

Leveraging Human Thinking Style for User Attribution in Digital Forensic Process

Ikuesan Richard Adeyemi^{#*}, Shukor Abd Razak[#], Mazleena Salleh[#], Hein S. Venter^{*}

[#]*Department of Computer Science, Universiti Teknologi Malaysia, Skudai, Malaysia*
E-mail: raikuesan2@live.utm.my, {shukorar, mazleena}@utm.my

^{*}*Department of Computer Science, University of Pretoria, Pretoria, 0082, South Africa*
E-mail: aikuesan@cs.up.ac.za, hventer@cs.up.ac.za

Abstract— User attribution, the process of identifying a human in a digital medium, is a research area that has received significant attention in information security research areas, with a little research focus on digital forensics. This study explored the probability of the existence of a digital fingerprint based on human thinking style, which can be used to identify an online user. To achieve this, the study utilized Server-side web data of 43-respondents were collected for 10-months as well as a self-report thinking style measurement instrument. Cluster dichotomies from five thinking styles were extracted. Supervised machine-learning techniques were then applied to distinguish individuals on each dichotomy. The result showed that thinking styles of individuals on different dichotomies could be reliably distinguished on the Internet using a Meta classifier of Logistic model tree with bagging technique. The study further modelled how the observed signature can be adopted for a digital forensic process, using high-level universal modelling language modelling process- specifically, the behavioural state-model and use-case modelling process. In addition to the application of this result in forensics process, this result finds relevance and application in human-centered graphical user interface design for recommender system as well as in e-commerce services. It also finds application in online profiling processes, especially in e-learning systems.

Keywords— Sternberg thinking style; online digital-signature; user attribution; online user identification; digital forensic process

I. INTRODUCTION

Human-computer interaction as a discipline entails the integration of human factor into computing systems to enhance its usability. Such human-centered approach to system development finds relevance in the development of effective online mediation processes. Myriads of studies have established the necessity as well as the promising potential of human cognitive styles in Internet technology, particularly on digital signatures in relation to human-computer interaction [1]–[6] and digital forensics [7]–[9]. Kozhevnikov (2007) define these cognitive styles as a psychological dimension representing consistencies in an individual manner of cognitive functioning with respect to acquiring and processing of information. The study further highlights cognitive style to be a pattern of adaptation that regulates individual cognitive ability capable of establishing a distinction in individual perception and personality. The habitual manner of task structure and organization, as well as the habitual manner of information representation, are the two dimensions of cognitive styles. Early cognitive researchers have established that individual differs in the

performance of the simple task, which is dependent on the ability to adjust to the environment [10], [11]. A study in [2] posits that web search pattern is one of such task that can reveal cognitive differences in individuals on the Internet. Cognitive styles differ from cognitive ability [12]. The former refers to individuals' preference for usage of cognitive abilities, while the latter refers to individuals' cognitive abilities in itself. The assertion in [13] posits that the Internet has an enormous capacity to facilitate communication which can reveal individual cognitive styles. Measurement of cognitive style as a construct includes learning style, personal style, tolerance ambiguity scale and thinking style. There are several measurement instruments for human thinking styles, amongst which, is the Sternberg Weigner Thinking Style Inventory (SWTSI). The SWTSI is the most explored theory that explains human cognitive style in the form of human thinking style [3], [6], [14]–[19]. SWTSI has been applied to study on online consumer behaviours [20], web search pattern and navigation behaviour [1], [3], [21], [22], online learning and education [18], [23]–[26] and online gaming pattern [27]. These studies have attempted to explain the thinking style of online

users from platform specific perspectives, the relationship between thinking and academic performance, the relationship between human thinking style and online shopping environment as well as the influence of thinking style on consumer buying habit. However, cognitive styles as expressed in thinking style suggest a mechanism that can control human actions, and it is independent of human intellectual capability. For example, [20] observed telepresence (individual's feeling of being in a virtual space; a feeling of existence in the computer generated surrounding where time and space are condensed) phenomenon using thinking style as mediator. Similarly, [3] observed that thinking style could be used to define linear and nonlinear browsing pattern of online users. This example thus suggests that an integral component of thinking style can be used to elucidate on the online behaviour of users. As a step in this direction, this study explores the probability of online signature of users based on their thinking style dichotomy. Thinking style as defined by the Sternberg Weigner's theory of mental self-government (SWTMSG) is considered in this study [28]. As a theory, the reliability, validity, and applicability of SWTMSG have been verified in both academic [29]–[31] and non-academic [17], [32]–[34] settings. This study is an extension of the study presented in [35]. Improvement on the experimental result, the integration of user attribution model for digital forensics, discussion of result and Conclusion of the findings of the improvement are discussed in this manuscript. the underlying logic and the process used to extract the thinking style behavioural signature from the network traffic of the respondents used in this study is presented in the next section.

II. MATERIAL AND METHOD

This section discussed the materials, method, and the process of data collection used for this study.

A. Digital Thinking Style Signature Process

The probability of the existence of a unique pattern of human thinking style on the Internet is based on the premise that humans typically translate their day-to-day characteristic behaviour in physical interactions onto the Internet [22]. The research question based on this assertion is as follows: given a dichotomous continuum of human thinking style and server-side network sessions of known users, can a supervised machine learning technique be modelled to develop a one-to-many identification process with an accuracy significantly greater than the baseline accuracy of the data and area under the receiver operating characteristic curve of ≥ 0.9 . To answer this research question, a mix of thinking style cognitive data and network traffic of online interaction of known users are collected. The proposed perspective is as follows:

- Collect cognitive data of known users using the SWTSI measurement instrument. Dichotomize the cognitive data based on the laypeople percentile of the SWTSI subscale as specified in [28].
- Collect network traffic of identified users in i-above. Identify and extract human-centric features from the

network data. Distribute the network traffic data based on the defined dichotomy in i-above.

- Explore applicable supervised machine learning technique on the distribution derived in ii above. Measure the accuracy and reliability of the classification process based on standardized machine learning metrics.

Based on this perspective, the probability of extracting unique signature of online users through thinking style cognitive data can then be defined. This phenomenon is hereinafter defined as a human psychosocial attribute. The methodology considered in this study is presented in the proceeding section.

B. Thinking Style Measurement Instrument And Participants

The SWTSI measurement instrument has been extensively adopted in the study on human thinking style [26]. Fig. 1 shows the depiction of the conventional subscale of the SWTSI measurement instrument. As shown in Fig. 1, the adapted thinking style continuum has six subscales. To recruit respondents for this study, two basic requirements were established. First, the respondent must be an actively employed staff member of an organization with the capability to effect frequent client-server communication using the Internet service. Second, each respondent must use login identity that is strictly used by only the respondent throughout the duration of the study. Based on these two criteria, the Research Management Centre of a research university in Malaysia was selected. Initial consent forms were distributed to all staff members of the organization. A total of 66-respondents returned a completed consent form indicating their readiness to participate in the study, amongst which a total of 55-respondents submitted a completely filled thinking style inventory measurement instrument. 43-respondents satisfied the criteria for inclusion in this study. Thinking style inventory, as developed by Sternberg [28], comprises 13-factors and 104-questionnaire items. Summary of the factors and its description is presented in Table 1.

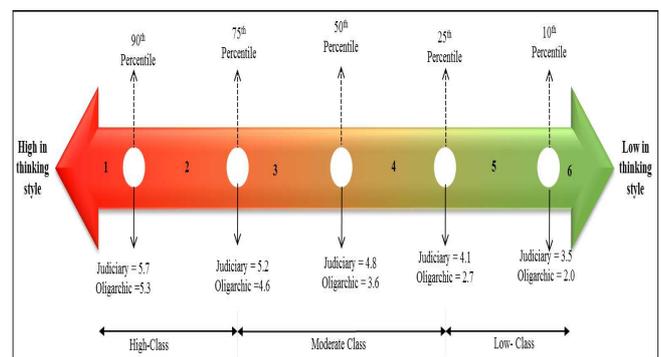


Fig. 1 Class creation for observed thinking styles

TABLE I
SUMMARY OF THINKING STYLE INVENTORY

Thinking style	Description
Legislative	One prefers to work on tasks that require creative strategies; one prefers to choose one's own activities.
Executive	One prefers to work on tasks with clear instructions and structures; one prefers to implement tasks with set guidelines
Judicial	One prefers to work on tasks that allow for one's evaluation; one prefers to evaluate and judge the performance of other people
Hierarchical	One prefers to distribute attention to several tasks that are prioritized according to one's valuing of the tasks.
Monarchical	One prefers to work on tasks that allow complete focus on one thing at a time.
Oligarchic	One prefers to work on multiple tasks in the service of multiple objectives, without setting priorities.
Anarchic	One prefers to work on tasks that would allow flexibility as to what, where, when and how one works
Global	One prefers to pay more attention to the overall picture of an issue and to abstract ideas.
Local	One prefers to work on tasks that require working with concrete details
Internal	One prefers to work on tasks that allow one to work as an independent unit.
External	One prefers to work on tasks that allow for collaborative ventures with other people.
Liberal	One prefers to work on tasks that involve novelty and ambiguity.
Conservative	One prefers to work on tasks that allow one to adhere to the existing rules and procedures in performing tasks

A back-to-back English language to Bahasa Melayu translation was employed for effective comprehension of each item in the Thinking style inventory. Given the relatively smaller number of respondents who completed the questionnaire items, the study adopted the dichotomy defined in Fig. 1. The dichotomy comprises three classes (High, Moderate, and Low). Each class corresponds to the combination of two subscales as depicted in Fig. 1. Five factors (External, Hierarchical, Legislative, Judiciary and Oligarchic thinking styles) were extracted from the distribution of the response from the respondents. The other factors were either skewed towards High-class, Moderate-class or Low-class. Due to the limitation in a sample size, this study considered only the Low and High class of the thinking style continuum. The process of extracting and pre-processing the network traffic of individual in each High and Low category is presented in the next section.

C. Network Features

Server-side network traffic of the 43-respondents was collected from April 2014 to December 2014. A heuristic methodology was developed to clean the raw log file of the requested URL and to extract relevant human-centric features. The heuristic considers web requests that originate as a result of human action, in contrast to requests initiated by system or network facility on behalf of the individual. The heuristic was then applied to individual requests, and the following human-centric features were extracted based on a 30-minutes session boundary. 30-minutes session boundary is the generally accepted session duration [36], [37]. Two categories of network traffic characteristics feature space are extracted. The first category is a unidimensional time series, which is adapted for extracting online vocabulary signature of each dichotomy. The second category comprises characteristics features, such as web request pattern, web page visitation pattern as well as session characteristics. The second category data is used to observe the probability of

distinction among all observed dichotomies. This is discussed in the subsequent section.

1) *Web Request Characteristics*: Individual request pattern is observed through the inter-request characteristics observed in each session. Inter-request time is the time difference between two consecutive requests within a session. Statistical properties of web request characteristics as defined in [38], which include mean, standard deviation, variance, kurtosis, and skewness of individual web request were extracted from each session. A total of 10 human-centric features were extracted, thus constituting the web request characteristics. The inter-request (also referred to as flight time) pattern of each user is processed as a unidimensional time series, which can be described as follows: suppose the time for initial request is t_i and the time elapsed between when a respondent submits a request and when it gets to the server is given as δt , the flight time between the requests is defined by Equation 1.

$$F_i \equiv t_{i+1} - t_i \quad (1)$$

2) *Visitation Pattern*: The University Centre operates two Servers, load-balancing client-server communication architecture. This implies that the possible number of the probable web page is bounded by the total web pages in the two servers as represented in Equation 2.

$$URL_{Total} = \sum_{i=1}^s \left[\sum_{j=1}^N URL_j \right] \quad (2)$$

where s = a total number of server, N = number of unique URL on each server.

This study assumes that individual web request pattern obeys power law as asserted in [39], [40]. The visit-characteristics considered in this study include aggregation of visit within a session, the rate of revisit per session and

session length with respect to visit aggregation as presented in Equations 3, 4, and 5 respectively.

The notion of rate of visit is in conformity with Equation 2, given the logic that the probable URLs that can be visited by an individual are limited to the observable URLs on the server.

$$V_{s,j} = \frac{\sum_{i=1}^n (\text{URL per session})_i}{\sum_j (\text{URL under observation})} \quad (3)$$

$$R_w = \frac{\sum_{i=1}^n (\text{URL per session})_i}{S_d},$$

$$S_d = \text{session duration} = \int_{t_{i-1}}^t t_i dt, \leq 30 \text{ minutes} \quad (4)$$

$$S_{agg} = \frac{\sum_j (\text{URL under observation})_j}{S_d} \quad (5)$$

In addition, this presupposes that the interest-driven model and priority queue model of probable request pattern [40] are captured in the bounded URL distribution such that all observed users share similar working conditions, and the major observable distinction can be revealed through observation of human behavioural composition. Three features were derived from the visitation pattern. In addition, session duration and a total number of requests per session were also derived. A total of 15-features were extracted from the network traffic dataset. The extracted features, 15-features in this case, formed the attribute of the data, each extracted session characteristics formed the instances while the High and Low dichotomies, formed the classes on which supervised machine learning classification process was performed. The next section discussed the process and the experimental procedure involved in the application of the supervised machine learning techniques.

3) *Classifier Exploration*: In order to observe the distinction among the observed dichotomies, two supervised machine-learning algorithms were explored. The choice of the two selected classifiers was based on an initial exploration of applicable classifiers on the extracted features. The two classifiers include a logistic model tree (LMT) and J48 decision tree. Discussion on these classification algorithms can be found in [41]–[43]. Ensemble classifiers have been acclaimed to proffer higher classification accuracy (highly improved predictive performance) than classical classifier [44], [45]. An ensemble classifier is developed based on different instantiation of same classifier or different classifiers. This study further explores the probability of accuracy optimization using different instantiation on LMT and J48 classifiers. The process adapted for classification exploration in this study is similar to the defined process in [42] as presented in Fig. 2. The exploratory process involves the search of applicable classifiers capable of establishing discriminative boundaries among classes in the dataset based on the informative structure of the feature sets. The process starts with the pre-processing of the data. The pre-processing involves data cleaning, the extracting sequence of request and sessionization of the request based on an adapted session threshold. The next stage involves splitting the dataset into training and testing samples. The classifier exploration process then follows this. The default baseline for the exploration process is based on the highest-class probability.

ZeroR algorithm in WEKA[®] workbench satisfied this condition. WEKA (Waikato Environment for Knowledge Analysis) software version 3.8 was adopted for the classifier exploration process in this study. WEKA software is a java-based open source software which has gained wider adoption in pattern classification and machine learning process due to its robustness [43] and its capacity to support within script automation [46]. The experimental process is based on the accuracy obtained using 10-fold cross validation and 10-iteration process to prevent overfitting. Standardized performance evaluation measures are adopted to evaluate the overall performance of each classifier. These include accuracy, Recall, precision, root mean square error (RMSE), Kappa statistics, F-measure and Area Under the receiver operating characteristic Curve (AUC).

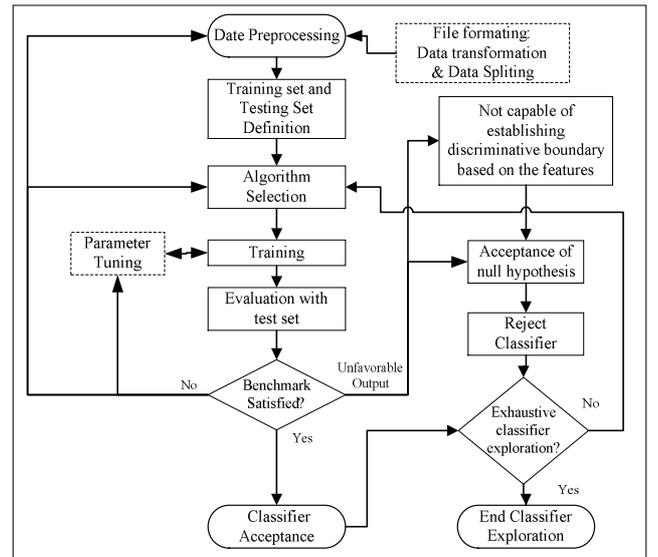


Fig. 2 Procedure for Classifier Exploration

III. RESULTS AND DISCUSSION

A. Experimental Result

The three classes defined in the continuum presented Fig. 1 was applied to the responses from the respondents. Summary of the analysis of the extracted thinking style factors is presented Table 2. The result of the standardized estimate factor loading expresses the reliability of the five factors extracted.

TABLE II
SUMMARY OF RELIABILITY OF MEASUREMENT INSTRUMENT

Thinking style Factor	Variance Extracted for 1-factor	KMO Sampling adequacy	Standardized estimate Factor Loading
Judiciary	57.20	0.82	≥0.55 for all items, excluding item-6
Oligarchic	53.11	0.82	≥0.51 for all items, excluding item-1
Legislative	54.29	0.81	≥0.56 for all items, excluding item-1
Hierarchical	62.28	0.83	≥0.71 for all items
External	63.50	0.88	≥0.73 for all items, excluding item-6 & item-8
Extraction Method: Principal Axis Factoring, 1 factor extracted, 6 iterations required.			

All the observed factors: Judiciary, oligarchic, legislative, hierarchical and external thinking styles reflected high conformity to a single-factor based on the value of the average variance extracted, 57.2, 53.11, 54.29, 62.28 and 63.5 respectively. Based on this reliability, categorization of the network data of each respondent was performed. Summary of the distribution of the network traffic is shown in Table 3.

TABLE III
SUMMARY OF NETWORK DATA DISTRIBUTION

Thinking Style Factors	Class	Respondent	Number of instances (sessions)
Judiciary	High	12	7118
	Low	7	
Oligarchic	High	4	6675
	Low	9	
Legislature	High	4	1209
	Low	5	
Hierarchical	High	6	2148
	Low	3	
External	High	8	3139
	Low	4	

As shown in Table 3, the representative respondents for the High:Low class of Judiciary and Oligarchic factor are 12:7, and 4:9 respectively. Similarly, the respondents for the High:Low classes of the External, Hierarchical and Legislative thinking styles are 8:4, 6:3, and 4:5 respectively. The corresponding number of sessions is also computed for each class as shown in Table 3. Due to the relatively small session size of each class, the experimentation process was carried using a 10-fold cross-validation, on a 10-iteration process. The result of the classification algorithms is presented in Table 4.

The result of J48 and LMT classifiers depict relatively stable consistencies, and reliability on all explored thinking styles. Based on the outcome of the Baseline classifier for each thinking styles, the J48 decision tree with Bagging showed higher accuracy and reliability across all the observed thinking styles relative to the other classifiers. The accuracy of J48 decision tree classifier is significantly higher than the baseline classifier for each thinking style, which confirms the existence of the explored thinking style. This is a slight deviation from the prior study, where LMT with bagging was observed to present a higher accuracy and reliability than other explored classifiers. Similarly, the tendency to AdaBoostM1 algorithm to overfit dataset to a model [44] which was observed in the prior study was not supported in the current study.

The value of the AUC, F-Measure, and Kappa Statistics further lend credence to the reliability of the obtained achieved accuracy. The AUC metric measures the effectiveness of a classification algorithm. Values ≥ 0.8 is generally considered to indicate a reliable discriminative performance. AUC is not biased towards instance distribution and class imbalance of a dataset. The results of the training and validation model of Bagging-LMT classifier demonstrate consistent performance on the Judiciary (0.995 and 0.997 respectively) and Oligarchic (0.999 and 0.958

respectively) thinking style datasets. Similarly, the value of the Kappa statistics affirms the reliability of the Bagging-LMT classifier. Values of Kappa statistics ≥ 0.7 are generally accepted to express the effectiveness of a classifier.

As shown in table 4, the result indicates reliable Kappa statistic value. Furthermore, the F-measure for Bagging-J48 on all observed thinking styles is very high. This further shows the effectiveness of the discriminative model based on Bagging of J48 classifier. The developed model is especially relevant in interpreting human signature in online interaction. Both J48 and LMT classifier are tree based human interpretable rules sets. However, LMT-Model is a composite model, which integrates the logistic model into decision tree pattern identification process. A decision tree is a rule-based discriminatory process which lends itself to ease of human interpretation. The applicability of this result is presented in the next section. This is based on the probability of attributing human users to an online session for the digital forensic process.

B. Digital Forensic Attribution Model

Digital forensics is the science of identifying, preserving and analysing digital evidence for onward litigation or strengthening of security systems. Attribution is a major process in the digital forensic analysis process [8], [47]. This section presents the attribution process for the analysis process in the proposed digital forensic attribution model. The model is developed using the unified modelling language (UML), as a process of formalization of the attribution process. UML is generally used to represent a behavioural high-level abstract structure of a phenomenon, system or process. UML is generally used in four non-exclusive processes: use-case model diagram, interaction diagram, state-chart diagram and activity diagram.

This study employs the Use-case diagram and State-chart diagram to depict the formalization process. The Use-case diagram depicts the typical investigation scenario while the State-chart diagram elucidates on the various state processes for digital forensic investigation, and specifically, the state processes in network forensic analysis process based on the integration of human thinking style into user attribution process. The proposed high-level attribution process is presented in Fig. 3. Evidence from a reliable source is defined as the input to the model, while a probable likelihood of the profile of the unknown user is generated, as output. The model is contingent on the initial (updateable) database of behavioural signature based on human thinking style. The attribution model covers areas of evidence preparation and evidence analysis. The process starts with the dissection of the obtained evidence into the reasonable and probable dataset.

Identification, definition, and extraction of behavioural features from the dataset form part of the fundamental basis for the development of the digital signature. The behavioural feature includes the human-centric features defined in the network section (**Error! Reference source not found.**) of this study. Other probable features could include web page complexity, the size of the document, behavioural biometric such as keystroke dynamics, mouse dynamics and a corresponding soft biometrics (such as gender, age, handedness, ocular-dominance).

TABLE IV
EXPERIMENTAL RESULT OF CLASSIFICATION ALGORITHMS

Evaluation Metrics	Data	Baseline Classifier	J48	J48 with Bagging	J48 with AdaBoostM1	LMT	LMT with Bagging	LMT with AdaBoostM1
Area Under ROC Curve (AUC)	1	0.50(0.00)	0.95(0.01) v	0.97(0.01) v	0.98(0.01) v	0.94(0.01)v	0.97(0.01) v	0.94(0.01) v
	2	0.50(0.00)	0.94(0.03) v	0.97(0.02) v	0.97(0.01) v	0.94(0.03)v	0.97(0.02) v	0.96(0.02) v
	3	0.50(0.00)	0.97(0.01)v	0.99(0.01)v	0.99(0.01)v	0.97(0.01)v	0.99(0.01) v	0.99(0.01) v
	4	0.50(0.00)	0.83(0.03) v	0.89(0.02) v	0.89(0.01) v	0.88(0.02)v	0.92(0.01) v	0.86(0.02) v
	5	0.50(0.00)	0.97(0.01) v	0.99(0.01) v	0.99(0.01) v	0.98(0.01)v	0.99(0.01) v	0.99(0.01) v
Accuracy	1	66.04(0.06)	90.68(1.81)v	93.10(1.40)v	92.85(0.95)v	90.24(1.80)v	92.09(1.04)v	88.22(1.64)v
	2	65.43(0.28)	91.03(2.95)v	92.50(2.38)v	91.87(2.68)v	90.37(3.07)v	91.76(2.65)v	90.56(2.71) v
	3	70.11(0.18)	95.60(1.46)v	96.99(1.16)v	96.06(1.25)v	95.78(1.42)v	96.54(1.22)v	96.90(1.25)v
	4	66.34(0.04)	80.38(1.60)v	83.09(1.63)v	82.83(1.19)v	83.62(2.01)v	86.50(1.26)v	81.20(1.79)v
	5	72.51(0.14)	95.62(1.12)v	96.79(0.97)v	96.19(1.08)v	96.01(1.01)v	96.58(0.99)v	95.92(1.28)v
F-Measure	1	0.00(0.00)	0.86(0.03) v	0.90(0.02) v	0.89(0.02) v	0.85(0.03)v	0.88(0.02) v	0.82(0.03) v
	2	0.79(0.00)	0.93(0.02) v	0.94(0.02) v	0.94(0.02) v	0.93(0.02)v	0.94(0.02) v	0.93(0.02) v
	3	0.82(0.00)	0.97(0.01)v	0.98(0.01)v	0.97(0.01)v	0.97(0.01)v	0.98(0.01) v	0.98(0.01) v
	4	0.80(0.00)	0.86(0.01) v	0.88(0.01) v	0.88(0.01) v	0.88(0.01)v	0.90(0.01) v	0.86(0.01) v
	5	0.00(0.00)	0.92(0.02) v	0.94(0.02) v	0.93(0.02)v	0.93(0.02)v	0.94(0.02) v	0.92(0.02) v
Kappa Statistics	1	0.00(0.00)	0.79(0.04) v	0.84(0.03) v	0.84(0.02)v	0.78(0.04)v	0.82(0.02) v	0.73(0.04) v
	2	0.00(0.00)	0.80(0.07)v	0.83(0.05)v	0.81(0.06)v	0.79(0.07)v	0.81(0.06) v	0.79(0.06) v
	3	0.00(0.00)	0.89(0.03)v	0.93(0.03)v	0.90(0.03)v	0.90(0.03)v	0.92(0.03) v	0.93(0.03) v
	4	0.00(0.00)	0.52(0.05) v	0.61(0.04) v	0.59(0.03)v	0.61(0.05)v	0.68(0.03) v	0.57(0.04) v
	5	0.00(0.00)	0.89(0.03) v	0.92(0.02) v	0.90(0.03)v	0.90(0.03)v	0.91(0.03) v	0.90(0.03) v
Root Mean Square Error	1	0.47(0.00)	0.27(0.02)	0.25(0.03)	0.24(0.01)	0.28(0.02)	0.26(0.01)	0.32(0.02)
	2	0.48(0.00)	0.27(0.04)	0.26(0.04)	0.25(0.03)	0.28(0.04)	0.26(0.03)	0.29(0.04)
	3	0.46(0.00)	0.19(0.03)	0.16(0.03)	0.18(0.02)	0.19(0.03)	0.16(0.02)	0.16(0.04)
	4	0.47(0.00)	0.38(0.02)	0.38(0.02)	0.35(0.01)	0.35(0.02)	0.32(0.01)	0.40(0.02)
	5	0.45(0.00)	0.19(0.03)	0.17(0.03)	0.17(0.02)	0.18(0.02)	0.16(0.02)	0.19(0.03)
Precision	1	0.00(0.00)	0.89(0.03)	0.91(0.02)	0.93(0.02)	0.89(0.03)v	0.92(0.02) v	0.85(0.03) v
	2	0.65(0.00)	0.91(0.03)	0.93(0.02)	0.91(0.03)	0.92(0.03)	0.91(0.03)	0.92(0.03)
	3	0.70(0.00)	0.97(0.02)	0.97(0.01)	0.96(0.01)	0.97(0.01)	0.97(0.01)	0.97(0.01)
	4	0.66(0.00)	0.81(0.02)	0.85(0.02)	0.83(0.01)	0.84(0.02)v	0.86(0.01) v	0.84(0.02) v
	5	0.00(0.00)	0.92(0.03)	0.95(0.02)	0.95(0.03)	0.93(0.03)	0.95(0.03)	0.94(0.03)
Recall	1	0.00(0.00)	0.83(0.05)	0.88(0.03)	0.86(0.03)	0.82(0.04)v	0.84(0.03) v	0.79(0.03) v
	2	1.00(0.00)	0.95(0.03)	0.96(0.02)	0.97(0.02)	0.94(0.03)	0.96(0.02)	0.94(0.02)
	3	1.00(0.00)	0.97(0.02)	0.98(0.01)	0.98(0.01)	0.97(0.01)	0.98(0.01)	0.98(0.01)
	4	1.00(0.00)	0.93(0.01)	0.90(0.02)	0.93(0.01)	0.93(0.01)	0.95(0.01)	0.88(0.02)
	5	1.00(0.00)	0.92(0.04)	0.93(0.03)	0.91(0.03)	0.93(0.03)v	0.93(0.03)	0.91(0.04)

1=Oligarchic-style 2=Legislative-style, 3=Hierarchical-style, 4=Judiciary-style, 5=External-style, v=statistical significance at p of 0.05

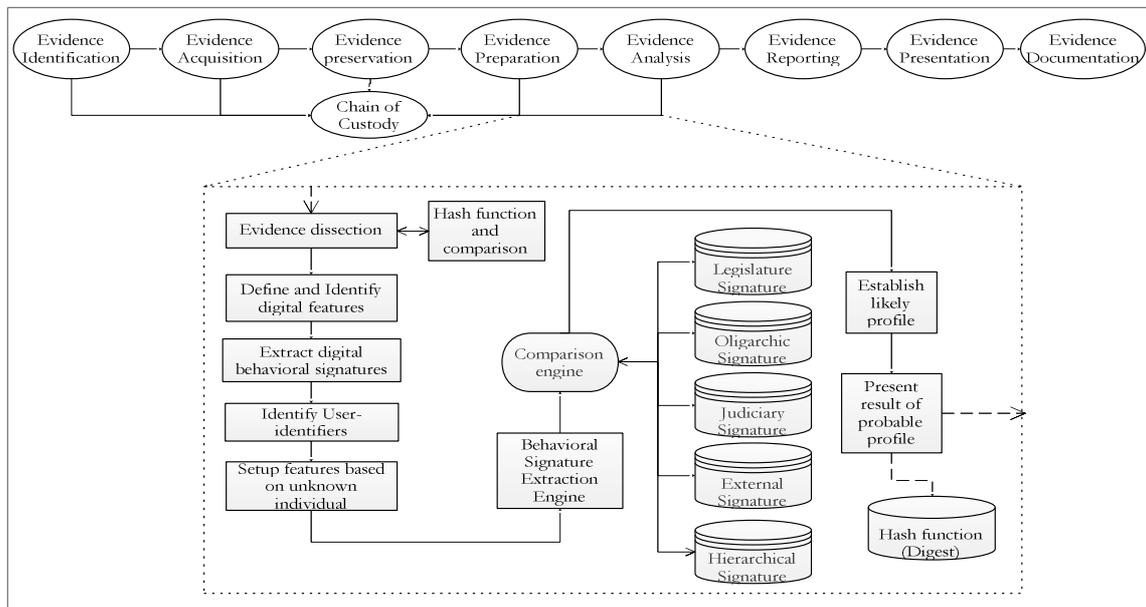


Fig. 3 State-chart formalization diagram for User attribution process in digital forensics

The comparison engine computes the (dis)similarity index between the processed behavioural attributes of the unknown user, and the digital thinking style signatures. A hash function is carried out on the input data and the output profile to ensure experimental repeatability and reliability, for admissibility in a competent court of jurisdiction. For every profile developed, a hash digest is stored, to ensure the preservation of the integrity of the profile.

The hash digest is computed from the developed profile, and tagged with a unique identifier, such that each completely developed profile is ensured against alteration and integrity compromise. A forensic process Use-case is depicted in Fig. 4. The proposed digital forensic Use-case defines a separation of the evidence analysis and evidence preparation process, as a distinct process, which requires the special expertise of digital forensic stakeholders such as a forensic technician, a forensic researcher, as well as a forensic analysis investigator. This distinction is necessarily for few reasons.

First, evidence from network-based sources contains unwanted non-human initiated information (network traffic generated by application on behalf of the user, HTTP in-line request, traffic from Bots, etc.), which could induce higher rate of false positives and false negatives occurrence, as well as a degradation of the discriminatory power of the extracted features (in some cases, feature extraction might not be feasible).

Second, the process of analysing probable digital evidence transcends traditional event correlation and usage profiling, into the integration of complementary behavioural biometrics into network and computer forensics. Such behavioural biometrics requires the expertise of researchers and technicians who can develop an adaptive process for analysis and interpretation of the result. Third, the proposed integration of psychosocial attributes into digital processing process presents a more complex mechanism, which requires the continual and frequent update of style databases. These processes require the technical expertise of digital technicians, researchers as well as a skilled forensic analyst.

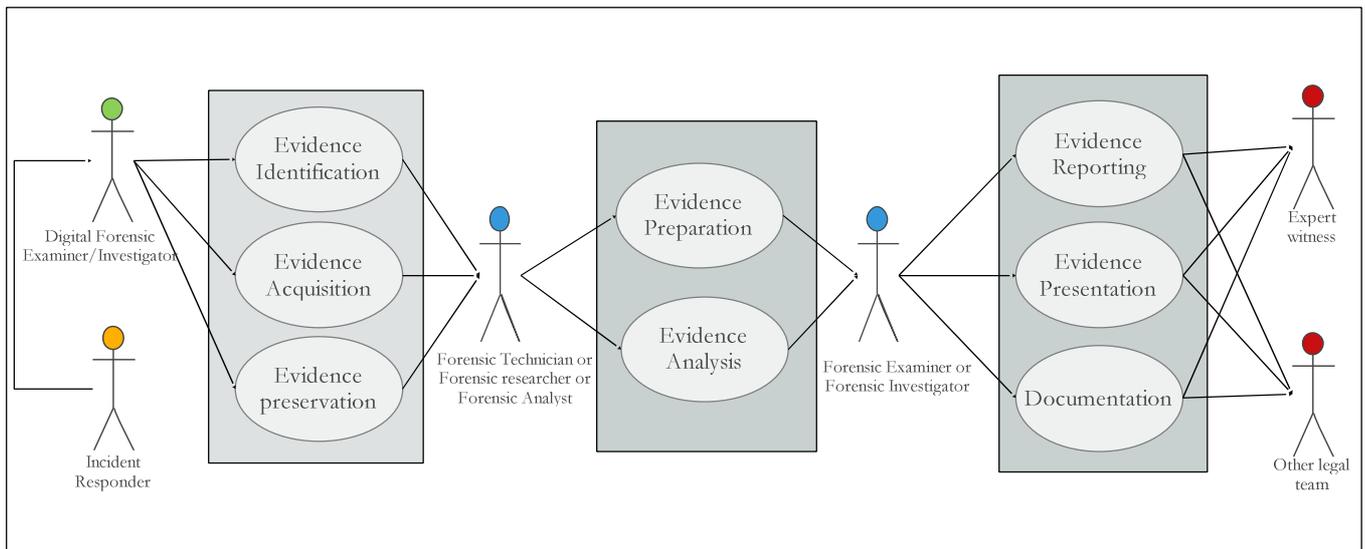


Fig. 4 Use-case formalization diagram for digital forensics process

The distinction of the stakeholders for the Evidence analysis process is essential, in conducting a forensic investigation. Furthermore, given the growing body of the pervasive device, cloud-based systems, and the frequency in privacy preservation and anonymity assurance mechanisms, this distinction becomes apparently necessary.

This study attempts to answer a fundamental research question on the probability of the existence of digital signature based on human thinking style on the Internet, as an extension of prior study [35]. The result obtained in Table 4 indicates online interaction can be reliably discriminated using Bagging-J48 decision tree classifier and Bagging-LMT classifier, similar to the prior finding. The results thus suggest a high probability of the existence of human thinking style signature –Oligarchic, Judiciary, Hierarchical, Legislative, and External- on the Internet. This finding closely aligns with the assertion in [2] which suggest the

incorporation of thinking style into online search engine interface for better search intention prediction.

Furthermore, the assertion suggests that incorporating thinking style into search engine interface would better help users to comprehend search engine results. The assertion in [13] lends further credence to this logic. The study argued that the incorporation of human cognitive styles into individual decision support system would significantly improve the effectiveness of decision-making process. In the review presented in [48], the incorporation of individual difference (based on cognitive style) into computer system design is defined as an aspect of HCI.

The existence of digital thinking style signature on the Internet is not surprising. This is because several studies have observed the tendency of individual preference and differences in information seeking behaviour based on different thinking styles. For example, individual who are

high on the global thinking styles have a preference for abstract reasoning. Conversely, individual high on the Local thinking style has a preference for details in tasks [2]. In addition, Executive and Legislative thinking styles manifest a significant distinction in learning transference over the Internet [21].

The finding presented in this study extends the body of literature on the individual difference for online system development, specifically for user attribution on the Internet. In addition, it opens the research community to study on digital signatures that are capable of complementing traditional authentication systems. In areas of HCI particularly on E-learning and E-commerce, the finding from this study presents a critical composition for the development of a framework that can improve user analysis and online interface design.

In the area of digital investigation, the developed UML model shows how the result from this study can be applied in identifying online users. The signature from each dichotomy can be used as a baseline benchmark for a 1-to-N user attribution process, which can then applied to reduce the suspect pool list during an investigation. Furthermore, this psychosocial attribute can be used to confirm an online user, by comparing the profile of a known behavioural template of the user to a profile under investigation. This is specifically relevant in insider misuse cases, as well as in cases, which involves non-repudiation. Non-repudiation cases attempt to prove or disprove plausible deniability.

This study has few limitations. First, the sample size used in this study is relatively small. Larger sample size can be used to explore more generalizable digital signatures. The study was able to extract only two subclasses for each of the extracted thinking styles. A larger sample size would present a wider range of subclasses as well as a higher number of thinking style factors. Second, the highest accuracy obtained in the study is still below perfection.

As part of ongoing research, the next phase of this study will consider the development of a system, and a customized graphical interface based on the extracted signatures for the digital investigation model. The graphical interface will be used to validate the reliability and operational accuracy of the existence of digital signatures based on thinking style. A further step in this direction can also include the incorporation of thinking style signature into a website for an e-profiling process. Such incorporation will provide a measure to extend the practicality of these findings and the implementation of the forensic model. Furthermore, other areas of probable incorporation include E-learning, a soft authentication mechanism for access control systems.

IV. CONCLUSION

This research investigated the probability of the existence of online digital thinking style signature. Thinking style inventory as proposed by [28] was adopted as a measurement instrument for human thinking style. The Judiciary, Oligarchic, Legislative, Hierarchical, and External thinking style were extracted from 43-respondents. Network traffic of the 43-respondents was collected and analysed for consistent and recurring patterns in communication. Various supervised machine-learning techniques were explored. J48 decision Tree classifier using on Bagging technique was

observed to achieve the highest accuracy in this study. Future works can further consider the integration of other psychosocial attributes, such as Personality trait, into online digital signature exploration. In addition, client-side dataset (mobile phone data, and or personal computer data) can be integrated with Server-side data to present a more comprehensive data.

ACKNOWLEDGMENT

This research was supported/partially supported by Universiti Teknologi Malaysia and Ministry of Higher Education Malaysia under the vote number: R.J130000.7813.4F804.

REFERENCES

- [1] R. LaRose, D. Mastro, and M. Eastin, "Understanding Internet usage a social-cognitive approach to uses and gratifications," *Soc. Sci. Comput. Rev.*, vol. 19, no. 4, p. 395–413., 2001.
- [2] G. Y. M. Kao, P. L. Lei, and C. T. Sun, "Thinking style impacts on Web search strategies," *Comput. Human Behav.*, vol. 24, no. 4, pp. 1330–1341, 2008.
- [3] L. Fiorina, A. Antonietti, B. Colombo, and A. Bartolomeo, "Thinking style, browsing primes and hypermedia navigation," *Comput. Educ.*, vol. 49, no. 3, pp. 916–941, Nov. 2007.
- [4] K. K. Khamsum, "Towards Modelling Web Search Behaviour: Integrating Users' Cognitive Styles," Queensland University of Technology, 2013.
- [5] M. Kozhevnikov, "Cognitive styles in the context of modern psychology: toward an integrated framework of cognitive style.," *Psychol. Bull.*, vol. 133, no. 3, pp. 464–81, May 2007.
- [6] A. S. Richmond, H. M. Krank, and R. Cummings, "A brief research report : thinking styles of online distance education students," *Int. J. Technol. Teach. Learn.*, vol. 2, no. 1, pp. 58–64, 2006.
- [7] E. S. Pilli, R. C. Joshi, and R. Niyogi, "Framework for Network Forensic Analysis," in *Communications in Computer and Information Science*, 2010, vol. 101, pp. 142–147.
- [8] A. Valjarevic, "Harmonised Digital Forensic Investigation Process Model," in *Information Security for South Africa*, 2012, pp. 1–10.
- [9] 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.
- [10] R. J. Riding and E. Sadler-Smith, "Cognitive Style and Learning Strategies: Some Implications for Training Design," *Int. J. Train. Dev.*, vol. 1, no. 3, pp. 199–208, Sep. 1997.
- [11] B. J. Zimmerman, "A social cognitive view of self-regulated academic learning.," *J. Educ. Psychol.*, vol. 81, no. 3, pp. 329–339, 1989.
- [12] R. J. R. Sternberg and L. Zhang, "Styles of Thinking as a Basis of Differentiated Instruction," *Theory Pract.*, vol. 44, no. 3, pp. 245–253, 2005.
- [13] A. H. Huang, D. C. Yen, and S. Y. Hung, "Research issues of internet-integrated cognitive style," *Comput. Human Behav.*, vol. 28, no. 5, pp. 1547–1554, Sep. 2012.
- [14] L. Zhang, "Thinking styles, self-esteem, and extracurricular experiences.," *Int. J. Psychol.*, vol. 36, no. 2, pp. 100–107, 2001.
- [15] L. Zhang, "Are Thinking Styles and Personality Types Related?," *Educ. Psychol.*, vol. 20, no. 3, pp. 271–283, 2000.
- [16] L. Zhang, "Thinking styles and the big five personality traits revisited," *Pers. Individ. Dif.*, vol. 40, no. 6, pp. 1177–1187, Apr. 2006.
- [17] L. Zhang and P. Higgins, "The predictive power of socialization variables for thinking styles among adults in the workplace," *Learn. Individ. Differ.*, vol. 18, no. 1, pp. 11–18, Jan. 2008.
- [18] M. Albaili, "Differences in thinking styles among low average and high-achieving college students," in *Proceedings of the The 13th International Conference on Thinking*, 2007, pp. 5–10.
- [19] M. a. Albaili, "Differences Among Low-, Average- and High-achieving College Students on Learning and Study Strategies," *Educ. Psychol.*, vol. 17, no. 1–2, pp. 171–177, 1997.
- [20] G. Vinitzky and D. Mazursky, "The effects of cognitive thinking style and ambient scent on online consumer approach behavior, experience approach behavior, and search motivation," *Psychol.*

- Mark.*, vol. 28, no. 5, pp. 496–519, 2011.
- [21] C. Lee and F. Tsai, “Internet project-based learning environment: the effects of thinking styles on learning transfer,” *J. Comput. Assist. Learn.*, vol. 20, no. 1, pp. 31–39, 2004.
- [22] G. Y.-M. Kao, P.-L. Lei, and C.-T. Sun, “Thinking style impacts on Web search strategies,” *Comput. Human Behav.*, vol. 24, no. 4, pp. 1330–1341, Jul. 2008.
- [23] J. Fan and L. Zhang, “The role of learning environments in thinking styles,” *Educ. Psychol.*, vol. 32, no. August, pp. 1–18, 2013.
- [24] L.-F. Zhang, “Thinking Styles: Their relationships with modes of thinking and academic performance,” *Educ. Psychol.*, vol. 22, no. March 2015, pp. 331–348, 2002.
- [25] C. Zhu and L. Zhang, “Thinking styles and conceptions of creativity among university students,” *Educ. Psychol.*, vol. 31, no. 3, pp. 361–375, 2011.
- [26] T. M. Yu and C. Chen, “Thinking styles and preferred teacher interpersonal behavior among Hong Kong students,” *Learn. Individ. Differ.*, vol. 22, no. 4, pp. 554–559, 2012.
- [27] H. Wang, H. Yang, and C. Sun, “Thinking Style and Team Competition Game Performance and Enjoyment,” *IEEE Trans. Comput. Intell. AI Games*, vol. 7, no. 3, pp. 243–254, 2015.
- [28] R. Sternberg, “Mental self-government: A theory of intellectual styles and their development,” *Hum. Dev.*, vol. 31, no. 4, pp. 197–224, 1988.
- [29] A. Al-Thani, T. Al-Thani, and Y. Semmar, “Investigating the Relationship between Students’ Thinking Styles, Self-Efficacy for Learning, and Academic Performance at Qatar University,” *Am. Int. J. Soc. Sci.*, vol. 3, no. 2, 2014.
- [30] J. Turki, “Thinking styles’ In Light of Sternberg’s Theory’ prevailing among the students of Tafila Technical University and its relationship with some variables,” *Am. Int. J. Contemp. Res.*, vol. 2, no. 3, pp. 140–152, 2012.
- [31] L.-F. Zhang and J. Sachs, “Assessing thinking styles in the theory of mental self-government: A Hong Kong validity study,” *Psychol. Rep.*, no. 1983, pp. 915–928, 1997.
- [32] G. Y.-M. Kao, P.-L. Lei, and C.-T. Sun, “Thinking style impacts on Web search strategies,” *Comput. Human Behav.*, vol. 24, no. 4, pp. 1330–1341, Jul. 2008.
- [33] L. Zhang and G. Postiglione, “Thinking styles, self-esteem, and socio-economic status,” *Pers. Individ. Dif.*, 2001.
- [34] L. Zhang, “Validating the theory of mental self-government in a non-academic setting,” *Pers. Individ. Dif.*, vol. 38, no. 8, pp. 1915–1925, Jun. 2005.
- [35] I. R. Adeyemi, A. S. Razak, and M. Salleh, “Individual Difference for HCI Systems: Examining the Probability of Thinking Style Signature in Online Interaction,” in *4th International Conference on User Science and Engineering 2016 (i-USER 2016)*, 2016, p. In Press.
- [36] Y. (Catherine) Yang and B. Padmanabhan, “Toward user patterns for online security: Observation time and online user identification,” *Decis. Support Syst.*, vol. 48, no. 4, pp. 548–558, 2010.
- [37] R. Kumar and A. Tomkins, “A characterization of online browsing behavior,” in *Proceedings of the 19th international conference World Wide Web*, 2010, pp. 561–570.
- [38] 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.
- [39] A. Barabasi, “The origin of bursts and heavy tails in human dynamics,” *Nature*, vol. 435, no. 7039, p. 207–211, 2005.
- [40] T. Zhou, X. Han, and B. Wang, “Towards the Understanding of Human Dynamics,” *Sci. matters Humanit. as complex Syst.*, no. 1, pp. 207–233, 2008.
- [41] T. T. T. Nguyen and G. Armitage, “A survey of techniques for internet traffic classification using machine learning,” *Commun. Surv. Tutorials, IEEE*, vol. 10, no. 4, pp. 56–76, 2008.
- [42] S. Kotsiantis, I. Zaharakis, and P. Pintelas, “Supervised machine learning: A review of classification techniques,” *Informatica*, vol. 31, pp. 249–268, 2007.
- [43] F. M. Othman, T. Moh, and S. Yau, “Comparison of Different Classification Techniques Using WEKA for Breast Cancer,” in *3rd Kuala Lumpur International Conference on Biomedical Engineering*, 2007, vol. 15, pp. 520–523.
- [44] Y. Freund and R. E. Schapire, “Experiments with a New Boosting Algorithm,” *Int. Conf. Mach. Learn.*, pp. 148–156, 1996.
- [45] L. Rokach, “Ensemble-based classifiers,” *Artif. Intell. Rev.*, vol. 33, no. 1–2, pp. 1–39, 2010.
- [46] Y. C. Yang, “Web user behavioral profiling for user identification,” *Decis. Support Syst.*, vol. 49, no. 3, pp. 261–271, 2010.
- [47] R. White, A. Hassan, A. Singla, and E. Horvitz, “From devices to people: Attribution of search activity in multi-user settings,” in *Proceedings of the 23rd International World Wide Web Conference Committee (IW3C2)*, 2014, pp. 431–441.
- [48] J. Kuurstra, “Individual differences in Human-Computer Interaction: A review of empirical studies,” University of Twente, 2015.