URKUND

Document Information

Analyzed document	21_VRJ_Project_Phase_1_Report-New.pdf (D90235099)
Submitted	12/19/2020 3:50:00 PM
Submitted by	Vishal Jaiswal
Submitter email	vrjaiswal@pict.edu
Similarity	6%
Analysis address	vrjaiswal.pict@analysis.urkund.com

Sources included in the report

SA	**Pune Institute of Computer Technology / Final Report (3).pdf** Document Final Report (3).pdf (D68239032) Submitted by: nvburadkar@pict.edu Receiver: nvburadkar.pict@analysis.urkund.com	3
W	URL: https://lists.linbit.com/pipermail/drbd-user/2010-February/013449.html Fetched: 12/19/2020 3:50:00 PM	1
W	URL: https://www.elprocus.com/linux-operating-system/ Fetched: 12/19/2020 3:51:00 PM	1
W	URL: https://sites.ualberta.ca/dept/chemeng/AIX-43/share/man/info/C/a_doc_lib/aixbman/c ... Fetched: 12/19/2020 3:51:00 PM	1
SA	**Thesis_luhtala_final_pdf_a.pdf** Document Thesis_luhtala_final_pdf_a.pdf (D14518752)	1

Entire Document

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module A Preliminary Project Report on "Distributed Replicated Block Device (DRBD) Network Packet Tracing Module"

95%	MATCHING BLOCK 1/7	SA	Final Report (3).pdf (D68239032)

Submitted to the
Savitribai Phule Pune University
In partial fulfillment for the award of the Degree of
Bachelor of Engineering
in
Information Technology
by

Bharat Kothari - 43110 Animesh Landge - 43132 Shrijan Vats - 43258 Amod Dhopavkar - 43304 Under

47%	MATCHING BLOCK 2/7	SA	Final Report (3).pdf (D68239032)

the guidance of
Prof. V. R. Jaiswal
Sponsored by
Veritas LLC, Pune

Department Of Information Technology
Pune Institute of Computer Technology

College of Engineering Sr. No 27, Pune-Satara Road, Dhankawadi, Pune - 411 043 2020-2021 PICT - INFORMATION TECHNOLOGY – 2020-21 1

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module SPONSORSHIP LETTER PICT - INFORMATION TECHNOLOGY – 2020-21 2

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module ACKNOWLEDGEMENT This project would not have been possible without the support of our friends and family. We would like to thank our guide from Veritas LLC, Mr. Swapnil Ujgare sir and our internal guide Prof. V. R. Jaiswal sir, and for their constant advice and support. Additionally, we also thank our external reviewers Prof. T. A. Rane sir and Prof. S. S. Mane sir whose constructive criticism and valuable inputs made us think in ways which would not have been possible. We would like to thank BE Project coordinator Prof. R. V. Kulkarni ma'am for providing guidance throughout the way. In the end we would like to thank Dr. A. M. Bagade sir, the HOD of IT department, and the entire IT department for their constant support. We would also like to thank the principal, Dr. P. T. Kulkarni sir for providing us with this opportunity. Bharat Kothari Animesh Landge Shrijan Vats Amod Dhopavkar 43110 43132 43258 43304 PICT - INFORMATION TECHNOLOGY – 2020-21 3

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module ABSTRACT Distributed Replicated Block Device (DRBD) is an open-source replication storage system. DRBD works by layering over logical block devices on participating nodes. The aim of the project is to write a packet tracer and logger for DRBD. DRBD is implemented as a loadable kernel module. We also want to write a loadable kernel module. This module will be responsible for communicating with the drbd module and preparing the log file. We will add an user land utility which will make system calls to our module and be responsible for analysing the data in the log file and display results and reports in graphical format. PICT - INFORMATION TECHNOLOGY – 2020-21 4

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module LIST OF TABLES LIST OF FIGURES Tab 1. List of Figures ABBREVIATIONS Tab 2. Abbreviations PICT - INFORMATION TECHNOLOGY – 2020-21 5 Sr. No. Table No. Table Name Page No. 1. Tab 1. List of Figures 5 2. Tab 2. Abbreviations 5 Sr. No. Figure No. Figure Name Page No. 1. Fig 1. Linux Architecture 10 2. Fig 2. Gantt Chart 15 3. Fig 3. Architecture Diagram 17 4. Fig 4. Character Driver Code 20 5. Fig 5.

Data Structure of Log 21 6. Fig 6. Cscope 21 7. Fig 7. System Tap 22 Sr. No. Abbreviations Full Form 1. DRBD Distributed Replicated Block Device 2. VM Virtual Machine 3. LKM Linux Kernel Module

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module CONTENTS PICT - INFORMATION TECHNOLOGY – 2020-21 6 Chapter Topic Page No TITLE 1 SPONSORSHIP LETTER 2 ACKNOWLEDGEMENT 3 ABSTRACT 4 List of Tables 5 List of Figures 5 List of Abbreviations 5 1 Introduction 8 1.1 Title 8 1.2 Overview 8 1.3 Motivation 8 2 Literature Review 9 2.1 Existing Methodologies 9 2.1.1 DRBD 9 2.2.2 DRBD vs RAID 9 2.2 Proposed Methodologies 9 2.2.1 Linux Architecture 9 2.2.2 Character Device Driver 10 2.2.3 Linux Kernel Module 11 3

100% **MATCHING BLOCK 3/7** SA Final Report (3).pdf (D68239032)

Requirement Specifications and Analysis 13
3.1 Problem Definition 13
3.2 Concept 13
3.3 Scope 13
3.4 Objective 13
3.5 Project Requirements 14

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module PICT - INFORMATION TECHNOLOGY – 2020-21 7 3.5.1 Functional Requirements 14 3.5.2 Non Functional Requirements 14 3.5.3 Hardware Requirements 14 3.5.4 Software Requirements 14 3.6 Project Plan 14 3.6.1 Modules Split-Up 14 3.6.2 Gantt Chart 15 4 System Analysis and Design 16 4.1 Architecture 16 4.2 Components of the System 17 5 Implementation 18 5.1 Stages of Implementation 18 5.1.1 Character Device Driver 18 5.1.2 Data Structure of Log File 21 5.2 Elaborate Implementation Issues Techniques/ Software Tools etc. 21 5.2.1 Cscope 21 5.2.2 System Tap 22 6 Conclusion 23 6.1 Conclusion 23 6.2 Drawbacks 23 7 References 24 8 Appendices 25 8.1 Project Weekly Log 25

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module 1. INTRODUCTION 1.1. TITLE Distributed Replicated Block Device (DRBD) Network Packet Tracing Module 1.2. OVERVIEW DRBD is an open source data replication software that is used to replicate the data at multiple nodes. It consists of a primary node, and one or several other secondary nodes. Our project is to build a packet tracer for DRBD which would maintain a log file of all the writes and reads at the primary and secondary nodes respectively. DRBD is implemented as a loadable kernel module. We also write a loadable kernel module. The main aim of this module is to create a global queue for storing all packets being sent and received by the node in the cluster. From this queue, a log file is created in the user space. We will also add a user space daemon. This daemon will be responsible to read and analyse the log file. This daemon will also display the data in a visual format (graphs). 1.3. MOTIVATION ● Opportunity to work on an open-source project. ● We got first hand industry exposure by working with Veritas LLC, Pune. ● This project also gave us the opportunity to work on an industry standard replication project. ● We had the opportunity to learn about cluster networks, as well as about system programming. PICT - INFORMATION TECHNOLOGY – 2020-21 8

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module 2. LITERATURE REVIEW 2.1. EXISTING METHODOLOGIES 2.1.1. DRBD DRBD stands for Distributed Replicated Block Device. It is an open source data replication storage solution. It consists of a primary node, and one or several other secondary nodes. DRBD works by layering over logical block devices on participating nodes. Primary node is responsible for functioning of the system. When a write operation occurs at the primary node, it is simultaneously propagated to secondary nodes. In case the primary node fails, then one of the secondary nodes is promoted to the designation of primary node. [1] DRBD consists of two independent modules - ● A kernel module, that implements a driver replicated across the primary and secondary nodes. ● User space applications, that are responsible for administration. 2.1.2. DRBD vs RAID ●

in a cluster-aware file system. 2.2. PROPOSED METHODOLOGIES 2.2.1. LINUX ARCHITECTURE The Linux System consists of four primary components - ● Hardware layer ● Kernel PICT - INFORMATION TECHNOLOGY – 2020-21 9

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module ● Shell ● User applications and utilities

computer and the hardware. The kernel is a separate entity as compared to the user space, and hence the programs written for the kernel are different as compared to the normal user programs. Fig 1. Linux Architecture 2.2.2. CHARACTER DEVICE DRIVER There are two types of device drivers - ● Character Driver ● Block Driver This division is made on the basis of speed and the organisation of data in the system. Character drivers are usually slow speed devices, which handle comparatively smaller amounts of data and do not require frequent seeks or writes from and to the system. Examples - Keyboards, mouse, joysticks etc. These devices generally involve sequential read and write, i.e, byte by byte. PICT - INFORMATION TECHNOLOGY – 2020-21 10

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module Block drivers are the ones which handle comparatively larger amounts of data. Devices such as hard disks, floppies and ram generally fall in this category. Reading and writing of data takes place at the block level in these devices. [2] The operating system relies on two unique and fixed identifiers for identifying each device - ● Major Number ● Minor Number Each unique identifier consists of two parts or halves. The first part is used to identify the type of the device (serial port, etc); and the second part is used to denote the device (1. disk, 2. serial port). The major number usually identifies the driver, and the minor number denotes the device that is using the particular driver. Generally each driver has a major number that is responsible for all the minor numbers associated with that major number. 2.2.3. LOADABLE KERNEL MODULE Kernel configuration primarily refers to selecting the files that we want, when the kernel is compiled and run. Appending code to the kernel basically refers to adding the required files to the source code and recompiling the kernel. [5] Loadable Kernel Module (LKM) refers to the code that is added to the kernel, without shutting it down completely. LKMs are generally of three types - ● Device Drivers ● Filesystem Drivers ● System Calls PICT - INFORMATION TECHNOLOGY – 2020-21 11

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module The kernel is designed so as to encapsulate the functions, so that new functions can be added to it without disturbing the basic functionalities. The new functions/ LKMs are added in complete isolation of the existing kernel code base. LKMs are primarily used for these six tasks - ● Device drivers - The operating system (kernel) establishes contact with the hardware so as to ensure its proper functioning. ● Filesystem drivers - These drivers are used to understand the hierarchy of a filesystem in an operating system. ● System calls - User space applications and programs make use of system calls in order to perform certain tasks such as reading a file, creating a new process etc. ● Network drivers - These drivers are used to identify the network protocols.. ● TTY line disciplines - A console command. ● Executable interpreters - Responsible to load and execute an executable file/ code. PICT - INFORMATION TECHNOLOGY – 2020-21 12

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module 3. REQUIREMENT SPECIFICATION AND ANALYSIS 3.1. PROBLEM DEFINITION Build a packet tracing module on DRBD to track the data packets transferred over a cluster in a network during data replication between multiple block devices and analyse the transfer of packets. 3.2. CONCEPT Data Replication is the process of storing data in more than one site or node. It is useful in improving the

availability of data. It is simply copying data from a database from one server to another server so that all the users can share the same data without any inconsistency.

100%	MATCHING BLOCK 6/7	W	https://sites.ualberta.ca/dept/chemeng/AI ...

Packet tracing is the process by which you can verify the path of a packet through the layers to its destination.

Using packet tracing it is possible to check the volume of data being transferred and various information related to the process of replication as at its core, replication occurs through the transfer of packets. 3.3. SCOPE ● Provide an enterprise grade solution to analyze the packet transfer in replication phase at multiple data nodes. ● Calculate the accuracy of data replicated at multiple data nodes in a cluster. ● Determine the bandwidth required over a network based on the accuracy of the transfer of packets. ● Determine the number of replication nodes required in a data cluster. ● Learning Scope : ○ Work on an open-source project with industry guidance ○ Learn Linux programming ○ Learn about 3.4. OBJECTIVE ● Understand how replication in DRBD works - code flow how/where we put the packet on air in code. ● Writing a kernel module to trace/keep track of logs in memory and on-disk. ● Enable tracing on the modules dynamically. PICT - INFORMATION TECHNOLOGY – 2020-21 13

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module ● Least or no performance impact is desirable. ● Writing utility/daemon to pull tracing information from kernel - parse the data for reporting/debugging problems ● Use a container to run the analyser tool. ● Eg: plot a graph, create sequential logs combining primary and secondary logs. 3.5. PROJECT REQUIREMENTS 3.5.1. FUNCTIONAL REQUIREMENTS ● Log data and acknowledgement packets passed between nodes in a DRBD cluster ● Analyse above mentioned logs ● Represent analysis in graphical format 3.5.2. NON-FUNCTIONAL REQUIREMENTS ● No performance impact on DRBD cluster 3.5.3. HARDWARE REQUIREMENTS ● System with i5 or higher ● System with 8 GB of RAM 3.5.4. SOFTWARE REQUIREMENTS ● Oracle VM VirtualBox ● Two Virtual Machines ● CentOS running on two VMs ● DRBD 3.6. PROJECT PLAN 3.6.1. MODULE SPLIT-UP ● Understanding and Identifying Data Flow Points in DRBD. ● Creating Global Queue for logs in the module. ● Calling our module functions from DRBD. PICT - INFORMATION TECHNOLOGY – 2020-21 14

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module ● Preparing a log file. ● Writing a user land utility to analyse data from a log file. ● Display analysis in graphical format. 3.6.2. GANTT CHART Fig 2. Gantt Chart PICT - INFORMATION TECHNOLOGY – 2020-21 15

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module 4. SYSTEM ANALYSIS AND DESIGN 4.1. ARCHITECTURE Being an open source software, we can review and make changes in the code base of DRBD. We can identify the points where data packets are sent and received in the code. We will implement our module as a loadable kernel module. This module will be responsible for communicating with the drbd module and preparing the log file. We will add an user land utility which will make system calls to our module and be responsible for analysing the data in the log file and display results and reports in graphical format. The kernel space consists of a global queue, where the corresponding details will be enqueued and dequeued as and when the read and write operations take place at the primary and secondary nodes respectively. We then plan to write a user space utility that will take this log data as the input, and plot different graphs accordingly. PICT - INFORMATION TECHNOLOGY – 2020-21 16

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module Fig 3. Architecture Diagram 4.2. COMPONENTS OF THE SYSTEM 1. K-thread 2. User space entity 3. Kernel module 4. Global queue PICT - INFORMATION TECHNOLOGY – 2020-21 17

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module 5. IMPLEMENTATION 5.1. STAGES OF IMPLEMENTATION 5.1.1. CHARACTER DEVICE DRIVER PICT - INFORMATION TECHNOLOGY – 2020-21 18

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module PICT - INFORMATION TECHNOLOGY – 2020-21 19

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module Fig 4. Character Device Driver Code PICT - INFORMATION TECHNOLOGY – 2020-21 20

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module 5.1.2. DATA STRUCTURE OF LOG Fig 5. Data Structure of Log 5.2. ELABORATE IMPLEMENTATION ISSUES/ TECHNIQUES/ SOFTWARE TOOLS ETC. 5.2.1. CSCOPE

Cscope is used to browse the code in the console. Cscope generally works along with Vim. Using Cscope we identified the main important functions in DRBD source code like drbd_send_dblock etc. [6] Fig 6. Cscope PICT - INFORMATION TECHNOLOGY – 2020-21 21

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module 5.2.2. SYSTEM TAP System Tap is generally used along with Cscope to find the flow data, in the functions identified using Cscope. [7] Fig 7. System Tap PICT - INFORMATION TECHNOLOGY – 2020-21 22

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module 6. CONCLUSION 6.1. CONCLUSION Thus in this project, we have studied about DRBD which is an open source, data replication technology. DRBD is used to replicate the data between a primary, and one or multiple secondary nodes. We also started the implementation of a network packet tracing module for DRBD, that will basically maintain a log of all the reads and writes at the primary and secondary data locations. This is done by using a global queue. We will also develop a user space daemon that will help represent all the data from the log in a visual format (in the form of graphs, etc). 6.2. DRAWBACKS DRBD is comparatively more difficult to set up as compared to RAID. DRBD also does not promise the performance benefits, as in the case of RAID (especially RAID 0). The packet tracing module will be written inside the DRBD send data function and called as and when the read or write operation is performed. Thus the entire DRBD code needs to be re-compiled in case of any error in the code. PICT - INFORMATION TECHNOLOGY – 2020-21 23

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module 7. REFERENCES [1] https://www.linbit.com/drbd/ [2] https://linux-kernel-labs.github.io/refs/heads/master/labs/device_drivers.html [3] https://www.elprocus.com/linux-operating-system/ [4] https://developer.ibm.com/tutorials/l-drbd/ [5] https://tldp.org/HOWTO/Module-HOWTO/x73.html [6] https://courses.cs.washington.edu/courses/cse451/12sp/tutorials/tutorial_cscope.html [7]

100% **MATCHING BLOCK 7/7** **SA** Thesis_luhtala_final_pdf_a.pdf (D14518752)

https://access.redhat.com/documentation/en-us/red_hat_enterprise_linux/5/html/ [8]

D. A. Patterson, P. Chen, G. Gibson and R. H. Katz, "Introduction to redundant arrays of inexpensive disks (RAID)," Digest of Papers. COMPCON Spring 89. Thirty-Fourth IEEE Computer Society International Conference: Intellectual Leverage, San Francisco, CA, USA, 1989, pp. 112-117, doi: 10.1109/CMPCON.1989.301912. [9] M. Wiesmann, F. Pedone, A. Schiper, B. Kemme and G. Alonso, "Understanding replication in databases and distributed systems," Proceedings 20th IEEE International Conference on Distributed Computing Systems, Taipei, Taiwan, 2000, pp. 464-474, doi: 10.1109/ICDCS.2000.840959. PICT - INFORMATION TECHNOLOGY – 2020-21 24

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module 8. APPENDICES 8.1. PROJECT WEEKLY LOG PICT - INFORMATION TECHNOLOGY – 2020-21 25

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module 8.2. REVIEW 1 CHECKLIST PICT - INFORMATION TECHNOLOGY – 2020-21 26

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module PICT - INFORMATION TECHNOLOGY – 2020-21 27

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module PICT - INFORMATION TECHNOLOGY – 2020-21 28

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module 8.3. REVIEW 2 CHECKLIST PICT - INFORMATION TECHNOLOGY – 2020-21 29

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module PICT - INFORMATION TECHNOLOGY – 2020-21 30

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module 8.4. RESEARCH REVIEW CHECKLIST PICT - INFORMATION TECHNOLOGY – 2020-21 31

Distributed Replicated Block Device (DRBD) Network Packet Tracing Module PICT - INFORMATION TECHNOLOGY – 2020-21 32

Hit and source - focused comparison, Side by Side

Submitted text As student entered the text in the submitted document.

Matching text As the text appears in the source.

1/7	SUBMITTED TEXT	28 WORDS	95%	MATCHING TEXT	28 WORDS

Submitted to the
Savitribai Phule Pune University
In partial fulfillment for the award of the Degree of
Bachelor of Engineering
in
Information Technology
by

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY,PUNE
IN THE PARTIAL FULFILLMENT FOR THE AWARD OF THE DEGREE
OF

BACHELOR OF ENGINEERING
IN
INFORMATION TECHNOLOGY

BY

SA Final Report (3).pdf (D68239032)

2/7	SUBMITTED TEXT	24 WORDS	47%	MATCHING TEXT	24 WORDS

the guidance of
Prof. V. R. Jaiswal
Sponsored by
Veritas LLC, Pune

Department Of Information Technology
Pune Institute of Computer Technology

THE GUIDANCE OF
Prof.
                      N.V. BURADKAR

DEPARTMENT OF INFORMATION TECHNOLOGY
PUNE INSTITUTE OF COMPUTER TECHNOLOGY

SA Final Report (3).pdf (D68239032)

URKUND

3/7	SUBMITTED TEXT	34 WORDS	100%	MATCHING TEXT	34 WORDS

Requirement Specifications and Analysis 13
3.1 Problem Definition 13
3.2 Concept 13
3.3 Scope 13
3.4 Objective 13
3.5 Project Requirements 14

REQUIREMENT SPECIFICATION AND ANALYSIS 14
3.1 PROBLEM DEFINITION 14
3.2 CONCEPT 14
3.3 SCOPE 14
3.4 OBJECTIVE 15
3.5 PROJECT REQUIREMENTS 15
3.5.1

SA Final Report (3).pdf (D68239032)

4/7	SUBMITTED TEXT	127 WORDS	90%	MATCHING TEXT	127 WORDS

RAID is designed for two or more disks connected locally whereas
DRBD is designed to replicate a block-device over a network.
● We use RAID (software or hardware) to increase the reliability of the
local storage over 2 or more disks.DRBD would sit on top of your
RAID and replicate that data to another server for failover purposes.
● DRBD can also allow two servers to access that data at the same time,
which you can't really do with RAID

RAID is designed for two or more disks connected locally.
<   DRBD is designed to replicate a block-device over a network.
<
< Long:
<
< You'd use RAID (software or hardware) to increase the reliability of
< your _local_ storage over 2 or more disks.
< DRBD would sit on top of your RAID and replicate that data to another
< server for fail-over purposes.
< DRBD can also allow two servers to access that data at the same time,
< which you can't really do with RAID (

W https://lists.linbit.com/pipermail/drbd-user/2010-February/013449.html

5/7	SUBMITTED TEXT	34 WORDS	57%	MATCHING TEXT	34 WORDS

The kernel is the central part of any operating system. It is responsible for the
overall functioning of the

The kernel is the core part of the operating system, which is responsible for all the major activities of the

W https://www.elprocus.com/linux-operating-system/

URKUND

6/7	SUBMITTED TEXT	37 WORDS	100%	MATCHING TEXT	37 WORDS

Packet tracing is the process by which you can verify the path of a
packet through the layers to its destination.

Packet tracing is the process by which you can verify the path of a packet through the layers to its destination.

W https://sites.ualberta.ca/dept/chemeng/AIX-43/share/man/info/C/a_doc_lib/aixbman/commadmn/tcp_pro ...

7/7	SUBMITTED TEXT	3 WORDS	100%	MATCHING TEXT	3 WORDS

https://access.redhat.com/documentation/en-us/red_hat_enterprise_linux/5/html/
[8]

https://access.redhat.com/documentation/en-US/Red_Hat_Enterprise_Linux/6/html/

SA Thesis_luhtala_final_pdf_a.pdf (D14518752)