a four-copy backup system.

2. Related Research

Digital files are vulnerable to corruption due to multiple reasons such as failed storage media, outdated backup, obsolete recording/reading devices, neglected human errors, and undesirable disasters. The longevity of digital storage media has been a subject of interest to librarians and archivists. In 2002, National Archives and Records Administration (NARA) directed a study of high density magnetic tapes life expectancy and revealed tapes can have a life expectancy of 50 -100 years [8][9]. The Library of Congress completed an unpublished report to study prerecorded compact discs (CD-ROMs). Both the National Institute of Standards and Technology (NIST) in 2004 [6] and Canadian Conservation Institute in 2005 published reports of life expectancies of recordable CDs (CD-Rs), rewriteable CDs (CD-RWs), and recordable DVD (DVD-Rs). All the studies show that higher deterioration for optical and magnetic media, when exposure to high temperature and humidity condition.

To establish a process to ensure long-term sustainability for digital collections, Research Library Group (RLG) and United States National Archives and Records Administration (NARA) released a report for evaluating a trusted digital repository. The report covers critical digital preservation issues, including physical and logical preservation for long term preservation. The report states that "D1.5 Repository has effective mechanisms to detect data corruption or loss" [3] and illustrates a sample reliability rate: "if the policy were the repository could not lose more than 0.001% of the collection per year…" [3] The quantitative data allows preservation institutions and certificate issuing organizations to measure the capacity of a trusted digital repository.

Since 1999, Lots of Copies Keep Stuff Safe (LOCKSS) [7] advances digital preservation research and receives tremendous success in libraries and publishers. LOCKSS is a peer-to-peer open source software to convert a PC into a digital preservation node, creating low cost, persistent, and accessible copies of web-based data. Since LOCKSS is a peer-to-peer system, it is an innovation to just show the concept of "the more, the better". However, LOCKSS might not be appropriate for close data.

In 2005, Constantopoulos, Doerr and Petraki (CDP) published a paper [1] to introduce a reliability model that uses the Continuous Time Markov Chain to measure the reliability of a computer preservation system.

3. Methodology

As more and more preservation institutions are involved with digitization, and at the same time anticipating growing needs of preserving born-digital materials, it is critical to have quantitative study on the reliability of a computer backup system so that preservation institutions can base on outputs from quantitative analysis to make decisions for long-term preservation. In the CDP's paper [1], it was calculated that a typical computer backup system with three-copy of data (two disks and one tape) has a reliability rate of 67.46% in 1000 years. Since this paper does not provide the unreliability rate of one year, we drew the system and calculated that the system's unreliability rate is 0.033%. This result obviously does not meet the 0.001% unreliability rate illustrated by the RLG-NARA report. Is the reliability modeling appropriate? If we develop a four-copy data model, will it fulfill the RLG-NARA's required 0.001% unreliability rate? Is it possible that the modeling can be easily extended to more copies of data and different storage media?

Continuous Time Markov Chain (CTMC) is used to analyze lifetime and reliability rate of a backup system. Computer system components such as disk, tape, and other forms of storage media could break down at any time due to depreciation of the components or some unexpected external factors such as earthquake or flooding, which can take place at a random time. On the other hand, the recovery process of a component is approximately a continuous process. Moreover, it is reasonable that the probability of a system's next state bases only on current state of the system. Therefore, CTMC is an appropriate methodology to analyze system continuous failure/recovery processes and state status of the whole system. Inspired by the RLG-NARA's report and the CDP's paper, we conduct further research on this topic.

3.1 Markov Modeling The preservation policy is: for each digital file we create one or more copies in disk, tape, or other forms of storage media, and if detecting a failure of disk, tape, or other forms of storage media, we replace them. We assume that the preservation policy is consistent over the time.

In this case, we analyzed a mirrored computer system with two copies of data in hard disks. We are interested in finding out the reliability of this system. The CDP's paper has already described the process of constructing the Markov chain for the two-copy system [1]. We had the same result. (See Figure 1)

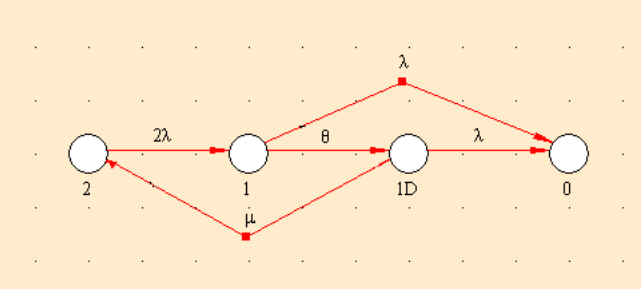

Figure 1: Modeling a two-copy system
Where
State 2: Both disks function properly.
State 1: One of the disks has failed, but has not detected yet.
State 1D: One of the disks has failed and the failure has been detected.
State 0: Both disks failed (absorbing state). Therefore, the data is not recoverable.

Initially the system starts at state 2 (2 copies function properly) and each disk has a failure rate; assuming that both have the same failure rate λ, the rate for the system going from State 2 to State 1 is 2λ, as shown on arc (2,1). There is a rate regarding the detection of the failure disk, which is θ shown on arc (1,1D). Moreover, there is a possibility that the other functioning disk fails even before failure of the failed disk has been detected, and the rate for the system going from State 1 to State 0 is λ as shown on arc (1, 0), which results in the failure of the whole system and the data never being recovered. Similarly, at State 1D, the failure of the disk has been detected and is repaired. There is a possibility that the system can fail (i.e. from State 1D to State 0) and the rate is λ. There is a possibility the system recovers to its initial state (2 disks) by recovering the failed disk as shown on arc (1D, 2) with rate μ.

3.2 Our Experience about Input Parameters and Storage Media. The CDP's paper [1] conducted experiments to study the above parameters such as mean time to failure (MTTFdisk), mean time to repair (MTTRdisk), and mean time to detect failure (MTTFDdisk) for their modeling. The University of Arizona Libraries had a few server disk failures and tape failures in the past. Our experience shows that it takes us

about 25 hours to restore 10TB data back to a storage (hard disks) server, if the backup policy requires systems administrators to recover the data as soon as possible. This process includes reinstalling Operating System (OS) and copying 10TB of data using 1000 Mbps network connection. It is true that less data takes less time to repair. Therefore, MTTRdisk depends on the amount of data and computer backup policies. For MTTFDdisk we can detect hard disk failure right away, because modern OS automatically sends emails/text messages to us and server vendor when detecting failed hard disks. Server vendors such as Dell offer 24x7 replacement service plan to deliver new hard disks to us, and we uses Dell's 4-hour replacement plan. If the storage server is critical, we can upgrade our service plan to get quicker service. Our MTTRdisk is 25 hours and MTTFDdisk is 4 hours, compared to the CDP's 50 hours of MTTRdisk and 14 days of MTTFDdisk. Our experience on MTTFDtape is different from MTTFDdisk. Currently we do not have a tape library and thus our systems administrators have to manually change tapes. This slows down time to detect and repair tapes. Using restoring 10TB of data as an example, our MTTRtape is 60 hours, and MTTFDtape is about 60 days, In addition, the costs and benefits of storage media and staffing should not be ignored. In practices, tapes require more staffing time to handle, more time to access data, and is less reliable, but they are easy to store offsite and cheap in terms of cost per GB. Compared to hard disks and magnetic tapes, CDs and DVDs are limited in storage size, require frequent human handling when reading data, and are usually not rewriteable. Due to the above disadvantages, in 2004 we made a decision to remove optical media for permanent storage at the University of Arizona Libraries.

The input parameters from the CDP's model are close to what we measured from real life experience except MTTFDdisk. To best comparing the differences between the CDP's model and our modified model, we use the same input parameters.

$MTTF_{disk} = 1/\lambda = 3$ years $\Rightarrow \lambda = 1/3$ per year
$MTTR_{disk} = 1/\mu = 50$ hours $\Rightarrow \mu = 175.2$ per year
$MTTFD_{disk} = 1/\theta = 14$ days $\Rightarrow \theta = 365/14$ per year

Let

$m_i = E[\text{time before absorption} \mid \text{the system starts from state } i]$

Based on the CTMC model, we have the following set of equations:

$$m_2 = 1/2\lambda + m_1$$
$$m_1 = [\lambda/(\theta+\lambda)] \times 1/\lambda + [\theta/(\theta+\lambda)] \times m_{1D}$$
$$m_{1D} = [\lambda/(\mu+\lambda)] \times 1/\lambda + [\mu/(\mu+\lambda)] \times m_2$$

Which are reduced to the following:

$$m_2 = 1/2\lambda + 1 \times m_1$$
$$m_1 = 1/(\theta+\lambda) + \theta/(\theta+\lambda) \times m_{1D}$$
$$m_{1D} = 1/(\mu+\lambda) + \mu/(\mu+\lambda) \times m_2$$

Solve these and we get:

$$m_2 = \frac{(\theta+\lambda)(\mu+\lambda) + 2\lambda(\mu+\lambda) + 2\lambda\theta}{2\lambda^2(\theta+\mu+\lambda)}$$

Petraki's paper has $m_2 = \frac{(\theta+\lambda)(\mu+\lambda) + 2(\mu+\lambda) + 2\lambda\theta}{2\lambda^2(\theta+\mu+\lambda)}$ [2], which might be a typographical error. The expected TTF_{system} m_2 is 106.46 years, which means that the two-copy system is expected to crash after 106.46 years. To verify the result, we used a software package called SHARPE to model this Markov Chain. The result is exactly the same as we got from the above formula: 106.46 years.

3.3 Modeling a three-copy model The CDP's paper [1] also described the process of extending the system by adding another backup copy. Their example is to add magnetic tapes for an additional copy of data. Other media such as CD and DVD can also be used with appropriate rates (λ_3 θ_3 μ_3).

Our Markov model shown in Figure 3 on a three-copy system (2 in disk and 1 in tape) is similar to that of CDP's model[1] shown in Figure 2, but we propose some modifications which we think are more realistic in real life situations and less risky in preventing a backup system from ending at the absorbing state. In the figures, we use (*DiskCopiesFunctioningDiskCopiesFailureDetected _TapeCopiesFunctioningTapeCopiesDetected*) notation to represent state status. *DiskCopiesFunctioning* represents the number of copies of data functioning, while status *DiskCopiesFailureDetected* can be either *NULL* or *D* (meaning failure detected). Figure 2 shows CDP's model for three-copy system.

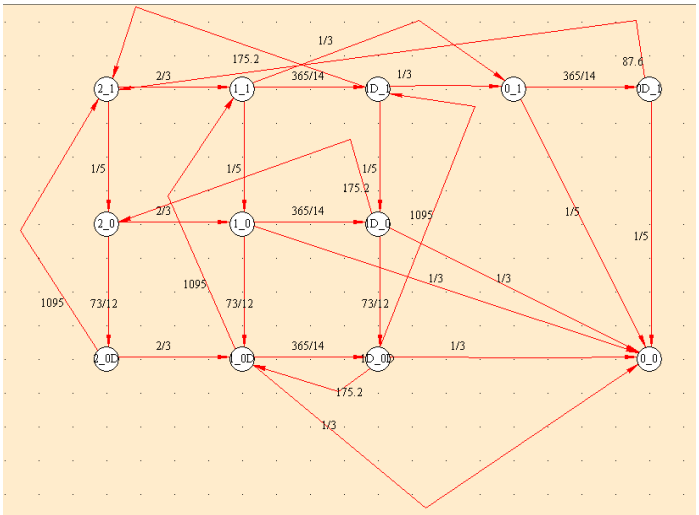

Figure 2: CDP's Markov model for a three-copy system (2 in disk and 1 in tape)

We believe that repairing failed copies one at a time is more realistic due to resource limitations. Therefore, we suggest that whenever there is/are failed copy/copies detected, only one failed copy will be repaired at a time to roll back to previous state (i.e. arc (0D_1, 1D_1)). The benefit is that repairing one copy at a time is always faster than working on multiple copies. In other words, it reduces the risk of the system ending at the absorbing state. This is more obvious when the system has only one functioning copy left, i.e. 0D_1. Under this situation, it would be wiser to repair one copy rather than repairing multiple copies at a time so that we can have one extra backup copy sooner.
In CDP's model, they considered repairing one failed copy at a time, but not all of them. For example, they did not consider a recovery from state 0D_1 (both disks have failed and the failures have been detected, 1 tape is functioning) to state 1D_1 and from state 1D_0D (1 disk is functioning and the other has failed and been detected, the tape has failed and been detected) to state 2_0D, which is possible and should be considered. On the other hand, we see that CDP's three-copy model allows the repair of multiple failed copies at the same time such as recovering the system from 0D_1 to 2_1, which means the current system has two failed disks and one functioning tape, and the failures of both disks have been detected. Connecting an arc from 0D_1 to 2_1 means that we allow simultaneous repairing of these two disks at the same time. Simultaneous repairing can be allowed if multiple distributed data centers are involved. In a real life situation, repairing a failed copy needs computing resources and staffing. Staffing and certain computing resources such as networking bandwidth in a data center can cause a bottleneck, because only certain amount of data can be transferred and a limited number of staff is available at a time.

We believe that repairing one copy at a time is more realistic in a real life situation for a data center. Therefore, there should be a transaction from state 0D_1 to 1D_1, which means an arc (0D_1, 1D_1). Similarly, there should be a transaction recovering from state 1D_0D (1 disk is functioning and the other has failed and been detected, the tape has failed and been detected) to state 2_0D. In addition, repairing multiple failed copies is unrealistic and risky as we have explained above for a data center. Therefore, we propose a model which merely repairs one failed copy of data (e.g. disk, tape or other forms of storage media) at a time and simultaneous repairs are not considered. Figure 3 shows the modified model for the computer backup system.

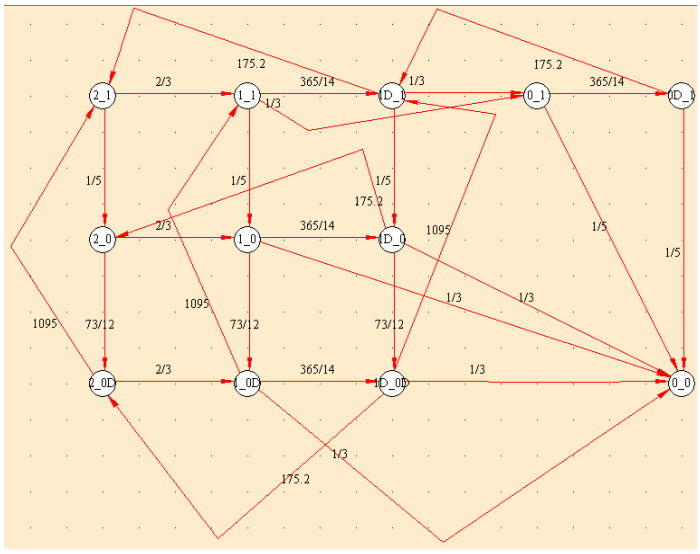

Figure 3: Modified Markov model for a three-copy system (2 in disk and 1 in tape)

As it can be seen that from Figure 3, arc (0D_1, 1D_1) and arc (1D_0D, 2_0D) have been added to our model, and the arc (0D_1, 2_1) has been removed as we have explained above. Note that we remove arc(s) representing the repair of multiple failed copies at the same time not only because it is unrealistic but also because it provides inaccurate information about the true life time and reliability of the backup system. From a computational point of view, it is always true that the more repairing arcs are added to the model, the outputs will always give longer life time and higher reliability of the system. However, these outputs do not reflect the true life time and reliability of the backup system.

3.3 Computational Comparisons The following is a discussion on the computational output of the CDP's model on a three copies of data and ours. We use the same input parameters as the CDP's paper [1] to illustrate computational differences.

$MTTF_{disk} = 1/\lambda_1 = 3$ *years* $\Rightarrow \lambda_1 = 1/3$ *per year*
$MTTR_{disk} = 1/\mu_1 = 50$ *hours* $\Rightarrow \mu_1 = 175.2$ *per year*
$MTTFD_{disk} = 1/\theta_1 = 14$ *days* $\Rightarrow \theta_1 = 365/14$ *per year*
$MTTF_{tape} = 1/\lambda_2 = 5$ *years* $\Rightarrow \lambda_2 = 1/5$ *per year*
$MTTR_{tape} = 1/\mu_2 = 8$ *hours* $\Rightarrow \mu_2 = 1095$ *per year*
$MTTFD_{tape} = 1/\theta_2 = 60$ *days* $\Rightarrow \theta_2 = 73/12$ *per year*

According to our computations, the mean time to failure ($MTTF_{system}$) of the system based on the CDP's model (Figure 2) should be 2565 years and the reliability rate is 67.72% after 1000 years (CDP [1] suggests that $MTTF_{system}$ = 2551 years and reliability rate is about 67.46% after 1000 years). The computational output of our model is: $MTTF_{system}$ is 2633 years and the reliability rate is 68.4% after 1000 years. One can see that both $MTTF_{system}$ and

reliability rate of our model are higher than that of CDP's, because our model allows repairing one copy of data at a time (i.e. arcs from 0D_1 to 1D_1 and 1D_0D to 2_0D). The unreliability rate in 1 year of our three-copy model is 0.0324%.

4. A CTMC model for a four-copy system

While the three-copy model does not fulfill the 0.001% unreliability yearly rate set forth by the RLG-NARA's report, we extend the model for four-copy of data (2 in disk and 2 in tape). The model is shown as follows (Figure 4). Again, as we have done for the three-copy model, the four-copy model allows repairing one failed copy at a time, but does not allow repairing multiple failed copies at a time. Using the same input parameters (e,g. $MTTF_{disk}$, $MTTF_{tape}$) as above, Computational output of the model for four-copy of data is as follows: $MTTF_{system}$ is about 4.238×10^4 years, reliability rate is 97.67% after 1000 years and the unreliability rate in 1 year is 0.001693% which nearly fulfills the RLG-NALA's requirement.

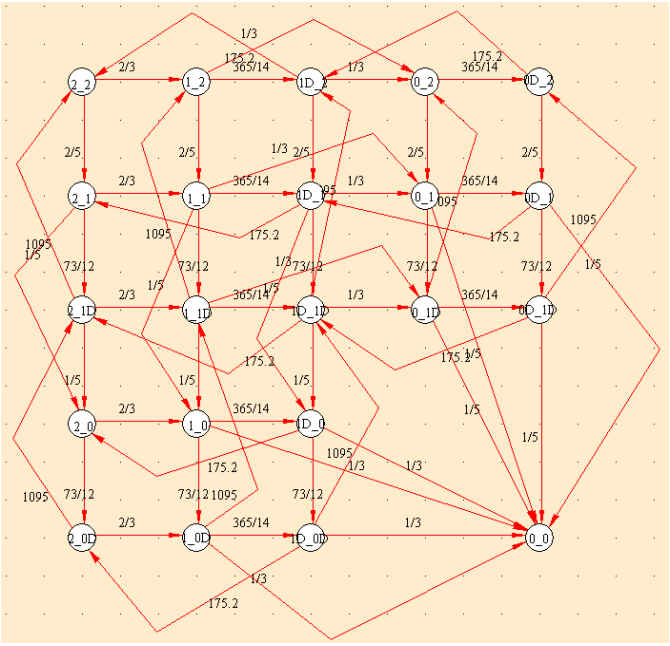

Figure 4: Our Markov modeling for a four-copy system (2 in disk and 2 in tape)

When feeding our input parameters (e.g. $MTTF_{disk}$, $MTTFD_{disk}$), the four-copy system fulfills the RLG-NALA's requirement. This makes sense because our $MTTFD_{disk}$ takes much less time to detect failures and our $MTTF_{disk}$ is 50% quicker to repair failed disks. One can also construct different four-copy systems such as 3-disk-1-tape and 4-disk. This of course proves the concept of "the more, the better", but the model gives a way to demonstrate how much better. Input parameters (MTTF, MTTR, and MTTFD) are critical to the reliability of a backup system. How much effort does each parameter play can be a following topic for research. The result can help an institution to tune its preservation policy.

5. Discussions

Inspired by the RLG-NARA's report and the CDP's paper [1], we have developed a modified CTMC model, which we think is more realistic in practice. We took a close look at the CDP's model and believe that the CDP's model is sound except handling repairing failed copies. We believe that repairing one copy at a time is more realistic in a real life situation.

Our experience shows that our MTTF, MTTR, and MTTFD in disk are different. We researched optical storage media and discuss pitfalls of CDs and DVDs, and recommend not to use them for permanent storage. Tapes also have limitations when considering dropping cost and growing capacity of disks. We are considering reducing using tapes for backup.

Based on the rationale to build the CTMC model for the three-copy backup system, we've also developed a model for a four-copy backup system to test whether it can fulfill the sample reliability rate set by the RLG-NARA paper. With CTMC technique, reliability of a computer preservation systems can be calculated so that preservation institutions can use quantitative basis to decide their preservation strategies (e.g. how many copies of data are needed, forms of storage media, preservation policies) to ensure readability of bit-streams.

6. Acknowledgements

We would like to thank Professor Kishor S. Trivedi of Department of Electrical and Computer Engineering at Duke University for offering software SHARPE. We also would like to thank Qu Miao (who was a Computer Science graduate student at the University of Arizona) for providing interesting derivation of m_2 (expected time to failure of the system) using conditional expectations.

References

[1] Constantopoulos, P., Doerr, M., and Petraki, M. 2005. *Reliability modelling for long term digital preservation*. http://delos-wp5.ukoln.ac.uk/forums/dig-rep-workshop/constantopoulos-1.pdf

[2] Petraki, M. 2005. *Evaluating the reliability of system configurations for long term digital preservation*, Master of Science Thesis, Dept. of Computer Science, University of Crete. http://www.ics.forth.gr/isl/publications/paperlink/Petraki.pdf

[3] RLG and NARA, 2005. *Audit Checklist for Certifying Digital Repositories*. http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf

[4] Ross, Sheldon M. 2003. *Introduction to probability models*, 8th ed. San Diego, CA: Academic.

[5] Ross, Sheldon M. 1996. *Stochastic Processes*, 2nd ed. New York: Wiley.

[6] Slattery, O., Lu, R., Zheng, J., Byers, F., and Tang, X. 2004. *Stability Comparison of Recordable Optical Discs – A Study of Error Rates in Hash Conditions.* Journal of Research of the National Institute of Standards and Technology, vol, 109. pp. 517-524.

[7] Stanford University Libraries. LOCKSS. http://www.lockss.org

[8] Judge, J.S., Shmidt, R.G., Weiss, R. D., and Miller, G. 2003. *Media Stability and Life Expectancies of Magnetic Tape for Use with IBM 3590 and Digital Linear Tape Systems*. Proceedings of the 20th IEEE/11th NASA Goddard Conference on Mass Storage Systems and Technologies (MSS'03). http://storageconference.org/2003/papers/15_Judge-Media.pdf

[9] Weiss, R.D. 2002. *Environmental Stability Study and Life Expectancies of Magnetic Media for Use with IBM 3590 and Quantum Digital Linear Tape Systems.* http://www.archives.gov/research/electronic-records/magnetic-media-study.pdf