2017

Insider Misuse Identification using Transparent Biometrics

Clarke, Nathan

10.24251/hicss.2017.487
Proceedings of the 50th Hawaii International Conference on System Sciences (2017)
Hawaii International Conference on System Sciences

Insider Misuse Identification using Transparent Biometrics

Nathan Clarke[1,2], Fudong. Li[1], Abdulrahman. Alruban[1,3], Steven. Furnell[1,2,4]
[1]Centre for Security, Communications and Network Research, Plymouth University, Plymouth, UK
[2]Security Research Institute, Edith Cowan University, Perth, Western Australia
[3] Computer Sciences and Information Technology College, Majmaah University Majmaah, Saudi Arabia
[4]School of Information and Communication Technology, Nelson Mandela Metropolitan University, Port Elizabeth, South Africa
N.Clarke@Plymouth.ac.uk

Abstract

Insider misuse is a key threat to organizations. Recent research has focused upon the information itself – either through its protection or approaches to detect the leakage. This paper seeks a different approach through the application of transparent biometrics to provide a robust approach to the identification of the individuals who are misusing systems and information. Transparent biometrics are a suite of modalities, typically behavioral-based that can capture biometric signals covertly or non-intrusively – so the user is unaware of their capture. Transparent biometrics are utilized in two phases a) to imprint digital objects with biometric-signatures of the user who last interacted with the object and b) uniquely applied to network traffic in order to identify users traffic (independent of the Internet Protocol address) so that users rather than machine (IP) traffic can be more usefully analyzed by analysts. Results from two experimental studies are presented and illustrate how reliably transparent biometrics are in providing this link-ability of information to identity.

1. Introduction

Insider misuse have become widespread in the last decade and has been considered an important security issue by many recent research studies [1-6]. Whilst much research has been undertaken into the detection of insider misuse and particularly information leakage, less focus has been attributed to linking this to the people who are committing the criminal activity. Assuming many will utilize stolen credentials from colleagues to perpetrate the misuse, there is frequently a reliance upon other security controls to verify the identity of the person responsible (i.e. through the use of logs, CCTV etc), if indeed this is possible at all.

This paper focusses upon the application of transparent biometrics to the problem of insider misuse in two differing scenarios that together provide robust identity verification, enabling incident analysts and/or forensic examiners a reliable, efficient and effective process for identifying the responsible individual. The first phase deals with the problem of understanding who is actually responsible for leaking the information. As a user interacts with a computer, biometric signals are capture covertly (enabling biometric techniques such as facial recognition, keystroke analysis, behavioral profiling, iris recognition) to be imprinted within the digital object being used. In this manner, all digital objects will be covertly imprinted with the biometric signature of the user who has last interacted with it. Upon detection of the misuse, the biometric signature can be recovered and this will identify the individual who sent it – negating the need for expensive and time consuming investigation of system. Having identified the user, an investigator will want to understand what else the user has been up to. Organizational network traffic logs provide an independent basis for understanding this – however, due to the mobile nature of the workforce and the nature of DHCP, it has long been understood that IP is not a reliable means of filtering traffic. The use of multiple devices further compounds the issue as the device will be given differing IP addresses upon each connection. The second phase, utilizes a novel feature extraction process and applies behavioral profiling to the network traffic to provide investigators with all the traffic associated with a particular user (rather than machine), enabling a far more efficient analysis of user activity.

The paper is structured with Section 2 providing an overview of the background literature in insider misuse and transparent biometrics. Section 3 describes the system architecture and processes that underpin the approach. Section 4 presents two sets of experiments to illustrate the feasibility of applying transparent biometrics focused around both phases. Section 5 presents a discussion with Section 6 providing the concluding remarks and direction of the future work.

URI: http://hdl.handle.net/10125/41647
ISBN: 978-0-9981331-0-2

2. Background

2.1. Insider Misuse

Typically, there are three main types of threats that insiders could commit, Internet technology sabotage, insider theft of intellectual property and insider fraud. The majority of research within the field has been focused upon the detection rather than the response. Several academic studies were published with the aim of combating the issue of insider misuse. Some studies have tried to use biometric information as a factor for detecting the malicious activities of an insider by measuring the individual's biometric signals and in particular when they change, for example Heart Rate Variability (HRV), core body temperature, and skin temperature [7]. The authors claim that there is a correlation between these signals change and individual's malicious intent. Clearly their experimental setup is rather controversial, and there is neither a distinct nor strong correlation between their hypotheses or the actual implementation to the data loss detection/prevention. Other researchers investigated the feasibility of linking persons with the leakage incidents by inserting a fake data (object) into the confidential data that they want to protect [8-11]. However, such an approach is not a simple task to achieve, even with methods such as the guilt identification model, which is designed precisely for the case when a data distributor has given sensitive information to a group of supposedly trusted agents. Other authors proposed monitoring systems that operate proactively for forensic and audit purposes [12-15]. While such techniques could provide comprehensive logs and detailed information about actions performed by the user, it fails to biometrically link an incident to those individuals who conducted them. Indeed, there is no current literature that has specifically sought to transparently embed biometric signatures in digital objects. Whilst the use of network traffic to identify individuals has been the focus of many studies, results have not been encouraging due to the level of noise and the uniqueness of features being utilized in the biometric system [16-19]. They have tended to use very raw features such as individual packet or flow-based information or even source IP – which is not appropriate. The approach in this paper is to apply the extensive knowledge of transparent biometrics to develop a unique set of features from raw network metadata that are focused specifically upon the user.

2.2. Transparent Biometrics

With the aim of providing continuous protection for IT services and their data, a number of research studies were conducted to investigate the use of biometrics for transparent and non-intrusive authentication over the last 15 years. For example, a thin-client based topology is proposed by [20] that utilized face recognition to enable transparent authentication on mobile devices; while DARPA proposed an Active Authentication research program aiming to address the problem of the point-of-entry technique by utilizing behavioral-based biometrics for desktop computing [21]. Based upon their characteristics (either physiological or behavioral based), details of existing studies on transparent biometrics are presented in the following subsections.

2.2.1. Physiological Transparent Biometrics. The physiological biometric based systems utilize the characteristics of a human body part (e.g. fingerprint and face) to identify individuals. In general, physiological biometric characteristics contain high levels of discriminatory information and are resistant to various factors (e.g. age, body fitness or the weather conditions). Traditionally, physiological biometric techniques are solely used as an alternative point of entry authentication method. With the advancement of sensors, several transparent authentication techniques started to emerge. [22] and [20] utilized face recognition to provide continuous and transparent authentication for desktops and mobile devices respectively. [23] proposed a fast and fully automatic ear recognition approach based on 3D local surface features for generic usage; while [24] investigated an implicit authentication technique based upon ear shapes during a call session on smartphones. By using touch screens, [25] and [26] demonstrated that fingerprint identification can be utilized to provide transparent and continuous protection for mobile devices. In addition, the use of iris recognition (one of the most accurate biometric technique) for real-time continuous authentication was also investigated [27].

2.2.2. Behavioral Transparent Biometrics. Behavioral biometrics identifies a person based upon their unique behavior, such as the way they type on a keyboard. In contrast with physiological biometric techniques, behavioral based methods are less unique as human behaviors tend to change over time due to various reasons (e.g. social environment or mood); however they are more flexible and user-friendly, hence a natural candidate for transparent authentication. Indeed, researchers have been carried out a number of studies. Initially, the two most common user behaviors with the computing

environment (i.e. the use of mouse and keyboard) were investigated for the purpose of continuously and non-intrusively authenticating users: mouse dynamics [28-32] and keystroke analysis [33-35]. Then, with the prevalence of mobile devices, several new behavioral biometrics were explored for the same purpose (i.e. transparent authentication), including voice verification [36, 37], gait recognition [38, 39] and behavioral profiling [40].

The field of transparent biometrics has become more established in recent years with a significant focus upon the development of more reliable modalities and the underlying management system that provides a degree of intelligent context support.

3. A Proactive and Reactive Approach

As illustrated in Figure 1, the core of the architecture is built around the capturing, processing and classification of biometric signals derived from built-in sensors within the computing technology that can be captured without the explicit interaction by the user. The transparent, continuous and multimodal capture of the biometric signals is key to mitigating against forgery [41]. Whilst a perpetrator might be able to steal password credentials or even forge a biometric, to do so continuously and against a wide-range of modalities is extremely challenging (there are few studies that have successfully hacked a multibiometric system let alone a continuous and transparent multibiometric system). This provides for a strong basis of ensuring it is the user who is using the system whose biometrics are captured rather than someone they might be masquerading to be.

The source of biometric signals differs depending upon the proactive or reactive approach. The proactive is focused upon the capture and analysis of signals captured via the client computer. The reactive approach is focused upon deriving a behavioral-based profile from the network traffic.

3.1. Proactive Approach – Biometric Imprinting

The proactive approach focuses upon establishing a reliable correlation between individuals and the digital object (e.g. an email, an image file or a document) that is being used. This correlation is achieved by mapping the hexadecimal representations of the extracted biometric features of the user with the hexadecimal representations of the object. This process generates a digital imprint file representing locations within the object, each of which corresponds to a respective portion of the extracted biometric features.

These corresponding locations are the main element of the resulted imprint file and with perhaps some extra information (e.g. object metadata, timestamp). Thereby, the imprint file does not carry any direct biometric signal but rather the locations that correspond to the biometric signal.

Since any file type can be represented in hexadecimal form, this makes the proposed method a generic approach that works with any object type—as long as it is in form of a digital object.

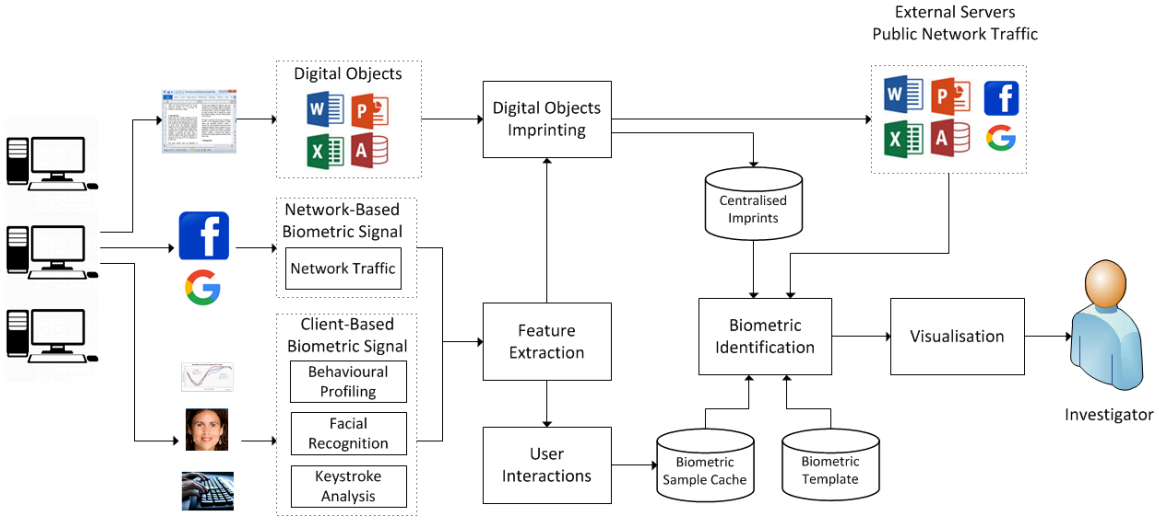

Figure 1: Insider Misuse Identification Architecture

The generated imprints are then stored in a centralized database for later use when needed. For instance, upon the detection of misuse such as data leakage, the object (whether it is posted on a public website or captured by the network) can be analyzed for extracting the imprinted biometric signals by reversing the mapping process. Hence, an investigator could establish the identity of the user who interacted

with the object by processing the extracted biometric feature with the biometric system and compared against the database of users.

3.2. Reactive Approach – Network Fingerprinting

The reactive part of the insider misuse identification architecture (i.e. network fingerprinting) aims to identify individual users via their interactions of various network-based services, assisting investigators to narrow down the pool of potential suspects when an incident occurs. In order for the network fingerprinting technique to work accurately, two critical processes need to be completed: the extraction from raw traffic of user interactions (application-level descriptions of user interactions rather than raw traffic) which form the template and the identification strategy used to classify and link network traffic.

During the template creation stage, users use their network services (e.g., Facebook, Google) and the Feature Extraction function extracts various metadata information, including the date and time stamps and various header information of the IP packet (e.g., sender and receiver's IP address and port numbers, the size of the packet) and the User Interactions function derives interaction signatures on different user activities within network based services. The novelty of this approach is unlike current research that focusses upon which service a user might user (e.g. Facebook) and for how long, these user interactions are able to discern more discreet information as to what the user was doing whilst using the service (e.g. messaging on Facebook, watching a video, reading the all). This additional granularity of user-specific information provides a richer source of discriminative information. In order to ensure the template quality, sufficient network data is required to be collected over a period time; also, due to the nature of the behavioral biometric technique, the template will be regularly renewed.

The user identification process is required when an incident occurs. During this process, all network metadata related to the incident will be analyzed. Similar to the template creation process, the network traffic metadata is utilized by Feature Extraction function and then the User Interactions function. These interactions are then compared to the template database to obtain a probabilistic identification.

Once the identification process is completed, the network traffic will be prioritized based upon the identification outcome and visualized according to identified users. In this way, the investigator can make a sensible decision upon which part of the network traffic should be focused first.

4. Experimental Analysis

In order to assess the feasibility of the utilizing transparent in this proactive and reactive manner, two experiments were devised to investigate the core functionality.

4.1. Proactive Biometric Imprinting

The ability to capture, extract and imprint a digital object with a transparent biometric signature is not a technically challenging task (and is already the subject of a patent by the authors [42]). The key research question with respect to the imprinting of the biometric signature is how robust is the approach given subsequent modification of the digital object. As such the following two experiments were undertaken that simulate two types of attacks that might occur to a digital object after the imprinting process has occurred:

- The first experiment simulates the case where the imprinted image was modified in several parts. This type of attack is very influential since various and random parts of the image are affected by such alterations. The modified parts are determined randomly (an example of which is shown in Figure 2).
- The second attack scenario simulates the case of when only part of the imprinted image is available and the rest is missing; for instance, the imprinted image could be resized or cropped. To simulate such alterations, a random section of the images in the dataset was cropped in different sizes after been imprinted. Figure 3 illustrates an example of some of these cropped samples.

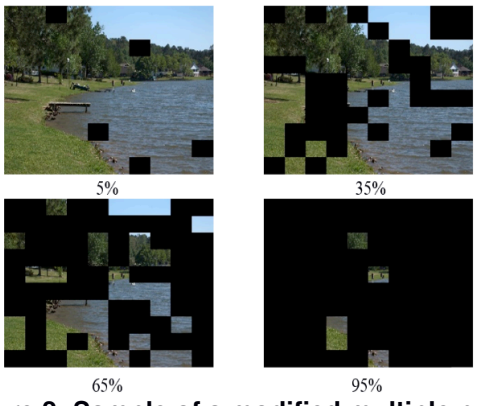

Figure 2: Sample of a modified multiple parts of an image

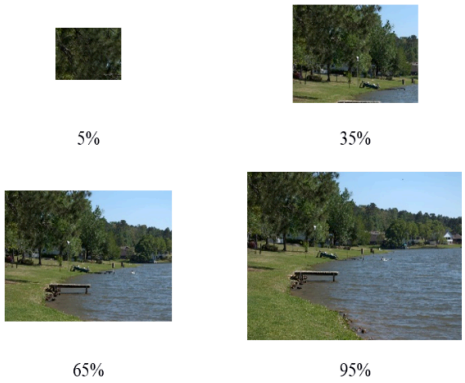

Figure 3: Samples of a cropped image in certain percentage

In the experiments, the feature vector used is a real facial feature vector sample with a length of 57 numeric characters, as illustrated in Figure 4. The length of the vector does vary but this is fairly representative of the type and length of information that would need embedding. In this study, Fisherfaces algorithm is used to generate the feature vector for the captured users' faces images [43]. The UCID image dataset version 2 is used for to simulate both attacks and evaluates the performance of the proposed technique [44]. For the purpose of this study, only the first 100 images are used from this dataset, since it is assumed that this number is enough for the purpose of evaluation. Each of the 100 images was modified from 0 to 100% and whether the feature vector could be retrieved noted.

[1679.2235398,-1555.40390834,-1140.07728186,-1999.85500108]

Figure 4: Facial feature vector

The experimental results show that the imprinted feature vectors are successfully retrieved, even when the images are altered in a significant manner. Figure 5 illustrates the averaged percentage of images where the feature vector was successfully retrieved. The printing of the black boxes (signifying modification) does consequently affect the mapped indexes' values. Therefore, many of the imprints become useless after such an attack. However, despite massive destruction on the image with the increased rate of the modification, it is possible to recapture the feature vector from some of those images, even up to 95% modification with some images.

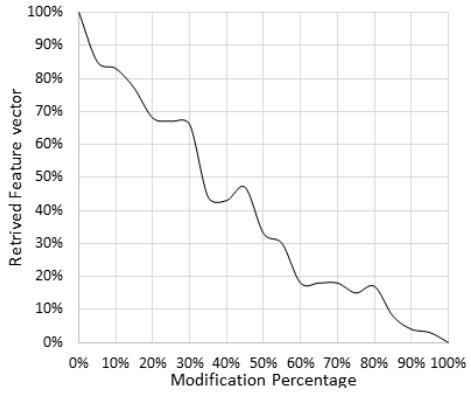

Figure 5: Number of images with successful retrieved feature vectors under multiple parts modification attack

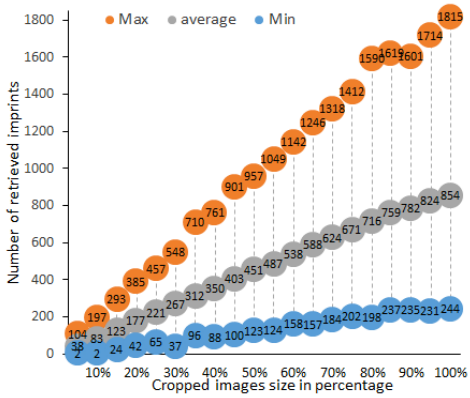

Figure 6: partial image attack

In the second experiment, the results show the feature vector is retrieved and reassembled 100% among all the tested images. This means that by giving only part of the original imprinted image (as small as 5% of the original file), it is possible to restore the feature vector to its original state. In addition, the results in Figure 6 shows that the average, maximum, and minimum numbers of a retrieved feature vector cross on all examined images (i.e. 100 images). However, these results were obtained by assuming that the preserved indexes of the hexadecimal values of interest are not changed after the cropping process. This means that all of the imprints in the database are correlated with the questioned samples as a part of the original images. In practice this is not always possible since the original object might not be accessible or available after the imprinting process took a place.

4.2. Reactive Approach - Network Fingerprinting

Users can perform a wide range of activities within different online services, such as reading news, sending

emails, editing documents, and chatting with friends. If users were "profilable" based upon their interactions with online services (i.e. not only knowing which service they are using but also what they are doing with the service such as uploading pictures, listening to audio, watching a video, instant messaging) rather than merely relying upon their IP addresses (as the current state of the art presents), it would be easier for network security analysts to track down involved individuals when an incident occurs.

The novel approach to feature extraction seeks to remove the noise and low-level information created at the packet-level and attempt to derive application – level user interactions. Based upon this hypothesis, 9 Internet services (i.e. BBC, Dropbox, Facebook, Google, Hotmail, Twitter, Skype, YouTube and Wikipedia) were selected due to their popularity [45].

Various user actions within each aforementioned service were performed and their corresponding network signals were collected via Wireshark. In order to increase the accuracy of the analysis, each user action was manually tested 20 times by three individual researchers to explore the relationship between user actions with their network signals. A list of network metadata parameters were utilized for the analysis: the date and time stamp, the IP addresses of source and destination, their port numbers, protocol ID (either TCP or UDP), the length of a datagram, and several TCP flags (e.g. SYN, FIN, ACK and PUSH). In general, user actions can be represented in three forms via their network metadata: a single packet, multiple packets and a stream of packets [46]. When a user action is carried out, 1) only one packet is sent to the server; 2) 2-4 packets which are sent to the server within less than a one-millisecond timeframe; 3) a stream of packets (with similar payload size) that are sent to the server within 100 microseconds one after another . Table 1 illustrates the results of this process and also describes in detail the form that each interaction pattern in each application.

In total, 46 users completed the data collection process during the period of 12 November 2014 to 11 January 2015. Also, ethical approval was obtained from the Faculty Research Ethics Committee of the authors' university. A total of 112 Gigabytes of IP header information was accumulated. Each user's data is stored in an individual SQLite database file; also, each record contains the following fields: a time and date stamp (e.g. 2015.01.15.22:38:05.587341), sender's IP address and port number, receiver's IP address and port number, the packet length, the type of traffic (e.g. TCP) and the flags (e.g. SYN and FIN).

Based upon the above observation from Table 1, a user's interaction databases were created for each of the 9 chosen Internet services (as illustrated in Table 2). In total, 5,285,384 activities records across 9 services were collected from the 46 users during the aforementioned experimental period.

Table 1: User Actions Derived from 9 chosen services

User Interactions	No. of packets	Total length (bytes)	Directions
BBC			
Page navigation	Stream	Various	Server > Client
Watching video	Stream	~MTU	Server > Client
Listening to audio	Stream	~MTU	Server > Client
Dropbox			
Download files	Stream	~MTU	Server > Client
Upload files	Stream	~MTU	Client > Server
Folder navigations	Two	155	Client > Server
Facebook			
Page loading	Stream	Various	Server > Client
Attach files	One	~MTU	Server > Client
Chat	2	2,625+	Client > Server
Typing	2	1,502	Client > Server
Google			
Page navigation	Multiple	268	Server > Client
Editing	Multiple	270	Server > Client
Hotmail			
File attachments	Stream	~MTU	Client > Server
Compose an email	One	981	Server > Client
Insert a recipient	One	917	Client > Server
Skype			
Text messages	One	794+	Client > Server
Audio calls	Stream	129-147	Both Clients
Video calls	Stream	1165-1365	Both Clients
File transfer	Stream	~MTU	Sending Client> receiving client
Click on contacts	One	731	Client> Server
Idle	One	572	Client > Server
Twitter			
Tweet	One	192+	Client > Server
Upload	Stream	~MTU	Client > Server
Click contact	One	747	Client > Server
Ideal / Reading	Multiple	625+	Server > Client
Wikipedia			
Viewing	Stream	~MTU	Server > Client
YouTube			
Watch videos	Stream	~MTU	Server > Client
Upload videos	Stream	~MTU	Client > Server

Table 2: An Overview of User Interactions with Each Service

Services	Users	Single	Multiple	Stream	Total
BBC	30	9.64%	15.60%	74.76%	36,994
Dropbox	31	71.21%	26.51%	2.28%	108,168
Facebook	46	63.53%	30.95%	5.52%	1,619,741
Google	46	60.09%	34.39%	5.52%	878,434
Hotmail	45	67.99%	28.20%	3.81%	227,469
Skype	31	73.62%	14.48%	11.89%	260,623
Twitter	46	56.70%	35.34%	7.96%	172,213
Wikipedia	44	12.00%	31.81%	56.19%	14,720
YouTube	45	44.39%	43.26%	12.35%	1,967,022

In order to successfully identify users through their service interactions, a classifier that can discriminate individual users based upon their behavioral activities is required. Several Artificial Intelligence (AI) techniques, including Feedforward Multi-Layer Perceptron (FF MLP) neural network, Radial Basis Function (RBF) neural network, and Self-Organising Map (SOM), can be utilized for this purpose. Amongst them, FF MLP was chosen as the default classifier based upon its previous performance within the transparent authentication domain [40, 47, 48].

With the aim of investigating the feasibility of identifying users via their network activities, a set of experiment was conducted upon the 9 Internet services of the 46 users' two month data. The experiment was carried out within the Matlab R2013a environment on a Windows 7 Enterprise 64-bit Operating System with Intel Core i7-2600 CPU (3.4 GHz) and 16 GB memory. Based upon the outcome of some initial testing, a network size of 40 neurons FF MLP with the *trainlm* training function was utilized. In order to ensure sufficient data for both training and validation processes, a minimum 200 of user interactions were set for each service; also, a fixed size (i.e. 60% of the predefined threshold) was set to the training data that was randomly selected; while if uneven data sizes were chosen, a bias would be introduced by the neural network towards those with more training samples. Moreover, all features apart from the user's IP address were used from each record as the input vector; obviously, a classifier would not be required if the user's IP address were utilized due to it is distinct uniqueness within the given dataset. As the classifier was configured in the identification mode, the True Positive Identification Rate (TPIR) is used to measure the performance accuracy. By using this setup, a set of experiments were conducted on all users and services and each configuration was repeated 5 times, and the average performance was calculated.

Once the experiment was completed, the result for each user's all services was accumulated to check the overall performance. In general, the average TPIR is 47.5%, respectively for all users' all services, suggesting almost half of the network traffic activities can be identified to individual user by using user's interaction. In addition, a breakdown of the results for each user is illustrated in Table 3. The Table contains the user IDs, the Total Number of User Interactions (TNUI) during the chosen period, and the performance. The table is sorted based upon the user ID. The best and worst performance was 86.3% and 12.6% (in terms of TPIR), achieved by users 40 and 2 respectively.

Table 3: Overall identification results for each user with all services

User ID	TNUI	TPIR (%)	User ID	TNUI	TPIR (%)
1	9,835	62.8	24	97,688	81.8
2	82,165	12.6	25	16,936	54.7
3	27,666	32.8	26	36,248	27.1
4	441,831	19.2	27	21,828	67.1
5	104,086	69.2	28	24,232	41
6	58,469	51.5	29	296,048	43.4
7	72,374	23.1	30	111,951	45.9
8	83,873	43.8	31	65,628	51.6
9	310,364	42.9	32	319,239	68.6
10	34,310	39.1	33	961,357	51.9
11	40,496	39.6	34	206,768	32.9
12	171,949	26.8	35	95,501	73.7
13	50,776	42.7	36	150,750	56.4
14	34,337	42.2	37	26,297	51
15	147,776	39.2	38	113,918	31.1
16	8,492	29.9	39	137,543	59.6
17	1,992	59.4	40	2,270	86.3
18	24,701	41.6	41	104,722	56.7
19	71,610	49.7	42	16,662	73.9
20	118,655	28.6	43	13,952	46
21	28,328	16	44	97,848	64.1
22	732	47	45	290,472	63.5
23	64,955	29	46	87,754	67.4

The objective of this study was not to obtain perfect identification but rather seek to reduce the volume of traffic an investigator would have to analyze – a reduction on average of half is excellent.

5. Discussion

The use of transparent biometrics to support the proactive and reactive identification of insider misuse is a flexible and robust approach to providing a strong link between information and the use of information by employees. The experiments performed into biometric imprinting within digital objects focused upon its ability to robustly embed and retrieve a biometric signature even after significant modification – an expectation in many cases of information leakage. Using stenographic approaches, the experiments have shown it is possible, even with 95% of the image having been modified. The experiments into the

reactive identification of users from network traffic has also shown that up to 86% of traffic can be successfully attributed – which provides for a huge reduction in traffic analysis. Used alongside IP address, which for small windows of time can be assumed to come from the same user, provides for a very effective approach reducing the volume of data to be analyzed. Furthermore, these techniques could also be utilized together to provide further verification that the digital object imprinted by a user was indeed sent by the same user (although albeit to a lower confidence).

That said, there a number of challenges that exist that require further research to investigate:

- the ability to utilise a wider range of digital objects. Differing objects have varying degrees of stability due to their structure. For example, word documents and their underlying data structure can change considerably given small alterations to a file. Therefore, alternative approaches need to be developed that overcome this (e.g. the use of higher-level objects whose orientation does not move)
- the use of multibiometric signals within the imprinting process. How much data can reliable be imprinted and what effect will this have when the file is modified. It is expected there will be a trade-off between the volume of imprinted data and the percentage a digital object can be modified.
- the identification performance that can be achieved within a multibiometric system – using multi-instance, multi-sample and in the proactive approach multimodal techniques.
- how to apply both classification and logic to user interactions to further reduce the volume of traffic that needs to be analyse
- visualisation techniques that can present, interrogate and interact with digital objects and network traffic to provide a responsive and user friendly forensic tool
- the scalability and performance characteristics of an operational system. How much processing is required to parse and present the data. Can the system achieve near real-time visualisation of user traffic? This begins to turn the reactive approach also into a more proactive SIEM based approach.

It is worth highlighting in both approaches, there is not a need (although potentially desirable) to verify the identity perfectly/correctly. The approaches exist to narrow the focus of an investigation to a subset of individuals who investigators can then focus upon in more upon. So even in cases where the perpetrator was not listed first in the identification process, as long as their name appears towards the top of the list, this would provide the capability for investigators to more quickly identify the guilty and build a case of evidence against them. In organizations with large volumes of employees this will be particularly useful.

6. Conclusions and Future Work

The paper has presented a holistic approach to aid an incident analyst in establishing and investigating a case of insider misuse – particularly with respect to information leakage, although the reactive approach would be useful in all cases. The resulting experiments performed to validate the fundamental approach have proved very encouraging.

The ability to identify an individual is however largely dependent upon the performance of the underlying biometric modalities utilized and further work is required on developing these individually and in the case of the proactive approach also multimodal/multibiometric models. Further work will also be undertaken into investigating the capabilities with regards to forgability and circumventabilty in order to ensure the system is robust against targeted attack.

7. Acknowledgement

This research was undertaken with the support of the Engineering and Physical Sciences Research Council (EPSRC) grant EP/K03345X/1 “Identifying and Modelling Victim, Business, Regulatory and Malware Behaviours in a Changing Cyberthreat Landscape”.

8. References

[1] A. Vance, P. B. Lowry, and D. Eggett “Increasing Accountability through User-Interface Design Artifacts: A New Approach to Addressing the Problem of Access-Policy Violations”, MIS Quarterly, Edition 2, Volume 39, Issue 2, Pages 345–366, 2015.

[2] M Siponen and A Vance “Neutralization: New Insights into the Problem of Employee Information Systems Security Policy Violations”, MIS Quarterly, Edition 3, Volume 34, Issue 3, Pages 487-502, 2010.

[3]A. Shabtai, Y. Elovici, and L. Rokach, *A Survey of Data Leakage Detection and Prevention Solutions*, 1st ed. Boston, MA: Springer US, 2012.

[4] M. L. Collins, D. Spooner, D. Cappelli, A. P. Moore, and R. F. Trzeciak, “Spotlight On : Insider Theft of Intellectual Property inside the U . S . Involving Foreign Governments or Organizations,” 2013.

[5] C. L. Huth, "The insider threat and employee privacy: An overview of recent case law," *Comput. Law Secur. Rev.*, vol. 29, no. 4, pp. 368–381, 2013.

[6] J.S. Valacich, J.L. Jenkins, J.F. Nunamaker, S. Hariri, J. Howie "Identifying Insider threats through monitoring mouse movements in concealed information tests" In: Proceedings of the HICSS-46 symposium on credibility assessment and information quality in government and business; 2013.

[7] S. Lee, M. Park, J. Eom, and T. Chung, "PDT-BI: Proactive Detection Technology based on the Biometric Information for Preventing Internal Information Leakage," *Int. J. Bio-Science Bio-Technology*, vol. 5, no. 5, pp. 187–196, Oct. 2013.

[8] P. Papadimitriou and H. Garcia-Molina, "Data Leakage Detection," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 1, pp. 51–63, Jan. 2011.

[9] S. Kale and P. S. V Kulkarni, "Data Leakage Detection," *Int. J. Comput. Sci. Inf. Technol.*, vol. 1, no. 9, pp. 668–678, 2012.

[10] R. Jadhav, "Data leakage detection," *Int. J. Comput. Sci. Commun. Networks*, vol. 03, no. 01, pp. 37–45, 2012.

[11] J. Chavan and P. Desai, "Relational Data Leakage Detection using Fake Object and Allocation Strategies," *Int. J. Comput. Appl.*, vol. 80, no. 16, pp. 15–21, 2013.

[12] C. Shields, O. Frieder, and M. Maloof, "A system for the proactive, continuous, and efficient collection of digital forensic evidence," *Digit. Investig.*, vol. 8, no. SUPPL., pp. 3–13, 2011.

[13] G. Magklaras, S. Furnell, and M. Papadaki, "LUARM – An Audit Engine for Insider Misuse Detection," *Int. J. Digit. Crime Forensics*, vol. 3, no. 3, pp. 37–49, Jan. 2011.

[14] M. I. Cohen, D. Bilby, and G. Caronni, "Distributed forensics and incident response in the enterprise," *Digit. Investig.*, vol. 8, pp. S101–S110, Aug. 2011.

[15] M. Rafique and M. N. A. Khan, "Exploring Static and Live Digital Forensics: Methods, Practices and Tools," *Int. J. Sci. Eng. Res.*, vol. 4, no. 10, pp. 1048–1056, 2013.

[16] A. Sperotto, G. Schaffrath, R. Sadre, C. Morariu, A. Pras, B. Stiller, "An Overview of IP Flow-based Intrusion Detection" In: IEEE Communications Surveys & Tutorials, 12 (3). pp. 343-356. 2010

[17] P. Winter, E. Hermann, M. Zeilinger "Inductive Intrusion Detection in Flow-Based Network Data Using One-Class Support Vector Machines," in New Technologies, Mobility and Security (NTMS), 2011 4th IFIP International Conference on , vol., no., pp.1-5, 7-10 Feb. 2011 doi: 10.1109/NTMS.2011.5720582

[18] Z. Jadidi, V. Muthukkumarasamy, E. Sithirasenan; M. Sheikhan "Flow-Based Anomaly Detection Using Neural Network Optimized with GSA Algorithm" in Distributed Computing Systems Workshops (ICDCSW), 2013 IEEE 33rd International Conference on , vol., no., pp.76-81, 8-11 July 2013 doi: 10.1109/ICDCSW.2013.40

[19] R. Hofstede; V. Bartos; A. Sperotto; A. Pras, "Towards real-time intrusion detection for NetFlow and IPFIX" in Network and Service Management (CNSM), 2013 9th International Conference on , vol., no., pp.227-234, 14-18 Oct. 2013 doi: 10.1109/CNSM.2013.6727841

[20] N.L. Clarke, S. Karatzouni, S.M. Furnell (2008) "Transparent Facial Recognition for Mobile Devices", Proceedings of the 7th Security Conference, Las Vegas, USA, 2nd-3rd June, 2008

[21] DARPA (2016) "Active authentication" http://www.darpa.mil/program/active-authentication

[22] K. Niinuma and A. K. Jain "Continuous user authentication using temporal information" in Proc. SPIE, 2010, vol. 7667, p. 76670L.

[23] S. Islam, R. Davies, A.S. Mian, M. Bennamoun "A fast and fully automatic ear recognition approach based on 3D local surface features". Advanced Concepts for Intelligent Vision Systems. Lecture Notes in Computer Science, vol. 5259, pp. 1081–1092. Springer, Berlin 2008

[24] P.N.A. Fahmi, E. Kodirov, D.J. Choi, G.S. Lee, A. Mohd Fikri Azli, S. Sayeed"Implicit authentication based on ear shape biometrics using smartphone camera during a call". In: 2012 IEEE InternationalConference on Systems, Man, andCybernetics (SMC), pp. 2272–2276.

[25] T. Feng, Z. Liu, K.A. Kwon, W. Shi, B. Carbunar, Y. Jiang, N. Nguyen "Continuous mobile authentication using touchscreen gestures" In: 2012 IEEE Conference on Technologies for Homeland Security (HST), pp. 451–456. IEEE (2012)

[26] P. Koundinya, S. Theril, T. Feng, V. Prakash, J. Bao, W. Shi "Multi resolution touch panel with built-in fingerprint sensing support". In: Design, Automation & Test in Europe Conference & Exhibition (DATE), pp. 1–6. IEEE Conference Publications, New Jersey (2014)

[27] K. Mock, B. Hoanca, J. Weaver, M. Milton "Real-time continuous iris recognition for authentication using an eye tracker". In: Proceedings of the 2012 ACM Conference on Computer and Communications Security, pp. 1007–1009.

[28] M. Pusara, C.E. Brodley "User re-authentication via mouse movements" In: Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security, pp. 1–8. ACM Press, New York, NY, USA

[29] N. Zheng, A. Paloski, H. Wang "An efficient user verification system via mouse movements", Proceedings of the 18th ACM conference on Computer and communications security pp139-150 ACM New York, NY, USA 2011

[30] S. Mondal, P. Bours "Continuous authentication using mouse dynamics" In: 2013 International Conference of the Biometrics Special Interest Group, pp. 1–12. IEEE (2013)

[31] Aksari, Y., Artuner, H.: Active authentication by mouse movements. In: ISCIS 2009, 24th International Symposium on Computer and Information Sciences, 2009, pp. 571–574. IEEE (2009)

[32] Lin, C., Chang, C., Liang, D.: A new non-intrusive authentication approach for data protection based on mouse dynamics. In: 2012 International Symposium on Biometrics and Security Technologies, pp. 9–14. IEEE (2012)

[33] R. Monrose, A. Rubin "Keystroke dynamics as a biometric for authentication". Future Gener. Comput. Syst. 16(4): 351–359, 1999

[34] P.S. Dowland, H. Singh, S.M. Furnell "A preliminary investigation of user authentication using continuous keystroke analysis". In: 8th IFIP Annual Working Conference on Information Security Management and Small System Security (2001)

[35] J. Roth, X. Liu, D. Metaxas "On continuous user authentication via typing behavior" IEEE Trans. IMAGE Process. 23, 4611–4624 (2014)

[36] M. Kunz, K. Kasper, H. Reininger, M. Möbius, J. Ohms "Continuous speaker verification in realtime" In: Proceedings of the Special Interest Group on Biometrics and Electronic Signatures, BIOSIG, vol. 2011, pp. 79–88 (2011)

[37] M. Abdullah, H. Bashier, S. Sayeed, I. Yusof, A. Azman, S.Z. Ibrahim, T.H. Liew "Answering incoming call for implicit authentication using smartphone" J. Theor. Appl. Inf. Technol. 61, 193–199 (2014)

[38] M.O. Derawi, D. Gafurov, P. Bours "Towards continuous authentication based on gait using wearable motion recording sensors" In: Traore, I., Ahmed, A.A.E. (eds.) Continuous Authentication Using Biometrics: Data, Models, and Metrics, pp. 170–190. IGI Global, Hershey (2012)

[39] H. Lu, J. Huang, T. Saha, L. Nachman "Unobtrusive gait verification for mobile phones". In: Proceedings of the 2014 ACM International Symposium on Wearable Computers—ISWC '14. pp. 91–98. ACM Press, New York, NY, USA (2014)

[40] F. Li, N.L. Clarke, M. Papadaki, P.S. Dowland "Active authentication for mobile devices utilising behaviour profiling", International Journal of Information Security, Volume 13, Issue 3, pp 229-244, ISSN:1615-5262, 2014

[41] N., Clarke (2011) "Transparent User Authentication", Springer, ISBN 978-0-85729-804-1, pp229

[42] A. Alruban and N. Clarke "Method Of Associating A Person With A Digital Object" GB Patent filing GB1609673.7 dated 2nd June 2016

[43] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 7, pp. 711–720, 1997.

[44] G. Schaefer and M. Stich "UCID: an uncompressed color image database" in SPIE 5307, Storage and Retrieval Methods and Applications for Multimedia 2004, pp. 472–480

[45] Alexa "Top Sites In United Kingdom", http://www.alexa.com/topsites/countries/GB

[46] F. Li, N.L. Clarke, G. Alotibi, D. Joy (2015) "Forensic Investigation of Network Traffic: A Study into the Derivation of Application-Level features from Network-Level Metadata", 6th Annual International Conference on ICT: Big data, Could and Security (ICT-BDCS 2015), 27-28 July, , 2015, ISSN: 2382-5669, pp68-73

[47] N.L. Clarke, S.M. Furnell "Authenticating Mobile Phone Users Using Keystroke Analysis" International Journal of Information Security, 6(1), 1-14, 2006

[48] H. Saevanee, N.L. Clarke, S.M. Furnell, V. Biscioneb "Continuous user authentication using multi-modal biometrics", Computer & Security, Vol. 53, pp 234-246, September 2015