

RepliWeb R-1 White Paper Large File Accelerator (LFA) Transport Engine

1. Overview

RepliWeb R-1 is a high-level, one-to-many file distribution system, engineered to perform in complex and dynamic IP networks (LAN, WAN, Internet, eCDN, satellite based, over Unicast and Multicast). R-1 provides unattended, Center-to-Edge content deployment and synchronization that scales in volume of data, number of processes and number of target machines, employing cross-platform server environments (Microsoft Windows™, UNIX, Linux). Each content deployment / synchronization job resolves the absolute minimum amount of data to be transferred or deleted between the distribution source and targets.

To achieve the required level of mission-critical reliability and security, R-1 deploys a robust transport layer allowing the use of the best transfer method for each network topology and characteristics, employing multiple transport engines.

The Large File Accelerator (LFA) transport engine provides a new, reliable & efficient transfer of large files between remote sites.

This document describes the Large File Accelerator (LFA) transport engine operation in R-1. LFA is available in R-1 version 3.6.4b and above as an optional, add-on.

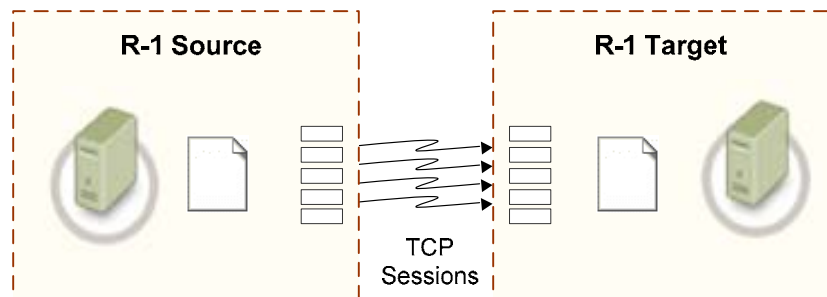


Figure 1 - One large file replicated with LFA set to use multiple sessions

1.	Overview	1
2.	Principles of Operation	3
2.1.	Multiple Sessions	3
2.2.	Differential Transfer	3
2.3.	Transfer Integrity	4
2.4.	Compression	4
2.5.	SSL	4
3.	LFA Features	5
4.	LFA in R-1	6
4.1.	Recovery	6
4.2.	Rollback	6
4.3.	Temporary Files	6
4.4.	Unicode support	6
5.	Use Cases Guidelines	7
5.1.	Selecting LFA as the Transport Engine	7
5.2.	Calculating Sessions	8
5.2.1.	Network Availability	8
5.2.2.	Replication using Multiple Sessions	8
5.2.3.	Replication using Multiple Sessions & Concurrent Transfers	9
5.2.4.	Distribution Considerations	9
5.3.	Using Compression	10
5.4.	Using Differential Transfer	10
6.	About RepliWeb	11

2. Principles of Operation

R-1's Large File Accelerator (LFA) transport engine is geared for speed and maximal throughput. Supporting concurrent sessions (by breaking files into multiple sessions) and parallel transfer of partial file lists (by joining smaller files into parallel groups & sessions) it employs the most out of the network capabilities.

R-1's LFA provides additional options including compression, differential transfer, file integrity, overwrite after completion, security and recovery.

2.1. Multiple Sessions

When multiple sessions are used, large enough files are divided into a number of sessions and transferred in parallel. For example, when configured to use 5 sessions, a 5MB file will be transferred on 5 simultaneous links, each transferring a different section of the file. Small files, 128k and below, will not be divided. Also, there will not be a session smaller than 128k.

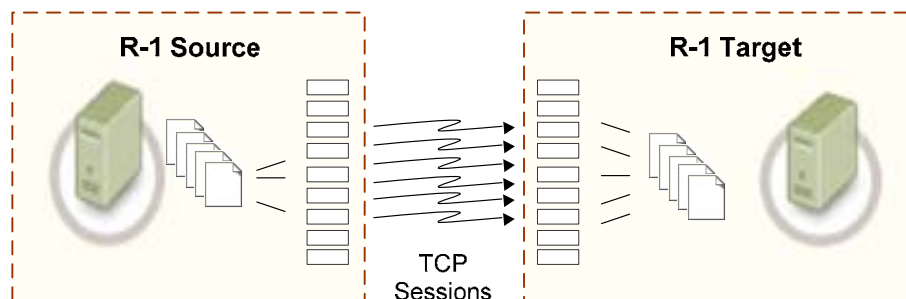


Figure 2 – Multiple large files replicated with LFA set to use multiple sessions

2.2. Differential Transfer

Differential transfer can be configured so that files are analyzed prior to the transfer and only the differences (if exist) are actually transferred to the target. This provides considerable performance improvements especially when the files typically only change in limited areas (for example, when files to be transferred are log files that are appended to, or BMP files).

Analysis works in the following manner:

- Prior to transfer, LFA queries the sizes of the source and target file. If the target file is missing, no comparison is attempted.

- LFA then calculates the common size (minimum of the two file sizes) and compares them. For example, if the file on one side is 50MB and the other file is 40MB (e.g. a log file that has grown), only the first 40MB will be compared.
- The common size is divided between the sessions; in the example above, if there are 4 sessions, each will compare 10MB of file.
- Each session divides its “slice” of the file into 96 parts (or less if the slice isn’t big enough); MD5 signatures of each sub-slice are calculated by the source and the target.
- The signatures are compared; only slices where the file has actually been modified will be transferred.

LFA can be configured to only analyze files larger than a certain size threshold; this is useful when the network configuration makes it more efficient to simply transfer small files instead of analyzing them first.

2.3. Transfer Integrity

LFA can compare the source file with the target file after transferring. If the files differ the transfer is considered to have failed. The comparison uses the same algorithm as Differential Transfer analysis above, but the effect is different: if any difference is found between the source and target after transfer, the transfer is considered to have failed.

2.4. Compression

LFA supports optional `deflate` (`gzip` algorithm) compression. Various configuration options can be used to optimize compression usage when compression might slow down the replication process:

- According to file types (do not try to compress files that are already in a compressed format, like images)
- According to file size (do not try to compress too small files as it will degrade performance).

2.5. SSL

SSL protocol can be used for Authentication and Encryption between the Center and the Edge while replicating files using the LFA transport engine.

3. LFA Features

R-1's LFA transport engine is a fast transport engine (using similar principles of R-1's LAN transport engine), with the additional advantages of its WAN transport engine and new features that further accelerates the transferring of very large files:

- **Speed** - Employs the most out of the network connection, geared for speed and maximal throughput.
- **Speed II - Multiple Sessions** – In order to accelerate transfer of large files, these files can be broken into smaller sessions and be transferred in parallel.
- **Speed III - Concurrent Transfers** – Transfer performance can be boosted when performing simultaneous parallel transfers of partial file lists.
- **Compression** – Transparent compression providing excellent performance.
- **Differential Transfer (block level)** – The data transfer can be limited to only the differences between the source and target files, reducing both transfer time and bandwidth costs.
- **Integrity** - R-1 can perform an integrity check on each and every file transferred to ensure the accuracy of the transfer.
- **Integrity II – Overwrite after completion** - files can be transferred to a temporary location on the target, leaving the original file on the target intact. Once the entire file has been transferred to the target, the original file is overwritten, and the temporary file removed. This will ensure that users will only access a complete file.
- **Security** – SSL protocol can be used for Authentication and Encryption between the source and target machines.
- **Recovery** - An interrupted transfer will continue from the exact point of failure within the file, down to the block level.

4. LFA in R-1

Using LFA as the transport engine in R-1 you gain all R-1's features and advantages as an integral part of the complete replication process.

LFA's replication integrity and speed are accompanied by R-1's replication features like scheduling, continuous updates, notifications, pre- and post-commands and cross-platform replication.

4.1. Recovery

If a network connection is dropped or power outage occur during the replication, transfer will resume from the by transferring the interrupted file. Successfully transferred files will not be transferred again.

4.2. Rollback

R-1's Rollback functionality provides a means of effectively undoing the effects of multiple replications on a target system to a user-specified 'point-in-time'. This allows system administrators to undo the changes carried out by a one-time, or scheduled R-1 job, providing Disaster Recovery and Rollback from unsuccessful publishing.

4.3. Temporary Files

R-1's LFA can be configured to create the files with a temporary name, and only to rename them after they have been transferred. The temporary file is created in a temporary directory and is renamed to the target's name after the file has been transferred completely.

4.4. Unicode support

As R-1 in general, LFA also completely supports transferring files whose names are in Unicode format.

5. Use Cases Guidelines

This section provides guidelines on when to choose LFA as your R-1 transport engine and how to set its properties.

Note that these guidelines are only rule-of-thumbs, and as conditions may vary significantly, depending on your files and network conditions, some testing should be performed to achieve maximal speed.

5.1. Selecting LFA as the Transport Engine

In general, the LFA transport engine will be more effective when:

- The files are larger;
- The files are compressible;
- You want to maximize the throughput on your network whatever the pipe is (from 512k up to Gig WAN)

A file is regarded “large” when its size is at least approximately 2MB. This number can vary, of course, depending on network availability and file format.

How well the data can be compressed should be considered while selecting the most appropriate transport engine. Obviously compressed formats (.zip, .rar, .mp3, etc.) cannot be compressed at all. XML, HTML and the likes can be compressed very well, while other document types (like Office documents) are somewhere in between.

“Stability” is an additional factor that should be taken into account while trying to determine which transport engine to use. Stability refers to how much a file tends to contain the same data at the same offset. The most stable files would be log files, which only grow, and the existing part remains unchanged. A camera dump that just takes different pictures is an example of a very unstable file. Human-edited documents, generated HTML etc. are not very stable, as a simple change like inserting a word can shift all the contents of the whole file.

The smaller and non-compressible the files are (like movies), and the better network connection is, it may be more efficient to use the LAN transport engine.

Consider using the LFA transport engine in the following cases:

- When replicating a small number of very large files, use the LFA transport engine breaking the files to multiple sessions.

- When replicating a mixture of files sizes, or many large files, use the LFA transport engine, breaking the files to multiple sessions and sending the files using concurrent transfers.
- When replicating any files over LAN (or other good performance network) and compression is required, use the LFA transport engine with compression.
- When replicating incremented files (like log files) or files which tend to change in small portions of the file use LFA with Differential Transfer.

5.2. Calculating Sessions

Setting R-1 and LFA to use multiple sessions and concurrent streams in an effective way is influenced by a number of parameters.

5.2.1. Network Availability

The ideal number of sessions depends on the bandwidth and latency. The purpose is to maximize the bandwidth available to speed up the transfer time.

As a rule of thumb, use the following formula:

$$\# \text{ of Sessions} = \text{BDP} / 64$$

where **BDP** is

$$\text{Bandwidth (in MB)} * \text{Latency (in ms)}$$

and **Sessions** is

$$\# \text{ of LFA Concurrent Sessions} * \# \text{ of Transfers}$$

Set LFA properties inline with the calculated number of sessions to use. Tweak the numbers according to the transferred files. If there are multiple files, it may be more effective to use a higher # of Transfers rather than Concurrent Sessions. When there are fewer large files, it may be more effective to use a larger number of Concurrent Sessions.

5.2.2. Replication using Multiple Sessions

While setting up a replication job, consider the amount of data that is replicated and the network availability while calculating the number of sessions to use.

5.2.3. Replication using Multiple Sessions & Concurrent Transfers

In addition to using multiple sessions, R-1 also allows transferring files in concurrent transfers. Transfer performance can be further boosted when performing parallel transfers of partial file lists.

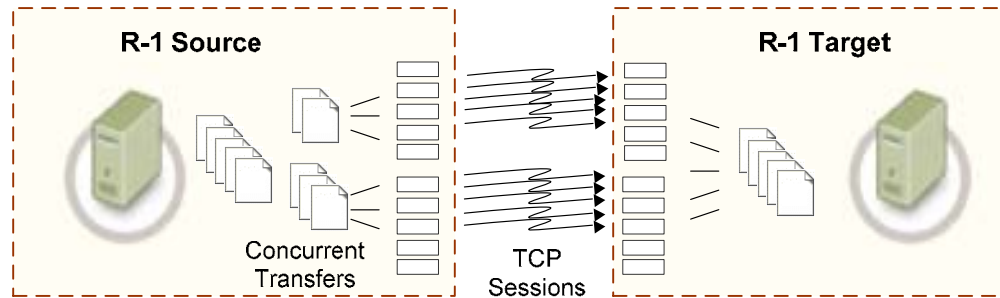


Figure 3 – Multiple large files replicated with LFA set to use multiple streams & concurrent transfers

5.2.4. Distribution Considerations

When distributing large files to multiple Edges with LFA set to use multiple sessions, LFA is applied to each target. R-1 will calculate what needs to be transferred to each target separately, depending on the network availability and transferred data.

For example, a distribution job distributing files to 5 different Edges, set to use 3 concurrent transfers and 4 concurrent sessions, distributing a very large file (such that will use all concurrent transfers and sessions), will use up to $3 \times 4 \times 5$ concurrent TCP sessions.

While setting a distribution job, consider using different settings for each Edge depending on the network availability to that Edge and the data that is going to be transferred to it (using the Edge Override option).

5.3. Using Compression

Consider the following table as basic general guidelines for using compression, and compression level according to “compressibility” and network performance while using the LFA transport engine.

	Low Compressibility (ZIP)	Mid Compressibility (DOC)	High Compressibility (XML)
WAN		Medium	High
LAN (100Mb)		Low	Low
LAN (1Gb)			

NOTE: The R-1 LAN Transport Engine does not support compression. As the above table shows, on a 100Mb LAN it would be generally better to use LFA with (low) compression when the files are compressible. This only applies to large files; for scenarios with predominately small files it is recommended to use the LAN transport engine.

5.4. Using Differential Transfer

Consider the following table as basic general guidelines for using differential transfer according to “stability” and network performance while using the LFA transport engine:

	1%-15% Stable	15%-40% Stable	40%-90% Stable	>90% Stable
WAN	√	√	√	√
LAN (100Mb)				
<i>Uncompressed</i>		√	√	√
<i>Compressed, single-stream</i>			√	√
<i>Compressed, multi-stream</i>				
LAN (1Gb)				
<i>Single stream</i>				√
<i>Multi-stream</i>				

NOTE: It is not recommended to specify Differential Transfer while using the LAN transport engine with Multiple Streams. While using Single Stream, use Differential Transfer with relatively stable files.

6. About RepliWeb

RepliWeb specializes in file transfer and content automation solutions. The company's products accelerate file transfer processes, simplify their management, and improve the security and reliability of related business processes.

###

For any additional information, contact us at support.repliweb.com