



Extraktion von Dateianhängen und Bildern aus PDF-Dateien

Extraktion mittels Kommandozeilentools

Dokumentversion 1.1

iQ.Suite für IBM Domino

Inhalt

1	Einführung	3
2	Beschreibung der ‚ExtractAttachFromPDF.zip‘	3
2.1	Beispielskript	3
2.2	Beispielkonfigurationen für den Entpacker	3
3	Vorbereitungen.....	5
4	Konfigurationen in der iQ.Suite.....	6
4.1	iQ.Suite-Job.....	6
4.2	Entpacker	7
5	Über GBS	11

1 Einführung

Um Komponenten von PDF-Dateien (Dateianhänge und/oder eingebettete Bilder) mit der iQ.Suite untersuchen zu können, müssen diese Komponenten zunächst aus der PDF-Datei extrahiert werden. Neben einem Job (Watchdog Mail Job oder Wall Mail Job Advanced) wird hierzu ein Entpacker benötigt, der Xpdf-Tools entweder direkt oder über ein Skript aufruft.

In der vorliegenden Dokumentation beschreiben wir die Vorgehensweise anhand eines Beispiels mit einem Watchdog-Job.

2 Beschreibung der ‚ExtractAttachFromPDF.zip‘

Extrahieren Sie den Inhalt der Datei **ExtractAttachFromPDF.zip** in ein beliebiges Verzeichnis. Diese ZIP-Datei enthält ein Beispielskript (jeweils für Windows und Unix) und Beispielkonfigurationen für den Entpacker.

2.1 Beispielskript

- Beispielskript für Windows: **pdfunpack.cmd**

Legen Sie dieses Skript in ein beliebiges Verzeichnis ab.

In unserem Beispiel legen wir ein Verzeichnis unter `<Domino-Verzeichnis>\iqsuite an`, z.B. `pdfunpack`. Danach kopieren wir das Skript in das neu angelegte Unterverzeichnis:

```
<Domino-Verzeichnis>\iqsuite\pdfunpack\pdfunpack.cmd
```

- Beispielskript für Unix: **pdfunpack.sh**

Legen Sie dieses Skript in ein beliebiges Verzeichnis ab.

In unserem Beispiel wird dieses Skript wie folgt abgelegt:

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

2.2 Beispielkonfigurationen für den Entpacker

- Datei **pdfunpack.gxl.zip**, die Entpacker-Beispielkonfigurationen für Windows und Unix (AIX, Linux und Solaris) enthält. Die `pdfunpack.gxl.zip` enthält die Datei `g_wdog.gxl`.

Folgende Beispielkonfigurationen sind vorhanden:

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

Damit wird das Tool `pdfdetach.exe` direkt (ohne Skript) aufgerufen.

Alle Anhänge werden auf auf einen Schlag aus dem PDF entpackt. Der Quarantäne-Report enthält neben dem Rückgabecode auch den Rückgabertext von *pdfdetach.exe*; unter Unix wird lediglich der Rückgabecode ausgegeben.

□ **Optional: WIN PDF decompressor images only (32-Bit und 64-Bit)**

Damit wird das Tool *pdfimages.exe* direkt (ohne Skript) aufgerufen.

Alle im PDF eingebetteten Bilder werden auf einen Schlag aus dem PDF entpackt. Der Quarantäne-Report enthält neben dem Rückgabecode auch den Rückgabertext von *pdfimages.exe*; unter Unix wird lediglich der Rückgabecode ausgegeben.

□ **WIN PDF Unpack (32-Bit und 64-Bit)**

Das Skript ruft die Tools *pdfdetach.exe* und *pdfinfo.exe* auf. *pdfimages.exe* kann optional aufgerufen werden.

Die Anhänge werden über das Skript einzeln aus dem PDF entpackt.

Der Quarantäne-Report enthält neben Rückgabecode und Rückgabertext weitere Informationen über das PDF, die über das Xpdf-Tool *pdfinfo.exe* ermittelt werden.

Optional: Das Beispielskript enthält einen vorbereiteten Aufruf von *pdfimages.exe*, um die in den PDFs eingebetteten Bilder zu entpacken. Der Aufruf ist auskommentiert. Wenn Sie *pdfimages.exe* ausführen lassen möchten, entfernen Sie die Kommentarzeichen, d.h. unter Windows in der CMD-Datei die `rem`-Zeichen bzw. unter Unix in der SH-Datei die `#` im Abschnitt `optional: extract the images (ppm-format)`.

3 Vorbereitungen

Vor der Konfiguration des Entpackers in der iQ.Suite führen Sie folgende Schritte aus:

1. Laden Sie die aktuelle Version von Xpdf herunter:

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

Beispiel: *xpdfbin-win-3.04.zip* (für Windows 32-Bit oder 64-Bit)

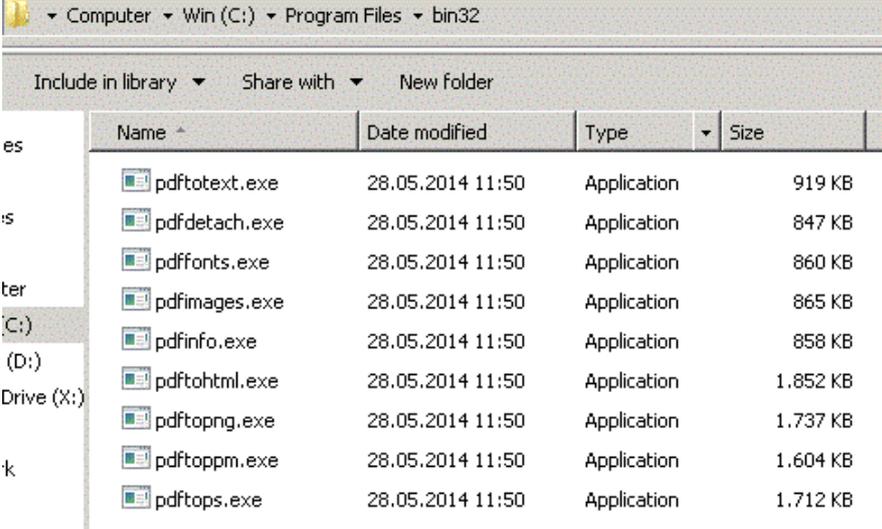
- Unter **AIX:**
<http://www.perz1.org/aix/index.php?n=Main.Xpdf>

Die dort angegebenen Abhängigkeiten müssen erfüllt werden.

Auf 64-Bit-Systemen kann auch die 64-Bit-Version von Xpdf verwendet werden.

2. Entpacken Sie aus der heruntergeladenen Datei das Verzeichnis `bin32` oder `Bin64`, für Windows z.B. nach `C:\Program Files\xpdf\bin32\`.

Dieses Verzeichnis wird nachfolgend **appdir** (Applikationsverzeichnis) genannt.



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. Legen Sie das Beispieldokument in ein beliebiges Verzeichnis ab. Beispiel:

- Unter Windows: `<Domino-Verzeichnis>\iqsuite\pdfunpack\pdfunpack.cmd`
- Unter Unix: `/opt/iqsuite/pdfunpack.sh`

4. Importieren Sie die Beispielkonfigurationen in Ihre iQ.Suite. Weiterführende Informationen zu den Beispielkonfigurationen finden Sie unter 2.2.

4 Konfigurationen in der iQ.Suite

In diesem Kapitel werden nur die Einstellungen beschrieben, die speziell für den hier dokumentierten Anwendungsfall gelten.

Weiterführende Informationen zur Konfiguration finden Sie im Administrationshandbuch der iQ.Suite (Download unter www.gbs.com) oder in der HILFE.

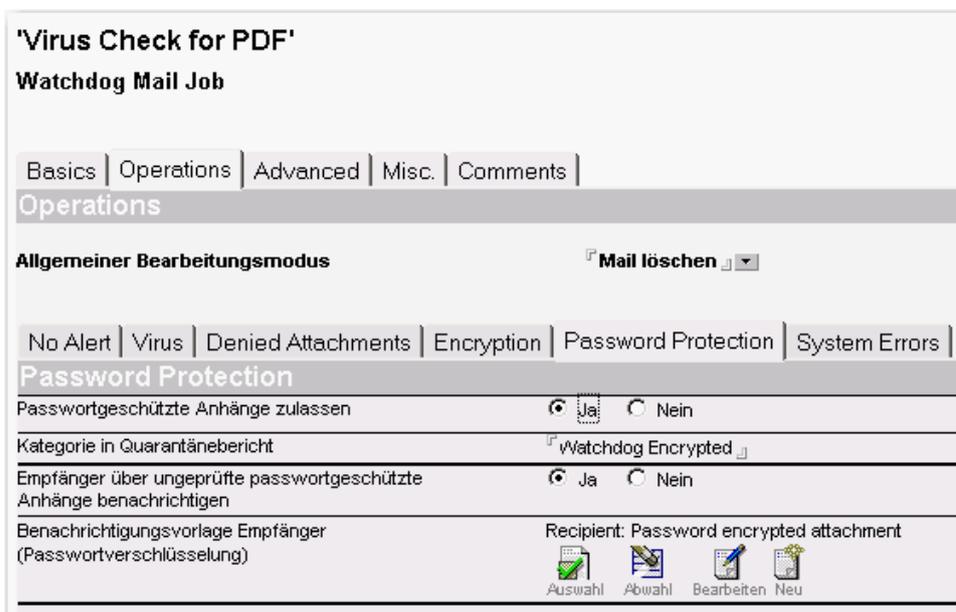
4.1 iQ.Suite-Job

Sie benötigen einen **Watchdog Mail Job** oder einen **Wall Mail Job Advanced**.

In unserem Beispiel konfigurieren wir einen Watchdog Mail Job, mit dem die entpackten Komponenten auf Viren geprüft werden sollen:

1. Kopieren Sie in der iQ.Suite-Administrationskonsole bei WATCHDOG -> MAIL-JOBS den Job **DEFAULT – Virus Check all Mails** oder legen Sie einen neuen Job an.
2. In der Registerkarte **Basics** geben Sie einen Jobnamen ein, z.B. **Virus Check for PDF**, und aktivieren Sie den Job (Status ‚Aktiv‘).
3. In der Registerkarte **Operations** -> **Password Protection** definieren Sie, ob passwortgeschützte Anhänge zugelassen werden sollen.

Hierbei geht es um Dateianhänge, die vom Entpacker als passwortgeschützt gemeldet werden. Notes-verschlüsselte Anhänge sind von dieser Einstellung nicht betroffen.



'Virus Check for PDF'
Watchdog Mail Job

Basics | Operations | Advanced | Misc. | Comments |

Operations

Allgemeiner Bearbeitungsmodus Mail löschen

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

Password Protection

Passwortgeschützte Anhänge zulassen Ja Nein

Kategorie in Quarantänebericht Watchdog Encrypted

Empfänger über ungeprüfte passwortgeschützte Anhänge benachrichtigen Ja Nein

Benachrichtigungsvorlage Empfänger (Passwortverschlüsselung) Recipient: Password encrypted attachment

- ‚Ja‘: Passwortgeschützte Anhänge sind erlaubt und werden ungeprüft zugestellt.
- ‚Nein‘: Passwortgeschützte Anhänge sind verboten und werden gemäß den Einstellungen in dieser Registerkarte behandelt.

Zur Konfiguration des Returncodes für passwortgeschützte Dateien: siehe Registerkarte **Settings** des verwendeten Entpackers.

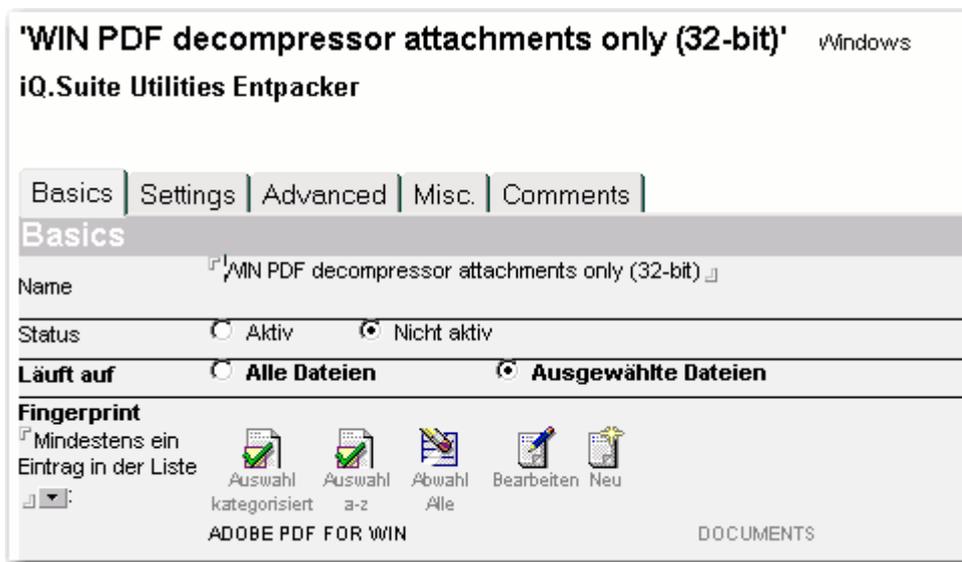
4.2 Entpacker

Als Entpacker wird das Beispielskript (Batchdatei) verwendet. Die Batchdatei ruft **Xpdf-Tools** auf, mit dem Text und Grafiken extrahiert werden können. Xpdf ist unter der GNU General Public License (GPL) lizenziert.

Um die Beispielkonfiguration eines Entpackers Ihren Bedürfnissen anzupassen, gehen Sie wie folgt vor:

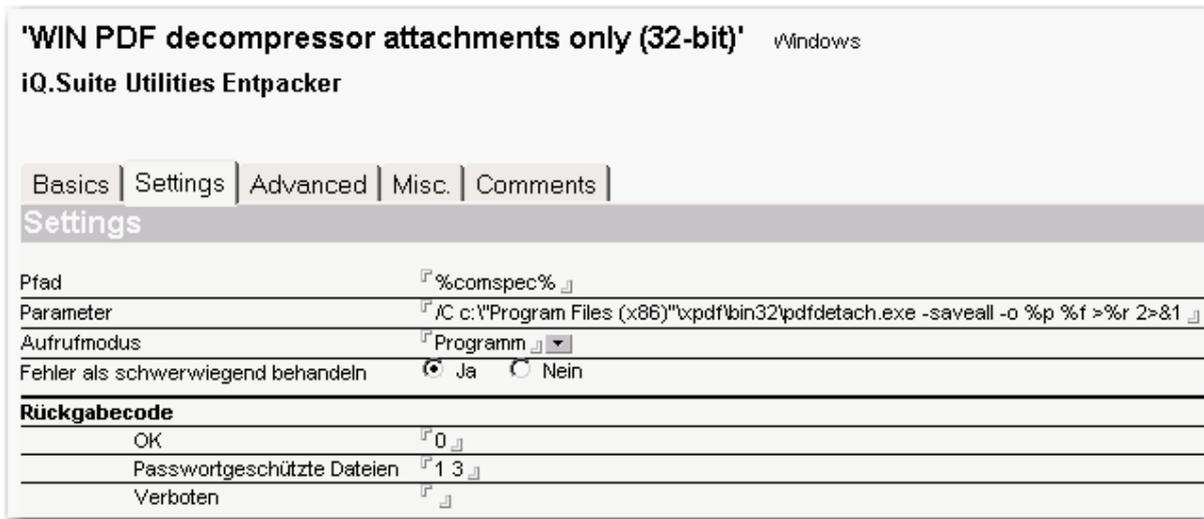
1. Öffnen Sie das gewünschte Entpacker-Dokument bei IQ.SUITE -> WATCHDOG oder WALL -> UTILITIES -> ENTPACKER.

Beispiel mit dem Entpacker unter 32-Bit-Windows für ‚attachments only‘ (nur Anhänge):



- a) Ändern Sie bei Bedarf den **Namen** des Entpackers.
- b) Der **Fingerprint** ADOBE PDF FOR WIN muss ausgewählt sein.
- c) Aktivieren Sie das Dokument.

2. Öffnen Sie die Registerkarte **Settings**:



- a) Geben Sie den **Pfad** zum Programm oder Tool an, das zum Entpacken ausgeführt werden soll:

Unter Windows:

Mit der Windows-Umgebungsvariable `%comspec%` geben Sie den Pfad zum Kommandozeilen-Interpreter an.

Beispiel unter 32-Bit-Windows: `C:\Windows\system32\cmd.exe`

Unter Unix:

Geben Sie den Pfad zu `pdfdetach` (für ‚attachments only‘) bzw. zu `pdfimages` (für ‚images only‘) an. Bei Verwendung des Skriptes (Entpacker ‚PDF Unpack‘) geben Sie den Pfad zur Batchdatei an.

Voreingestellte Pfade nach Entpacker und Betriebssystem:

	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 (mit Skript)	<code>/opt/iqsuite/pdfunpack.sh</code>	<code>/opt/iqsuite/pdfunpack.sh</code>	<code>/opt/iqsuite/pdfunpack.sh</code>

- b) Geben Sie bei **Parameter** die Kommandozeilenparameter ein, die an das bei **Pfad** angegebene Programm übergeben werden sollen:

Für Windows: Voreingestellte Parameter nach Entpacker:

	Windows
Decompressor attachments only	<code>/C c:"Program Files (x86)"\xpdfbin32\pdfdetach.exe -saveall -o %p %f >%r 2>&1</code> Fett-markiert ist der Pfad zu der auszuführenden Datei.
Decompressor images only	<code>/C c:"Program Files (x86)"\xpdfbin32\pdfimages.exe %f %p\img >%r 2>&1</code> Fett-markiert ist der Pfad zu der auszuführenden Datei.
PDF Unpack	<code>/C c:"Program Files (x86)"\IBM\Lotus\DominoliQSuite\pdfunpack\pdfunpack.cmd %f %p c:"Program Files (x86)"\xpdfbin32 >>%r 2>&1</code> Fett-markiert ist zunächst der Pfad der Batchdatei (Skript) und dann der Pfad zu dem Verzeichnis, der die auszuführenden Xpdf-Tools enthält.

Für Unix: Voreingestellte Parameter nach Entpacker und Betriebssystem:

	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>

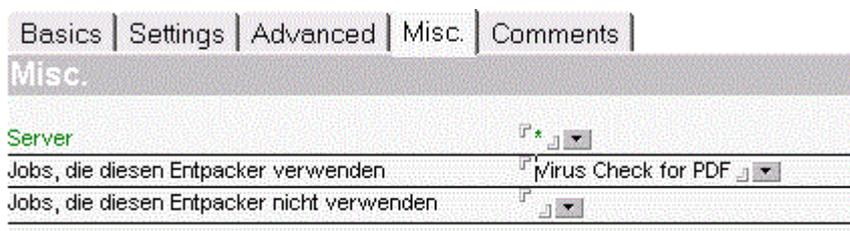
Beschreibung der Variablen:

<code>%f</code>	Dateiname der zu entpackenden Datei (hier: PDF-Datei).
<code>%p</code>	Pfad zum Arbeitsverzeichnis, in dem sich die zu entpackenden Dateien befinden (temp-Verzeichnis der iQ.Suite).
<code>%r</code>	Name der Reportdatei, die die Ausgabe des Aufrufs enthält.

- c) Wählen Sie den **Aufrufmodus** ‚Programm‘.
- d) Bei **Fehler als Schwerwiegend behandeln** wählen Sie ‚Ja‘, wenn alle nicht genannten Rückgabewerte als Fehler im Job bearbeitet werden sollen. Anderenfalls wählen Sie ‚Nein‘.
- e) Definieren Sie die **Rückgabecodes**:
Die voreingestellten Rückgabecodes können Sie Ihren Bedürfnissen anpassen.

Rückgabecodes	Bedeutung
0	Das PDF konnte erfolgreich bearbeitet werden.
1	Der Entpacker konnte das PDF nicht öffnen. Im Regelfall war das PDF passwortgeschützt, aber auch Macintosh-generierte PDFs können vom Entpacker nicht gelesen werden.
2	Der Entpacker konnte die Ausgabe nicht schreiben, z.B. weil zu wenig Speicherplatz verfügbar war.
3	Der Entpacker kann auf die PDF-Datei oder deren Inhalt nicht zugreifen. Mögliche Gründe: Berechtigungen fehlen, entweder auf dem Dateisystem (Betriebssystem-Zugriffrechte) oder im PDF selbst.
99	Alle unbekanntes Fehler, z.B. wenn die Parameter %p, %f und %r vertauscht wurden.
Alle anderen Returncodes	Konfigurationsfehler in der iQ.Suite, z.B. Hochkommata wurden im Feld Parameter falsch gesetzt.

- 3. In der Registerkarte **Misc** wählen Sie bei **Jobs, die diesen Entpacker verwenden** den Watchdog-Job aus, der diesen Entpacker verwenden soll:



- 4. Speichern Sie das Dokument.

5 Über GBS

GROUP Business Software ist führender Anbieter von Lösungen und Services in den Bereichen Messaging Security und Workflow für die IBM und Microsoft Collaboration Plattformen. Weltweit vertrauen mehr als 5.000 Kunden und 4 Millionen Anwender auf die Expertise von GBS. Der Konzern ist in Europa, Nordamerika sowie Asien tätig.

© 2016 GROUP Business Software GmbH

Die Produktbeschreibungen haben lediglich allgemeinen und beschreibenden Charakter. Sie verstehen sich weder als Zusicherung bestimmter Eigenschaften noch als Gewährleistungs- oder Garantieerklärung. Spezifikationen und Design unserer Produkte können ohne vorherige Bekanntgabe jederzeit geändert werden, insbesondere, um dem technischen Fortschritt Rechnung zu tragen.

Die in diesem Dokument enthaltenen Informationen stellen die behandelten Themen aus der Sicht der GROUP Business Software GmbH (nachfolgende ‚GBS‘ genannt) zum Zeitpunkt der Veröffentlichung dar. Da GBS auf sich ändernde Marktanforderungen reagieren muss, stellt dies keine Verpflichtung seitens GBS dar und diese kann die Richtigkeit der hier dargelegten Informationen nach dem Zeitpunkt der Veröffentlichung nicht garantieren. Dieses Dokument dient nur zu Informationszwecken.

GBS schließt für dieses Dokument jede Gewährleistung aus, sei sie ausdrücklich oder konkludent. Dies umfasst auch Qualität, Ausführung, Handelsüblichkeit oder Eignung für einen bestimmten Zweck.

Alle in diesem Dokument aufgeführten Produkt- oder Firmennamen können geschützte Marken ihrer jeweiligen Inhaber sein.

Website: <https://www.gbs.com/de>

E-Mail-Adresse: info@de.gbs.com

Standorte: <http://www.gbs.com/de/standorte>

