

ThreatFactor NSIA - Feature #72

Create Archive of Scanned Pages

04/08/2010 11:58 PM - Luke Murphey

Status:	New	Start date:	04/08/2010
Priority:	Normal	Due date:	
Assignee:		% Done:	0%
Category:	Scan Engine	Estimated time:	0.00 hour
Target version:			
Description			
NSIA should create an archive of pages scanned so that the page data can be viewed by the Analyst reviewing the findings.			

History

#1 - 04/08/2010 11:59 PM - Luke Murphey

A system to display the findings should be either (or both):

1. Displayed the actual bytes seen (hex editor style) or as colorized HTML
2. The page should be shown pre-rendered (as an image)
3. Don't display the resource at all, let the user download it and open it in any editor they wish

Note that the content of the pages must be escaped so that exploits aren't relayed to the user. Method 1 above avoids this issue since the actual contents are displayed. Method 2 is inherently dangerous since all of the content must be preprocessed. This pre-processing may not filter the contents correctly and may actually cause the application itself to be exploited.

A warning should be presented when the page contents are truncated.

Data that should be contained in the archive includes:

- URL and Filename
- Content-Type
- Date First Archived
- Hash (this will be used as a unique signature, perhaps even the database primary key if it is random enough)

It is highly possible that the same file will exist in the archive multiple times for different websites (likely with shared libraries such those provided with Ruby on rails) or on the same website (URLs may be different but point to the same file on the web-server drive). Therefore, files should be stored in such a way that the file data does not have to be stored more than once (since it is large). To do this, two tables will be used:

- Meta-Data Table: Contains the URL, Content-Type, Date Archived
- Data-Archive Table: Contains the file contents, size, isTruncated bit

#2 - 04/09/2010 12:07 AM - Luke Murphey

The scan report should have links to the content that was scanned. The pages should load in a separate window or be displayed in frame.

#3 - 11/02/2010 01:18 PM - Luke Murphey

- *Category set to Scan Engine*