ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

IMPROVING COLLECTION OF DATA TYPE EVIDENCE AND THE INTEGRITY OF EVIDENCE COLLECTED USING SHA-256 HASHING ALGORITHM FOR WEB BROWSERS

[1]NUR HAZIERAH MOHD SHAH, [2]AZIAH ASMAWI, [3]SHARIFAH MD YASIN
[1]Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Serdang, Malaysia
E-mail: [1]GS64768@student.upm.edu.my, [2]a_aziah@upm.edu.my, [3]ifah@upm.edu.my

ABSTRACT

This study introduces a method to enhance web browser evidence collection in digital forensic investigations. The focus of this study specifically operating 3 evidence collection software forensic toolkits in one developed forensic toolkit called ForenWebSight (FWS). Data is collected from 4 most popular web browsers, Google Chrome, Mozilla Firefox, Microsoft Edge, and Opera in the Windows 11 environment in the context of evidence collection, emphasizing the significance of 35 data types such as history visits, history search, search keyword, cookies, cache, file, session, bookmarks, downloaded files and many more in digital forensic investigations. The existing tools for evidence collection primarily rely on SHA-1 hashing and using older version windows and software toolkits version. Therefore, this study proposes the addition in toolkits implementation, the latest software tools version and the latest solution, an improvement proof-of-concept utilizes SHA-256 hashing algorithm to improve the collection of evidence and enhance integrity. The use of the SHA-256 hash algorithm currently considered secure and resistant to collision attacks. It offers a higher level of security than SHA-1. The evaluation involves comparing the ForenWebSight (FWS) with previous study shows the importance of robust evidence collection tools and methodologies in combating cybercrimes.

Keywords: *Digital Forensics, Web Browsers Evidence Collection, SHA-256 Hashing*

1 INTRODUCTION

Digital forensic investigations are vital for addressing cybercrime and digital incidents, focusing on the collection, analysis, and preservation of digital evidence. This process is essential for uncovering illicit activities and preventing future occurrences. Digital forensics, defined by scientifically derived methods, involves key stages such as evidence collection, validation, analysis, interpretation, and presentation. Each stage is crucial, with evidence collection being the most critical as it ensures the identification and preservation of data from various sources, including web browsers. Validation ensures the authenticity of the evidence, while analysis and interpretation help in forming a coherent narrative of the incident. Finally, the findings are presented clearly to stakeholders.

Web browser evidence collection is particularly significant in digital forensics due to the extensive use of web browsers in everyday online activities. Web browsers serve as a primary gateway for internet access, making them a potential tool for criminals to exploit. Therefore, collecting, analyzing, and preserving web browser evidence is critical in digital forensic investigations. The widespread use of web browsers and their potential for misuse highlight the importance of robust evidence collection methodologies to combat cybercrime effectively.

2 LITERATURE REVIEW

2.1 Web Browser Forensic

The summary of related work encompasses a comparative analysis of several forensic toolkits, namely BrowStEx from Mendoza et al., [14], WEFA from Dissanayake et al., [6] and Vidya et al., [30], Hetman Internet Spy from Mugisha et al., [17], FTK Imager from Arshad et al., [3], Autopsy from Rasool et al., [26] and WBEC from Dafiqah et al., [15] and many more. The objective of this project is to enhance data extraction capabilities based on prior research efforts. The literature review predominantly focuses on comparing forensic toolkits for evidence collection, particularly within web browsers, as observed in related research papers. The findings reveal that many toolkits exhibit limitations in collecting a comprehensive range of digital evidence, prompting the selection of three tools which are Hetman Internet Spy and Browser History Examiner from Dafiqah et al., [15] WBEC Toolkit and Autopsy from Rasool et al., [26] which known for their reliability output in data extraction in the Windows operating system.

The chosen browsers for evaluation are Google Chrome, recognized for gathering the most data type evidence, Mozilla Firefox which acknowledged for storing the least residual data artifacts, allowing data recovery even in private mode and other two additional browser which are Microsoft Edge and Opera will be included for this study. This study specifically aims to collect 35 web browser data type evidence from target web browsers, Google Chrome, Mozilla Firefox, Microsoft Edge, and Opera. Notably, the approach to ensuring the integrity of collected evidence has evolved from using the MD5 and SHA-1 hash algorithm to the more secure SHA-256, resulting in increased resistance to brute-force attack, albeit with longer processing times.

The proposed toolkit's development follows the waterfall model and employs the python programming language, utilizing a personal computer as the host. A synthetic data sample is created to simulate a scenario involving suspicious activities, including searching for potentially criminal content, bookmarking articles and downloading images. The research plan incorporates a comparison of hash algorithms, specifically SHA-1 and SHA-256, evaluating key aspects such as definition, output length, hexadecimal format length, time required for brute-force attack and processing speed using the latest Hashcat software tool. This comprehensive overview of the related work sets the stage for the proposed research, highlighting gaps in existing toolkits and delineating the path for advancements in web browser forensic analysis.

2.2 Integrity

Integrity is a key component of the CIA Triad in information security, alongside Confidentiality and Availability. Ensuring data integrity is crucial for maintaining information security and one common method for verifying data integrity is through cryptographic hash functions. Various hash algorithms, such as MD5, SHA-1, and SHA-256, are commonly used for this purpose.

Table 1 shows the comparison of data type collection and integrity of six existing tools, BrowStEx, WEFA, Hetman Internet Spy, FTK Imager, Autopsy and WBEC. WBEC Toolkit stands out by collecting 16 different types of evidence while ensuring data integrity with the SHA-1 hash algorithm. BrowStEx, though only collecting one type of evidence, secures data integrity using MD5. WEFA, Hetman Internet Spy, FTK Imager, and Autopsy gather between 5 to 12 types of evidence but do not provide evidence integrity measures. All tools are compatible with multiple web browsers, but WBEC offers the most effective balance between the variety of evidence collected and maintaining data integrity.

Table 1: Existing Works on Data Type Collection and Integrity

Toolkits	BrowStEx (Mendoza, et al., 2015)	WEFA (Dissanayake et al., 2021 & Vidya et al., 2022)	Hetman Internet Spy (Mugisha, 2018)	FTK Imager (Arshad et al., 2021)	Autopsy (Rasool, 2020)	WBEC (Dafiqah, 2022)
Number of types of evidence collected	1	10	12	5	11	16
Integrity of the evidence	/	X	X	X	X	/
Hash Algorithm	MD5	X	X	X	X	SHA-1
Web Browser Compatibility	Multiple	Multiple	Multiple	Multiple	Multiple	Multiple

2.3 Problem Statement

The continual global expansion of internet usage has transformed it into a virtual repository of information, becoming an essential service in today's world. However, this ubiquity also presents challenges, as individuals, including criminals, exploit the internet for information gathering and planning illicit activities. In the domain of digital forensics, evidence handling, specifically preservation, emerges as a pivotal aspect [8]. Preservation entails isolating and safeguarding digital evidence in its original state for subsequent processing.

Diverse solutions have been proposed to ensure the relevance and legal admissibility of collected evidence in court trials. Numerous tools have been developed to provide reliable and comprehensive evidence types. In evidence collection, the quantity of collected evidence significantly impacts the depth of investigations. Web browsers, essential software applications for internet access, inherently collect various information, including browsing history, cookies, searches, cache files, and downloaded files [23].

Despite the availability of numerous web browser evidence collection tools, there remains a deficiency in the evidence gathered. Based on Mugisha et al., [17], while 18 data types of evidence have been identified, existing tools only manage to extract 12 of them. Moreover, the use of the SHA-1 hash algorithm in current tools for preserving extracted evidence is deemed insufficient for data security when compared to the more advanced SHA-256 [15], [20], [24], [25]. High risk probability of crimes happens in digital platform. Collection digital evidence from digital sources is essential to perform digital investigation. Web browsers are one of the crucial sources of digital evidence within operating systems (OS) as it has significant potential to uncover clues and apprehend criminal acts which provide interaction and potential involvement. The current approach only compares MD5 and SHA-1 hash algorithm which has limitation in integrity compared to SHA-256 [14-15]. Therefore, there is an imperative need for a method capable of gathering a more extensive range of evidence data types, and implementing a more secure hashing algorithm, such as SHA-256, to enhance data security in digital forensics, particularly in web browser evidence collection.

2.4 Objective

This study aims to address the shortcomings in web browser forensics by proposing an innovative acquisition method through the development of the forensic evidence toolkit called ForenWebSight (FWS). Focusing on the four widely used web browsers, Google Chrome, Mozilla Firefox, Microsoft Edge and Opera within the Windows 11 environment, the toolkit seeks to enhance the collection of data types crucial for forensic investigations. Additionally, the study aims to elevate the integrity of collected evidence by transitioning to the more secure SHA-256 hash algorithm.

The primary objective is to structure a comparative analysis, with the aim to validate results against existing forensic toolkits. The research activities encompass proving the integrity of collected evidence using the SHA-256 hash algorithm, enhancing the collection of data types from web browsers, extracting evidence data types effectively and conducting a thorough comparison and validation of results with other established forensic toolkits. Hence, this study must improve the number of sample web browsers, improve the amount of data type evidence collection from existing forensic toolkits, compare and validate results with other web browsers existing forensic toolkits and improve the integrity of the evidence collected using SHA-256 hash algorithm.

This research aligns with the growing importance of advancing forensic methodologies, as demonstrated by previous studies [3], [6], [14], [15], [17], [26], [30]. The proposed objectives are designed to contribute to the field by providing a more comprehensive and secure approach to web browser forensics, offering a valuable toolkit for forensic practitioners and researchers.

3 METHODOLOGY

The research methodology used in this study is focused on digital forensic investigation and web browser evidence collection. The methodology includes problem formulation, proposed mechanisms, and toolkit development. The Waterfall Model was chosen as the development model for the toolkit due to its ease of management and understanding as framework structure presents in Figure 1.

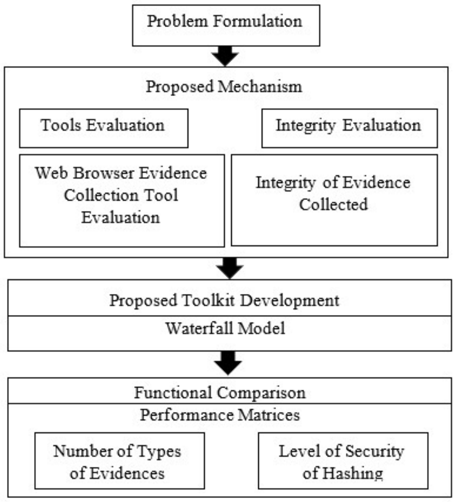

Problem Formulation
Proposed Mechanism
Tools Evaluation
Integrity Evaluation
Web Browser Evidence Collection Tool Evaluation
Integrity of Evidence Collected
Proposed Toolkit Development
Waterfall Model
Functional Comparison
Performance Matrices
Number of Types of Evidences
Level of Security of Hashing

Figure 1 Research Framework [15]

Figure 1 outlines the research process, starting with problem formulation to identify the need for improved web browser evidence collection and integrity verification. It then moves to evaluating existing tools and the integrity of the collected evidence. A new toolkit is developed then compared based on two key performance metrics which are the number of evidence types collected and the security level of hashing algorithms, ensuring both quantity and data integrity improvements.

3.1 Hardware and Software Requirement

Table 2 provides a comprehensive overview of the hardware and software requirement for conducting the proposed web browser evidence collection process, ensuring that the system meets the necessary specifications for successful execution of the forensic tasks.

Table 2: Hardware and Software Requirement

Hardware	Software
Personal computer with a configuration of: • 500 GB SSD (C:): NVMe WDC PC SN530 Disk Drives • 500 GB SSD (D:): Seagate Barracuda SSD Disk Drives • 16 GB RAM • 11th Gen Intel(R) Core (TM) i5-1135G7 @ 2.40GHz 2.42 GH • Intel ® Iris ® Xe Graphics	• Operating System: Windows 11 • Visual Studio Code 1.91.1 • Forensic Tools: Hetman Internet Spy 3.8, Browser History Examiner 1.20.6 and Autopsy 4.21.0 • Web Browsers: Google Chrome 126.0.6478.127, Mozilla Firefox 128.0.2, Microsoft Edge 127.0.2651.74 and Opera 112.0.5197.30 • Hashing Tool: HashMyFiles 2.44 • Hashcat 6.2.6

The FWS Toolkit was written in Python and designed to work on a personal computer with Windows 11. The system used for this study is equipped with a 500 GB SSD, 16 GB RAM, an Intel Core i5-1135G7 processor, and Intel Iris Xe Graphics. The software includes latest version of web browsers such as Google Chrome, Mozilla Firefox, Microsoft Edge, and Opera. Forensic tools used in the analysis were Hetman Internet Spy, Browser History Examiner, and Autopsy which also with latest version as in Table 2, while HashMyFiles and Hashcat were employed for hashing tasks.

3.2 Data Sample

Figure 2 illustrates the simulated process flow of a case scenario attempt within a web browser environment for data collection purposes. In this study, a case scenario was created involving a fake urgent employment email sent to a target user. A synthetic dataset was employed to represent various types of evidence, likely including web browser data which are history visits, history search, search keywords, cookies, cache data, files, sessions, images, favicons, bookmarks, passwords, login data, top sites, thumbnails, emails, social media, downloads, URLs, metadata, timestamps, SQLite database files, autofill data, form data, browser settings, cache images, cached web pages, site settings, site storage, domain, shortcuts, extensions, IndexedDB and local storage data, snapshots, preferences and web data, totalling 35 data types from the criminal account. The use of a synthetic dataset allowed the researchers to control and manipulate the data, ensuring a consistent and reproducible experiment. The criminal sets up fake email and Facebook accounts to impersonate a legitimate company. These accounts are used to track potential targets and collect their email addresses. The criminal then creates a scam email template, using Google search and images, to impersonate a legitimate entity. The fraudulent job offer emails are sent to the targets through Gmail. To ensure consistency, this scenario is executed across different web browsers to confirm that each browser sample yields the same output environment.

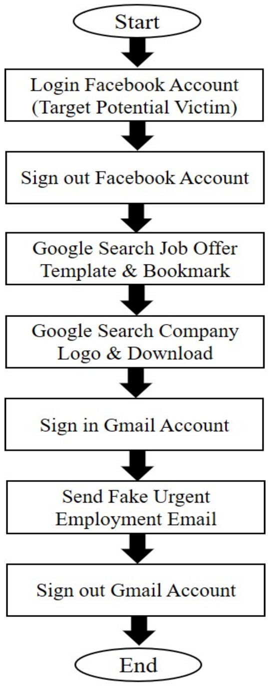

Start
Login Facebook Account (Target Potential Victim)
Sign out Facebook Account
Google Search Job Offer Template & Bookmark
Google Search Company Logo & Download
Sign in Gmail Account
Send Fake Urgent Employment Email
Sign out Gmail Account
End

Figure 2 Case Scenario Attempt

An integrity check was performed using two types of synthetic data samples, each with different character lengths, 'qwerty' with six characters and 'hazierah' with eight characters, as shown in Table 3, instead of actual data collection. This approach was chosen to avoid the lengthy process of performing a brute-force attack on large real data sets. These synthetic samples were designed to test the resilience of the SHA-256 hash algorithm against brute-force attack. By varying the character count in the data samples, the study aimed to assess the time required to brute-force the hash file, providing insights into the algorithm's strength and potential vulnerabilities. All extracted data is expected to use the hashing algorithm for integrity verification, with the synthetic data sets using only lowercase letters.

Table 3: Synthetic Data Sample for Hashing

Character	Hashing Algorithm	
	SHA-1	SHA-256
qwerty (6 char)	b1b3773a05c0ed0176787a4f1574ff0075f7521e	65e84be33532fb784c48129675f9eff3a682b27168c0ea744b2cf58ee02337c5
hazierah (8 char)	70a1feca3cc2bbddd061fe6a6c46ddf6b5cb7dd4	bb3dc575125ddb32bb2b5b988b16e46412da939f1703fe068499e618464db3cc

3.3 Experiment Setup

Data from the browsers was collected using forensic tools, and the integrity of the evidence was verified by applying the SHA-256 hashing algorithm through HashMyFiles. To assess the resilience of the collected evidence, brute-force attacks were conducted using Hashcat, comparing the security strength of SHA-1 against SHA-256.

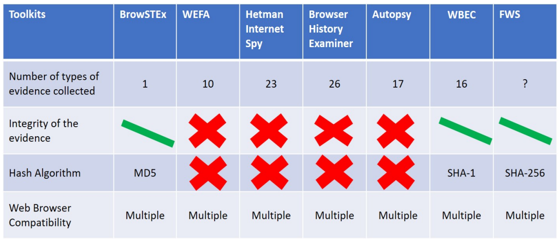
3.4 Comparison and Evaluation

The evidence collected by the FWS Toolkit was compared to that obtained from other existing tools, including WBEC, BrowStEx, WEFA, and Hetman Internet Spy. The evaluation criteria included the number of data types collected, the integrity of the evidence verified through hash algorithms, and the processing time for brute-force attacks, which serves as a measure of security.

4 IMPLEMENTATION

This research scope is based on statistical insights from Market Share, recognizing the prominence of Google Chrome, Mozilla Firefox, Microsoft Edge, and Opera as the most frequently used browsers for Internet technologies [29]. As shown in Table 4, the top three tools that collected the most data artifacts are Hetman Internet Spy, Browser History Examiner, and Autopsy based on their original standard interface. Consequently, this study focuses on utilizing these top three tools and applying the WBEC concept to enhance data integrity and collection.

Table 4: Comparison of Data Type Evidence Collection of FWS Toolkit and Existing Toolkits



Toolkits	BrowSTEx	WEFA	Hetman Internet Spy	Browser History Examiner	Autopsy	WBEC	FWS
Number of types of evidence collected	1	10	23	26	17	16	?
Integrity of the evidence	✓	✗	✗	✗	✗	✓	✓
Hash Algorithm	MD5	✗	✗	✗	✗	SHA-1	SHA-256
Web Browser Compatibility	Multiple	Multiple	Multiple	Multiple	Multiple	Multiple	Multiple

The FWS Toolkit enhances the WBEC Toolkit by Dafiqah et al., [15] incorporating additional samples and creating scenarios on various web browsers. Using latest version software applications, web browsers as evidence collection source, Google Chrome version 126.0.6478.127, Mozilla Firefox version 128.0.2, Microsoft Edge version 127.0.2651.74, and Opera version 112.0.5197.30, evidence is collected through Hetman Internet Spy version 3.8, Browser History Examiner version 1.20.6, and Autopsy version 4.21.0. The collected data is then hashed with SHA-256 using HashMyFiles version 2.44 and subjected to a brute-force attack with Hashcat version 6.2.6 to ensure integrity and reliability of collected data. The research framework of FWS Toolkit is illustrated as in Figure 3.

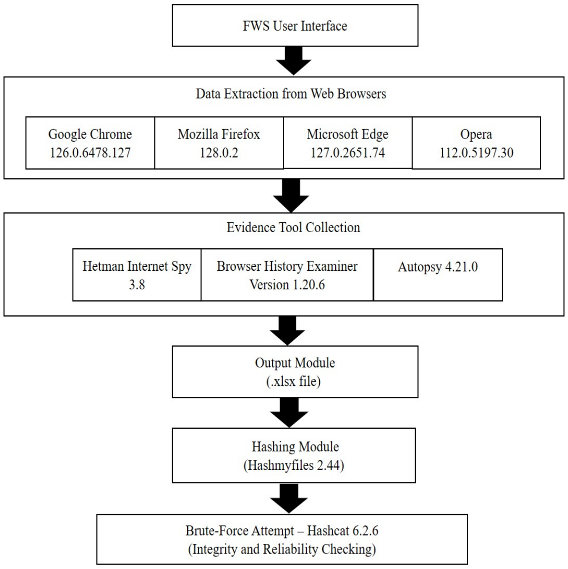

FWS User Interface
Data Extraction from Web Browsers
Google Chrome 126.0.6478.127
Mozilla Firefox 128.0.2
Microsoft Edge 127.0.2651.74
Opera 112.0.5197.30
Evidence Tool Collection
Hetman Internet Spy 3.8
Browser History Examiner Version 1.20.6
Autopsy 4.21.0
Output Module (.xlsx file)
Hashing Module (Hashmyfiles 2.44)
Brute-Force Attempt – Hashcat 6.2.6 (Integrity and Reliability Checking)

Figure 3 Research Framework

5 RESULTS AND DISCUSSION

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 5 compares the evidence data types collected by the proof-of-concept FWS Toolkit with those gathered by other existing tools. For example, the WBEC toolkit collects 16 data types from Google Chrome and 10 from Mozilla Firefox, BrowStEx gathers only 1 data type, WEFA collects 10, Hetman Internet Spy can collect 23, Browser History Examiner collects 26 and Autopsy collects 17. In contrast, the proof-of-concept FWS Toolkit collects 33 out of 35 data types from Google Chrome, 27 from Mozilla Firefox, 29 from Microsoft Edge and 13 from Opera. This demonstrates that FWS Toolkit improves data collection, gathering 17 more data types than some existing tools and ensuring that none of the 35 data types are entirely absent across the four web browsers, suggesting the potential for even broader applicability in other scenarios.

Table 5: Comparison of Data Type Evidence Collection of FWS Toolkit and Existing Toolkits

No	Data Type Evidence	WBEC Toolkit (Dafiqah, 2022)		BrowStEx (Mendoza, et al., 2015)	WEFA (Dissanayake et al., 2021)	Hetman Internet Spy	Browser History Examiner	Autopsy	Proposed (Proof of Concept FWS (ForenWebSight) Toolkit)			
		Google Chrome	Mozilla Firefox						Google Chrome	Mozilla Firefox	Microsoft Edge	Opera
1	History Visits	/	/	/	/	/	/	/	/	/	/	/
2	History Search	/			/	/	/	/	/	/	/	/
3	Search Keyword	/	/		/	/	/	/	/	/	/	/
4	Cookies	/			/	/	/		/	/		
5	Cache Data	/			/	/	/		/	/	/	/
6	File	/	/		/	/	/		/	/	/	/
7	Session	/			/	/	/	/	/	/	/	/
8	Images	/	/			/			/	/	/	
9	Favicon	/				/	/	/	/	/	/	
10	Bookmark	/	/		/	/	/	/	/	/	/	/
11	Password	/				/			/	/	/	
12	Login Data	/	/			/	/	/	/	/	/	
13	Top sites		/			/	/		/	/	/	
14	Thumbnails	/					/		/	/		
15	Email	/	/			/	/		/	/	/	
16	Social Media	/	/			/			/	/	/	
17	Downloads	/	/		/	/	/	/	/	/	/	
18	URLs					/	/	/	/	/	/	/
19	Metadata Database Information					/		/	/	/	/	/
20	Timestamp					/	/	/	/	/	/	/
21	SQLite Database Files					/			/	/		
22	Autofill Data				/	/		/	/	/	/	
23	Form Data					/	/		/	/	/	
24	Browser Settings						/		/	/		
25	Cache Images						/		/	/	/	
26	Cached Web Pages						/				/	
27	Site Settings						/		/			
28	Site Storage						/		/			
29	Domain					/	/	/	/	/	/	/
30	Shortcuts							/	/		/	
31	Extension							/	/	/	/	/
32	IndexedDB and Local Storage Data						/	/	/		/	/
33	Snapshots							/			/	
34	Preferences						/		/		/	
35	Web Data						/		/		/	
TOTAL		16	10	1	10	23	26	17	33	27	29	13

In conclusion, the data collection percentages are as follows shown in Table 6, WBEC Toolkit achieves 45.71%, BrowStEx 2.86%, WEFA 28.57%, Hetman Internet Spy 65.71%, Browser History Examiner 74.29% and Autopsy 48.57%. The FWS Toolkit stands out with the highest score at 94.29%. When compared to the WBEC Toolkit's 45.71%, the proposed FWS Toolkit shows a significant improvement, with a 48.58% increase, nearly doubling the effectiveness of the previous work.

Table 6: Number and Percentage of Data Type Evidence between FWS Toolkit and Existing Tools

Toolkits	WBEC Toolkit (Dafiqah, 2022)		BrowStEx (Mendoza, et al., 2015)	WEFA (Dissanayake et al., 2021)	Hetman Internet Spy	Browser History Examiner	Autopsy	Proposed (Proof of Concept FWS (ForenWebSight) Toolkit)			
	Google Chrome	Mozilla Firefox						Google Chrome	Mozilla Firefox	Microsoft Edge	Opera
Number of types of evidence collected	16/35	10/35	1/35	10/35	23/35	26/35	17/35	33/35	27/35	29/35	13/35
Percentage	45.71%	28.57%	2.86%	28.57%	65.71%	74.29%	48.57%	94.29%	77.14%	82.86%	37.14%
Total Percentage	**16/35 = 45.71%**		**1/35 = 2.86%**	**10/35 = 28.57%**	**23/35 = 65.71%**	**26/35 = 74.29%**	**17/35 = 48.57%**	**33/35 = 94.29%**			

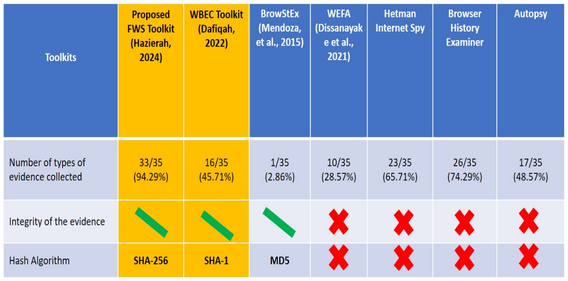
Table 7 shows the time taken for SHA-1 and SHA-256 hash algorithm to crack using brute-force attack. For 6 characters, the time taken to brute-force the SHA-1 hash algorithm is 4 seconds while the SHA-256 hash algorithm is 8 seconds. Hence, SHA-256 takes more 4 seconds to crack the ciphertext than the SHA-1. For 8 characters, the time taken to brute-force the SHA-1 hash algorithm is 4 minutes and 34 seconds while the SHA-256 hash algorithm is 9 minutes and 12 seconds. SHA-256 takes more time to crack with a difference of 4 minutes and 38 seconds. Thus, SHA-256 takes a higher time to crack for both 6 and 8 characters and it is proved that SHA-256 is more secure than the SHA-1 hash algorithm.

Table 7: Hashing Comparison

Brute-Force Attack / Hashing	**qwerty (6 Character)**		**hazierah (8 Character)**	
	SHA-1	**SHA-256**	**SHA-1**	**SHA-256**
Time Started (a)	20:11:17	20:25:55	20:14:32	20:27:45
Time Stopped (b)	20:11:21	20:26:03	20:19:06	20:36:57
Time Taken (b-a)	4 sec	**8 sec**	4 min 34 sec	**9 min 12 sec**
Level of Security	low	**high**	low	**high**

The functional comparison is conducted between proof-of-concept FWS Toolkit and existing tools. Table 8 shows the functional tool comparison. Compared with WBEC Toolkit from Dafiqah et al., [15] work, FWS Toolkits increased by 17 data type evidence is 48.58%, FWS Toolkits have better integrity using SHA-256 than the other existing tool using MD5 and SHA-1 hash algorithm [14-15].

Table 8: Functional Tools Comparison between FWS Toolkit and Existing Tools



Toolkits	Proposed FWS Toolkit (Hazierah, 2024)	WBEC Toolkit (Dafiqah, 2022)	BrowStEx (Mendoza, et al., 2015)	WEFA (Dissanayake et al., 2021)	Hetman Internet Spy	Browser History Examiner	Autopsy
Number of types of evidence collected	33/35 (94.29%)	16/35 (45.71%)	1/35 (2.86%)	10/35 (28.57%)	23/35 (65.71%)	26/35 (74.29%)	17/35 (48.57%)
Integrity of the evidence	✓	✓	✓	✗	✗	✗	✗
Hash Algorithm	SHA-256	SHA-1	MD5	✗	✗	✗	✗

The FWS Toolkit significantly improves both the quantity and quality of evidence collection in browser forensics. It addresses the limitations of tools that either collect limited data or have weaker integrity measures. The study achieves its primary goals, with a 48.58% increase in evidence types compared to the WBEC Toolkit, and the adoption of SHA-256 enhances data security. While some studies note that SHA-256 is slower than SHA-1 [13] and [23], this can be justified in digital forensics that the enhanced security and integrity provided by SHA-256 are more important than the processing time required.

6 CONCLUSION

The study introduces the ForenWebSight (FWS) Toolkit, a web browser evidence collection system designed to enhance both the quantity and integrity of digital evidence by utilizing the SHA-256 hash algorithm. Tested on Google Chrome, Mozilla Firefox, Microsoft Edge, and Opera, FWS Toolkit outperforms previous toolkits like WBEC Toolkit by increasing the number of data types collected from 16 to 33 out of a possible 35, marking a 48.58% improvement. The toolkit also demonstrates superior security, as SHA-256 is significantly harder to crack than SHA-1, ensuring the integrity of the evidence. The study's objectives highlight on expanding browser samples, increasing data collection, validating results through comparison, and enhancing evidence integrity with SHA-256 are reflected in its contributions. The FWS Toolkit has improved evidence collection and strengthened data integrity, demonstrating significant advancements over existing forensic tools. This study advances the field of digital forensics by creating a toolkit that not only expands the range of collected evidence but also ensures that it is securely hashed for future investigations.

7 LIMITATION AND FUTURE WORK

The primary limitation of this research is that evidence extraction using FWS Toolkit, was only conducted on samples from four web browsers which are Google Chrome, Mozilla Firefox, Microsoft Edge, and Opera, where evidence extraction varies across different browsers. Future enhancements could include extending support to private browsing modes, improving cross-platform compatibility, and incorporating even more secure hash functions, thereby increasing the toolkit's effectiveness and reliability in digital forensic investigations.

8 ACKNOWLEDGEMENT

This research was not funded by any grant. This work is supported by Universiti Putra Malaysia under Faculty of Computer Science and Information Technology.

REFRENCES:

[1] Agarwal, Ankit, Ms Gupta, Mr Gupta, Yatendra Gupta, and Chandra Gupta. "Systematic Digital Forensic Investigation Model." *Gupta International Journal of Computer Science and Security*, 2011.

[2] Ahmed, Adnan, Abdul Rehman Javed, Zunera Jalil, Gautam Srivastava, and Thippa Gadekallu. "Privacy of Web Browsers: A Challenge in Digital Forensics." In *Proceedings of the 2022 International Conference on Digital Forensics and Cyber Security*. Springer, 2022.

[3] Arshad, M. R., M. Hussain, H. Tahir, S. Qadir, F. I. Ahmed Memon, and Y. Javed. "Forensic Analysis of Tor Browser on Windows 10 and Android 10 Operating Systems." *IEEE Access* 9 (2021): 141273–94.

[4] Devi, Dhruwajita, Dhrubajyoti Pathak, and Sukumar Nandi. "Vulnerabilities in Web Browsers." *ResearchGate*, 2012.

[5] Dija, S., J. Ajana, V. Indu, and M. Sabarinath. "Web Browser Forensics for Retrieving Searched Keywords on the Internet." In *Proceedings of the 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, Greater Noida, India, 1664–68. IEEE, 2021.

[6] Dissanayake, D., S. Rajakaruna, D. Ranasinghe, A. Wijesooriya, A. Jayakody, and S. Rajapaksha. "Platform Independent Browser Forensic Tool for Advanced Analysis of Artifacts and Case Management." In *Proceedings of the 2021 3rd International Conference on Advancements in Computing (ICAC)*, Colombo, Sri Lanka, 383–88. IEEE, 2021.

[7] Gros, T., R. Dirauf, and F. Freiling. "Systematic Analysis of Browser History Evidence." In *Proceedings of the 2020 13th International Conference on Systematic Approaches to Digital Forensic Engineering (SADFE)*, New York, NY, USA, 1-12. IEEE, 2020.

[8] Hagan, A. "Digital Forensic Process-Preservation / Collection." *DriveSavers Data Recovery Blog*, 2018.

[9] Hariharan, M., A. Thakar, and P. Sharma. "Forensic Analysis of Private Mode Browsing Artifacts in Portable Web Browsers Using Memory Forensics." In *Proceedings of the 2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS)*, Kochi, India, 1–5. IEEE, 2022.

[10] Iqbal, Salman, and Soltan Alharbi. "Advancing Automation in Digital Forensic Investigations Using Machine Learning Forensics." In *Digital Forensics: Threatscape and Best Practices*. IntechOpen, 2019.

[11] Jadoon, Abid, Waseem Iqbal, Muhammad Amjad, Hammad Afzal, and Yawar Bangash. "Forensic Analysis of Tor Browser: A Case Study for Privacy and Anonymity on the Web." *Forensic Science International* 299 (2019).

[12] Johnson, Chapin, Sharveen Paramiswaran, and Akalanka Mailewa. "Discovering Vulnerabilities in Web Browser Extensions Contained by Google Chrome." *ResearchGate*, 2023.

[13] Majeti, Ganesh, Sai YVL, Sai Ulichi, Sachi Mohanty, and Sudha V. "Digital Forensic Advanced Evidence Collection and Analysis of Web Browser Activity." *ICST Transactions on Scalable Information Systems*, 2023.

[14] Mendoza, Abner, Avinash Kumar, David Midcap, Hyuk Cho, and Cihan Varol. "BrowStEx: A Tool to Aggregate Browser Storage Artifacts for Forensic Analysis." *Digital Investigation* 14 (2015): 63–75.

[15] Mior Rayman, D., A. Asmawi, and N. Mohd Ariffin. "WBEC: A Web Browsers Evidence Collection Toolkit for Web Browsers Usage in Windows 10." *International Journal of Technology Management and Information System* 4, no. 1 (2022): 1–15.

[16] Mohammed, Monem. "A Performance Comparative on Most Popular Internet WebBrowsers." *ResearchGate*, 2023.

[17] Mugisha, David. "WEB BROWSER FORENSICS: Evidence Collection and Analysis for Most Popular Web Browsers Usage in Windows 10." *International Journal of Cyber Criminology* 54, no. 12 (2018): 12.

[18] Müller, J. "Number of Internet Users in Malaysia from 2017 to 2023." *Statista*, 2019. https://www.statista.com/statistics/553752/number-of-internet-users-in-malaysia/.

[19] Nalawade, Apurva, Bharne, and Vanita Mane. "Forensic Analysis and Evidence Collection for Web Browser Activity." *ResearchGate*, 2021.

[20] Nakamura, Kazuki, Koji Hori, and Shoichi Hirose. "Algebraic Fault Analysis of SHA-256 Compression Function and Its Application." *Information* 12 (2021): 433.

[21] Ntonja, Morris, and Moses Ashawa. "Investigating Google Chrome 66.0.3359 Artefact: Internet Forensics Approach." *International Journal of Science and Research (IJSR)* 7, no. 7 (2018): 112–22.

[22] Oh, Junghoon, Seungbong Lee, and Sangjin Lee. "Advanced Evidence Collection and Analysis of Web Browser Activity." *Digital Investigation* 8, Supplement (2011): 62–70.

[23] Pereira, Murilo. "Forensic Analysis of the Firefox 3 Internet History and Recovery of Deleted SQLite Records." *Digital Investigation* 5 (2009): 93–103.

[24] Prasanna, S. R., and B. S. Premananda. "Performance Analysis of MD5 and SHA-256 Algorithms to Maintain Data Integrity." In *Proceedings of the 2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)*, Bangalore, India, 246–50. IEEE, 2021.

[25] Ramdani, Ficry, Alam Rahmatulloh, and Rahmi Shofa. "Implementation of JSON

Web Token on Authentication with HMAC SHA-256 Algorithm." *SISTEMASI* 12 (2023): 194.

[26] Rasool, Aamir, and Zunera Jalil. "A Review of Web Browser Forensic Analysis Tools and Techniques." *Researchpedia Journal of Computing* 15, no. 21 (2020).

[27] Rathod, Digvijaysinh. "Web Browser Forensics: Google Chrome." *International Journal of Advanced Research in Computer Science* 8 (2017): 896.

[28] Shayau, Y. H., A. Asmawi, S. N. M. Rum, and N. A. M. Ariffin. "Digital Forensics Investigation Reduction Model (DIFReM) Framework for Windows 10 OS." In *Proceedings of the IEEE 9th International Conference on System Engineering and Technology (ICSET)*, Shah Alam, Malaysia, 459–464. IEEE, 2019.

[29] Similarweb LTD. "Browser Market Share in July 2024." *Similarweb*, 2024. https://www.similarweb.com/browsers/.

[30] Vidya, V., K. Saly, and C. Balan. "Forensic Acquisition and Analysis of Webpage." In *Proceedings of the 2022 2nd International Conference on Intelligent Technologies (CONIT)*, Hubli, India, 1–6. IEEE, 2022.

[31] Yu, Hongbo, Yonglin Hao, and Dongxia Bai. "Evaluate the Security Margins of SHA-512, SHA-256, and DHA-256 Against the Boomerang Attack." *Science China Information Sciences* 59 (2016).