Disseminating Location Privacy Threats and Protection Techniques through Capture-the-Flag Games (Vision Paper)

Gabriel Ghinita
UMass Boston
100 William T. Morrissey Blvd
Boston, MA 02125
Gabriel.Ghinita@umb.edu

Mihai Maruseac
UMass Boston
100 William T. Morrissey Blvd
Boston, MA 02125
mmarusea@cs.umb.edu

ABSTRACT

Recent years witnessed a tremendous growth in the area of mobile computing. Users with mobile devices are able to access services customized to their geographical coordinates, and to engage in complex interactions with other users in their proximity. However, in addition to its many benefits, sharing location with service providers and other users also introduces serious privacy threats. If not properly addressed, the loss of location privacy can bring significant harm to mobile users. Currently, there is a low level of awareness among mobile users, and even among computer professionals, with respect to the contingent threats on location privacy, and to the approaches available to mitigate such threats. We envision *GeoCTF*, an educational capture-the-flag (CTF) - style game designed to raise the level of awareness about the dangers of uncontrolled sharing of location data, and to illustrate prominent location protection techniques. The game-based approach of *GeoCTF* makes it an effective and engaging educational tool, suitable for high-school and University students, as well as computer-literate general population mobile users.

Categories and Subject Descriptors

H.2.8 [Database Applications]: Spatial databases and GIS

General Terms

Design, Security, Human Factors.

Keywords

Location Privacy, Trajectory Protection, Education, Capture-the-Flag (CTF) Games.

1. INTRODUCTION

The increased popularity of mobile communication devices with embedded positioning capabilities (e.g., GPS) has generated

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

© 2015 ACM. ISBN 978-1-4503-3969-8/15/11...\$15.00

DOI: <http://dx.doi.org/10.1145/2830834.2830836>

unprecedented interest in the development of location-based applications. Users are able to ask queries about points of interest in their proximity, and they can generate their own content with geo-spatial tags. For instance, many online photo sharing services allow geo-spatial annotations for images, and browsing photos based on their location is made possible. In addition, location-based social networks allow users to share their whereabouts with friends, to find nearby contacts, and to provide/search for recommendations about points of interest in their proximity.

Many exciting applications benefit from the availability and potential for sharing of location information. However, uncontrolled location sharing can have dire consequences, when location data falls in the hands of malicious entities. With knowledge of user locations, an adversary can stage a broad spectrum of attacks against individuals, from physical surveillance and stalking, to identity theft, to inferring sensitive information, such as the individual's health status, alternative lifestyles, political and religious affiliations, etc.

Unfortunately, there is currently little awareness among the general population about the threats to location privacy. In fact, even among computer-literate users and computer professionals, the level of awareness about location privacy threats is rather low, not to mention the level of familiarity with protection techniques available to counter these threats. One of the main reasons is that mobile computing evolved at such an accelerated pace that the existing educational tools could not keep up with this development. The course curricula in computer-oriented programs (Computer Science and Engineering, Information Technology) have not kept up-to-date with the fast growth of the mobile applications domain. Furthermore, a significant population segment with high exposure to mobile technology is that of high-school students. At this level, a novel and effective mechanism to deliver information is required, as conventional computer-oriented courses are not appropriate for this age segment.

We propose to disseminate fundamental concepts about location privacy threats and defense techniques through a mechanism that is both accessible from an early age, as well as very effective and engaging: computer-based educational games. Specifically, we envision *GeoCTF*: a “*Capture-the-Flag (CTF)*” – style game focused on geospatial topics.

CTF games [8] are a specific type of cybersecurity competition that has gained tremendous popularity in recent years. The objective of participants in such games is to acquire secret tokens

of information, generically referred to as flags, that can be passwords, cryptographic keys, or string-formatted access codes that allow users to accede to a more difficult level of the game. Many such CTF games have been proposed, but their focus is on conventional security problems, such as cryptography, network or database security, operating system security, etc. None of the existing CTFs, however, focuses on attacks on location data.

A number of recent cybersecurity education initiatives [6-8], including several funded by federal agencies, highlighted the high effectiveness that games and competition-style training has on increasing students' interest and their ability to absorb cybersecurity concepts and techniques. In the case of CTF games, one interesting aspect is that they may combine both offensive and defensive tasks, making them an ideal vehicle for learning both the existing threats, as well as how to mitigate those threats.

We are confident that a CTF-style game focused on geospatial topics is a very promising approach to address the current lack of awareness and educational tools available for location privacy topics. Such an approach has the advantage of exposing participants to a large set of attacks and defenses, and is accessible to a broad spectrum of users, such as high-school and University students, as well as general population individuals that have a reasonable level of computer literacy.

In the rest of the paper, we outline the envisioned *GeoCTF* educational game. We present an overview of the multiple stages of the game in Section 2, followed by a brief review of related work in Section 3. We conclude and outline future work directions in Section 4.

2. GeoCTF OVERVIEW

A typical organization of a CTF game is to have multiple stages (or levels), with increasing level of difficulty. To accede to the next level, the player needs to collect a certain number of flags (in our case, flags could be for instance coordinates of mobile users). The presence of multiple stages has two benefits: first, the game is more exciting, as the players are incentivized to gather sufficient flags to accede to the next level; second, it allows players with diverse sets of abilities to participate in the game. A strong player is able to quickly accede to higher levels. On the other hand, a weaker player is also able to enjoy the game, by solving multiple challenges at a lower level.

The envisioned *GeoCTF* game will have four stages, three offensive and one defensive: (1) **Recon** (short for reconnaissance) is the easiest level, and challenges players to recover GPS coordinated from unsanitized content, such as geo-tagged images; (2) the **Data Capture and Forensics** stage requires players to analyze a packet capture obtained from one of the locations recovered in Stage 1. The additional coordinates obtained from the capture allow the player to unlock the following stage. (3) **Background Knowledge Attacks**: players combine the coordinates from Stage 2 with external public databases to derive the identity of mobile users. Finally, (4) **Data Protection** is a defensive stage where the player needs to sanitize the data using an appropriate algorithm. The algorithm will already be implemented in *GeoCTF*, the player only needs to understand the high-level details of the protection, and how to execute the sanitization algorithm using a GUI. After sanitization, the player will re-attempt the attack at Stage 3, and witness that the attack is no longer successful.

In terms of implementation, *GeoCTF* will be packaged in the form of a portable, self-contained Linux-based Virtual Machine (VM) that can be downloaded and run on any commodity PC or laptop. There will be minimal effort overhead in installing and running the VM, and each stage will consist of a set of tasks with rich visual features (e.g., GUIs, geospatial data visualization), resulting in an intuitive and exciting game environment.

Stage 1: Recon.

This stage will allow players to train their ethical hacking skills, by acting as an adversary who performs reconnaissance activities seeking to find a suitable victim who does not properly protect their geolocation data. Within the *GeoCTF* VM, we will run a web server that hosts several web pages, mimicking the content and look-and-feel of a social network application. The web pages will be rich in multimedia content, such as images and videos, some of which have geo-tags.

The player will have to inspect the contents of the web pages using simple tools, for instance a viewer of the page source, which is present in all commercial web browsers. Some of the geo-tags will be directly embedded inside the page source as hidden HTML or XML elements, as illustrated in Figure 1. In other cases, the player needs to save the image or video to a file, and then use another widely available tool (e.g., Windows Photo Viewer or GIMP) to inspect the metadata of the image/video, which may contain the geographical coordinates of the location where the image/video was shot.

Figure 1. Extracting GPS Coordinates from HTML/XML source code and/or Images/Videos.

To cater to players with diverse sets of skills, some geo-tags will be relatively straightforward to find, whereas others may require more investigation, for instance they may be encoded as base64 strings. After the player has acquired a certain amount of geo-tags (e.g., ten), then Stage 2 of the game will be unlocked.

Stage 2: Data Capture and Forensics.

At this level, players will first have to place the coordinates retrieved in the first stage on the map. For this purpose, a mock-up version of a widely-available tool such as GoogleMaps will be used, with several points of interest hard-coded in the database. The player is instructed that some of the coordinates retrieved represent places that the victim user visits frequently. By selecting a specific point of interest on the map (e.g., a popular coffee shop), the player is virtually transported to that location, simulating the scenario where the adversary is now present at that

location and is eavesdropping on the victim's communication. This can be done either by assuming a wireless access point that does not use appropriate encryption, or by having the player create his/her own access point that spoofs the coffee shop's one, and acting as a man-in-the-middle on all communication between the victim user and service providers (the player will not actually have to go through all the details of setting up an access point, these options will be presented to the player through a high-level GUI).

The player will then start a packet capture and analysis tool, such as Wireshark [9], which also has modules for interception and analysis of wireless communication, e.g., 802.11 protocol family. The capture, for which a mock-up snapshot is provided in Figure 2, will contain communication from several users, and the player must be able to isolate the data stream from each user, and then parse the stream for data items that look like geographical coordinates. To accommodate players with weaker network knowledge, we will also build a simplified Wireshark-lookalike GUI, which shields the user from the network protocol complexity, and presents the data to the user in a simplified interface.

We will use a HTTP stream that will allow the player to visually inspect and identify coordinates. For more advanced players, we will also include encoded coordinates (e.g., base64, or coordinated encrypted with a key that is also present somewhere in some other stream). The central idea of this stage is that, by intercepting communication between the victim and its location-based service (e.g., a social network provider), the adversary can retrieve a considerably larger set of historical (and possibly future) coordinates on the victim's trajectory. In addition, since often times mobile users share their coordinates with other users through social media, the player can also acquire the whereabouts of other users in the victim's contacts list.

Figure 2. Recovering Trajectory Data from Packet Captures with Wireshark.

After the player has acquired several different sets of coordinates, the next level will be unlocked.

Stage 3: Background Knowledge Attacks.

This scenario will illustrate the serious privacy breaches that occur when performing more complex types of attacks. Armed with the sets of coordinates from the previous level, the player will again use the mapping tool to visualize the coordinates on the

map. Following this step, the player will be able to identify user trajectories based on the relationships between individual coordinates (for illustration purposes, the coordinates at Stage 2 will be chosen to allow clear separation of trajectories).

Once individual trajectories are identified, the player will have to inspect each individual location snapshot, and identify within each trajectory a point that is not a public location (e.g., an address in a residential area). Using a simple web search, e.g., by running a mock-up Google query against a server hosted within the *GeoCTF* VM, which has the addresses of the trajectory snapshots present inside its database, the player will be able to retrieve a number of relevant hits, for instance, mock-up ads on a real estate website where the address is present together with a name of the property owner and his/her phone number.

Once the identity of the victim has been established, the player will be able to check other locations on the victim's trajectory, and find that some of these locations correspond to sensitive features, such as a hospital, the location of an "Alcoholics Anonymous" meeting, a pawn shop, etc. This association will showcase to the players the power of the background knowledge attacks, and the amount of significant sensitive information that a motivated adversary is able to discover about an unsuspecting victim, due to the careless sharing of location data.

The inference process can be repeated for each of the identified trajectories. Following the completion of this stage, the player will accede to the fourth and final stage, which is covering data protection techniques.

Stage 4: Defense/Appropriate Data Protection.

After the player is introduced in incremental fashion to the increasingly complex attacks that can be staged to breach location privacy, the fourth stage focuses on defense, and illustrates how to protect location data appropriately. The objective of this stage is to familiarize the player with several prominent protection techniques, and provide a visual interface that helps the player apply the protection transformations. Subsequently, the player is presented with the view from Stage 3, but this time with sanitized data, which makes it impossible to perform the privacy breaches conducted earlier.

Although some of the protection mechanisms are complex algorithms that require deep understanding of Computer Science concepts to implement, it is possible to capture the main idea of the approach and present it to the user in a visually intuitive manner. The protection algorithms will already be implemented in *GeoCTF*, and the player needs only to access a GUI and execute the algorithms with a minimal effort. Short demos in the form of videos or animated presentations will be created to illustrate the main concepts of the protection techniques. Specifically, we will focus on three approaches: the k -anonymous cloaking region technique from [5], the private information retrieval approach from [3] and the trajectory sanitization solution from [1].

The approach focused on cloaking regions [5] will briefly introduce the system architecture, and provide some intuitive visual examples of how the location of an individual is generalized, such that an adversary cannot distinguish among k users enclosed by a cloaked region. Furthermore, the cloaked regions must be constructed in such a way that an adversary is not able to associate the users' locations with sensitive regions such as hospitals, nightclubs, bars, etc. A sample example of

showcasing this protection technique is outlined in Figure 3. A GUI will let the players choose the privacy parameter k , and construct for the player the anonymized cloaking regions. When visualizing the regions, the player will notice that trajectory reconstruction is either impossible, or made significantly more difficult.

Figure 3. Showcasing Spatial Cloaking Techniques.

Another prominent approach that will be showcased is the private information retrieval technique from [3]. Instead of sending user coordinates, this approach builds a cryptographic token that is sent to the location-based server. No plaintext location information whatsoever is included in the communication. The player will interact with this protection technique using a GUI, and the system will generate the cryptographic tokens encapsulated in a packet trace that is being sent to the location-based server (LBS). The player can analyze the capture the same way it was done in Stage 2, and find that no information whatsoever can be derived from the tokens, without knowledge of the secret encryption key. This scenario is illustrated in Figure 4.

Figure 4. Showcasing Cryptographic Techniques.

Finally, *GeoCTF* will illustrate how a number of trajectory traces can be published using differential privacy, as shown in [1]. This scenario is suitable for applications that need statistical data about movement, e.g., traffic analysis. A GUI will illustrate to the player how the location traces change compared to the original data when differentially private noise is added, such that an adversary can no longer identify individual trajectories.

3. RELATED WORK

Location privacy has been identified as a serious threat, and extensively studied in [4], [5], [3], [1], [2] etc., where diverse techniques for protection are proposed, ranging from spatial cloaking to cryptographic techniques and differential privacy. However, to the best of our knowledge, no work exists that showcases these threats and protection techniques in a manner that is accessible to a broad mobile user population. Capture-the-flag (CTF) games have gained tremendous popularity in recent years, and offer an excellent tool for education with respect to conventional cybersecurity topics like data, network and operating system security. However, to the best of our knowledge, no CTF game focusing on geospatial data and location privacy exists.

Hence, our envisioned *GeoCTF* game fills an important niche in educating a broader audience about the dangers of uncontrolled sharing of location data.

4. CONCLUSIONS

We introduced *GeoCTF*, a capture-the-flag style game that we envision with the objective of educating high-school and University students, as well as a broader computer-literate audience, on the threats to location privacy when participating in mobile applications, as well as the protection techniques available to mitigate these threats. In the future, we will continue development of the system, and we will also consider providing some of the protection techniques implemented in *GeoCTF* as a mobile app that can inform users in real-time about the dangers they expose themselves to, and about venues of achieving location protection.

Acknowledgment: This work has been supported from National Science Foundation Grant 1523101.

5. REFERENCES

- [1] R. Chen, G. Acs, T. Leek, and C. Castelluccia. "Differentially private sequential data publication via variable-length n-grams". In *Proceedings of the ACM Conference on Computer and Communications Security (CCS)*, 2012.
- [2] G. Cormode, C. Procopiuc, E. Shen, D. Srivastava, and T. Yu. "Differentially private spatial decompositions". In *Proceedings of International Conference on Data Engineering (ICDE)*, 2012.
- [3] G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K.-L. Tan. "Private queries in location based services: anonymizers are not necessary". In *Proceedings of ACM Conference on Management of Data (ACM SIGMOD)*, 2008.
- [4] M. Gruteser and D. Grunwald. "Anonymous Usage of Location-Based Services through Spatial and Temporal Cloaking". In *Proceedings of International Conference on Mobile Systems, Applications, and Services (MobiSys)*, 2003.
- [5] M. F. Mokbel, C.-Y. Chow and W. G. Aref. "The new Casper: query processing for location services without compromising privacy". In *Proceedings of International Conference on Very Large Databases (VLDB)*, 2006.
- [6] M. Olano, A. Sherman, L. Oliva, R. Cox, D. Firestone, O. Kubik, M. Patil, J. Seymour, I. Sohn, and D. Thomas. "SecurityEmpire: Development and Evaluation of a Digital Game to Promote Cybersecurity Education". In *Proceedings of USENIX Summit on Gaming, Games and Gamification in Security Education*, 2014.
- [7] J. Ryoo, A. Techatassanasoontorn, D. Lee, and J. Lothian. "Game-based InfoSec Education Using OpenSim". In *Proceedings of 15th Colloquium for Information Systems Security Education (CISSE)*, 2011.
- [8] J. Werther, M. Zhivich, T. Leek, and N. Zeldovich. "Experiences in cyber security education: the MIT Lincoln laboratory capture-the-flag exercise". In *Proceedings of the 4th conference on Cyber security experimentation and test (CSET)*, 2011.
- [9] Wireshark Network Protocol Analyzer. <https://www.wireshark.org>.