



Extraction of File Attachments and Images from PDF Files

Extraction via Command Line Tools

Document Version 1.0

iQ.Suite for IBM Domino

Inhalt

1	Introduction	3
2	Description of 'ExtractAttachFromPDF.zip'	3
2.1	Sample Script	3
2.2	Sample Configurations for the Decompressor	3
3	Preparations	5
4	Configurations in iQ.Suite.....	6
4.1	iQ.Suite Job	6
4.2	Decompressor	7
5	About GBS	11

1 Introduction

To use iQ.Suite for analyzing PDF files (file attachments and/or embedded images), these components must first be extracted from the PDF file. In addition to a job (Watchdog Mail Job or Wall Mail Job Advanced), a decompressor is needed which calls Xpdf tools either directly or by using a script.

In this documentation, an example with a Watchdog job is used to describe the proceeding.

2 Description of 'ExtractAttachFromPDF.zip'

Extract the contents of the **ExtractAttachFromPDF.zip** file to any directory. This ZIP file contains a sample script (one for Windows and one for Unix) and sample configurations for the decompressor.

2.1 Sample Script

- Sample script for Windows: **pdfunpack.cmd**

Save this script in any directory.

In our example, we create a directory under `<Domino directory>\iqsuite`, e.g. `pdfunpack`. Then, we copy the script to the newly created subdirectory:

```
<Domino_directory>\iqsuite\pdfunpack\pdfunpack.cmd
```

- Sample script for Unix: **pdfunpack.sh**

Save this script in any directory.

In our example, the script is saved as follows:

```
/opt/iqsuite/pdfunpack.sh
```

2.2 Sample Configurations for the Decompressor

- **pdfunpack.gxl.zip** file containing the Decompressor sample configuration for Windows and Unix (AIX, Linux and Solaris). The *pdfunpack.gxl.zip* contains the *g_wdog.gxl* file.

The following sample configurations are available:

- **WIN PDF decompressor attachments only (32-Bit and 64-Bit)**

This is used to call the *pdfdetach.exe* tool directly (without script).

All attachments are unpacked from the PDF at once. The quarantine report contains the

return code and the return text from *pdfdetach.exe*; under Linux, only the return code is returned.

□ **Optional: WIN PDF decompressor images only (32-bit and 64-bit)**

This is used to call the *pdfimages.exe* tool directly (without script).

All images embedded in the PDF are unpacked from the PDF at once. The quarantine report contains the return code and the return text from *pdfimages.exe*; under Linux, only the return code is returned.

□ **WIN PDF Unpack (32-bit and 64-bit)**

The script calls the tools *pdfdetach.exe* and *pdfinfo.exe*. *pdfimages.exe* can be called optionally.

The attachments are unpacked from the PDF through the script one by one.

The quarantine report contains the return code, the return text and additional information about the PDF. This information is determined by the *pdfinfo.exe* Xpdf tool.

Optional: The sample script contains a prepared call of *pdfimages.exe* in order to unpack the images embedded in the PDFs. The call is commented out. When you want *pdfimages.exe* to be executed, remove the comment characters, i.e. under Windows the `rem` characters in the CMD file, under Unix in the SH file remove the `#` from the section `optional:`
`extract the images (ppm-format).`

3 Preparations

Before configuring the decompressor in iQ.Suite, follow these steps:

1. Download the current version of Xpdf:

- Under **Windows, Linux, Solaris:**
www.foolabs.com/xpdf/download.html

Example: *xpdfbin-win-3.04.zip* (for Windows 32-bit or 64-bit)

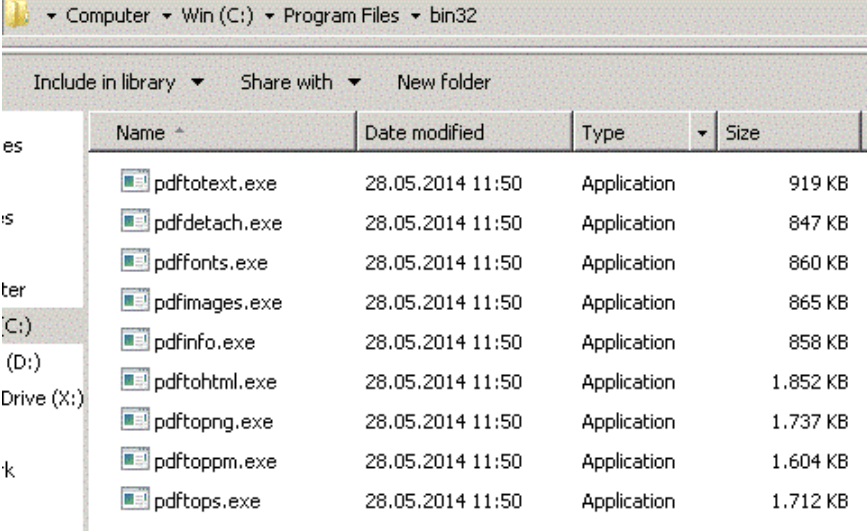
- Under **AIX:**
<http://www.perz1.org/aix/index.php?n=Main.Xpdf>

The dependencies mentioned there must be met.

On 64-bit systems, the 64-bit version of Xpdf can also be used.

2. From the downloaded file, unpack the `bin32` or `Bin64` directory, for Windows to `C:\Program Files\xpdf\bin32\` for example.

Hereinafter, this directory is called **appdir** (application directory).



Name	Date modified	Type	Size
pdftotext.exe	28.05.2014 11:50	Application	919 KB
pdfdetach.exe	28.05.2014 11:50	Application	847 KB
pdffonts.exe	28.05.2014 11:50	Application	860 KB
pdfimages.exe	28.05.2014 11:50	Application	865 KB
pdfinfo.exe	28.05.2014 11:50	Application	858 KB
pdftohtml.exe	28.05.2014 11:50	Application	1.852 KB
pdftopng.exe	28.05.2014 11:50	Application	1.737 KB
pdftoppm.exe	28.05.2014 11:50	Application	1.604 KB
pdftops.exe	28.05.2014 11:50	Application	1.712 KB

3. Save the sample script in any directory. Example:

- Under Windows: `<Domino_directory>\iqsuite\pdfunpack\pdfunpack.cmd`

- Under Unix: `/opt/iqsuite/pdfunpack.sh`

4. Import the sample configuration into your iQ.Suite. For further information on the sample configurations, please refer to 2.2.

4 Configurations in iQ.Suite

In this chapter, only those settings are described that refer to the use case documented here.

For further information on the configuration, please refer to the iQ.Suite Administration Manual (download on www.gbs.com) or to the HELP.

4.1 iQ.Suite Job

You require a **Watchdog Mail Job** or a **Wall Mail Job Advanced**.

In our example, we will configure a Watchdog Mail Job that is used to check the extracted components for viruses:

1. In the iQ.Suite administration console under WATCHDOG -> MAIL JOBS, copy the **DEFAULT – Virus Check all Mails** job or create a new job.
2. In the **Basics** tab, enter a job name, e.g. **Virus Check for PDF**, and activate the job (status ‚Active‘).
3. In the **Operations -> Password Protection** tab, define whether password-protected attachments are allowed.

This relates to attachments that are reported by the decompressor as being password-protected. Notes-encrypted attachments are not affected by this setting.

'Virus Check for PDF'
Watchdog Mail Job





Basics | **Operations** | Advanced | Misc. | Comments

Operations

General modify option Delete mail ▾

No Alert | Virus | Denied Attachments | Encryption | **Password Protection** | System Errors

Password Protection

Allow password protected attachments	<input checked="" type="radio"/> Yes <input type="radio"/> No
Category in Quarantine report	Watchdog Encrypted ▾
Notify recipient about unchecked password protected attachments	<input checked="" type="radio"/> Yes <input type="radio"/> No
Notification template recipient (password protection)	Recipient: Password encrypted attachment <div style="display: flex; justify-content: space-around; font-size: small;">  Select  Deselect  Edit  New </div>

- 'Yes': Password-protected attachments are allowed and will be delivered without being checked.
- 'No': Password-protected attachments are prohibited and will be handled according to the settings in this tab.

For configuring the return code for password-protected files, please refer to the **Settings** tab of the used decompressor.

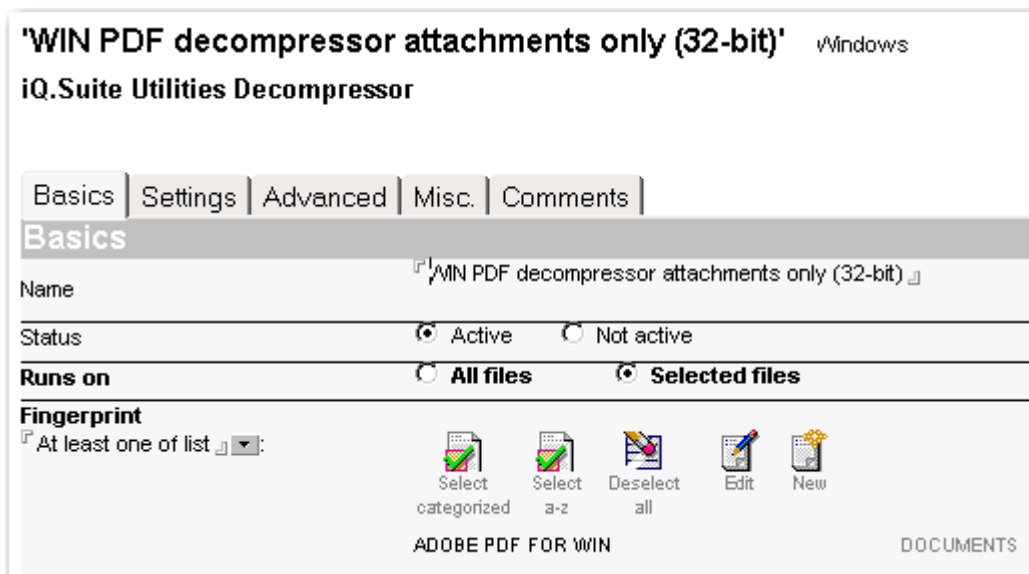
4.2 Decompressor

The sample script (batch file) is used as the decompressor. The batch file calls **Xpdf** tools that can be used to extract text and images. Xpdf is licensed under the GNU General Public License (GPL).

To adjust the sample configuration of a decompressor to your needs, proceed as follows:

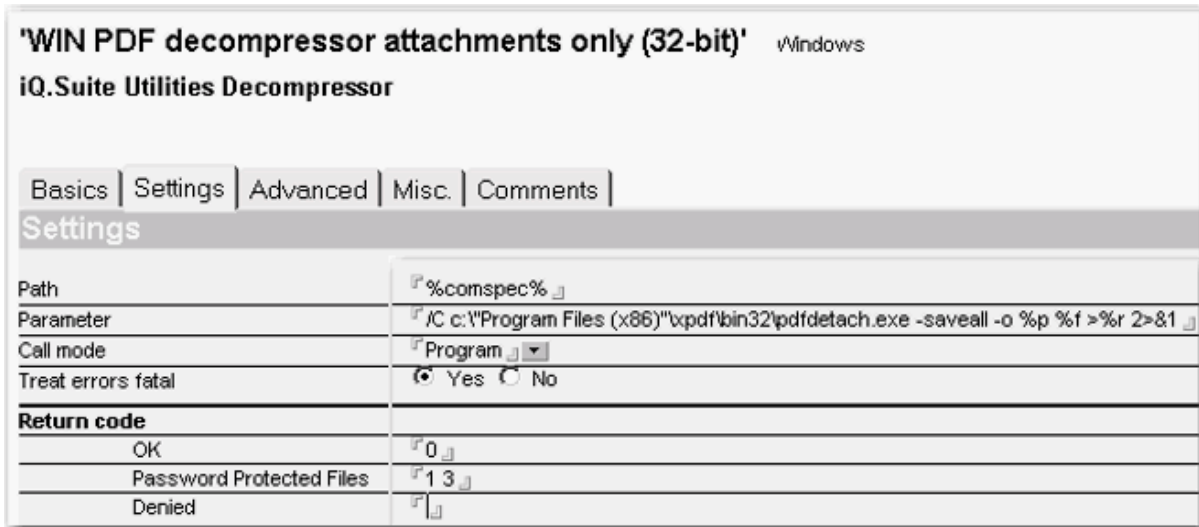
1. Open the desired decompressor document under IQ.SUITE -> WATCHDOG or WALL -> UTILITIES -> DECOMPRESSOR.

Example with the decompressor under 32-bit Windows for 'attachments only':



- a) If needed, change the **Name** of the decompressor.
- b) The **Fingerprint** ADOBE PDF FOR WIN must be selected.
- c) Activate the document.

2. Open the **Settings** tab:



a) Specify the **path** of the program or tool to be used for unpacking.

Under Windows:

Use the `%comspec%` Windows environment variable to specify the path to the command-line interpreter.

Example under 32-bit Windows: `C:\Windows\system32\cmd.exe`

Under Unix:

Specify the path to `pdfdetach` (for 'attachments only') resp. to `pdfimages` (for 'images only'). If the script (Decompressor 'PDF Unpack') is used, specify the path to the batch file.

Preset paths by decompressor and operating system:

	Solaris	Linux	AIX
Decompressor attachments only	<code>/usr/sfw/bin/pdfdetach</code>	<code>/usr/local/bin/pdfdetach</code>	<code>/bin/pdfdetach</code>
Decompressor images only	<code>/usr/sfw/bin/pdfimages</code>	<code>/usr/local/bin/pdfimages</code>	<code>/bin/pdfimages</code>
Unpack (with script)	<code>/opt/iqsuite/pdfunpack.sh</code>	<code>/opt/iqsuite/pdfunpack.sh</code>	<code>/opt/iqsuite/pdfunpack.sh</code>

b) Under **Parameter**, enter the command-line parameters that are to be passed to the program specified under **Path**:

For Windows: Preset parameters, sorted by decompressor:

	Windows
Decompressor attachments only	<code>/C c:"Program Files (x86)"\xpdf\bin32\pdfdetach.exe -saveall -o %p %f >%r 2>&1</code> Shown in bold: path to the file to be executed.
Decompressor images only	<code>/C c:"Program Files (x86)"\xpdf\bin32\pdfimages.exe %f %p\img >%r 2>&1</code> Shown in bold: path to the file to be executed.
PDF Unpack	<code>/C c:"Program Files (x86)"\IBM\Lotus\Domino\iQSuite\pdfunpack\pdfunpack.cmd %f %p c:"Program Files (x86)"\xpdf\bin32 >>%r 2>&1</code> Shown in bold: at first, the path of the batch file (script) and then the path to the directory that contains the Xpdf tools to be executed.

For Unix: Preset parameters, sorted by decompressor and operating system:

	Solaris	Linux	AIX
Decompressor attachments only	<code>-saveall -o %p %f</code>	<code>-saveall -o %p %f</code>	<code>-saveall -o %p %f</code>
Decompressor images only	<code>%f %p/img</code>	<code>%f %p/img</code>	<code>%f %p/img</code>
Unpack	<code>%f %p /usr/sfw/bin %r</code>	<code>%f %p /usr/local/bin %r</code>	<code>%f %p /bin %r</code>

Description of the variables:

%f	Filename of the file to be unpacked (here: PDF file).
%p	Path to the working directory that contains the files to be unpacked (temp directory of iQ.Suite).
%r	Name of the report file that contains the output of the call.

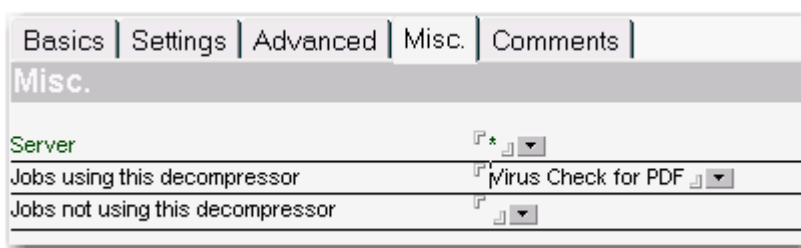
- c) Select the **Call mode** 'Program'.

- d) Under **Treat errors fatal**, select 'Yes', if you want all not mentioned return values to be processed as errors in the job. Otherwise, select 'No'.
- e) Define the **Return codes**.

You can adjust the preset return codes to your needs.

Return codes	Meaning
0	The PDF could be processed successfully.
1	The decompressor could not open the PDF. Usually, the PDF was password protected, but Macintosh-generated PDFs can also not be read by the decompressor.
2	The decompressor could not write the output, e.g. because not enough disc space was available.
3	The decompressor cannot access the PDF file or its contents. Possible reasons: Authorizations are missing, either on the file system (access rights to the operating system) or in the PDF itself.
99	All unknown errors, e.g. when the parameters %p, %f and %r were inverted.
All other return codes	Configuration errors in iQ.Suite, e.g. single quotation marks were misplaced in the Parameter field.

- 3. In the **Misc** tab under **Jobs using this decompressor**, select the Watchdog job that shall use this decompressor.



- 4. Save the document.

5 About GBS

GROUP Business Software is a leading vendor of solutions and services in the fields of messaging security and workflow for the IBM and Microsoft collaboration platforms. Over 5,000 customers and more than 4 million users worldwide trust in GBS expertise. The company operates in Europe, North America and Asia.

© 2016 GROUP Business Software GmbH

Our product descriptions are of a general and descriptive nature only. They do not stipulate any specific features nor do they represent any form of warranty or guarantee. We reserve the right to change the specifications and design of our products without notice at any time, in particular in order to keep abreast of technical developments.

The information contained in this document presents the topics from the viewpoint of GROUP Business Software GmbH (hereafter 'GBS') at the time of publishing. Since GBS needs to be able to react to changing market requirements, this is not an obligation for GBS and GBS cannot guarantee that the information presented in it is accurate after the publication date.

This document is intended for information purposes only. GBS does not extend warranty for this document, in either explicit or implied form. This also applies to quality, execution, standard commercial practice or suitability for a particular purpose.

All the product and company names that appear in this document may be trademarks of their respective owners.

Web site: www.gbs.com
Email address: info@de.gbs.com
Locations: www.gbs.com/en/locations

