# A Web-Based Mouse Dynamics Visualization Tool for User Attribution in Digital Forensic Readiness

Ernsberger Dominik[1], Ikuesan R. Adeyemi[2], Venter S. Hein[2], and Alf Zugenmaier[1]

[1]Department of Computer Science and Mathematics,
Munich University of Applied Science, Germany
[2]Department of Computer Science, Faculty of EBIT, University of Pretoria, South Africa
ernsberger.dominik@gmail.com,{aikuesan,hsventer}@cs.up.ac.za,
alf.zugenmaier@hm.edu

**Abstract.** The Integration of mouse dynamics in user authentication and authorization has gained wider research attention in the security domain, specifically for user identification. However, same cannot be said for user identification from the forensic perspective. As a step in this direction, this paper proposes a mouse behavioral dynamics visualization tool which can be used in a forensic process. The developed tool was used to evaluate human behavioral consistency on several news-related web pages. The result presents promising research tendency which can be reliably applied as a user attribution mechanism in a digital forensic readiness process.

**Keywords:** Mouse-dynamics; event-visualizer; digital forensic readiness; user identification and attribution; behavioral dynamics.

## 1 Introduction

A substantial aspect of Human-Computer interaction is based on pointing devices, either with the mouse, touch screens or other forms of pointing devices. The study of the behavioral components of human-mouse movement is generally referred to as mouse dynamics [1]–[3]. Mouse dynamics have been widely applied in user identification through authentication [3]–[7] or authorization [1], [8]. The integration of mouse behavioral dynamics as a biometrics for continuous and one-time authentication has gained wider attention in the recent years. This is generally attributed to the relatively cheap requirement specification, ease of data collection, and the high probability of individual uniqueness in mouse dynamics. In terms of the requirement specification, the study of mouse dynamics relies on the existing device, without a need for a specialized device. Furthermore, it does not require any specific positioning or intrusive setting for data acquisition.

Given this flexibility and robustness, the mouse-dynamics is gradually being considered as a suitable forensic mechanism [3], [9] through which human identification can be evaluated in a human-computer interaction. User attribution, as a mechanism for identification of the actual user in an interaction/event in digital forensics [10][11], relies on the reliability of the underlying identification mechanism. User attribution is generally referred to as the process of identifying a user on a digital device; the act of appending a given action/activity to a known user without ambiguity. The underlying mechanism implemented for continuous authentication based on mouse dynamics can, therefore, be adopted as a forensics attribution mechanism. However, the current reliability in the existing studies on continuous mouse dynamics falls below the 0.001 false acceptance rate, and 1.00 false rejection rate, of the European Standard for

Commercial biometric technology [3]. As a step towards the actualization of this reliability, this study, a part of an ongoing behavioral biometrics for user attribution, aims to explore other probable underlying intuition of mouse dynamics for user attribution. To achieve this aim, this study developed a tool that can be used to track and visualize the behavior of human mouse actions on different websites. Various news websites were used as a means to conduct this research study. However, the integration of such mechanism into user attribution for forensics purpose can be feasible, through a digital forensic readiness framework. A digital forensic readiness framework is defined in this context in accordance with the findings from different stakeholders as presented in [12]. Digital forensic readiness is the *proactive* process of collecting, reliably storing, preprocessing and preservation of digital information which would otherwise be unavailable in a postmortem forensic process. A digital forensic readiness framework (DFRF) is therefore defined as a structural capability designed by an organization to maximize the usage of the available digital information in the event of an incident whilst minimizing the eventual cost to such an organization [13]. Given that the DFRF provide a reliable platform for user attribution, a forensic investigation process can significantly benefit from a behavioral biometrics profiling mechanism that is based on either a greater sample size of users (≤ 205-users), or a relatively smaller sample size of user (≥ 4 users) [14], [15]. This further implies that the performance of a mouse-based behavioral-biometrics is not necessarily dependent on the sample size under consideration, rather, on the capability of the mechanism adopted.

The remainder of this paper is organized as follows: a review of related works on continuous authentication and forensics, based on mouse dynamics, is shown in Section 2. This is then followed by the methodology employed to develop the tool and evaluation of human action. In addition, the exploratory process of other feasible intuitions which can be used to observe individual uniqueness during interaction with a mouse (or any other pointing device) is presented in Section 3. Analysis of result is presented in section 4. Discussion and limitation of the findings of this study are presented in Section 5 of this paper, while the conclusion is presented in Section 6.

## 2 Related Work

Research works on behavioral biometrics (a nonintrusive method of identifying a user) in relation to human-computer interaction is gaining wider attention from the research community. Keystroke dynamics [16] and mouse dynamics are the two major focus of research in user identification. While some research focused on either mouse dynamics or keystroke dynamics, a few have attempted to integrate both mechanisms for user identification or authentication using a multimodal approach [1], [17]. A study in [18], which builds on the works in [19]–[21], investigated the probability of adopting mouse dynamics as a behavioral biometrics which can be used for user authentication. Raw mouse data was aggregated into high-level actions such as point-and-click, drag-and-drop. This is then characterized by action type, distance, angle, frequency, speed, duration, and direction. The aggregation process resulted in a segmentation of the raw mouse event into sessions of mouse strokes. A total of 39 mouse-action features, were further computed. Evaluation metrics include the false acceptance rate (FAR), false rejection rate (FRR), equal error rate (EER), accuracy, and pattern verification time. With an additional aim of verifying the influence of mouse device type, the study dichotomized mouse dynamics based on device. Findings from the study suggest that the type of mouse device used by a user can influence the behavior of the user. The dichotomy based on the type of device hardware yielded an accuracy of 96.7% and 97.8% respectively. Similarly, a study in [3] which builds on [22], explored the

probability of user authentication based on mouse dynamics. Additional features such as single-click, double-click were included in the study. The study asserted that authentication can be performed in 11.8seconds of mouse action, with a FAR and FRR of 8.74% and 7.69% respectively. This was based on 5550-data samples on 37 respondents. Recent findings in [1] observed that the multimodal approach based on C4.5 decision tree algorithm, LibSVM and Bayes Net classifiers can be used to improve the identification performance of mouse dynamics. The findings from the study showed that when authentication was based solely on mouse dynamics, the C4.5, LibSVM, and BayesNet resulted in an average accuracy of 74.26 ±5.55, 85.53 ±4.26, and 82.77 ±2.96 respectively. The study was also anchored on the earlier studies of [22] and [23]. All identified existing still suffers from the limitation of poor error rate, and classification accuracy. In addition, these studies are targeted at user authentication, which does not cover some forensic processes. Whilst user authentication can be integrated into forensics, there is a need for a forensic perspective on mouse dynamics. This perspective includes the ability to visualize individual mouse paths, correlate individual mouse action from the different timeline (consistency checker), generate individual mouse dynamics for storage, subsequent analysis, as well as correlation with other users. Research targeted at improving these evaluation parameters remains a major focus, especially for usage in digital forensic readiness.

### 2.1 Purpose and contribution of this study

In addition to the development of a tool for mouse dynamics visualization and analysis, this study differs from existing studies on mouse dynamics in terms of its aim and the fundamental unit of measurement. A pixel-based single path property is considered as the fundamental unit of measurement of mouse dynamics in this study. This is intuitively distinct from the click-based [18], [21], [22], and stroke-based [1], [3] approach which is aggregated over sessions, as observed in existing studies. The current approach is based on the observation of individual path, and their corresponding behavioral characteristics. Based on the structural characteristics of an individual path in a given mouse dynamics data, this study explored the behavioral consistencies in users. This consistency will thereafter be integrated into a forensic readiness framework. As an illustration of the forensic application, an illicit behavior can be mapped to an unknown subject within an organization based on the pre-defined template of each user gathered through a reliable forensic readiness process, within the organization. Furthermore, a deviation from the known behavioral consistency of a user can be used as a trigger for incident response and investigation. Such can also be applied to sniff out a malicious insider in an organization, by surreptitiously monitoring a triggered malicious-flag on a system. The digital forensic readiness approach defined in [24] identified event logs as a major aspect of the technology-enabled forensic process. This approach to forensics is also supported by the recommendation in [18] on the application of mouse dynamics in user attribution. User attribution in this context refers to the process of identifying a user based on their mouse dynamics. The methodology used to achieve this aim is presented in the next section.

## 3 Research methodology

The approach employed to address the aim of this study is detailed in this section. The overall design process of the proposed path-pattern visualization is divided into four main parts, as depicted in **Figure Figure 1**.

1. Tracking and recording of the computer cursor and the corresponding web page elements during surfing on news web pages. This includes the mouse click, scrolling as well as the cursor movement. The extraction of the HTML objects which the user clicked or hovered over during the recording.
2. Extracting relevant information and calculating human behavior attributes based on the data captured. Furthermore, arrange and store the results in a way to provide easy access and evaluate it afterward.
3. Visualize the stored data with re-drawn trajectories, tables, and timelines such that an investigator can quickly search and compare different paths.
4. Identification of user patterns based on the extracted features from the mouse actions of the user.

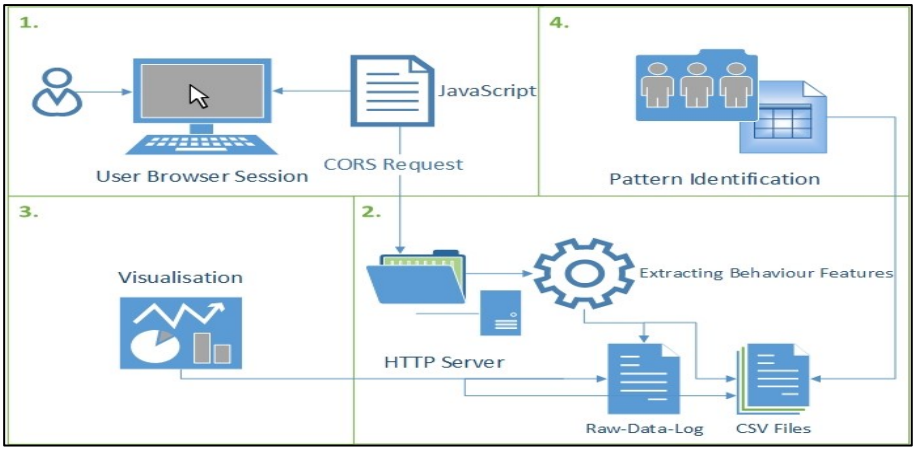



**Figure 1**. Overall design Approach

### 3.1 Mouse Navigation Tracking Process

To achieve the first goal, a client-side JavaScript is needed. It is embedded in the header of the loaded HTML page while running in the background. Two respective third-party browser extensions, i.e. Chrome: RunJS [25] and Firefox: Custom Style Script [26], which includes the stored JavaScript in the loaded pages, was used in this study. It would also be possible to use a specific proxy, which inserts the JavaScript code on the fly, into every page accessed through it, as implemented in [27]. As soon as an event (Mouse-Click-Down, Mouse-Click-UP, Scroll-Up, Scroll-Down, or Mouse-Move) occurs, the corresponding event listener is evoked. It captures the coordinates of the cursor, the precise timestamp, the HTML object of the page where the cursor is currently located, the delay (flight) between Mouse-Click-Down and Mouse-click-Up as well as the Uniform Resource Locator (URL) of the current web page. On the first Mouse-Click-Down event, it captures an additional basic user-agent information like the resolution of the page, time stamp, the type of browser and the browser version. In terms of scrolling, it captures, besides the location and timestamp, the number of scrolled pixels in the y-direction.

The coordinates, resolution and scrolled pixels are captured with *clientX, clientY, clientWidth, clientHeight* and *pageYOffset* methods. These methods return the value in Cascading Style Sheets (CSS) pixels. A CSS pixel is a software pixel which forms the unit of measurement, whereas a hardware pixel is an individual dot of light on the screen.

A CSS pixel can contain a few hardware pixels and is designed to be the same size across different devices. Therefore, CSS pixels are generally used for web pages to define uniform size irrespective of the hardware pixel resolution. We considered these characteristics as an added advantage to ensure the uniformity of pixels across all devices on which data is being captured. In addition, the coordinates are relative to the upper-left edge of the content area of the browser and do not change even if the user is scrolling. This was used as a measure to distinguish between a mouse movement and scrolling.

The captured information is transmitted directly afterward to the main Java program via *XMLHttpRequests* to a local HTTP Server running on *localhost:8080/ EventListener*. Given that the security model of a web browser (known as *same-origin-policy),* prevents the feasibility of sending web requests from one location to another outside the same domain, a Cross-Domain request with Cross-Origin Resource Sharing (CORS) [28], was implemented. The same-origin-policy of a web application is a security mechanism. This mechanism states that inter-access data is permitted from one web page to another if and only if both web pages have the same origin constrained by the same uniform resource identifier (URI) scheme, hostname, and port number. One of the downsides of bypassing this mechanism is server-flooding, a situation in which the server has no control over which packet to receive [29]. To prevent server flooding, the study implemented a threshold for the movement of the cursor. Whenever the *EventListener* for the Mouse action is triggered, it calculates the distance between the former (position of the last data transmission request) and the new position of the mouse cursor. The data transmission request is considered acceptable if the distance of the mouse cursor is greater than the pre-defined threshold of 10-CSS Pixels, otherwise, it is rejected.

### 3.2 Data Pre-Processing and Feature Extraction.

The raw data dumped from the web browser is parsed through a preprocessing module as shown in phase 2 of Figure 1. Feature extraction is based on the individual mouse path. A path is defined as a sequence of mouse events delineated by a time delay threshold, and/or any two consecutive mouse event clicks without the delimited threshold. A new path always starts from the last event of the preceding path, as shown in Figure 2. A time delay threshold is defined as the idle time that satisfies the condition confined by equation 1.

$$Path \stackrel{\text{def}}{=} \begin{cases} Delay \begin{cases} \min \geq 3\ seconds \\ \max \leq 10\ seconds \end{cases} \\ 2\ consecutive\ clicks \end{cases} \quad (1)$$

Based on the mouse events, four different types of path attributes can be extracted as shown in **Table 1**. These attributes are consistent with features in existing studies [1], [3], [30]. A mouse click ends a current path because a click symbolize a new intention of the user (e.g. clicking on a link to open a new page). Furthermore, a movement delay (silent time) of more than 10 seconds between two points is interpreted as a new user intention, and consequently, starts a new path. Preliminary observation of the mouse movement showed that two consecutive mouse movement have delays ≥10-seconds. Existing studies considered aggregation of mouse sequence, which neither indicates a path as the fundamental unit of mouse measurement nor defined the delay between mouse actions. For instance, the exposition in [3] defined the minimum, average and maximum mouse operation task as 6.2s, 11.8s, and 21.3s respectively. This does not show the actual delay between the mouse operations. However, it is logical to consider

a fundamental unit of mouse movement measurement, through which pattern observation can be measured. A mouse movement path presents such an intuition.

**Table 1.** Labels of Path

| Number | Actions of a path begin | Actions of a path end | Label |
|---|---|---|---|
| 1 | Click | Click | cc |
| 2 | Click | Movement (>10sec) | cm |
| 4 | Movement | Click | mc |
| 5 | Movement | Movement (>10sec) | mm |

A path is stored as a trajectory which contains several sequences. It includes the x-coordinate, y-coordinate, timestamp, angle of inclination, speed, mouse-click-up and mouse-click-down events, HTML object, weight, silent time, scrolled Pixel as well as the time delay between mouse-click-down and mouse-click-up events [1], [3], [31]. Furthermore, it contains the overall delay, direct distance between the start- and endpoint, a distance of the path (length), average speed, overall weight, overall direction, label, and URL. Description of the relevant features and human behavioral attributes adapted in this study are explained in more detail in the proceeding subsections as.

### 3.2.1 Speed

The speed of mouse movement is computed for every distance between two points of movement, as well as for the scrolled pixel. For the average speed, the study excludes the scrolling points, to separate the movement speed. The speed for the $i^{th}$ mouse-point is described by Equation 2. The intuition upon which speed is computed is based on existing studies [1], [3]. The average speed for the $i^{th}$ mouse path with $n$-points is defined with the expression presented in Equation 3, where $x$ and $y$ represents the coordinates, and $t$ represents the timestamp at that coordinate.

$$\Delta v_i = \frac{\sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}}{t_i - t_{i-1}} \tag{2}$$

$$\Delta v_{i\ average} = \frac{1}{n}\sum_{k=2}^{n} \Delta v_k \tag{3}$$

### 3.2.2 Distance or Path Length

This study considered the shortest distance between two points, based on the general definition of slope (Euclidean distance). This can also be referred as the direct distance between two points. The shortest distance between two points ($i_{th}$ and $i_{th\text{-}1}$) in a path is given by the expression presented in Equation 4.

$$\Delta d_{i\ direct} = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2} \tag{4}$$

Logically, a mouse path length can be defined as the summary of the distance between all points in the path: The length of the $i^{th}$ path with $n$-points is defined by the expression in Equation 5. This expression considers the transition from the point of path beginning to the point of path ending. This thus implies that the path length is a vector quantity. A path direction is considered with respect to the expression is Equation 8.

$$\overrightarrow{\Delta d_{\iota\, path}} = \sum_{k=2}^{n} \Delta d_{k\, direct} \quad (5)$$

### 3.2.3 Time Delay / Silent Time / Click delay

The delay is calculated for every point in the path. The silent time is the number of milliseconds within which the cursor was not moved. It captures the duration of all connected scrolling events. Furthermore, the click delay (flight) is computed for every click (start and end). This measures the time (*t*) in milliseconds between the mouse down and mouse up event. The delay at the $i^{\text{th}}$ point between time $t_i$ and $t_{i-1}$ is depicted by the expression shown in Equation 6.

$$\Delta t_i = t_i - t_{i-1} \quad (6)$$

### 3.2.4 Angle of Inclination

The angle of inclination (the arctangent of the slope between two points) is calculated for every distance between two points. It is the angle between the horizontal axes of two the points, with the *x*-axis, measured in a counterclockwise direction from $0° \leq \theta < 180°$ . It is defined by the expression in Equation in 7.

$$\Delta\theta_i = \tan^{-1} \frac{\Delta y_i}{\Delta x_i} \quad (7)$$

### 3.2.5 Direction and Weight

The direction of a mouse is calculated for the whole path as well as for the distance between two points. To compute this direction, the angle is logically assumed to have a right and a left quadrant. A left direction covers the negative left axis of a quadrant, while the converse is the right. For the direction, a path which ends with the Left = -1; Right = 1; Neutral=0 is defined by Equation 8, where $x$ = the x-coordinate, $n$ = start point, k = end.

$$\Delta direction_i = \begin{cases} -1, & x_{\text{k}} < x_n \\ 1, & x_{\text{k}} > x_n \\ 0, & x_k = x_n \end{cases} \quad (8)$$

The weight of a path is calculated, from the intuition in kinematics [30], for every distance between two points as well as for the entire path length. The weight for the $i^{\text{th}}$ path is defined by the expression in Equation 9. The overall weight for the $i^{\text{th}}$ path with *n* points is defined by the expression in Equation 10.

$$\Delta w_{i\, point} = \Delta v_i * \begin{cases} \Delta d_{i\, direct} * \sin(\theta_i), & direction = 1 \\ \Delta d_{i\, direct} * \cos(360 - \theta_i), & direction = -1 \\ \Delta d_{i\, direct}, & direction = 0 \end{cases} \quad (9)$$

$$\Delta w_{i\, path} = \sum_{k=2}^{n} \Delta w_{k\, point} \quad (10)$$

### 3.2.7 Skewness and Kurtosis

Higher order moments as defined in [32], are statistical properties that can provide representative properties of a distribution. Skewness and kurtosis are described in this section. However, first, and second order of moment are computed using the generalized expression. Skewness is calculated for the silent time, angle of inclination and speed of a path. The skewness of the $i^{\text{th}}$ path is defined by Equation 11 where $n$ = count of values, $\bar{x}$ = mean and $x_i$ = the $i^{\text{th}}$ value.

$$skew_i = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{\left(\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)^3} \quad (11)$$

Similarly, the kurtosis is computed for the silent time, angle of inclination and speed of a path. The kurtosis of the $i^{\text{th}}$ path is represented by the expression in Equation 12: Where $n$ = count of values, $\bar{x}$ = mean and $x_i$ = the $i^{\text{th}}$ value.

$$kurt_i = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2} - 3 \quad (12)$$

A total of 37 unique features for every path, as shown in **Table 2**, are generated. In order to provide the desired systematic visualization process (as shown in Figure 2), the raw data of the mouse action is segmented through preprocessing, into different files. The data from these files are then used for subsequent data analysis processes. The raw data includes the captured page resolution, user-agent information, x-coordinates, y-coordinates, HTML objects, mouse events, time stamps, click delay (flight) as well as the URL. Summary of the overall features used in this study is presented in **Table 2**.

**Table 2.** Human Behaviour Attributes

| Number | Features / Human Behavior Attributes |
|---|---|
| F1 | Number of the Path |
| F2 | Duration of the Path |
| F3 | Number of Points in the Path |
| F4-F6 | Properties of Scrolls (number of scrolls, scroll up and scroll down) |
| F7 | Number of Clicks in the Path |
| F8 | Number of Movement in the Path |
| F9-F16 | Statistics of Silent Points (number of silent periods, Mean, Std. Deviation, Min, Max, Variance, Skewness, and Kurtosis) |
| F17 | Flight of first Click (duration between. MouseDown and MouseUp) |
| F18 | Flight of the last Click |
| F19 | Length of Path |
| F20 | Overall Weight of the Path |
| F21 | Direction of the Path |
| F22-F29 | Statistics of Angle of Inclination (Mean, Std. Deviation, Min, Max, Variance, Skewness, Kurtosis, and Mode) |
| F30-F37 | Statistics of speed movement (Mean, Std. Deviation, Min, Max, Variance, Skewness, Kurtosis, and Mode) |

### 3.3 Visualization

To achieve the third design goal, it is necessary to read the stored data and visualize them. The developed tool accepts an input from a CSV file. The features and attributes of the corresponding file are loaded into the tool for visualization. The GUI offers the option to display an overview of the whole mouse action capture, as shown in Figure 2. Furthermore, it is possible to choose a path directly in the drop-down menu to see the corresponding details. When a new path is selected and added, it creates a new internal frame in the main window frame. These internal windows are adjustable, resizable as well as closable, as shown in Figure 3. From this, a comparison can be made among any number of paths. The layout of the internal path window in Figure 3 is as followed. On the bottom right side is a zoom-able area where the selected path is drawn from the recorded data. It is possible to zoom in and out on every drawn path, by scrolling the mouse wheel, to magnify or minimize individual points. For scroll events, it displays

the scrolled number of pixels. Furthermore, on the bottom of the window, the GUI provides a table which displays the overview of the data. A tabular display of the individual features and values of each point in the path is also provided. The drawings of the paths are scaled based on the size of the window. The stored page resolution of the browser during the recording is also provided. This visualization process can be instrumental in the reconstruction of user-event which can be used to observe user activity. This can be particularly useful in tracing the action of a user in the event of insider misuse and investigation. In addition to the probability of attributing a user, the inclusion of the visualization process can be used to trace the exact path, within a specified period of an event.

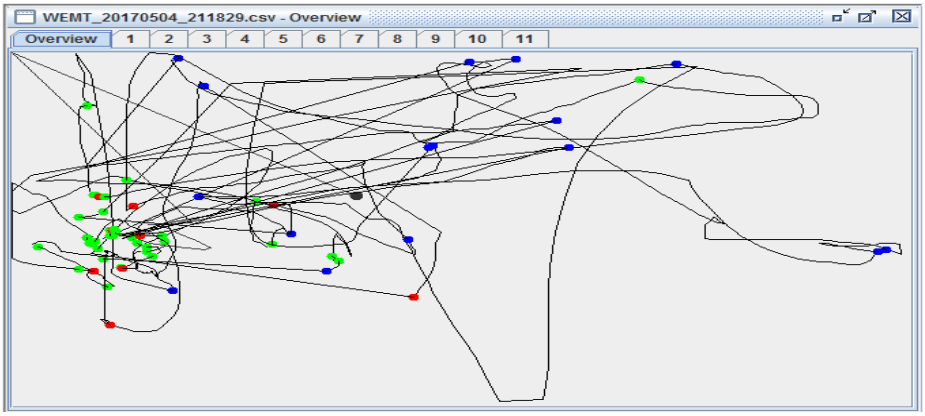
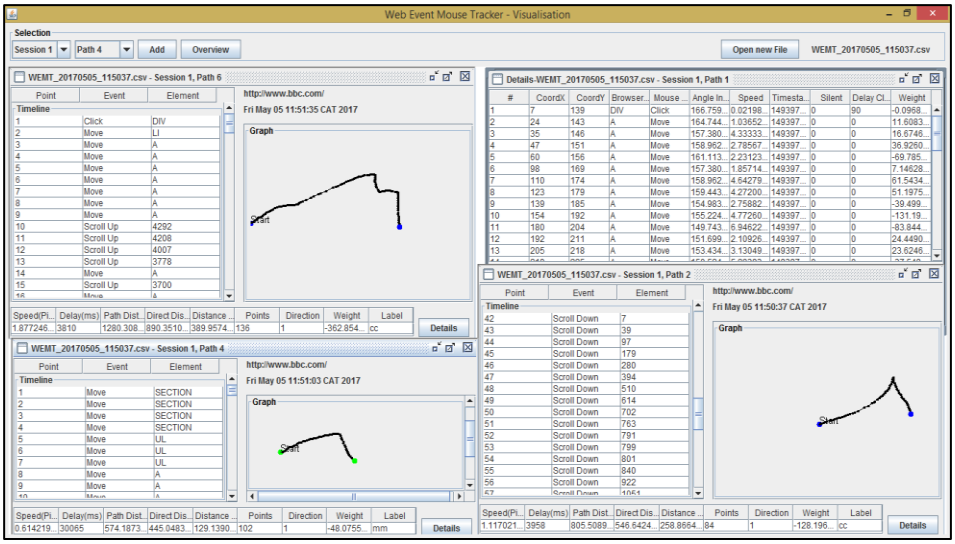



**Figure 2**. Overview of Visualization of all paths in a recording. Every colored dot displays a start and end point of a path. The bigger black dot (in the center) displays the start point of the whole recording. The blue points are displaying a click, red is for an expired session and green is for a movement.

**Figure 3.** Working space of the forensic-visualization tool with four open windows to compare movement paths of a user. Three (left and bottom right) displaying the overall path information




and one (right top) displays a detailed table of the first path.

### 3.4 Experimental Set-up

To further validate the feasibility of the developed tool, an experimental process was set up in a computer laboratory. Eleven volunteers were recruited for this purpose. The laboratory comprises numerous workstations, each with the same configuration of hardware, software, and each operates a Deep Freeze enterprise software, which restores the workstation to a pristine state, upon workstation reboots. The forensic-tool developed for this research was installed on the workstation for three consecutive days. Initial evaluation of the capability of forensic-tool was assessed. Users were monitored for action taken and the resultant output from the forensic-tool was evaluated. The result showed consistency between the observation and the action of the users. Three users participated in the lab section, while the other eight users installed the forensic-tool on their personal computers. Each user was asked to freely surf the web using either a Chrome or Mozilla Firefox browser, based on a given list of news websites. The tool works on all operating systems. Free web surfing was encouraged so as to mimic, as nearly as possible, a real life browsing behavior. This is in contrast to a fully controlled experimental environment. A controlled environment is asserted to prevent the influence of extraneous variables. The notion of the introduction of extraneous variables, as suggested by [3], is deemed non-practicable in the behavioral analysis of human action. In practice, human actions are generally guided by self-interest and discretion which cannot be limited to a controlled environment. Using the behavioral features defined in **Table 2**, feature extraction was performed on the dataset from all users, followed by a pattern observation process. Summary of the data description is presented in **Table 3**. For each path, the feature summarized in **Table 2** were extracted to generate individual datasets.

**Table 3**: Summary of the data

| Users | Duration (Minutes) | Number of instance (path) |
|---|---|---|
| 1 | 30 | 31 |
| 2 | 90 | 158 |
| 3 | 60 | 67 |
| 4 | 120 | 173 |
| 5 | 90 | 147 |
| 6 | 150 | 407 |
| 7 | 150 | 434 |
| 8 | 150 | 259 |
| 9 | 250 | 309 |
| 10 | 150 | 321 |
| 11 | 90 | 977 |

The pattern identification mechanism for a user attribution process was carried out in two phases. In the first phase, pattern consistency; intra-user pattern consistencies, was observed for all users (excluding user-1, who had only one session of data) based on the daily activity, using a non-supervised machine learning method: the X-Means (an extension of K-means) clustering algorithm. In order to perform the cluster analysis, feature selection was carried out on the 37-features defined in **Table 2**. Based on this dimension reduction process, 8-base-features were observed to provide a significant discriminatory factor for the intra-user analysis. These include the duration, number of points, flight, length, and weight of path. Thereafter, inter-user variation (through dissimilarity in the pattern) observation based on a supervised classification process was carried out on the three laboratory users (hereinafter referred to as the Tier-2

dataset, using each user as the class). The study explored a nonlinear support vector machine (LibSVM), artificial neural network, C4.5 decision tree and RandomForest classifiers. These classifiers were selected due to their general usage in mouse dynamics studies [3], [18]. This exploration was carried out in the entire feature space as described in **Table 2**, with the assertion that such process can harness the semantic relation in the data, in addition to its potential to harness the syntactic relation in the data.

## 4. Results and Analysis

On the evaluation process, an optimization process was performed on the clustering algorithm. A Manhattan-distance (also referred to as the taxicab geometry) was used for the cluster optimization. X-means successfully generated a single cluster (0% error rate) for one user (user 3: U-3), while other users show significant intra-cluster similarity as shown in **Table 4**. The table shows the total number of observed cluster for each user. The number of the class represents the different dataset for each user.

**Table 4**: Result of Intra-user Similarity

| **User** | U-2 | U-3 | U-4 | U-5 | U-6 | U-7 | U-8 | U-9 | U-10 | U-11 |
|---|---|---|---|---|---|---|---|---|---|---|
| **No. of class** | 3 | 2 | 4 | 3 | 5 | 5 | 5 | 7 | 5 | 3 |
| **No. of observed cluster** | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| **Cluster Similarity (%)** | 87 | 100 | 83 | 83 | 87 | 83 | 80 | 81 | 93 | 89 |

In order to study the inter-user dissimilarity, the supervised classifiers were applied to the laboratory users. The choice of the laboratory users was based on the uniform experimental condition across device and operating condition. RandomForest was subsequently observed to perform relatively better than the other explored classifiers on the Tier-2 dataset, extracted from the laboratory users. The true acceptance rate for the users are 0.935, 0.938 and 0.439 for users U-1, U-2 and U-3 respectively. Conversely, the false acceptance rate of 0.034, 0.361, and 0.035 was obtained for Users U-1, U-2 and U-3 respectively. Based on class distribution, the highest class prior probability of 53.81% and an average accuracy of 78.1% was obtained. The analysis was carried out on a 10-fold cross-validation process. The obtained result of the classification process falls below the European standard for commercial biometrics. However, this result shows a promising technique through which user attribution can be established.

## 5. Discussion

The result from the experimental process shows that the forensic-tool was able to capture every mouse action of each user. Furthermore, the visual representation shown in Figure 3, presents a very flexible process of visualizing the mouse activity of a user. A graphical plot of the features can also be carried out on the user-interface of the tool. These characteristics further extend the tool in examining individual difference and similarity, at a higher abstraction. On a lower abstraction, the tools support the preprocessing and generation of mouse dynamics features. The features considered in this study attempt to expand the repository of mouse dynamics attributes. More specifically, the specific features considered include the path characteristics, flight duration, and the overall weight of the path.

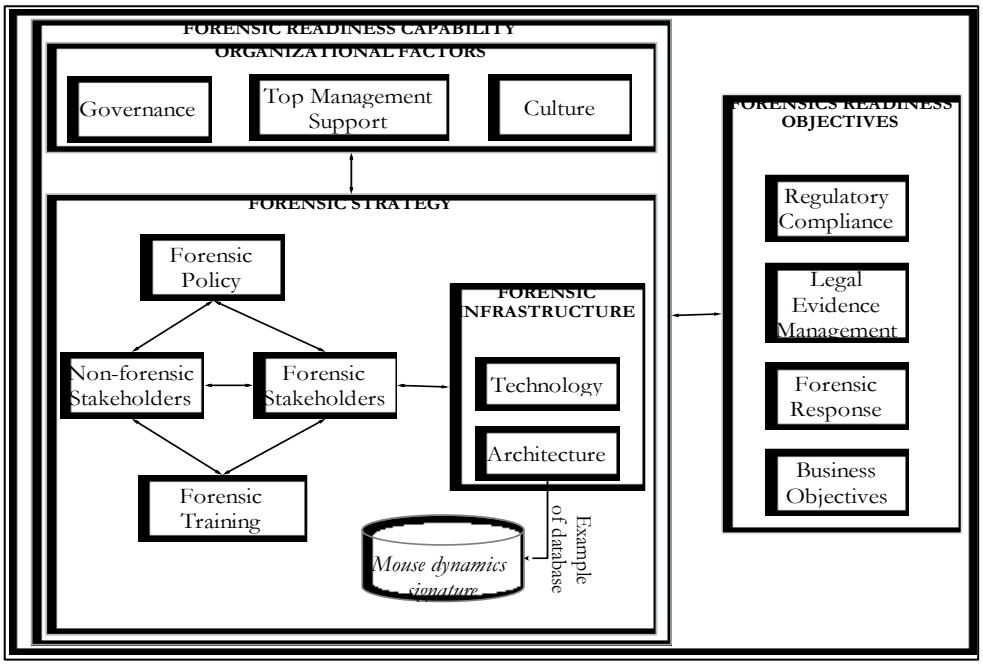



Figure 4: Digital Forensic Readiness framework (adapted from [12])

These features were observed to significantly influence the observed accuracy of the classifiers. In terms of the behavioral characteristic feature, which can be adapted for user attribution, the path characteristics present a measurable and reliable feature. The result from the unsupervised learning process shows a very high probability of the existence of a unique behavioral signature for each user. Such signature could represent the principal component needed for user attribution based on mouse dynamics. The result of the unsupervised learning approach also debunks the assertion that an uncontrolled experimental environment is not suitable for user authentication research based on mouse dynamics. Based on the empirical assertion and fundamental assumption on variables that could induce experimental bias on mouse dynamics study, the current study heeded several recommendations from [18] on the extraneous variables that could influence mouse behavior.

This includes the type of mouse device, screen resolution, acceleration setting of a computer system, the perpetual delay caused by the load on the CPU, and properties of the surface area on which the mouse is placed. The psychological state of the user was not considered in this study. However, the users were not subjected to any experimental pressure. In addition, the study assumed that the list of the website used in this study will not inject any negative psychological episode on the respondents. To prevent data loss due to encryption protocols, the experimental websites considered in this study were all HTTP-based websites. This is because the HTTPS does not work with the developed JavaScript of the forensic-tool.

The application of the findings of this study in a digital forensic readiness framework falls within the architecture sub-module of the forensic infrastructure in Figure 4, as asserted in [12]. A mouse dynamics signature database was introduced as an addition to the initial framework as shown in Figure 4. The integration of the mouse dynamics signature database into the framework will complement other existing forensic architectures. This could include the installation of the forensic-tool on the existing hardware of an organization. The preparation of such contingency policy remains a viable complementary process to a postmortem forensic mechanism.

### 5.1 Limitation and Future Works

Given that the baseline for FAR and FRR are 0.001 and 1.00 respectively [3], it is obvious that the obtained accuracy based on the Tier-2 dataset is relatively low. This can be attributed to the relatively smaller sample size of respondents, shorter experimental duration, and smaller number of experimental sections. In terms of features, the study could integrate discriminative features such as double click, drag & drop, event thresholding, and other probable behavioral attributes. Considering that the HTTPS-based website is gaining wider adoption in typical client-server communication, the non-inclusion of an HTTPS server to capture a secure-web-page-response is one of the major limitations of this study. In defining the path delimiter, the study utilized a 10-seconds threshold. An adaptive threshold could be developed in future works. In terms of the development of behavioral signature and the eventual development of an updateable database for DFR, future works will explore modalities towards the extraction of unique behavioral fingerprints based on mouse action which can be adapted for user attribution. A reliable user attribution model will be considered in future works. Models that aim to establish a reliable mechanism for a user identification process is a critical component in this area of forensic analysis.

## 6. CONCLUSION

On a general note, mouse dynamics satisfy the underlying characteristics – reasonably permanent, easy to collect and easy to measure– of biometric modalities for user identification. Studies on biometric verification, whether on physiological or behavioral topics, require sufficient sample sizes for the effective evaluation of their parameters and of their performance. The tool developed in this study presents a step towards the actualization of the goal of establishing mouse dynamics research for user identification. This, in turn, will create a platform for an effective user-attribution process in the digital forensic analysis. The findings presented in this manuscript are part of an ongoing research which aims to provide a reliable model for the user attribution process based on mouse dynamics.

## References


[1] K. O. Bailey, J. S. Okolica, and G. L. Peterson, “User identification and authentication using multi-modal behavioral biometrics,” *Comput. Secur.*, vol. 43, pp. 77–89, 2014.

[2] D. Chudá, P. Krátky, and J. Tvarožek, “Mouse Clicks Can Recognize Web Page Visitors!,” *Proc. 24th Int. Conf. World Wide Web*, pp. 21–22, 2015.

[3] C. Shen, Z. Cai, X. Guan, Y. Du, and R. A. Maxion, “User authentication through mouse dynamics,” *IEEE Trans. Inf. Forensics Secur.*, vol. 8, no. 1, pp. 16–30, 2013.

[4] P. Kasprowski and K. Harezlak, “Fusion of eye movement and mouse dynamics for reliable behavioral biometrics,” *Pattern Anal. Appl.*, 2016.

[5] A. A. Khalifa, M. A. Hassan, T. A. Khalid, and H. Hamdoun, “Comparison between mixed binary classification and voting technique for active user authentication using mouse dynamics,” *Proc. - 2015 Int. Conf. Comput. Control. Networking, Electron. Embed. Syst. Eng. ICCNEEE 2015*, pp. 281–286, 2016.

[6] I. Traore, I. Woungang, M. S. Obaidat, Y. Nakkabi, and I. Lai, “Combining mouse and keystroke dynamics biometrics for risk-based authentication in web environments,” *Proc. - 4th Int. Conf. Digit. Home, ICDH 2012*, pp. 138–145, 2012.

[7] I. Traore, I. Woungang, M. S. Obaidat, Y. Nakkabi, and I. Lai, “Online risk-based authentication using behavioral biometrics,” *Multimed. Tools Appl.*, vol. 71, no. 2, pp.

575–605, 2014.
[8] C. Bevan and D. S. Fraser, “Different strokes for different folks? Revealing the physical characteristics of smartphone users from their swipe gestures,” *Int. J. Hum. Comput. Stud.*, vol. 88, pp. 51–61, 2016.
[9] A. Alzubaidi and J. Kalita, “Authentication of smartphone users using behavioral biometrics,” *IEEE Commun. Surv. Tutorials*, vol. 18, no. 3, pp. 1998–2026, 2016.
[10] M. S. Olivier, “On metadata context in Database Forensics,” *Digit. Investig.*, vol. 5, no. 3–4, pp. 115–123, 2009.
[11] I. R. Adeyemi, S. A. Razak, and N. A. N. Azhan, “A Review of Current Research in Network Forensic Analysis,” *Int. J. Digit. Crime Forensics*, vol. 5, no. 1, pp. 1–26, 2013.
[12] M. Elyas, A. Ahmad, S. B. Maynard, and A. Lonie, “Digital forensic readiness: Expert perspectives on a theoretical framework,” *Comput. Secur.*, vol. 52, pp. 70–89, 2015.
[13] A. Valjarevic and H. S. Venter, “Towards a Digital Forensic Readiness Framework for Public Key Infrastructure systems,” *2011 Inf. Secur. South Africa*, pp. 1–10, 2011.
[14] F. Anjomshoa, M. Aloqaily, B. Kantarci, M. Erol-Kantarci, and S. Schuckers, “Social Behaviometrics for Personalized Devices in the Internet of Things Era,” *IEEE Access*, pp. 1–1, 2017.
[15] A. Alsultan and K. Warwick, “Keystroke Dynamics Authentication: A Survey of Free-text Methods,” *Int. J. Comput. Sci.*, vol. 10, no. 4, pp. 1–10, 2013.
[16] P. H. Pisani and A. C. Lorena, “A systematic review on keystroke dynamics,” *J. Brazilian Comput. Soc.*, vol. 19, no. 4, pp. 573–587, Jul. 2013.
[17] H. Saevanee, N. Clarke, S. Furnell, and V. Biscione, “Continuous user authentication using multi-modal biometrics,” vol. 53, pp. 234–246, 2015.
[18] Z. Jorgensen and T. Yu, “On mouse dynamics as a behavioral biometric for authentication,” in *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security - ASIACCS ’11*, 2011, pp. 476–482.
[19] H. Gamboa and a Fred, “A behavioural biometric system based on human computer interaction,” *Proc. SPIE - Int. Soc. Opt. Eng.*, vol. 5404, no. i, pp. 381–392, 2004.
[20] M. Pusara and C. E. Brodley, “User re-authentication via mouse movements,” *Proc. 2004 ACM Work. Vis. data Min. Comput. Secur. VizSECDMSEC 04*, pp. 1–8, 2004.
[21] H. Gamboa, A. L. N. Fred, and A. K. Jain, “Webbiometrics: User verification via web interaction,” *2007 Biometrics Symp. BSYM*, 2007.
[22] A. A. E. Ahmed and I. Traore, “A New Biometric Technology Based on Mouse Dynamics,” *IEEE Trans. Dependable Secur. Comput.*, vol. 4, no. 3, pp. 165–179, 2007.
[23] M. Pusara and C. E. Brodley, “User re-authentication via mouse movements,” *Proc. 2004 ACM Work. Vis. data Min. Comput. Secur. VizSECDMSEC 04*, pp. 1–8, 2004.
[24] D. Barske, A. Stander, and J. Jordaan, “A digital forensic readiness framework for South African SME’s,” *Proc. 2010 Inf. Secur. South Africa Conf. ISSA 2010*, 2010.
[25] Bell Global Technologies, “RunJS - Run Javascript on Page Load.” .
[26] R. Noe, “Execute JS.” .
[27] U. Sedlar, J. Bešter, and A. Kos, “Tracking mouse movements for monitoring users’ interaction with websites: Implementation and applications,” *Elektrotehniski Vestnik/Electrotechnical Review*, vol. 74, no. 1–2. pp. 31–36, 2007.
[28] “HTTP access control (CORS).” .
[29] K. Lakshminarayanan, D. Adkins, A. Perrig, and I. Stoica, “Taming IP Packet Flooding Attacks,” *ACM SIGCOMM Comput. Commun. Rev.*, vol. 34, no. 1, pp. 45–50, 2004.
[30] D. Mart??n-Albo, L. A. Leiva, J. Huang, and R. Plamondon, “Strokes of insight: User intent detection and kinematic compression of mouse cursor trails,” *Inf. Process. Manag.*, vol. 52, pp. 989–1003, 2015.
[31] Z. Jorgensen and T. Yu, “On mouse dynamics as a behavioral biometric for authentication,” in *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security - ASIACCS ’11*, 2011, pp. 476–482.
[32] I. R. Adeyemi, A. S. Razak, and M. Salleh, “A psychographic framework for online user identification,” in *International Symposium on Biometrics and Security Technologies (ISBAST)*, 2014, pp. 198–203.