iRODS and BitCurator

BitCurator project brings in a series of open source digital forensics tools and techniques to collecting institutions, to preserve their born-digital collections [1]. iRODS (Integrated rule-oriented data system) is a data-grid software system, where users can build sharable collections from data distributed across file systems and tape archives[4]. This project attempts to bring the two technologies together, by allowing a user of iRODS to run the BitCurator tools in an iRODS environment and copy the resulting reports into the iRODS grid. This document lists the BitCurator tools that are integrated into iRODS and a overview of each tool along with a description on how to use it.

Guymager

The prerequisite for running the Bitcurator tools on a media or any set of files is to use the tool "Guymager" (http://guymager.sourceforge.net/) and generate an image in the .aff or .E01 format.

Generate Digital Forensics XML file

This utility uses the BitCurator **Fiwalk** tool, takes an image in the .aff or E01 form and generates an XML file. As per [2], "Digital Forensics XML (or DFXML) is a metadata schema designed to facilitate the sharing of structured information produced by forensic tools. DFXML is an attempt to standardize abstractions by providing a formalized language for describing forensic processes". Refer to [2] for more details.

Command to be executed is located in directory irods/server/bin/cmd/fiwalk. This rule Invokes the Fiwalk tool to generate the XML output of the given disk image.

Input Parameter is: Image File path
Output Parameter is: XML File path

Command Structure:
irule -F rulemsiBcGenerateXml.r "*outXmlFile='/Path/to/xmlfile'" "*image='/path/to/image.aff'"

Rule Location:
iRODS/clients/icommands/test/rules3.0/rulemsiBcGenerateXml.r

Command examples:

1. irule -F rulemsiBcGenerateXml.r
default parameters can be modified by changing the following line with appropriate values:
*INPUT *outXmlFile="/AstroZone/home/pixel/bcfiles/xmlfile", *image="/AstroZone/home/pixel/bcfiles/charlie-work-usb-2009-12-11.aff"*

*2. irule -F rulemsiBcGenerateXml.r "*outXmlFile='/home/xmlfile'" "*image='/home/test.aff'"*

Files:

- Local File System:
  The following file resides on the Local File System:
  $iRODS/server/bin/cmd/fiwalk
- iRODS Grid:

Executing this rule creates the following file on the grid:
$iRODS_grid/<xmlfile>

Implementation notes:
The fiwalk tool, an executable file, is copied to iRODS/server/bin/cmd directory:
cp /usr/local/bin/fiwalk iRODS/server/bin/cmd/fiwalk

Bulk Extractor

"bulk_extractor is a computer forensics tool that scans a disk image, a file, or a directory of files and extracts useful information without parsing the file system or file system structures. The results can be easily inspected, parsed, or processed with automated tools." [3]
This tool takes the disk image (the .aff file) as an input and generates an output directory in the specified location, containing a text file for each of the features located in the input image.

For more information on Bulk Extractor scanners, refer to the following URLs:
http://www.forensicswiki.org/wiki/Bulk_extractor
http://wiki.bitcurator.net/index.php?title=Bulk_Extractor_Scanners

Command to be executed is located in directory irods/server/bin/cmd/bulk_extractor

```
bulk_extractor  <image.aff> -o <output directory>
```

Input Parameter is: Image File path
Output Parameter is: File Path for Feature Files

Command Structure:
*irule -F rulemsiBcExtractFeatureFiles "*image='/path/to/image.aff'" "outFeatDir='/path/to/outdir'"*

Rule Location:
iRODS/clients/icommands/test/rules3.0/rulemsiBcExtractFeatureFiles.r

Command examples:

1. *irule -F rulemsiBcExtractFeatureFiles.r*
   Default parameters can be modified by changing the following line:
   *INPUT *image="/AstroZone/home/pixel/bcfiles/charlie-work-usb-2009-12-11.aff", *outFeatDir="/AstroZone/home/pixel/bcfiles/BeOutFeatDir"*

2. *irule -F rulemsiBcExtractFeatureFiles.r "*image='<image>.aff'" "*outDir='/home/be_feature_dir'"*

Files:

- Local File System:
  The following file(s) resides on the Local File System:
  $iRODS/server/bin/cmd/bulk-extractor
- iRODS Grid:
  Executing this rule creates the following file on the grid:
  $iRODS_grid/be_feature_dir
  The actual list of files within this directory depends on the features identified within the image file. Examples:
  $iRODS_grid/be_feature_dir/domain.txt

$iRODS_grid/be_feature_dir/telephone.txt

Implementation notes:
The following file is copied to iRODS/server/bin/cmd directory:
cp /usr/local/bin/bulk_extractor iRODS/server/bin/cmd/bulk_extractor)

Generate Annotated Files (identify_filenames)

This tool takes the output files generated by bulk_extractor and the disk image file (.aff or E01 format) as the inputs and creates the annotated versions of each of the feature files generated by the bulk_extractor.
Input Parameters are:
  Image File path
  Bulk_extractor directory
Output Parameter is:
  Output directory annotatedFilesDir to store the annotated files.

Tool: identify_filenames --all –imagefile "path/to/imagefile.aff" "Path/to/beFeatDir" "Path/to/outAnnDir"

Command Structure:
*irule -F rulemsiBcAnnotateBeFiles.r "*image='/path/to/image.aff'" \
 "*beOutDir='/path/to/beDir'" "*annotateFilesDir='/path/to/newdir'"*

Rule Location:
iRODS/clients/icommands/test/rules3.0/rulemsiBcAnnotateBeFiles.r

Command examples:

1. *irule -F rulemsiBcAnnotateBeFiles.r*
   The default parameters can be modified by changing the following lines appropriately:
   *INPUT *image="/AstroZone/home/pixel/bcfiles/charlie-work-usb-2009-12-11.aff", *beFeatDir="/AstroZone/home/pixel/bcfiles/beFeatDir", *outAnnDir="/AstroZone/home/pixel/bcfiles/outAnnDir"*

2. *irule -F rulemsiBcAnnotateBeFiles.r "*image='/home/test.aff'" "*beOutDir='/home/beDir'" "*annotateFilesDir='/home/annotated_dir'"*

Files:

- Local File System:
  The following file(s) resides on the Local File System:
  $iRODS/server/bin/cmd/identify_filenames
- iRODS Grid:
  Executing this rule creates the following file on the grid:
  $iRODS_grid/annotated_dir
  The actual list of files within this directory depends on the features identified within the image file. Examples:
  $iRODS_grid/annotated_dir/annotated_domain.txt
  $iRODS_grid/annotated_dir/annotated_telephone.txt

Implementation Notes:
The following files are copied to iRODS/server/bin/cmd directory:
~/Research/Tools/bulk_extractor/python/fiwalk.py

~/Research/Tools/bulk_extractor/python/dfxml.py
~/Research/Tools/bulk_extractor/python/bulk_extractor_reader.py
~/Research/Tools/bulk_extractor/python/identify_filenames.py as **identify_filenames**

Generate BitCurator Reports

This tool takes the xml output of the Fiwalk tool and the annotated files created by identify_filenames as the inputs and produces various reports in Excel and PDF formats in the specified output directory. The Python script is located in irods/server/bin/cmd/bc_generate_reports

Input Parameters are:
  Annotated Files Directory (Generated by the rule *rulemsiBcAnnotateBeFiles.r)*
  XML file generated by fiwalk tool (using the rule: *rulemsiBcGenerateXml.r)*
  Configuration file
Output Parameter is:
  Output directory newBcReportsDir where the reports are generated.

```
Tool: bc_generate_reports --fiwalk_xmlfile </path/to/xmlfile/> --annotated_dir </path/to/annotatedDir/ \
                  --outdir </path/to/outdir/> --conf </path/to/configfile/>
```

Command Structure:

```
irule -F rulemsiBcGenerateReports.r "*fiwalkXmlFile='/Path/To/Xmlfile'" \
             "*annotatedDir='/Path/To/annotated_directory'" \
             "*outReportsDir='/Path/To/output_Reports_directory'" \
             "*conf='/Path/To/Config_file'"
```

Rule Location:
/home/sunitha/PER/iRODS/clients/icommands/test/rules3.0/rulemsiBcGenerateReports.r

Command examples:

1. *irule -F rulemsiBcGenerateReports.r*
   The default parameters can be modified by changing the following line with appropriate parameters:
   *INPUT *fiwalkXmlFile="/AstroZone/home/pixel/bcfiles/bcTestFiwalkXmlfile.xml",*
   **annotatedDir="/AstroZone/home/pixel/bcfiles/bcTestBeAnnDir",*
   **outReportsDir="/AstroZone/home/pixel/bcfiles/outReportsDir",*
   **conf="/AstroZone/home/pixel/bcfiles/bcTestConfigFile"*

2. *irule -F rulemsiBcGenerateReports.r "*fiwalkXmlFile='/home/xmlfile'" "*annotatedDir='/home/annotated_directory" "*outReportsDir='/grid/output_directory'" "*conf=/home/config_file"*

Files:

- Local File System:
  The following file(s) resides on the Local File System:
  $iRODS/server/bin/cmd/generate_report
- iRODS Grid:
  Executing this rule creates the following directories/files on the grid:
  $iRODS_grid/outReportsDir:
  $iRODS_grid/outReportsDir/BeReport.pdf
  $iRODS_grid/outReportsDir/FiwalkDeletedFiles.pdf

$iRODS_grid/outReportsDir/FiwalkReport.pdf
$iRODS_grid/outReportsDir/bcTestFiwalkXmlfile.xml.xlsx
$iRODS_grid/outReportsDir/bc_format_bargraph.pdf
$iRODS_grid/outReportsDir/format_table.pdf
$iRODS_grid/outReportsDir/bcfiles/outReportsDir/features

The files under the features directory depends on the image. Examples are:
$iRODS_grid/outReportsDir/bcfiles/outReportsDir/features/domain.xlsx
$iRODS_grid/outReportsDir/bcfiles/outReportsDir/features/telephone.xlsx
$iRODS_grid/outReportsDir/bcfiles/outReportsDir/features/domain.pdf
$iRODS_grid/outReportsDir/bcfiles/outReportsDir/features/telephone.pdf

Implementation notes:
The following files are copied to iRODS/server/bin/cmd directory:

$BitCurator/python/bc_reports_tab.py as **bc_reports_tab**
$BitCurator/python/generate_report.py as **bc_generate_reports**
$BitCurator/python/bc_utils.py
$BitCurator/python/bc_config.py
$BitCurator/python/bc_pdf.py
$BitCurator/python/bc_graph.py
$BitCurator/python/bc_regress.py
$BitCurator/python/bc_genrep_dfxml.py
$BitCurator/python/bc_genrep_text.py
$BitCurator/python/bc_genrep_xls.py
$BitCurator/python/bc_gen_feature_rep_xls.py
$BitCurator/python/bc_config_file

Bitcurator GUI

BitCurator supports a Graphical User Interface using which users can launch the tools explained above. A rule is written to launch this GUI. But more work needs to be done to make the GUI to appear on the client screen rather than on the server.

Rule for the Gui-based tool:
/home/sunitha/PER/iRODS/clients/icommands/test/rules3.0/rulemsiBcGenerateReportsGui.r

Command example:
irule -F rulemsiBcGenerateReportsGui.r

References

[1] BitCurator: http://www.bitcurator.net/
[2] DFXML: http://wiki.bitcurator.net/index.php?title=Fiwalk_and_DFXML
[3] Bulk Extractor: http://www.forensicswiki.org/wiki/Bulk_extractor
[4] iRODS: https://www.irods.org