CJTEC

VERIFICATION REPORT
Toolkit for Selective Analysis & Reconstruction of Files (FileTSAR)

CJTEC Verification Report **July 2022**

PURPOSE

Perform a verification assessment of Purdue University's Toolkit for Selective Analysis & Reconstruction of Files (FileTSAR).

- Office of Justice Programs. (2020). *FileTSAR final summary overview*. Retrieved from https://www.ojp.gov/pdffiles1/nij/grants/254635.pdf

CJTEC

INTRODUCTION

FileTSAR captures data flows and provides a mechanism to selectively reconstruct multiple data types, including documents (e.g., doc, docx, pdf), images (e.g., jpg, png, gif), email (based on SMTP, IMAP, IMP), and Voice over Internet Protocol sessions, for large-scale computer networks. FileTSAR attempts to address the challenges faced by digital forensic examiners when investigating cases involving large-scale computer networks by using hashing for authentication and indexing of each carved file to maintain the forensic integrity of probative data.

FileTSAR acquires and analyzes data from enterprise-scale networks using multiple processes: packet capture (i.e., recording the packet traffic on a network), protocol parsing (i.e., parsing out the different network protocols and fields), search and analysis, and data visualization. A suite of four modules performs these processes:

- **Collector module:** Captures and stores network traffic in a data repository.
- **Indexer module:** Processes, parses, and indexes data stored within the data repository by source address, destination address, and hashing function.
- **Analyzer module:** Identifies indexed data by interlinkage of files, packets, users, and timelines and reconstructs the data, which are sent back to the collector for storage.
- **Visualizer module:** Enables the viewing of data by the forensic investigator and allows for the identification of trends, patterns, or repetitions that may be pertinent to investigations.

This verification effort assessed the system's capabilities as described in the *FileTSAR Final Summary Overview* and the feasibility of the tool for law enforcement agency adoption and use. Furthermore, any notable performance or conceptual gaps that may influence implementation of this tool are addressed in this report.

CJTEC

Methodology

FileTSAR was installed on networked digital forensic servers and computers. The network environment comprised two servers specifically built to perform digital forensic science on all data formats and two forensic laptop computers.

Server Unit Specifications

Dual Intel Core i9 12 core CPU

(16) fully populated hard drive bays:

- (2) 16tb RAID 5: (5 x 4tb SATA hard drives in 5 drive chassis x2)
- (1) Hot Swap Shock Mounted 1tb SATA removable hard drive
- (1) 1tb NVMe M.2 solid state OS drive
- (4) 1tb Solid State drives (SSDs) in each SATA bays 1 through 4

128gb DDR4 RAM

(2) Gigabit LAN Controllers

Windows 10 Pro 64-bit OS

Laptop Unit Specifications

Intel Xeon 2.80 GHz CPU

(3) SSDs:

(2) 2tb drives

(1) 1tb drive

(1) 500gb Ultra-Fast SD drive

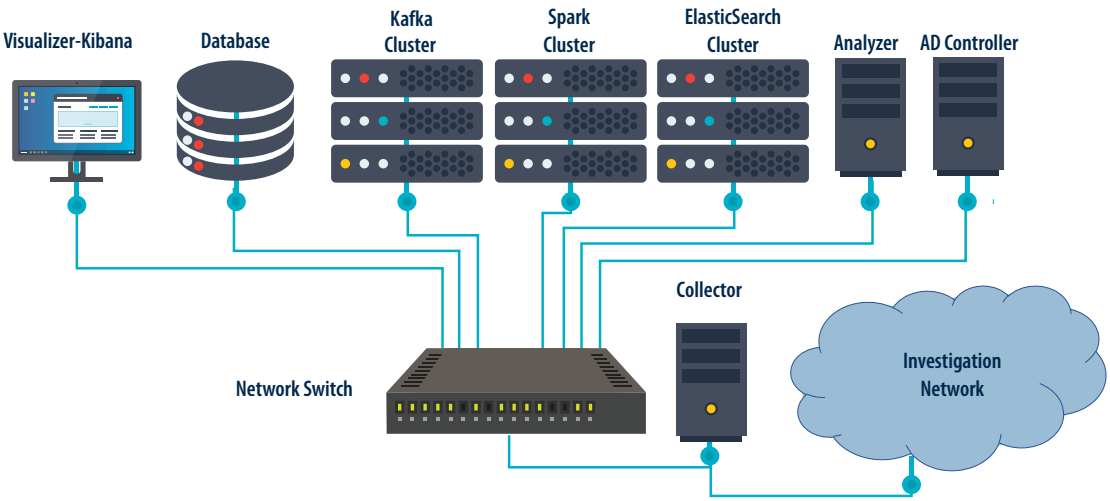
96gb RAM

(1) Gigabit LAN Controller

Windows 10 Pro 64-bit OS

For operation, the FileTSAR system uses multiple virtual machines (VMs) (guest OS or application environments) that run on the aforementioned physical machines. **Figure 1** illustrates the recommended FileTSAR architecture, which identifies the requirement for eight servers to operate the system. For this testing, Purdue University owned and operated the eight-server operating architecture; thus, to sufficiently test FileTSAR's proposed capabilities, this verification required remote access to their system via the two-server network testing environment.

Figure 1. Illustrates the recommended architecture for FileTSAR. Source: Purdue University. (2020). *FileTSAR Final Summary Overview*. National Criminal Justice Reference Service. Retrieved from https://www.ojp.gov/pdffiles1/nij/grants/254635.pdf



The FileTSAR components all require Linux Ubuntu 16.04 (64-bit) operating systems, which are identified in the installation guide as:

- Server 1—(Database) (p. 7) (with 1 required software package):
  - MySQL database
- Server 2— (Kafka Cluster) (p. 8) (with 2 required software packages):
  - Kafka server1
  - Kafka server2
  - Kafka server3
  - Zookeeper1
- Server 3—(Collector) (p. 9) (with 3 required software packages):
  - Wireshark/Tshark (Version: 2.6.6)
  - Python 3.6.0
  - Libpcap 1.7.4-2
- Server 4— (ElasticSearch Cluster) (p. 10) (between 16gb and 64gb RAM, 2 CPUs, and dependent disk space):
  - Elasticsearch-master
  - Elasticsearch-data-1
  - Elasticsearch-data-2
  - Elasticsearch-data-3
  - Elasticsearch-coordinator
- Server 5— (Spark Cluster) (p. 11) (with 2 required software packages):
  - Spark-master
  - Spark-slave1
  - Spark-slave2
  - Spark-slave3
- Server 6—(Analyzer) (p. 12) (with 44 required software packages) (32gb RAM)
- Server 7—(Visualizer-Kibana) (p. 14) (with 1 software package)
- Server 8— (Web server) (p. 15) (with 44 required software packages) (32gb RAM)

Following the installation guide directions, Linux Ubuntu 16.04.7 and associated updates were installed on the testing environment as expected. The eight VMs were configured via VMWare, and the related VMWare tools were successfully installed. Additionally, the MySQL database was successfully installed and verified as complete; however, after installing the MySQL database server, the host computer began to lag in performance, and the VM components of FileTSAR failed to reconnect in many instances. Thus, the successful installation and configuration of the remaining servers (2–8) proved challenging and could not be completed.

CJTEC

Results

The Deliverables and Expected Scholarly Products section of the National Institute of Justice (NIJ) solicitation Developing Improved Means to Collect Digital Evidence Eligibility - OMB No. 1121-0329 states that the deliverable would be *"[a]n exemplar of any tool or method resulting from research and development activities funded under this solicitation will be delivered to NIJ at the end of the award for third-party evaluation, along with detailed implementation instructions."*

To conduct this verification, access to the FileTSAR VMs that Purdue used in their validation and training was requested from the developers. In addition, a request was made for access to the Purdue FileTSAR environment to perform testing and evaluation. Ultimately, attempts to gain access to an operational instance of FileTSAR either using remote access to the Purdue environment, using copies of the Purdue VMs within the test lab environment, or completely building and configuring the VM environment in the test lab were unsuccessful. Therefore, a simulated test environment for a functional verification could not be completed.

As a result of the complicated design and configuration of FileTSAR and the lack of access to either the Purdue FileTSAR environment or copies of the FileTSAR VMs, the successful installation and testing of FileTSAR operation or functionality were not completed. Thus, this effort cannot confirm that FileTSAR performs as reported.

Further, forensically sound data collection refers to the process by which data are collected without any changes to the data or its metadata. To be forensically sound, a data collection process must be consistent, repeatable, and well documented. However, the FileTSAR final report does not conclusively state that the resulting network captures, indexes, parses, and/or performs other processes on the data to preserve it as forensically sound, and there are no references to the format of the forensically sound containers in which the captured data are stored. Because the data are captured in motion, the collections cannot be replicated to confirm the forensic soundness; therefore, this process cannot be conclusively determined to be consistent, and there is no information to confirm the collections are well documented.

The collection of data in motion by government and/or law enforcement agencies is an intercept and therefore is subject to court authorization before collection or capture.[1] Agencies performing these collections without acquiring court orders can lead to suppression of evidence and/or criminal penalties. Throughout the FileTSAR Final Summary Overview, references are made to collecting packet captures as well as indexing and parsing the following protocols: HTTP, FTP, SMTP, IMAP, IMF, SIP, and RTP. These are all internet protocols, which would require court authorization for intercept.

1. Goodison, S., Davis, R., & Jackson, B. (2015). *Digital evidence and the U.S. criminal justice system: Identifying technology and other needs to more effectively acquire and utilize digital evidence*. National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. Retrieved from https://www.ojp.gov/pdffiles1/nij/grants/248770.pdf

CJTEC

Conclusion

FileTSAR does not meet the requirements of the NIJ solicitation because the solicitation specifies that *"[t]his program furthers the Department's mission by sponsoring research to provide objective, independent, evidence-based knowledge and tools to meet the challenges of crime and justice, particularly at the State and local levels."* Moreover, the solicitation requires *"[a]n exemplar of any tool or method resulting from research and development activities funded under this solicitation will be delivered to NIJ at the end of the award for third-party evaluation, along with detailed implementation instructions."* The materials provided by Purdue consisting of a final report, installation guide, and user manual do not constitute an exemplar of the tool.

The information referenced from the testing performed by the developers, the training courses, and demonstrations does not address the interception of data in motion or the authority required to legally intercept, capture, and collect that communication data. Additionally, the materials presented by Purdue do not caution users that interception of internet traffic without proper authority is illegal. The majority of state and local law enforcement agencies do not apply for or conduct communications intercepts because they are labor intensive and consume more human, hardware, and financial resources than most agencies have available. Thus, it is posited that FileTSAR is not a practical solution for most state and local law enforcement agencies in the United States and fails to meet the fundamental criteria of the solicitation. Use of FileTSAR for internal testing and the training class is questionable. If FileTSAR worked as reported in the final report and training videos, no information was provided that the internet intercepts performed in published testing and training were performed with legal authority.

After several hours of installation, configuration, review of the documentation, and communication with the developers, efforts were unsuccessful at re-creating a functionally operational FileTSAR environment. Ultimately, FileTSAR is not a deliverable that should be released for use by the criminal justice community in its current state.

The FileTSAR system developed under NIJ award 2016-MU-MUK091 demonstrated to be successful when tested under controlled conditions and with guidance from the developers during the FileTSAR workshop on September 11–13, 2018. During this event, end users expressed interest in an updated version that would be more user friendly and amenable to smaller agencies with limited budgets and storage capabilities. The CJTEC verification of FileTSAR began in September 2020; however, the research team at Purdue University was funded for a follow-on effort to further refine the capabilities of FileTSAR (award no. 2020-DQ-BX-0008) in fiscal year 2020. Therefore, it is posited that the developers are continuing to develop and refine the FileTSAR platform to align with the needs of the law enforcement community.

Suggested Citation

O'Leary, R., Parsons, M. N., Planty, M., Ropero-Miller, J. (2022). *Verification Report: Toolkit for Selective Analysis & Reconstruction of Files (FileTSAR).* National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. https://cjtec.org/

CJTEC would like to thank Robert O'Leary, Digital Forensics & Investigations SME, for his expertise and insights in developing this document.

This publication was made possible by Award Number 2018-75-CX-K003, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect those of the Department of Justice.