Network Working Group
Internet-Draft
Intended status: Informational
Expires: January 7, 2009

D. Brasher
Interlinux LTD
July 6, 2008

Distributed Internet Archive Protocol (DIAP)
draft-brasher-diap-02

Status of this Memo

Abstract

A de-centralised, self-contained and managed storage protocol. A system to provide strong storage fail over by using existing resources over networks distributing vital data evenly. Rapid deployment and high redundancy for small to medium organisations as well as individuals. Designed to reduce dependency on tape backup systems. The protocol also has implications for long term archiving. By classifying data vitality values the limitations in physical space due to bandwidth constrictions can be overcome and the usefulness of DIAP maximised.

Table of Contents

1. Introduction 3
2. Architecture 3
3. Prototype Design 5
4. DVTL 6
5. Hyper Virtual auto-changer 6
6. Data Vitality 6
7. DIAP Rule of Thumb 6
8. Community Project and UK Trademark 7
9. DPA 7
10. Conclusion 7
11. Security Considerations 7
12. Acknowledgements 7
13. Change Log 8
14. Informative References 8
Author's Address 8
Intellectual Property and Copyright Statements 9

1. Introduction

Three nodes either between sites say between offices, homes, on a campus or over WAN's, which could be dedicated to storage or used for existing services, have a round robin synchronisation of FULL - differential backup pools where the source of data ranges from a personal laptop to a file store over unused band-width where the data rate is dynamically controlled, including compression, according to load and availability. Three for simplicity and because the probability of failure beyond three is so small the extra coding to accommodate more nodes would be self-defeating. In real life use of three nodes for a DIAP pool is strong enough. Chaining together DIAP pools to extend data retention periods is a future aim of the project. Also designed, as project maturity is reached, to help reduce an organisations carbon footprint, the extent to which is unknown at this stage.

2. Architecture

The system reduces single point of failure by creating a single FULL copy on each node at the beginning of the month the storing the differentials in a distributed manner. Use a program such as Bacula set to use a monthly FULL - differential. If a copy fails then the system will retry the next day but you loose the day of failure. Using rsync logs you can trace / track the successful copies. The copies are staggered so that each rsync copy list is made before new files are put into each directory by a few minutes. I.e. bd9-cd9 starts before ad9-bd9 and ad0-bd0 is last. Redundancy is split across three nodes, no duplicate days apart from the first FULL copy. DIAP pool is possibly equivalent to 30 tapes every month but stored at three locations. 10 days every three days, at each. You are advised to have some knowledge of the average differential size.

Slots

Slots	A	B	C
(Dirs)	aFull	bFull	cFull
	ad0	bd0	cd0
	ad1	bd1	cd1
	ad2	bd2	cd2
	ad3	bd3	cd3
	ad4	bd4	cd4
	ad5	bd5	cd5
	ad6	bd6	cd6
	ad7	bd7	cd7
	ad8	bd8	

Table 1: Slots

Calculations:-

LBM = Lowest Maximum Bandwidth between any three nodes NB: actual max transfer will vary so test transfers are recommended for accuracy.

LBM assumes all available bandwidth is allocated to running DIAP.
Max aFULL01 = LMB x 6 hrs This assumes no transfer interruptions and that the maximum bandwidth is constant.

Ave. Diff = (Sum 29 (or a month) Daily Differentials) / 29.

Ave Differential is variable depending on your storage growth, this represents a trend and can be an estimate to start with, but by watching the trend of Differential growth more accurate calculations can be made. It is assumed your Differentials are always less the the initial FULL copy.

Min DIAP Dir size (node a) = (Max aFULL01 x 2) + (29 x Ave. Diff) + (1 x Ave. Diff) Plus 1 x Ave. Diff to account for ad0.

Min DIAP Dir size (node b/c) = (Max aFULL01 x 2) + (29 x Ave. Diff)
You can include transfer log files in the Min DIAP Dir size, for simplicity they have been omitted.

Example System

Example System	LMB x 6 hrs	Ave. Diff	Max aFULL01
LBM occurs between b->c	1Mbit/Sec	Est. 500 MiB	2.6 GiB

Table 2: Example System

Min DIAP Dir size (node b-c) (2.6 x 2) + (29 x 0.5) = 19.7 GiB

DIAP Dir: this is the working directory used on each node and contains all DIAP configuration working and storage directories.

If a copy fails then the system will retry the next day but you loose the day of failure. Using rsync logs you can trace / track the successful copies

Flow of data.

Cron Jobs	Daily	Node A	B	C
t=0	Special	aFull-bFull (1st of Mnth)	bFull-cFull (2nd of Mnth)	
Start 00:00	2	ad0-bd0 t=30	bd0-cd0	cd0-ad1
End	3	ad1-bd1	bd1-cd1	cd1-ad2
07:30	4	ad2-bd2	bd2-cd2	cd2-ad3
t=0	5	ad3-bd3	bd3-cd3	cd3-ad4
(00:00)	6	ad4-bd4	bd4-cd4	cd4-ad5
	7	ad5-bd5	bd5-cd5	cd5-ad6
t=0-30	8	ad6-bd6	bd6-cd6	cd6-ad7
(00:30)	9	ad7-bd7	bd7-cd7	cd7-ad8
	10	ad8-bd8	bd8-cd8 t=0m	

Table 3: Data Flow

3. Prototype Design

The prototype is built of several components and uses the Linux Operating system. Bash scripts are used to deploy DIAP on three POSIX user accounts using expect and ssh. Ssh certificates are setup between three POSIX accounts. A single configuration file is use to set environment variables.

The system requires a series of directories used to store the data fed into ad0 and aFull:

```
mkdir aFull ad0 ad1 ad2 ad3 ad4 ad5 ad6 ad7 ad8 && touch log_a
```

Cron jobs are used to implement table 2 architecture:

```
0 1 0 * * rsync -az -e ssh --timeout=1800 --numeric-ids \ --log-
file=/home/diap/log_b --ignore-errors --bwlimit=128 \ ~/aFull/ diap@
$IP_ADD_B:bFull
```

4. DVTL

(VTL) The virtual tape Library is a device located in one location whereas DIAP enables a Distributed virtual Tape Library to exist.

5. Hyper Virtual auto-changer

This term is derived from the term virtual auto-changer. A virtual auto-changer still requires hardware tape drives, 'Hyper' takes this one stage further by emulating the virtual auto-changer in software.

6. Data Vitality

Data vitality is a measure of the organisation subjective view of the value of particular data types.

7. DIAP Rule of Thumb

Observing an email archive, at 272MBytes, having never deleted an email permanently and the file, ../mail, has been in use for 4 years. During this time available xDSL line Bandwidth has increased, 2004 500MBits/sec to 1GBit/sec, 2008 1GBit/sec to 6GBits/Sec this is about a 150% yearly increase whereas the mailbox has increased yearly by about 50%. It is this difference which DIAP attempts to use classing email record as 'mission critical' - Other record types will increase at different rates, as will bandwidth depending on location, but probably less than the average yearly bandwidth increase. This idea needs expanding but forms the foundation for the usefulness of DIAP, describing a DIAP rule of thumb. DIAP can also be viewed as a technique.

8. Community Project and UK Trademark

A community project resides at http://www.diap.org.uk to facilitate the development of working implementations. The current working prototype is released here under GPL licence rules. A UK Trade Mark has been applied for to protect the acronym DIAP for use by the wider Open Source community.

9. DPA

DPA compliance and awareness.

10. Conclusion

The incremental data retention tuned to the needs of an organisation so that some data is always available from any node in the backup pool quickly to within a certain time frame and perhaps tape storage stations strategically places in various secure locations for older data retention. This system would avoid using prohibitively expensive packages by reusing resources, building on Open Source technologies and have a coherent strategy across many sites increasing the level of redundancy to a high degree. A three tier strategy involving DIAP as the bottom layer, file collection uppermost and use of pre-existing mid-term infrastructure could make up a disaster recovery plan.

With layers of indexing, accounting and management facilities. An assumption is that individual file encryption the responsibility of the file owner, this does not rule out hard drive or partition encryption of individual nodes considered to reside at insecure locations. If used for these locations physical security automatic fail-safe measures to trigger archive deposits useless upon theft can be deployed. Similar fail-safe techniques deployed for attempted network security breaches. Virus scanners would be set to scan existing archives periodically and on entry to the archive pool.

11. Security Considerations

Open root access is not recommended for SSH. Using ports other than the default 22 is advised.

12. Acknowledgements

Thanks are due to Myles McClelland and a number of individuals from

various groups.

13. Change Log

06 July 08 - Architecture - arithmetic. Acknowledgements.

May 08 - Address.

14. Informative References

[RFC4810] Wallace, C., Pordesch, U., and R. Brandner, "Long-Term Archive Service Requirements", RFC 4810, March 2007.

[DIAP]   Brasher, D., "Distributed Internet Archive Protocol (DIAP)", April 2008, <http://www.diap.org.uk>.

Author's Address

Damian Brasher
Interlinux LTD
PO Box 1623
Southampton, Hampshire SO15 9AE
United Kingdom

Email: dbrasher@interlinux.co.uk

Full Copyright Statement

Intellectual Property