Memory and Disk Forensics in WhatsApp: Analyzing RAM, Hard Drive, and Snapshot Integrity

Alfredo Madrigal
University of North Georgia
amadr8905@ung.edu

ABSTRACT

This paper presents a comprehensive forensic analysis of WhatsApp Web interactions captured from Random Access Memory (RAM), hard drives, and system snapshots within a virtual machine environment. The study explores the extraction and analysis of forensic artifacts, including chat messages, images, user details, and deleted messages, using tools such as Autopsy 4.21.0 and Magnet RAM Capture. RAM analysis proved effective in recovering dynamic session data like chats, poll captions, and user "About" information, while hard drive forensics highlighted the persistence of static artifacts such as real names, emails, phone numbers, deleted messages, and documents. Snapshot forensics, capturing a point-in-time system state, uniquely enabled the recovery of transient media like images. These findings emphasize the complementary value of RAM, disk, and snapshot forensics in uncovering WhatsApp user activities, even after application sessions have ended. Future work will focus on integrating these methods with advanced data integrity techniques to enhance the reliability and scope of digital forensic investigations

.Categories and Subject Descriptors

H.3.5 [Information Storage and Retrieval]: Online Information Services – Web-based services.

I.4.9 [Image Processing and Computer Vision]: Applications – WhatsApp Web forensic analysis.

Keywords

WhatsApp forensics, RAM capture, virtual machine forensics, Autopsy, Magnet RAM Capture, forensic tools, forensic artifact extraction.

1. INTRODUCTION

WhatsApp has become a dominant platform for communication, making it a focal point for digital forensic investigations involving cybercrime or other illicit activities. The increasing adoption of WhatsApp Web introduces unique challenges for forensic analysts, as much of its data is managed transiently in volatile memory (RAM) rather than stored permanently on a device's hard drive. Consequently, understanding how WhatsApp Web handles and stores data across different system components is critical for uncovering evidence in forensic contexts.

This research explores the retrieval of forensic artifacts from RAM, hard drives, and virtual machine snapshots, providing a holistic approach to investigating WhatsApp Web activities. While RAM captures volatile session data like chats, poll captions, and user metadata, hard drives preserve more static and persistent information such as real names, emails, phone numbers, deleted messages, and documents. Additionally, snapshots offer a unique advantage by capturing the system's complete state at a specific point in time, enabling the recovery of transient artifacts like images that might not persist in memory or disk.

The study employs a controlled virtual machine environment with forensic tools such as Autopsy 4.21.0 and Magnet RAM Capture to analyze data generated from WhatsApp Web sessions. By comparing the capabilities and limitations of these methods, the research underscores the importance of integrating RAM, disk, and snapshot forensics to provide a comprehensive understanding of user activities, even when conventional storage mechanisms fail to preserve data.

This paper contributes to the field of digital forensics by detailing the methods, findings, and implications of analyzing WhatsApp Web artifacts across RAM, hard drives, and snapshots, providing valuable insights for both practitioners and researchers in digital investigations.

Our method[1] combines edge/boundary detection with a trained back-propagation neural network that classifies the shape of the object in the photograph, followed by corner detection if required, and completed by a geometric object decomposition, reconstruction, and texturing, with output to VRML or WebGL. Believable, visually compelling results have been obtained for boxes of widely varying dimensions, initial results for spheres demonstrate the proof of concept as of this writing, and cylinders and other shapes require additional work, hence the short paper format of this article.

2. RELATED WORK

Forensic analysis of social media activities has been explored in multiple contexts. Frady's research on Instagram Forensics (Instagram Forensics) provides a significant contribution to understanding how RAM forensics can be applied to capture and analyze data from social media platforms. Using tools like MAGNET RAM Capture and Bulk Extractor, the study outlined methods to retrieve sensitive information such as login credentials, real names, photos, and post captions from Instagram sessions. These techniques demonstrate the effectiveness of RAM capture in social media forensics, especially in environments where volatile memory contains critical evidence that may not be stored permanently on disk. In line with Frady's methodology, our research adopts similar RAM forensic techniques, applying them specifically to WhatsApp Web. While the tools and processes remain largely consistent—using virtual machines and forensic software to capture and analyze RAM—the focus shifts to

WhatsApp's unique data artifacts, such as chat messages, polls, and multimedia files. This further emphasizes the potential of volatile memory for social media forensics, reinforcing the importance of RAM analysis in uncovering user activity on web-based platforms.

3. APPROACH

3.1 Methodology.

The forensic investigation was conducted in a controlled and systematic virtual machine environment using VMware Workstation Pro. Three virtual machines were deployed to ensure data isolation and precise analysis: two machines (VM1 and VM2) for simulating WhatsApp Web user sessions and one dedicated forensic workstation (FW) for data extraction and analysis. The experiments involved capturing data at different stages of user interaction, focusing on three critical components of the system: volatile memory (RAM), snapshots, and hard drive storage.

The study leveraged industry-standard tools such as Magnet RAM Capture for volatile memory acquisition and Autopsy 4.21.0, a comprehensive digital forensic analysis platform, for parsing and analyzing data artifacts. Each method targeted specific types of data:

-RAM Analysis: Focused on volatile session data stored temporarily in unallocated memory.

-Snapshot Analysis: Combined data from memory and disk using merged snapshot files (.vmsn and .vmdk) to recover transient and persistent artifacts.

-Hard Drive Analysis: Aimed to identify long-term stored data by analyzing JSON and MFT files extracted from disk images.

Each analysis method was conducted under strict conditions to maintain the integrity of the data and ensure reproducibility. This approach allowed the comparison of volatile, semi-volatile, and persistent data across different forensic methods, revealing the strengths and limitations of each.

3.2 Data collection.

To simulate realistic WhatsApp Web usage, controlled sessions were created using distinct user accounts. Activities included exchanging messages, sharing documents, conducting polls, and deleting content. These activities ensured a broad range of data types and behaviors were captured.

Data collection protocols were structured as follows:

-Session Termination: After completing each activity, sessions were deliberately terminated to analyze the residual data left in memory and disk.

-Snapshot Creation: Snapshots were taken at specific intervals during interactions to preserve the system's state, including transient and persistent data.

-Disk Image Extraction: Hard drive data was extracted from the virtual machines for subsequent analysis of persistent storage.

Key tools employed:

-Magnet RAM Capture: Enabled precise memory extraction for analysis of transient data.

-Autopsy 4.21.0: Analyzed data from RAM, snapshots, and hard drive images, focusing on metadata, file carving, and keyword searches to retrieve artifacts.

3.3 RAM capture analysis.

RAM analysis was central to uncovering volatile data generated during active WhatsApp Web sessions. Using Magnet RAM Capture, memory data was extracted from VM1 and VM2 and analyzed with Autopsy. This approach highlighted the ephemeral nature of certain artifacts stored only in volatile memory during live interactions.

Findings:

Data Types Found:

-Chat messages, poll captions, and user "About" information, reflecting session-specific data.

-Real names, emails, phone numbers, documents, and partially deleted messages.

Most Common File Types:

-Unallocated Files: These files contained remnants of WhatsApp session data. Fragments of messages and metadata were retrieved through keyword searches and metadata analysis.

However, no images, audio files, or group chat messages were recovered, suggesting these are either encrypted or managed differently by WhatsApp Web in RAM.

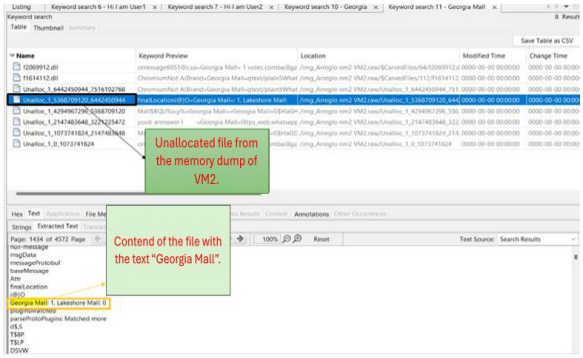

Image showing the finding in RAM memory for the poll vote.

3.4 Snapshot Analysis

Snapshot forensics were critical for bridging the gap between memory and disk-based analysis. By merging .vmsn (snapshot memory) and .vmdk (disk) files, Autopsy reconstructed the system's state at specific points in time, enabling the recovery of both transient and persistent artifacts.

Findings:

Data Types Found:

-Real names, emails, phone numbers, deleted messages, and documents, consistent with hard drive analysis.

-Images: Recovered exclusively through snapshots, demonstrating their transient presence during active sessions.

Data Types Not Found:

-Poll captions, chats, group messages, audio files, and "About" information, suggesting these artifacts are either too transient or encrypted beyond recovery.

Most Common File Types:

-JSON Files: Contained structured data, including metadata related to users and activities.

-MFT Files: Offered low-level traces of file activity, assisting in metadata reconstruction.

Snapshot analysis proved invaluable for retrieving transient artifacts like images that were absent in both RAM and hard drive investigations. It highlighted the unique value of snapshots in preserving ephemeral data.

3.5 Hard drive Analysis

Hard drive analysis focused on identifying persistent data stored on the virtual machine's disk. By extracting and analyzing disk images with Autopsy, a detailed picture of WhatsApp Web's long-term storage was obtained. Notably, the analysis was enhanced by merging .vmsn and .vmdk files to simulate a comprehensive storage environment.

Findings:

Data Types Found:

-Real names, emails, phone numbers, deleted messages, and documents. These artifacts reflect persistent user-related data stored on the disk.

Data Types Not Found:

-Poll captions, chats, group messages, "About" information, images, and audio files, indicating that these are transient or encrypted.

Most Common File Types:

-JSON Files: Captured structured data and metadata.

-MFT Files: Provided disk-level traces of file creation, modification, and deletion.

Hard drive forensics reveal primarily static data types, underscoring its role in analyzing persistent storage while highlighting its limitations in retrieving dynamic or transient session data.

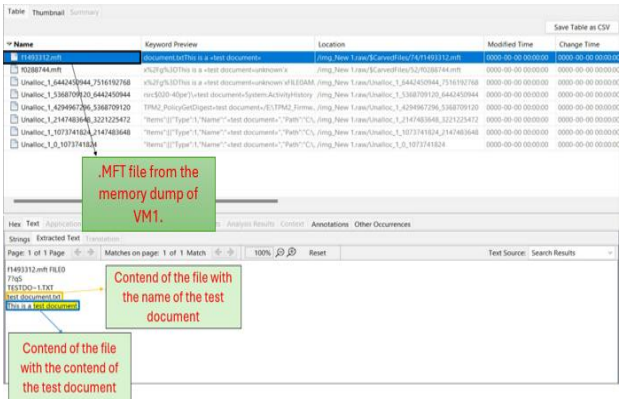

Image showing the finding in Hard drive memory for the document file.

3.6 Summary

The forensic analysis of WhatsApp Web data, conducted across RAM, snapshot, and hard drive environments, underscores the need for a multi-layered approach to digital investigations. Each method provided unique insights into different aspects of data management, contributing to a comprehensive understanding of how artifacts are stored, preserved, and retrieved in a virtual machine environment.

RAM analysis was highly effective for capturing transient session data that only exists during live interactions. Artifacts such as chat messages, poll captions, and "About" information were recoverable due to their ephemeral nature. Additionally, unallocated memory proved invaluable, storing fragments of user data and metadata that could be retrieved using keyword searches. However, RAM's limitations were evident in its inability to store or retain media files such as images and audio, which likely have a minimal memory footprint or are encrypted in memory.

Snapshot analysis filled the gap between volatile memory and persistent storage by combining data from .vmsn (memory) and .vmdk (disk) files, offering a holistic view of the system state at specific points in time. This method uniquely enabled the recovery of images, which were absent in both RAM and hard drive analyses. The ability to merge memory and disk data in snapshots highlighted the importance of capturing the entire system environment to understand transient and persistent behaviors. The findings also emphasized the value of snapshot analysis in reconstructing user activity beyond what is possible with memory or disk analysis alone.

Hard drive analysis demonstrated its strength in recovering static and persistent artifacts, such as real names, emails, phone numbers, documents, and deleted messages. These artifacts reflect the long-term storage behavior of WhatsApp Web and provide a stable source of evidence for forensic investigations. The prevalence of JSON files and MFT traces confirmed the role of the hard drive as a repository for structured data and metadata. However, the absence of session-specific artifacts such as chat messages, poll captions, and group messages underlined the limitations of disk-based analysis for dynamic interactions.

By integrating findings from all three methods, the study paints a complete picture of WhatsApp Web's data lifecycle. The complementary strengths of each method highlight the need for a layered forensic strategy:

RAM analysis is essential for recovering volatile session data critical to understanding user activity during interactions.

Snapshot analysis serves as a bridge, capturing both transient and persistent data, and is particularly effective for media recovery.

Hard drive analysis anchors the investigation with its focus on static, long-term data storage, providing a foundation for metadata and user information recovery.

The study's findings have significant implications for digital forensics. The ability to recover different data types across RAM, snapshot, and hard drive environments demonstrates the importance of tailoring forensic approaches to the unique characteristics of the application being investigated. For WhatsApp Web, this means focusing on memory for transient session artifacts, snapshots for a holistic view of activity, and hard drives for long-term data storage.

Future research should build upon these insights by exploring advanced methods for integrating these three approaches. Techniques such as automated artifact correlation, improved encryption handling, and snapshot integrity validation could enhance the robustness and accuracy of forensic investigations. Additionally, applying this multi-layered approach to other web-based applications could provide valuable comparisons and extend the utility of these findings across broader forensic contexts.

In conclusion, the layered analysis of RAM, snapshots, and hard drives provides a robust framework for investigating WhatsApp

Web and similar platforms. This approach ensures a comprehensive recovery of artifacts, supporting the needs of forensic analysts in uncovering critical evidence while addressing the complexities of modern web-based applications

4. RESULTS

The forensic investigation of WhatsApp Web data across RAM, snapshot, and hard drive environments revealed the strengths and limitations of each method in recovering user artifacts. RAM analysis was effective in capturing transient session data, such as chat messages, poll captions, and user metadata, but lacked media file recovery. Snapshot analysis provided a holistic view by merging memory and disk data, uniquely enabling the recovery of images, which were absent in RAM and hard drive findings. Hard drive analysis was most effective for recovering persistent data, such as real names, emails, phone numbers, and documents, but failed to capture transient artifacts like chats and poll captions.

The findings demonstrate that each forensic method complements the others:

-RAM excels at retrieving volatile session data.

-Snapshots bridge the gap, capturing transient media alongside persistent data.

-Hard Drives focus on long-term data storage and metadata.

By integrating these methods, a comprehensive reconstruction of user activity on WhatsApp Web is achievable, ensuring critical evidence is preserved across diverse data storage mechanisms.

Conclusion Table

Scenario	VM1	VM2	Extra Information
Real Names	Found	Found	
Emails	Found	Found	
Phone Number	Found	Found	
About User 1	Found	Found	
About User 2	Found	Found	
Poll captions	Found	Found	The vote was also found in the analysis represented with a 1 next to the name of the mall
Chats	Found	Found	
Image	Not Found	Not Found	
Audio	Not Found	Not Found	
Deleted messages	Found	Not Found	
Group Chat messages	Found	Found	
Document	Found	Found	The content of the document was also found in the analysis

RAM analysis conclusion table.

Conclusion Table

Scenario	VM1	VM2	Extra Information
Real Names	Found	Found	
Emails	Found	Found	
Phone Number	Found	Found	
About User 1	Not Found	Not Found	
About User 2	Not Found	Not Found	
Poll captions	Not Found	Not Found	
Chats	Not Found	Not Found	
Image	Not Found	Not Found	
Audio	Not Found	Not Found	
Deleted messages	Found	Found	
Group Chat messages	Not Found	Not Found	
Document	Found	Found	The document was found

Hard Drive analysis conclusion table.

Conclusion Table

Scenario	VM1	VM2	Extra Information
Real Names	Found	Found	
Emails	Found	Found	
Phone Number	Found	Found	
About User 1	Not Found	Not Found	
About User 2	Not Found	Not Found	
Poll captions	Not Found	Not Found	
Chats	Not Found	Not Found	
Image	Found	Found	
Audio	Not Found	Not Found	
Deleted messages	Found	Found	
Group Chat messages	Not Found	Not Found	
Document	Found	Found	The document was found

Snapshoot analysis conclusion table.

5. CONCLUSIONS

This study presented a comprehensive forensic investigation of WhatsApp Web data using three complementary approaches: RAM analysis, snapshot analysis, and hard drive analysis. Each method offered unique insights into how data is managed, stored, and preserved within a virtual machine environment, underscoring the importance of a multi-faceted forensic approach.

The methodology section detailed the structured approach to analyzing WhatsApp Web interactions, leveraging tools such as Magnet RAM Capture and Autopsy 4.21.0. Virtual machine environments provided a controlled setting to isolate artifacts and ensure data integrity. This foundation allowed precise data collection and artifact retrieval.

In the data collection phase, WhatsApp Web activities were simulated under controlled conditions to generate a diverse range of artifacts, including chat messages, polls, documents, and deleted content. Snapshots and disk images captured at critical points

ensured a comprehensive dataset, accommodating transient, persistent, and session-specific data.

The RAM analysis revealed the volatile nature of WhatsApp Web's session data. It successfully recovered chat messages, poll captions, user metadata, and documents stored temporarily in unallocated memory. However, it highlighted limitations in recovering media files like images and audio, likely due to their transient storage mechanisms or encryption.

The snapshot analysis bridged the gap between memory and disk by capturing the entire system state at a specific time. This approach uniquely enabled the recovery of images, a critical artifact absents in other methods. Snapshots also preserved transient and persistent artifacts, providing a more holistic view of WhatsApp Web activity.

The hard drive analysis focused on identifying persistent data stored in JSON and MFT files. This method proved effective for recovering user-related data, such as real names, emails, phone numbers, and documents. However, it failed to capture dynamic session data like chats, poll captions, and media, emphasizing the static nature of hard drive storage.

The results demonstrated the strengths and limitations of each approach. RAM analysis was invaluable for capturing volatile data during active sessions, while snapshot analysis provided a comprehensive state view, uniquely recovering images. Hard drive analysis contributed to long-term data storage insights, particularly for persistent user information and metadata.

In summary, the findings underscore the importance of integrating RAM, snapshot, and hard drive forensic methods to ensure a complete reconstruction of user activity on WhatsApp Web. Each method contributes a vital piece to the puzzle, addressing different layers of data storage and retrieval. Future work could explore advanced techniques for integrating these methods, such as automated artifact correlation and enhanced encryption handling, to further refine digital forensic investigations. This multi-layered approach provides a robust framework for investigating web-based platforms, supporting both researchers and practitioners in uncovering critical evidence.

6. FUTURE WORK

This study lays the foundation for several extensions that could broaden its impact and applicability. Future work can explore the application of these forensic techniques to other social media platforms, such as Facebook, Instagram, Twitter, and Telegram. These platforms share similar characteristics with WhatsApp Web, managing a mix of transient and persistent data, and could benefit from the methods used here to recover and analyze artifacts. Expanding the scope of including additional platforms would provide a comparative understanding of how data is stored and managed across different web-based services.

Additionally, this research can be extended to analyze the desktop version of WhatsApp. Unlike the web version, WhatsApp Desktop interacts more directly with the local system, potentially offering access to additional artifacts or different storage behaviors. Investigating these differences would provide a more complete view of WhatsApp's data management practices across platforms.

Another promising direction is the exclusive analysis of WhatsApp mobile applications. Smartphones, being the primary device for WhatsApp use, store data differently compared to web and desktop environments. Exploring how volatile and persistent data is managed on mobile devices would complement the findings from this study and provide deeper insights into WhatsApp's overall ecosystem.

By extending these techniques to other platforms and WhatsApp environments, future research can build on this study's findings to address a broader range of forensic challenges and enhance the ability to recover critical evidence across various digital contexts.

7. REFERENCES

[1] ACM, New York, NY, USA, 139-147. http://doi.acm.org/10.1145/988834.988859.

[2] J. Frady. 2019. Instagram Forensics: Analyzing RAM Forensics Artifacts in Virtual Machines. University of North Georgia. Available upon request.

[3] [4] J. Frady. 2019. Instagram Forensics: Analyzing RAM Forensics Artifacts in Virtual Machines. University of North Georgia. Available upon request.

[4] D. Keskin. 2019. Facebook Forensics: Volatile Memory Forensic Artifact Analysis. University of North Georgia. Available upon request.

[5] Carrier, B. (2005). **File System Forensic Analysis**. Addison-Wesley Professional.

[6] Casey, E. (2011). **Digital Evidence and Computer Crime: Forensic Science, Computers and the Internet** (3rd ed.). Academic Press.

[7] NIST. (2006). Guide to Integrating Forensic Techniques into Incident Response (Special Publication 800-86). National Institute of Standards and Technology. Available at: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-86.pdf

[8] Volatility Foundation. (2023). The Volatility Framework. Available at: https://www.volatilityfoundation.org/

[9] Reith, M., Carr, C., & Gunsch, G. (2002). **An examination of digital forensic models**. *International Journal of Digital Evidence, 1(3)*, 1-12.

[10] Richards, J. C., & Roussev, V. (2005). **The role of memory forensics in detecting advanced malware**. *Digital Investigation, 3(3)*, 142-147.