

The Rembo Wizard 2.0



System Snapshots

How To

by

Petri Mäkijärvi

January 10, 2003

Abstract:

The Rembo Wizard 2.0 is a free plug-in module for the Rembo Toolkit 2.0, a PXE-enabled, Pre-OS platform for the system hard disk management for Windows and Linux PC-computers.

This document explains the concepts behind the System Snapshot images and gives configuration tips to the system administration for better usage of The Rembo Wizard with the user managed backups.

©2002 European Synchrotron Radiation Facility, Grenoble, France

All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without written permission of the European Synchrotron Radiation Facility, Grenoble, France.





Introduction.....	5
System Snapshot is a full image.....	5
Using the local hard disk cache.....	6
Distributing and Backing Up Shared Files Archives.....	7
Protecting administrator's System Snapshots.....	8





Introduction

[The Rembo Wizard](#) is a free plug-in module for the [Rembo Toolkit](#).

The default operating mode of The Rembo Wizard is to take System Snapshots of a system once a base image has been taken from its system partition.

“System Snapshot” is a term of The Rembo Wizard defining that

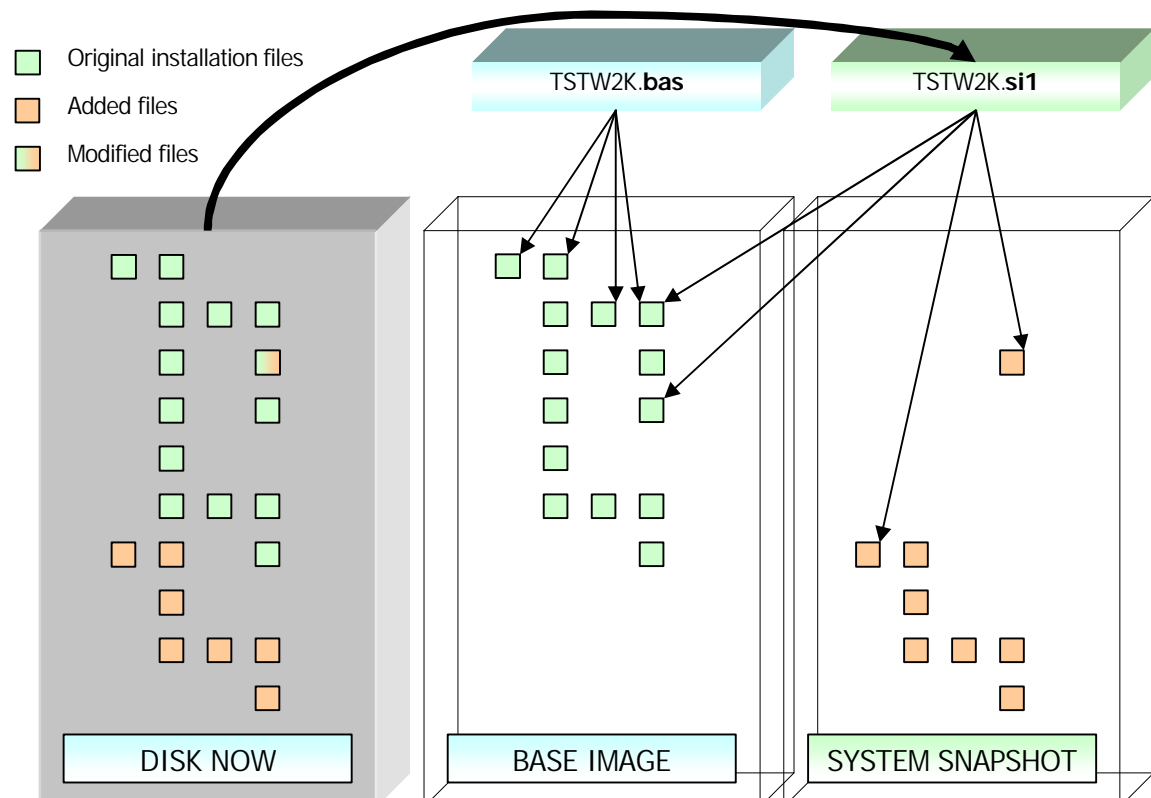
1. The image has been taken by the end-user of the system (or at the user level in general)
2. The image of the system is taken once there exists a base image for the system

Starting from the Rembo Toolkit 2.0, both the local hard disk’s cached Rembo file system and the server’s Rembo file system are using a concept which is called “*shared files*” for archives. Instead of the 1.1 version’s monolithic archive file, the archive file is rather a shared file system index file. The actual contents of the disk, the directories and the files are stored on separate structures, that can be shared between the archives.

The file sharing between the archives is the key of the System Snapshot images of The Rembo Wizard. Following chapters explain the concept in details.

System Snapshot is a full image

Logically, a System Snapshot is a full image of the system disk. Technically, it is a differential image compared to the system’s base image. This is due to the Rembo Toolkit’s shared files concept where it actually stores only the files that are different or new on the disk when compared to the base image. Below picture illustrates the operation principle.

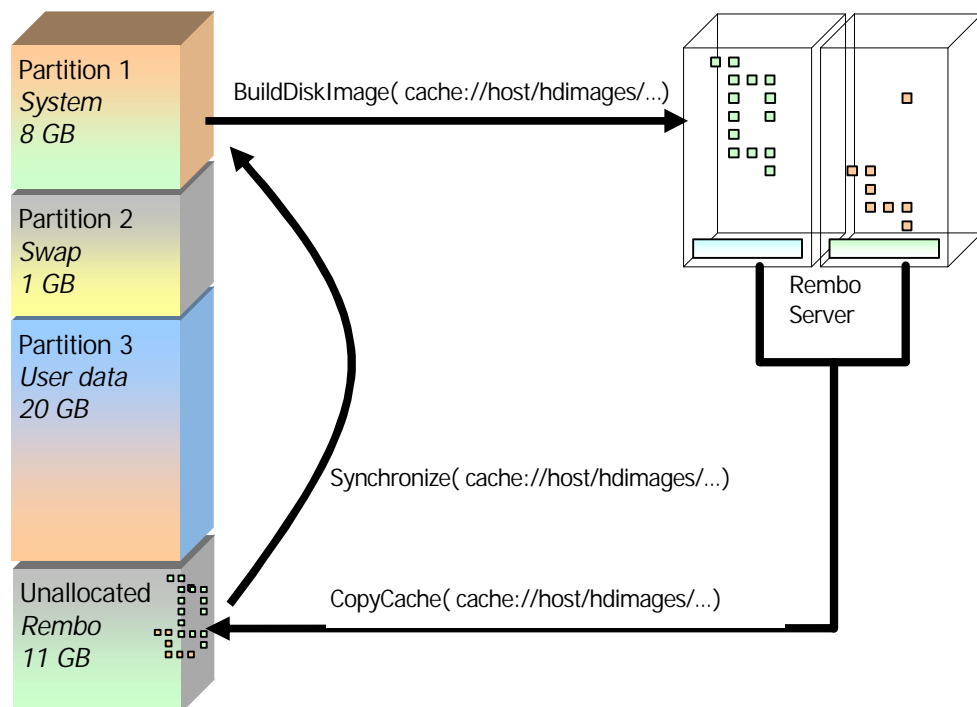




The above picture illustrates an example where a software package has been added on the system after the base image has been taken. This will naturally add some files on the system partition but also it will modify some existing files, such as registry files or such. When a System Snapshot image is taken with The Rembo Wizard, it will create an archive file on the server. The archive file does not contain the actual files that have been stored on the server but pointers to the files stored in the shared file structure. More precisely, the archive file is pointing to a list of MD5-checksums of the files that the archive is containing. The MD5 checksum does not only contribute to the reliability of the archived files but it also allows distinguishing files that have been changed on the installation.

Using the local hard disk cache

The operations with the shared file system are complicated and they can be time consuming if all the shared files are located server while they are needed on the client computer. The Rembo Toolkit allows a free disk space at the end of the disk to be used as a cache for the shared file system operations. Allowing The Rembo Toolkit to use a cache partition will **significantly** improve the restoration performance and the performance of The Rembo Wizard's System Snapshots. Only the base image creation will be slower, about the double of the time without the local cache partition. This is because all the shared files have to be uploaded and built on the Rembo Toolkit server prior to downloading them back on the cache partition.



In the above picture we have a 40 GB hard disk with a Linux installation on it. The entire system is installed on the first partition of 8 GB. The installation is typically around 3 GB but we expect it to grow up to 5 GB with the user's future software installations. We will leave further 3 GB for the /tmp, for log files and other live, fast growing files. Other Linux partitions follow. We do not allocate the entire disk but we leave an 11 GB (big) chunk at the end of the disk unallocated. The Rembo Toolkit will automatically use the unallocated space for the network operations when `cache://` URIs are used for file operations.

When The Rembo Wizard creates a base image for the system it first cleans the cache partition and then take the base image. The Rembo Wizard will ask Rembo to download the shared files of the archive



back to the cache partition of the local hard disk. This operation is time consuming but it will increase all the successive operation speeds considerably. Let's take few examples:

1. Supposing that no further images are taken. Later the system has to be restored to the state of the base image. This will take only a few minutes, since all the necessary files are already in the local hard disk's cache partition.
2. The system administrator hands over the system to the end user and allows her to take her own backup images, called System Snapshots. The end user has added, say 500 files in the system and modified 100 existing. Now only 600 files has to be stored on the Rembo server and copied back to the local hard disk's cache partition. This is considerable less than the about 135,000 files of the full Linux system that we are currently installing. Therefore both further System Snapshots and user level restorations will benefit from the local hard disk cache partition.

In some cases the automatic reading back of the shared files can be undesirable because of the extra time it takes. You can turn off this feature in the configuration dialog of the Administrator's Menu. In the *Backup strategy* section, uncheck the *Read Back Shared* checkbox.

NOTE: If you turn of the automatic reading back of the shared files they will be read back anyway to the local cache during the next restoration or backup procedure. If this will happen to the end user she may find the required time long, get confused and reboot the system before the operation ends. This would leave the local cache in a sorry state and you would have to clean it manually. Therefore it is advisable to keep the cache consistent on the machines where Rembo can be visible to the end users, such as on the public access workstations and such.

If the hard disk fails, The Rembo Wizard will switch automatically to use the copy on the Rembo server while the system is restored. It will also copy the files used in the restored image back to the new disk's cache partition.

The Rembo Wizard supports the file operations with the hard disk cache by default. To turn this feature off, you have to go to the configuration dialog. In the *Disks and partitions* section, select *Use cache ?* to *No* if you do not want to use the local hard disk's cache partition. There are two drawbacks with this, however:

1. The Rembo Wizard will switch to use *net://* URIs. This will switch the underlying Rembo OS to use NetFS protocol, which is much less effective for file transfers than the UCAST-protocol.
2. All the requests for shared file system operations should be executed on the server instead of the client computer. This will create more network traffic and further slow down the operations.

Distributing and Backing Up Shared Files Archives

Rembo Toolkit 2.0 and greater do have the advantage that they are using the local file system of the server. You can jump into the server's file system and actually delete, add and modify files. But for all practical purposes, shared archives are far too complicated to be manipulated manually with their sometimes over 100,000 of files. Here we do not talk about backing up the entire contents of the Rembo Toolkit server's file system, which is a quite straight forward task, since there is no need even to stop the server. We will discuss how to individually distribute and backup specific image archives, such as the Base Image and The System Snapshot.

The recommended distribution format for the Rembo shared files archives is RAD-format files (probably from Rembo's other product, Rembo Auto-Deploy). You can consider it as a tar-file (or zip-



file) that contains all the shared files needed by an archive, together with their path descriptions and MD5 checksums.

Building and distribution of the RAD-format files is done with a Rembo Toolkit utility program called *netclnt*.

- ✍ On Windows, *netclnt*-utility can be used through the right click menu of the Rembo Server Management Console (*sconsole.exe*).

Following is an example of the usage of the *radget*-operation in the *netclnt*-utility.

```
freak_petri$ /usr/local/rembo/misc/netclnt
Netclnt 2.0 (c) Rembo Technology Sarl
NETFS> open localhost
Password: ****
NETFS> cd hosts
Current directory is /hosts
NETFS> ls
0002b31a5f16          84440/75          <DIR>    12/10/02 08:05:45
0002b3267f13          28946/34          <DIR>    11/14/02 13:01:32
NETFS> cd 0002b31a5f16/hdimages
Current directory is /hosts/0002b31a5f16/hdimages
NETFS> radget TS1W2K.bas.rad TS1W2K.bas
TS1W2K.bas: 1392543 bytes read in 0.013 secs (104608.10 kb/s)
22 % completed,
```

The resulting file, *TS1W2K.bas.rad* in our example, can be

- ✍ stored for backup purposes
- ✍ distributed on other Rembo Toolkit servers using the *radput*-operation of the *netclnt*-utility.

If you do not want to use RAD-format for some reason, *netclnt*-utility provides two other operation types that can be used to distribute the shared file system and other files related to an archive:

1. *getshared* and *putshared*-operations should be used within a data directory that is initially empty. *getshared* extracts the directory structure of the shared files related to the archive and puts them into the data directory. The data directory can be tar'ed or zipped, moved to an other machine and the uploaded on the server with the *putshared*-operation.
2. *sync*, *msync* and *rsync*-operations allows the synchronization of the two Rembo Toolkit servers with a point-to-point connection. If the intention is to build identical servers for a control system, or redundant servers for a cluster of computers, for example, synchronization of the servers is the most efficient way to accomplish the task.

Protecting administrator's System Snapshots

The Rembo Wizard allows the system administrator to have additionally to the base image a protected System Snapshot image or images and still allowing the user to take her own System Snapshots. This is best explained with an example.

Let's suppose that you have to install an ordinary Linux workstation for a gang of mad rocket scientists that would start to hack the system as if there would be no tomorrow. You do not mind, since you have the corporate level base image of the system. In case of trouble you can restore the system to a known state using the base image.



One day your boss asks you to go and install a new graphics card on the system, urgently, of course! When you go to see the machine, you will be told that the graphics card installation is “temporary anyway” and that they (the rocket scientists) will install some software that would use the new graphics card. They will also inform you in their kind manner that you should not touch any of the currently installed software or else...

There you go. You cannot use the base image to get the system to a known state before installing the graphics card and its drivers. Also you know that in few weeks time you would be called to install the system to the state it was “before that bl*% new graphics card was installed”. Ask yourself the following questions:

1. How to keep the base image?
2. How to allow the user to take System Snapshots?
3. How to return the system to the state it was before the graphics card was installed?
4. How to return the system to the state it was after the graphics card was installed?

The answer is to use system administrator’s, protected System Snapshot images.

In our example, the mad rocket scientists are allowed to take three System Snapshot images and they have actually done that for the all three possible images allocated for them. Go to the configuration dialog and there into the *Backup strategy*-section. There you will see the selection *After the baseimage, take max. 3 Snapshot images*. Increase the maximum number of snapshot images to **5** and click *OK*. Take now a System Snapshot image *before* you install the new graphics card and an other one once you have installed successfully the new graphics card. The System Snapshot images will be numbered as **4** and **5**, respectively. Now go back to the configuration dialog and set back the maximum number of snapshot images to **3**.

Your System Snapshot images are now protected.

Of course, it may occur in our example that user has not taken any System Snapshot images at all (you may say that it is even more likely). Your System Snapshot images will be numbered as **1** and **2**, respectively. On the client, you can use FileMan(“cache://host/hdimages”); utility to rename your System Snapshot images to **4** and **5**, respectively. Or back in your office, use the Rembo Toolkit’s *netclnt*-utility, or on Windows, use the Rembo Server Management Console (*sconsole.exe*).

✍