Geo Forensics: Classes of Locational Data Sources for Embedded Devices

Peter Hannay

***Abstract*—A number of devices, web services and applications are being released with, or updated to be locationally aware. The use of location data can be used for a wide variety of purposes, including navigation, social networking, data mining and providing localised content. This location data has potential for establishing a locational history for these devices. The sources of this location data exceed the global positioning system (GPS) based data, including pre- and post-incident triangulation of mobile and cell towers, images and network histories. This paper proposes a classification framework ranking the reliability of potential evidence. This ranking is dependent on the intended purpose of the mechanism involved the generation of such. The classification classes proposed are implicit, connectivity based and metadata, each representing a different level of confidence and identifying features.**

***Index Terms*—Embedded forensics, locational forensics, digital forensics, forensics, GPS, embedded computing**

I. Introduction

The rise of location aware devices within the consumer market has been a gradual process over the last decade. Location aware devices range from cameras, GPS devices, video game consoles and phones through to automotive, control systems and general purpose computers [1]. The availability of locational information in these devices is not always an intentional design choice, for example a device wirelessly connected to a Wi-Fi access point (AP) or cellular tower may record the service set identifier (SSID) or cellular id of the connected network [2]. Such recordings can be compared against available databases such as those offered by SkyHook and Google[2] which match SSIDs and cellular IDs against physical locations. Other devices use locational information as part of their core design, these include navigation devices, mobile phones and tracking systems. Devices in both of these categories will be discussed in the relevant section of this paper.

Social media services handle an ever-increasing amount of personal data, which often includes locational data either within the social media data format or embedded in an associated artifact such as an image captured by a digital camera. This locational data can be in the form of message content, metadata associated with images or other postings. In the case of images the data is present as specific geolcational Exchangeable Image File Format (EXIF) tags containing co-ordinates of where the image was captured [3]. Alternately, cached social media data may include textual descriptions of locations provided by the user or metadata previously submitted to an online service. The presence of information of this type can provide locational information even where no traditional locational logging exists. For each of the devices examined, the presence of data from social media services will be noted, but not analysed in depth.

In this paper the author aims to classify the types of locational artifacts available to digital forensics investigators based on the properties of the functionality, which lead to their creation. The author aims to show that these properties provide an ideal classifier due to the impact that these properties have on confidence in locational history. In order to demonstrate the proposed classifications, several case studies will be provided to demonstrate the application of these classifiers for selected devices.

The privacy impacts of social media services have been widely commented on in the media, academic literature and Internet publications. These implications will not be discussed here as they are out of scope for this publication.

II. Proposed Classifications

A number of classes of locational data need to be defined in order to adequately classify the forms of information available from devices. Once defined, these classes may allow for better understanding of the evidentiary value of a particular artefact based on said class. The classes defined below are based on the intended purpose of the functionality which generated the associated artefacts. It is proposed that these purposes be used to classify any resulting locational data. A summary of these classifications is shown below in TABLE .

TABLE I: A Summary of the Proposed Classes

Class	Identifying Features	Confidence
Implicit	- Locational by design - Locational information and confidence can be determined without significant external information	High
Connectivity Based	- Locational information dependent on external data sources - Requires additional information to determine location	Variable
Metadata	- Embedded within an artifact as part of secondary functionality - Requires additional information to determine confidence	Limited

A. Implicit

The implicit class is used to classify artefacts produced by the device after an action is performed by an application or algorithm. Such applications and algorithms will have locational awareness tied into their core functionality. Items

Manuscript received December 20, 2012; revised March 5, 2013.

P. Hannay is with the Edith Cowan University Security Research Institute, Perth, WA 6027 Australia (e-mail: p.hannay@ecu.edu.au).

DOI: 10.7763/IJET.2013.V5.555

of potential evidence originating from features of devices such as navigation, localised search results, mapping, etc. would fall under this classification. As the required accuracy for these individual applications can be determined through post-incident analysis we can place a high level of confidence in determined accuracy of this data, limited of course by conditions and instrumentation. In the case of Navstar GPS, additional confidence may be able to be obtained through the use of almanac data to determine the number of satellites overhead at the time the artifact was generated [4].

B. Connectivity Based

External connectivity has become a pervasive feature of embedded systems. The characteristics of this connectivity can be utilised in order to determine location. Wi-Fi enabled devices potentially store a history of Wi-Fi networks that have been seen or connected to previously. This data can be combined with information from databases such as those provided by Google and Skyhook to determine historic location [5]. In the same manner devices with cellular radios may carry a history of CellIDs, which can be cross-referenced with databases such as OpenCellID to achieve a similar outcome [6]. There are other examples of connectivity mechanisms, which provide some potential locational history; these often depend on the availability of external databases or manual intelligence gathering. The confidence level in data gathered from these sources is dependent on the range of the radio hardware contained within the device, as well as the accuracy of historic reference data.

C. Metadata Based Location

Metadata based locations are those which are located within other media generated by the device as part of its functionality. An example of this would be locational data embedded within the data that is sent as a “post” to a social media website. Whilst not always visible on the site itself, this data is often available via Application Programming Interfaces (APIs) when present. Another example of such data is that present in the geolocational EXIF data of images when taken with a locationally aware camera, mobile phone or other device. It is common for such images to have location data embedded within as metadata [7]. Confidence in data of this class is often limited due to the “instant” nature of such artefacts, where immediate availability is far more critical to operation than accuracy. As a result of these limiting factors it is often the case that coarse or last known locations are often used in lieu of current or accurate data [8].

III. Case Studies

The devices selected for case studies were chosen as they each have a combination of features allowing for each classification to be demonstrated. In each of these the features of the device will be discussed and the details of potential locational evidence outlined. The information contained within this section is primarily the result of original research, in which the devices mentioned were examined and analysed in a forensic manner.

A. TomTom One Satellite Navigation Unit

The TomTom One is an automotive satellite navigation unit such as mentioned below. The device has a SirfStar III GPS receiver, which is used to determine location for navigational purposes. In addition to this the One contains a Bluetooth radio which allows for synchronisation of contacts, text messages and hands free functionality from compatible mobile phones. From these pieces of functionality we can see that there is potential for both implicit and connectivity based locational data to be discovered. In this case, metadata based artefacts have been disregarded as there is no secondary functionality allowing output from the device.

Historic locational data is stored on the device to facilitate navigation to recent or favourite locations, for travel history that can be viewed by the user (either on the device or via a desktop application) or for anonymous feedback mechanisms providing the manufacturer details on road conditions. The data storage schemes for the TomTom One GPS devices will be outlined below.

There are two main information sources within the TomTom navigation units, the first is a proprietary formatted binary file named MapSettings.cfg, the second is a series of encrypted files named triplog-YYYY-MM-DD.dat (where YYYY-MM-DD is the date that the file was created).

The MapSettings.cfg file contains favourite locations, home locations and recent destinations. This file is structured in fairly simple fashion. It is broken down into a number of records, with each of these records containing a specific set of items. Each item is a specific piece of data. Each record includes seven significant items: a precision identifier, a value indicating the type of record, a textual description of the record and four sets of latitude/longitude pairs.

The trip log data files are created by the TomTom device if the device is configured to collect usage information that will later be sent to TomTom. These files are encrypted with a public key prior to being written to the disk. The matching private key is held by TomTom. As a result there is limited ability to decode these files without the cooperation of the TomTom Corporation. It should be noted that there is precedent of TomTom assisting in law enforcement matters and providing decryption of these files.

In both of the above cases we can classify the evidence as being implicit, in that it is derived from functionality, which is core to the functionality of the device. From this we know that in order for the device to function correctly the accuracy of the signal needs to be of significant quality.

B. Nintendo Wii

The Wii video game console is limited in terms of potential for location data due to its atypical stationary nature. The locational elements discovered exist in terms of image storage, Wi-Fi BSSIDs, IP networking and messaging applications. The image storage mechanism is used to view images on the console as well as send/receive these from other Wii users. It is possible that the images stored on the local console may contain EXIF data showing where they were captured, this does not necessarily tie these images to the owner or primary user of the console, however. Any data harvested from these images simply provides indication of where the image itself may have been captured. This is a prime example of metadata-based data that is embeeded in artefacts. In these cases we have to examine the originator of these images and determine a level of confidence based on this additional information.

The Wii and other devices that make use of cellular or Wi-Fi connections may contain data such as the most recent network used, saved networks or a historical list of networks that have been seen. In each of the aforementioned cases there is potential to determine the historical location of these devices by looking up the associated ID in a number of databases. Rigour is essential when using this type of information as databases may be out of date, the associated access point may have moved since the device was within range or inaccuracies in the database may be present due to ID spoofing / errors in data collection. In order to increase the accuracy of findings, multiple databases should be compared and where it is feasible the site should be visited for manual validation.

A number of the devices examined contain Media Access Control (MAC) addresses of Wi-Fi access points and/or mobile Cell IDs. Whilst these do not contain significant location data within themselves (aside from the Cell ID which contains a prefix indicating the country and network) there are a number of databases which link the two. The two largest such databases are owned by Skyhook Wireless and Google, additionally projects such as Open Cell ID catalog cellular tower IDs only.

From the above it can be determined that the Wii has the potential to deliver evidence which would be classified in the category of "connectivity based" and as such has a variable confidence rating. These results are relevant as without a determination in confidence for the third party geolocation data sources used, it is not possible to make a firm statement with regards to the accuracy of any location extracted.

C. *Kindle e Book Reader (Kindle Paper White with 3G)*

The Kindle is an eBook reader which features a cellular modem and Wi-Fi capabilities. These features are primarily used for content delivery, however there are some native locational features. Locational information is used to determine regions for content licensing purposes as well as providing basic mapping functionality with location provided by Assisted Global Positioning System (AGPS) receivers [9]. As a result of this functionality the potential locational history obtainable from the device is based on AGPS, cellular towers and Wi-Fi SSIDs.

AGPS allows for location to be retrieved through a carrier assisted mechanism, in which the client device requests a location and the mobile carrier provides this based on triangulation using the cell sites within rage. As this mechanism relies on the carrier as well as terrestrial based triangulation, the confidence in this must consider the physical parameters of the surrounds, multipath effect and atmospheric conditions. As a result, evidence generated by this mechanism would be considered to be connectivity based, in which confidence is variable. The other available mechanisms of utilising cellular ID and Wi-Fi BSSID databases have already been discussed and were also to be attributed a variable confidence, however in this case the presence of multiple independent connectivity based sources of data may lead to a higher confidence level when investigated.

D. *Lumix DMC-TS3 (GPS Enabled Camera)*

The Lumix DMC-TS4 is a ruggedized point and shoot camera with inbuilt GPS functionality. The main purpose of this device is to record images and video; the inclusion of GPS functionality for tagging images is considered a secondary function. It is arguable that for the purposes of classification this data is "locational by design." However on examination of the device it is apparent that locational history is only stored within image metadata. As the only source of locational data is metadata and the purpose of the functionality is to generate the same, we can see that the appropriate item in the classifier is metadata.

The aforementioned classification appears to be apt, as with many devices that produce metadata the information provided is a 'best effort' based on the information available at the time of image creation. In the case of this device, if a current GPS signal is not available the device writes the last known location. This condition can be verified by comparing the GPS time embedded in the image with the local device time. In many cases it can be seen that hours or days pass without a valid GPS signal. As such there can only be limited confidence in the data written to image/video metadata.

E. *iPhone 3GS*

The iPhone 3GS has locational capability spanning all of the categories defined so far: AGPS, GPS, Wi-Fi and cellular. We can see that there is ample potential source for locational information, however we need to analyse the functions that may generate locational data if we are to classify the output. In terms of functionality we have mapping/navigation, location aware search, image geotagging and the potential for significant additional functionality through third party applications or future upgrades to the operating system.

In the case of this device it is required that classification be performed on a per item basis, considering the purpose of the specific functionality which has generated the data being analysed, see TABLE for examples. For example if we were to consider artefacts created through the use of navigation functionality we would classify this as implicit and allocate a high degree of confidence.

TABLE II: EXAMPLE CLASSIFICATIONS FOR IPHONE3GS FUNCTIONS

Function	Intended Function	Classification
Maps	Navigation	Implicit
Twitter Application	Sharing social media information	Metadata
Wi-Fi Connection	Providing data connectivity	Connectivity Based
Cellular Connection	Providing data & voice connectivity	Connectivity Based
Geotagging	Providing locational information in image EXIF records	Metadata

IV. LIMITATIONS

The devices outlined demonstrate the application of the classification system proposed. From this we can see that the proposed classes and related confidence ratings fit the cases provided. It is recognised that the major limitations of this classification system is the lack of granularity and application to a wide number of devices. As such significant further study and modification to this classification would be required prior to mainstream adoption; this classification is intended as an initial proposal only. It is anticipated that through input and continued research it will be possible to provide a more refined, granular and useful classification system that will apply to additional devices and scenarios. For these reasons input and further proposals in this regard are welcomed.

V. Conclusion

As general purpose computer use continues its decline in relative to increasing use of consumer grade embedded devices such as tablet computers, media players, video game consoles, ebook readers and other devices, we need to consider the additional evidentiary potential provided by these devices. Location aware technology in this class of embedded device is widely employed and produces data artefacts of significant evidentiary value. However in making use of this data for evidentiary purposes it is necessary to classify extracted artefacts in such a way as to define the confidence that should be placed in this data.

The proposed classification is based on the concept of intended functionality determining a minimum level of confidence that can be placed in any generated artefact. The proposed classifications are “implicit” a level granted where the functionality requires accurate locational data for core functionality, navigation is one example of this. The second “connectivity based” is identified in cases where functionality leaves identifiers tying the device to a specific point of connection, such as a Wi-Fi access point. The final classification “Metadata” is granted to functionality, which uses locational data as additional functionality, such as geotagging a photo.

It is important to that such a classification mechanism exist so that a reliable and usable set of criteria for reliability of such data can be established. Such a classifications system would prove essential for use in the legal process where digital location information is concerned.

The author has proposed a classification system for this purpose, which meets these needs in a coarse manner and is undergoing constant refinement. Through collaboration, ongoing research and practical examination of devices these outcomes will be achieved. Input during this process is invited and encouraged.

References

[1] S. A. Ahson and M. Ilyas, *Location-based services handbook: applications, technologies, and security*: *Taylor and Francis Group*, 2011.

[2] X. Wang, A. K. S. Wong, and Y. Kong, "Mobility tracking using GPS, Wi-Fi and Cell ID," in *Proc. Information Networking (ICOIN), 2012 International Conference on*, 2012, pp. 171-176.

[3] C. Valli and P. Hannay, "Geotagging Where Cyberspace Comes to Your Place," presented at the Proceedings of the 2010 International Conference on Security and Management, Las Vegas, 2010.

[4] J. J. Spilker, *The global positioning system: theory and applications* vol. 2: Aiaa, 1996.

[5] C. C. Post and S. Woodrow, *Location is Everything*, 2008.

[6] S. A. Shad and E. Chen, "Precise Location Acquisition of Mobility Data Using Cell-id," *arXiv preprint arXiv:1206.6099,* 2012.

[7] K. Cohen, "Digital still camera forensics," *Small Scale Digital Device Forensics Journal,* vol. 1, pp. 1-8, 2007.

[8] H. S. Lallie and D. Benford, "Challenging the Reliability of iPhone Geo-tags," *Forensic Computer Science Ijofcs,* vol. 6, pp. 59, November 1 2011.

[9] J. Stern. (2008, May 22). *Amazon Kindle Tips and Tricks*. [Online]. Available: http://www.laptopmag.com/advice/tips/amazon-kindle-tips-and-tricks.aspx?page=2

Peter Hannay is a Ph.D. student, researcher and lecturer based at Edith Cowan University in Perth Western Australia. His PhD research is focused on the acquisition and analysis of data from small and embedded devices. In addition to this he is involved in smart grid research and other projects under the banner of the ECU Security Research Institute.