

Enhancing IP Address Geocoding, Geolocating and Visualization for Digital Forensics

Mohammad Meraj Mirza^{1,2} and Umit Karabiyik¹

¹Department of Computer and Information Technology, Purdue University, West Lafayette, IN 47907, USA
{mimirza, umit}@purdue.edu

²Department of Computer Science, College of Computers and Information Technology, Taif University, Taif 21944, Saudi Arabia
mimirza@tu.edu.sa

Abstract—Internet Protocol (IP) address holds a probative value to the identification process in digital forensics. The decimal digit is a unique identifier that is beneficial in many investigations (i.e., network, email, memory). IP addresses can reveal important information regarding the device that the user uses during Internet activity. One of the things that IP addresses can essentially help digital forensics investigators in is the identification of the user machine and tracing evidence based on network artifacts. Unfortunately, it appears that some of the well-known digital forensic tools only provide functions to recover IP addresses from a given forensic image. Thus, there is still a gap in answering if IP addresses found in a smartphone can help reveal the user’s location and be used to aid investigators in identifying IP addresses that complement the user’s physical location. Furthermore, the lack of utilizing IP mapping and visualizing techniques has resulted in the omission of such digital evidence. This research aims to emphasize the importance of geolocation data in digital forensic investigations, propose an IP visualization technique considering several sources of evidence, and enhance the investigation process’s speed when its pertained to IP addresses using spatial analysis. Moreover, this research proposes a proof-of-concept (POC) standalone tool that can match critical IP addresses with approximate geolocations to fill the gap in this area.

Index Terms—Digital Forensics, Geocoding, Geolocation, IP geolocation, Visualization, Mobile Forensics

I. INTRODUCTION

According to the National Institute of Justice (NIJ) [1], digital devices (e.g., Internet of things (IoT), smartphones, wearable devices, drones, autonomous vehicles, and robotics) are extremely valuable for the identification and conviction of some types of crimes due to the fact that they are everywhere and they hold a large amount of data. Many of these devices are increasingly connected to the Internet and use it for many functions. Therefore, IP addresses (i.e., Internet Protocol version 4 (IPv4)), which have a specific formatting schema, are used as the main protocol to connect these devices to the Internet. In digital forensic cases that deal with the network layer, such as network forensics, IP addresses are extremely important as it is one of the primary artifacts responsible for routing data. Moreover, it is important in these cases to determine from where and who’s IP address is being investigated.

A recent study projected the number of mobile users to reach 290 million users by 2024 in the United States [2]. Moreover, another forecast estimated the number of IoT devices to reach around 24 billion by the year 2030 due to the rapid increase in technology associated with such devices [3]. IP addresses can be found in other digital devices, but they are treated as supplementary data and not of great importance in many cases. However, they hold valuable data even if they are looked at outside the network layer and more as located physically on the Earth. Currently, in a single case, there could be many technical details that need comprehensive or complicated procedures, which may cause some difficulties to practitioners and digital forensics investigators if they are not aware of these procedures. To overcome some of these challenges, many frameworks and tools have been developed to close the technology gap that might limit the outcome of any given case. However, most digital forensics tools lack IP address geolocating capabilities even though they are now supporting recovering them from many artifacts.

In the POC discussed in this paper, we demonstrate IP addresses, geocoding and geolocating techniques on the recovered artifacts. Moreover, geotagged pictures that hold geolocation information were used for spatial analysis to help highlight important IP addresses. In our POC experiment, we analyzed two different digital forensic images to map critical IP addresses that could assist investigators in their decision making process. Unfortunately, current digital forensics tools provide limited capabilities when dealing with recovered IP addresses. To the best of our knowledge, none of the digital forensics tools (i.e., open-source like Autopsy, and proprietary like Magnet AXIOM) can perform spatial analysis on recovered artifacts. In addition, we discuss the processing methods using the proposed tool developed to help geolocate IP addresses that are recovered from a given digital forensic image. Then describe practical visualization techniques and some geospatial analysis that can be used to create better spatial and situational awareness for digital forensics investigators. Finally, we explore how this can help investigators plan their investigations for enhanced geolocation exploration in general cases that deal with geolocation events.

This paper aims and focuses on extending the current capabilities of digital forensics tools and digital forensics

investigators by visualizing the insightful information gathered from digital forensics images by proposing a simple tool that geocodes IP addresses and then performs one type of spatial analysis. Moreover, we summarize the main contributions of the paper as follows:

- Extend digital forensic capabilities when dealing with IP addresses.
- Introduce a novel IP address visualization into the digital forensic process.
- Improve the awareness and ability of practitioners to think of IP address as an important element that can be visualized with other related artifacts.
- Highlight some spatial analysis techniques that can be used.
- Illustrate the importance of using IP geolocating utilizing the proposed tool.

This paper is organized as follows: Section II discusses the related work on IP addresses and how it has been used in cybersecurity and different forensic fields. In Section III we introduce and discuss the methodology we used to conduct the POC, while in Section IV we go over the proposed tool. Section V demonstrates the proposed tool POC in the given case studies. Finally, we discuss the findings in Section VI and concluding remarks and future work in Section VII.

II. RELATED WORK

Although the literature lacks studies that have introduced geolocating IP addresses in digital forensic, it is widespread to find real-time network monitoring services using IP to geolocate traffic by using IP-based filtering. According to [4], IP-based filtering has proven to be of great use to block and identify scammers. Furthermore, the authors highlighted that IP can help understand some patterns regarding the geographical extent they hold and discovered that some regions may have more incoming scams than others. On the other hand, e-mail header IP address identification works the same way.

In recent research [5], researchers have discussed previous concepts and designs for an edge management system that was developed by Oriwoh in [6] and highlighted how potential external and internal IP addresses can be recovered from IoT devices and how they can be an expected body of evidence. In addition, in an interesting work in [7], the researcher found log files that store the local and wireless local IP address of a smart cam with a mobile device application (App). In addition, the author found in his IoT forensics work a smart light bulb to store IP address, which can be recovered from the chip-off acquisition.

Although there are around 4 billion possible IPv4 addresses that are presented in human-readable notations [8], a newer protocol named Internet Protocol version 6 (IPv6) is effectively introduced and adopted by many devices to replace the limited IPv4 [9]. There is a demand for this change as there are more devices are connected due to the increase of IoT devices and handheld devices [10]. IP addresses can be classified as personal, which uses many technologies to be able to identify people in the physical world with their respected location. A

recent digital forensics framework by [11] that concentrates on investigating and reviewing cyber-attacks has drawn the importance of IP addresses. Moreover, a proof of concept conducted [12] has demonstrated how the use of consecutive IP addresses by hackers can help investigators in understanding geospatial temporal patterns to uncover their fingerprints which lead to their identity. On the other hand, the anonymity of the IP address is a challenge to most of the digital forensics cases and tools that deal especially with network forensics [13], [14]. Moreover, this challenge is currently creeping toward other digital forensics investigations.

In addition, [15] has illustrated the importance of IP addresses in a web hacking example. The author discussed how IP addresses can be of great importance in getting the location of the attacker and has highlighted how it is crucial from a legal perspective to help investigations ask Internet Service Providers (ISPs) for a specific address. However, researchers in [16] have discussed how crucial but challenging using the IP address as digital evidence in different investigations that deal with not only network investigation due to the complexity that they hold.

According to [17], [18] the literature shows that mapping IP addresses to their geolocation has been comprehensively studied. However, the use of current knowledge has not yet reached the digital forensics tools. Therefore, the amount of work in mobile forensics is not fair when compared with the rapid development of technology.

III. METHODOLOGY

As in any digital forensics work, this research follows the best practices and guidelines for device preparation, data population, acquisition, analysis, and examination that are introduced by the National Institute of Standards and Technology (NIST) [19]. Fig. 1 demonstrates an overview of the methodology that we performed.

In this research, we have used two forensically acquired mobile images as our POCs that were populated and acquired by Joshua Hickman, who have provided them to the public for research [20], [21]. Both images were populated using the same device (Pixel 3) operating Android; however, the first image is populated using Android 10 operating system (OS) and the second is using the latest available version Android 11. For both images, Joshua has considered rooting the device, which is an essential step to be able to recover as much information as possible from devices. The research also followed best practices in the population and acquisition processes. Table I underlines the smartphone used with its specifications (i.e., model, storage, RAM, and OS) and the acquisition tools used to create the digital forensic images along with each image creation date [20], [21].

To maintain integrity, the programs and tools used for this research were carefully chosen. In the acquisition phase, *Cellebrite UFED 4PC* and *Magnet Acquire* were used to acquire the images as seen in table I. To examine and analyze both images in this research, we have used two well-known digital forensics software. First is open-source, and

- 3) From the *Keyword Hits* panel, we extracted all IP addresses into an CSV file format.
- 4) We finally extracted pictures that contain latitude and longitude (i.e., geotag) from the case into a folder.

Now we have two elements that will be used as inputs for the developed tool. The first is the cleaned-up version of the CSV file containing a list of recovered IP addresses from the digital forensic image and the number of hits. The second is a folder that contains all recovered geotagged pictures. Therefore, in the proposed tool, we developed and used different functions as follows:

- **Geocoding:** It is the process of taking the text-based address of a location, then trying to return the latitude and longitude [26]. For instance, addresses are usually in text-based formats that can be directly read or converted to geographic coordinates; however, in the case of IP addresses, there is a need for specialized databases that contain a large amount of IP addresses with their relative locations on the Earth.
- **Multiple ring buffer:** Creating circle buffers of a certain distance around specific locations or features (e.g., points, polygons, and lines).
- **Intersect Operation (Analysis):** This operation takes multiple features and performs a geometric intersection calculation to determine if these features are intersected.
- **Exporting Keyhole Markup Language (KML) file:** The KML is a file format in which spatial details are presented using many available Earth browsers.
- **ArcGIS geodatabase:** This is a specialized proprietary database used by ESRI products that holds different kinds of geographical data.

The tool starts with geocoding all possible IP addresses and creates a new CSV file that contains three additional columns to hold information for latitude, longitude, and city name for each of the IP addresses. This new CSV file that stores the geolocated IP addresses will be used in the second phase for performing spatial operations, using the geoprocessing framework provided by ArcGIS Pro software on both the geolocated IP addresses and geotagged pictures recovered from the forensics image.

To be able to find geolocated IP addresses that are near the geotagged pictures recovered, the tool creates buffers around the pictures with set distances (i.e., 5, 10, and 20 miles) and then performs an intersect operation. This intersection operation will point out IP addresses that are within some geographical distance from a geotaged picture, which might lead to highlight that this IP address is important to be looked at and the artifacts that are associated with.

Finally, all results will be exported into a geodatabase supported by ESRI products and as KML files that can be visualized using many Earth browser software's (e.g., Google Earth Pro browser and online Google Maps). Fig. 3 demonstrates the interface for the developed tool where the user will be indicating all input and output file names, and Fig. 4 illustrates the workflow of the developed tool as discussed.

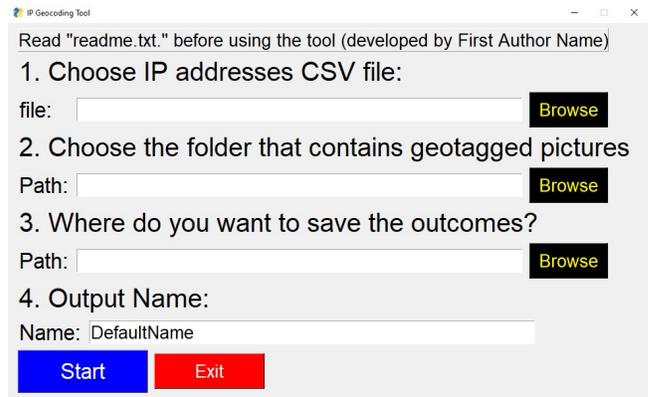


Fig. 3. Graphical user interface for the developed tool.

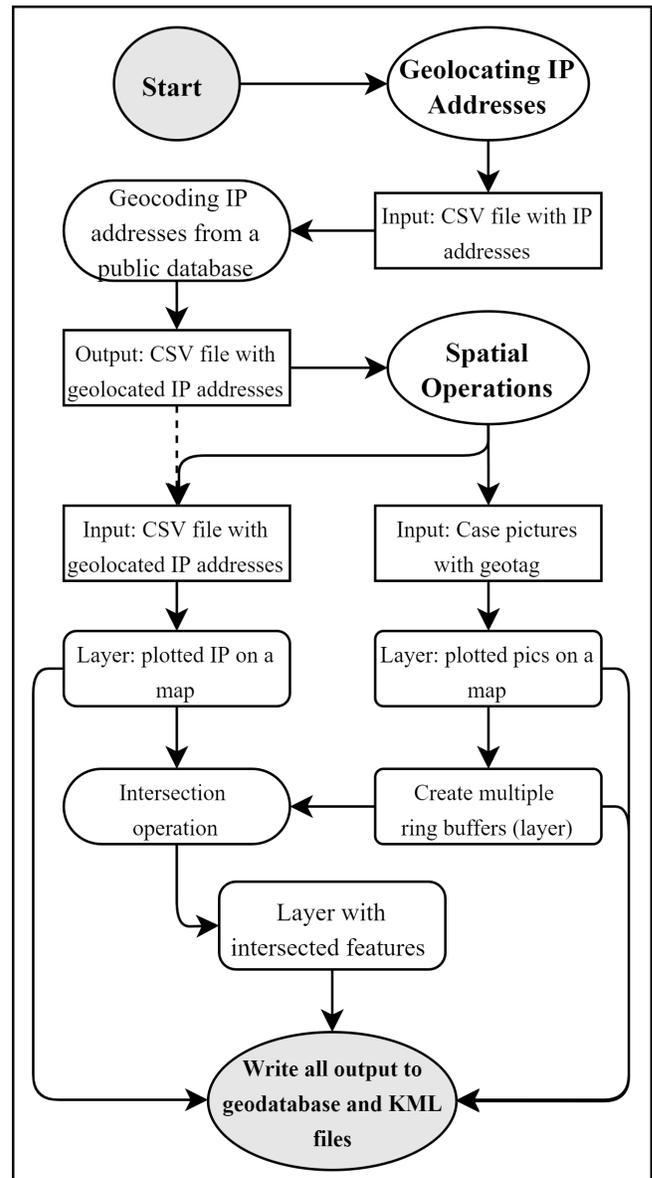


Fig. 4. Developed tool workflow

V. THE PROOF-OF-CONCEPT

We started by taking both forensic images into the digital forensics tools to examine them and to prepare the data for the developed tool as discussed in Section IV. After preparing the inputs, we ran the tool two times; first with the inputs from the first image and then with the second image. In addition, different output names were chosen to distinguish between the first image with Android 10 and the second image with Android 11.

As a result, the tool creates *KML* files and an ArcGIS geodatabase as outputs. We used these two outputs as appropriate for visualization methods to aid our investigation. First, we used ArcGIS software to deal with the geodatabase to utilize the best geographic and cartographic representations of the results to create useful maps. Fig. 5 demonstrates the top 80 geocoded IP addresses by the number of appearances that

are grouped by city name for each forensic image. On the other hand, Fig. 6 highlights the usefulness of the intersection spatial analysis performed to find geolocated IP addresses that intersect with the ring buffers around the geotagged pictures found in each case.

Our analysis drew important conclusions by identifying crucial IP addresses from two different cases, confirming that both users were within the same geolocation zone. This is due to the fact that these hypothetical cases were populated by the same individual. To this end, using our approach can aid the investigators in localizing the approximate location of the user via IP address geocoding.

Although these maps are clear, concise, and would serve well for documentation and/or reporting, they require some skills to be created because they are created with ArcGIS Pro software. Therefore, more accessible visualization options

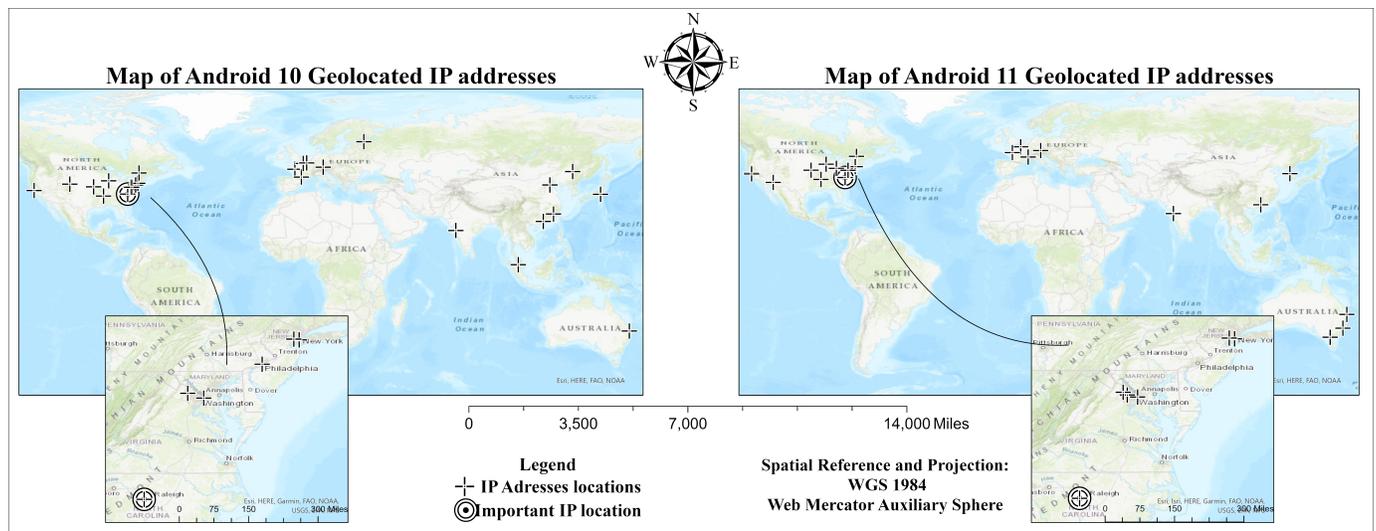


Fig. 5. Maps of the geocoded IP addresses found in each of the acquired mobile forensic images using ArcGIS Pro software.

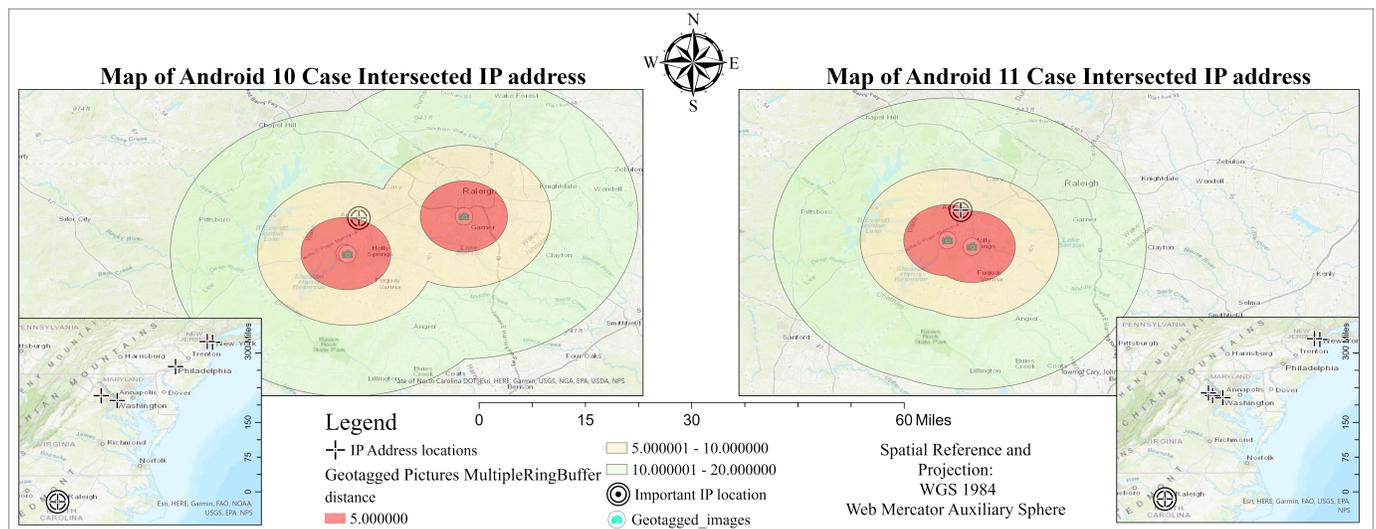


Fig. 6. Maps of the geolocated IP address intersected with geotagged pictures using ArcGIS Pro software.

using the KML file are available. For example, Google Earth Pro browser provides uploading KML files to be displayed. Fig. 7 demonstrates the combined results using Google Earth Pro browser to display KML outputted files.

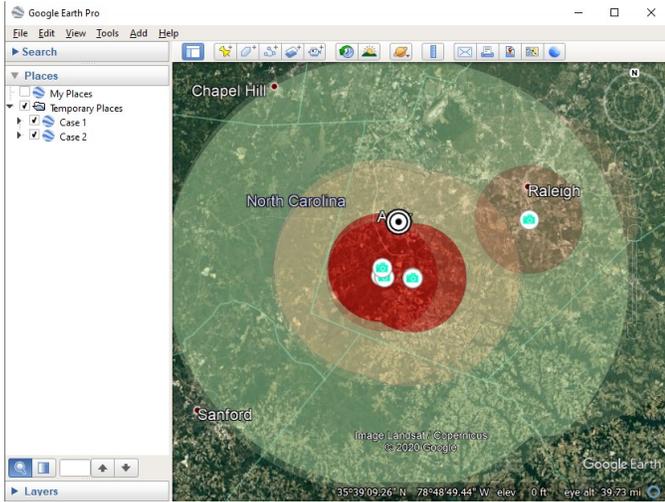


Fig. 7. KML files that the developed tool created for both cases (i.e., Android 10 Android 11) illustrated using Google Earth Pro browser.

VI. DISCUSSION AND SUMMARY OF FINDINGS

It is clear that there are apps in both investigated digital forensic images that preserve IP addresses, and they are easily recovered by Autopsy using its regular expression search module. However, current digital forensics tools provide limited capabilities and only provide and return found IP addresses in a given forensic image. In addition, the research illustrates the lack of IP address visualization in both digital forensics tools that were used. Although the demonstrated spatial analysis techniques allowed to adequately assess and forecast our knowledge to interpret and understand human activity trend patterns (i.e., pattern of life analysis) and its spatial representation, it is not widely used to help investigators make decisions and draw conclusions. Interestingly, our research shows that investigators need to be aware of the different types of artifacts and metadata used to harvest important locations, where our research has demonstrated one type that is not considered by many tools. Alternatively, the first concern is whether an IP address recovered from the digital forensic images can help investigators to locate where the user was or at least minimize the search area. Although IP addresses are essential to any given case, they usually need more advanced techniques. Therefore, after highlighting and discovering important IP addresses, investigators need to order a subpoena from ISPs to get precise details (e.g., owner information, address/home location, and internet records), which might complement other information in the investigation. Therefore, the proposed techniques in our work emphasize the importance of identifying user's related IP addresses. To the best of our knowledge, the recovered IP address is critical to this investigation because

the home address of the two digital forensics images is within close proximity of the geolocation of the IP address.

Our findings show promising results as many of the recovered IP addresses were recovered from certain apps that pertained the user's activity, such as WhatsApp, Telegram, TextNow, Instagram, Web browsers, etc. These applications store and digitally fingerprint the user's IP addresses for geographical identification purposes (i.e., city and/or country location). Therefore, even if the user does not permit the app to use GPS services on their mobile device, these apps can localize the user's approximate location without the owner's authorization. To this end, some apps deploy these technologies for targeted advertisement. However, from a security perspective, these practices can lead to the exposure of the user's privacy (e.g., identity and location), making it vulnerable to attackers such as malvertisement [27], [28].

With respect to geocoding, we found that the margin of error depends on the public IP addresses database provided. In the preformed POC, we were able to get decent results; however, this might not be true in other more difficult cases, such as using virtual private network (VPN) services to hide identifiable information. On the one hand, not all recovered IP addresses are recovered are important to the investigation due to the IPv4 subnetting. Therefore, it is wise to create a list of common IP addresses found in many digital devices related to known app servers, services, and even local address. On the other hand, IP signature-based detection could enhance the process of whitelisting and/or blacklisting IP addresses. This practice can actively empower digital forensics tools to distinguish between known, critical or malicious IP addresses in any given forensic case.

VII. CONCLUSION AND FUTURE WORK

In this paper, we found that geocoding and performing spatial analysis of IP addresses can contribute to the current digital forensic process by highlighting important artifacts and locations that can be associated with the user activity, which can technically improve investigative-based decision making. This can enhance digital forensic tools and investigative techniques to demystify the geographic coverage of an investigation. In addition, digital forensics investigations and investigators may benefit from enhanced spatial analysis to draw better conclusions.

As there are more devices that are connecting to the internet, there is a demand to change to the newer IPv6. Therefore, more work needs to be done to adapt accurate geolocating for devices using IPv6 as more devices are using it. For future work, we plan to incorporate validation techniques to measure and evaluate the admissibility of IP address visualization. Moreover, we aim to propose an IP address classification model for the digital forensics process.

Finally, our work emphasizes the importance of IP addresses due to their physical value in geography, which when available, can be used to localize IP addresses to identify important locations in a digital forensic case.

REFERENCES

- [1] S. E. Goodison, R. C. Davis, and B. A. Jackson, *Digital Evidence and the U.S. Criminal Justice System: Identifying Technology and Other Needs to More Effectively Acquire and Utilize Digital Evidence*. RAND Corporation, 2015.
- [2] Statista. (2019) Number of smartphone users in the united states from 2018 to 2024 (in millions). Accessed: 2020-12-24. [Online]. Available: <https://www.statista.com/statistics/201182/forecast-of-smartphone-users-in-the-us/>
- [3] ——. (2019) Number of internet of things (iot) connected devices worldwide in 2019 and 2030. Accessed: 2020-12-22. [Online]. Available: <https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide-2019-2030/>
- [4] Y. Hu, C. Guo, E. Ngai, M. Liu, and S. Chen, "A scalable intelligent non-content-based spam-filtering framework," *Expert systems with applications*, vol. 37, no. 12, pp. 8557–8565, 2010.
- [5] U. Karabiyik and K. Akkaya, "Digital forensics for iot and wsns," in *Mission-Oriented Sensor Networks and Systems: Art and Science*. Springer, 2019, pp. 171–207.
- [6] E. Oriwoh and P. Sant, "The forensics edge management system: A concept and design," in *2013 IEEE 10th International Conference on Ubiquitous Intelligence and Computing and 2013 IEEE 10th International Conference on Autonomic and Trusted Computing*. IEEE, 2013, pp. 544–550.
- [7] F. E. Salamh, "A forensic analysis of home automation devices (fahad) model: Kasa smart light bulb and eufy floodlight camera as case studies," *International Journal of Cyber Forensics and Advanced Threat Investigations*, pp. 1–1, 2020.
- [8] W. R. Parkhurst, *Routing first-step*. Cisco Press, 2004.
- [9] W. Stallings, "Ipv6: the new internet protocol," *IEEE Communications Magazine*, vol. 34, no. 7, pp. 96–108, 1996.
- [10] O. Babatunde and O. Al-Debagy, "A comparative review of internet protocol version 4 (ipv4) and internet protocol version 6 (ipv6)," *arXiv preprint arXiv:1407.2717*, 2014.
- [11] A. Dimitriadis, N. Ivezic, B. Kulvatunyou, and I. Mavridis, "D4i-digital forensics framework for reviewing and investigating cyber attacks," *Array*, vol. 5, p. 100015, 2020.
- [12] B. Gokaraju, R. Agrawal, D. A. Doss, and S. Bhattacharya, "Identification of spatio-temporal patterns in cyber security for detecting the signature identity of hacker," in *SoutheastCon 2018*, 2018, pp. 1–5.
- [13] M. Wazid, A. Katal, R. Goudar, and S. Rao, "Hacktivism trends, digital forensic tools and challenges: A survey," in *2013 IEEE Conference on Information & Communication Technologies*. IEEE, 2013, pp. 138–144.
- [14] D. Lillis, B. Becker, T. O'Sullivan, and M. Scanlon, "Current challenges and future research areas for digital forensic investigation," *arXiv preprint arXiv:1604.03850*, 2016.
- [15] R. S. Jeong, "Forza – digital forensics investigation framework that incorporate legal issues," *Digital Investigation*, vol. 3, pp. 29 – 36, 2006, the Proceedings of the 6th Annual Digital Forensic Research Workshop (DFRWS '06). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1742287606000661>
- [16] P. Sokol, L. Rózenfeldová, K. Lučivjanská, and J. Harašta, "Ip addresses in the context of digital evidence in the criminal and civil case law of the slovak republic," *Forensic Science International: Digital Investigation*, vol. 32, p. 300918, 2020.
- [17] P. Winter, R. Padmanabhan, A. King, and A. Dainotti, "Geo-locating bgp prefixes." IFIP, 2019, pp. 9–16.
- [18] E. Katz-Bassett, J. John, A. Krishnamurthy, D. Wetherall, T. Anderson, and Y. Chawathe, "Towards ip geolocation using delay and topology measurements," ser. IMC '06. ACM, 2006, pp. 71–84.
- [19] "Mobile devices," May 2017. [Online]. Available: <https://www.nist.gov/itl/ssd/software-quality-group/computer-forensics-tool-testing-program-cftt/cftt-technical/mobile>
- [20] B. Hick, "Android 10 image now available!" Feb 2020. [Online]. Available: <https://thebinaryhick.blog/2020/02/15/android-10-image-now-available/>
- [21] —, "New android image available. this one goes to 11!" Oct 2020. [Online]. Available: <https://thebinaryhick.blog/2020/10/07/new-android-image-available-this-one-goes-to-11/>
- [22] "Autopsy," Jul 2020. [Online]. Available: <https://www.basistech.com/autopsy/>
- [23] "Magnet axiom - digital investigation platform," Sep 2020. [Online]. Available: <https://www.magnetforensics.com/products/magnet-axiom/>
- [24] ESRI Inc., *ArcGIS Pro*, ESRI Inc., Redlands, CA, 2021. [Online]. Available: <https://www.esri.com/en-us/arcgis/products/arcgis-pro>
- [25] G. Inc., "Earth versions – google earth pro (windows)," <https://www.google.com/earth/versions/#download-pro>, (Accessed on 02/15/2021).
- [26] Statista. Geocoding api. Accessed: 2020-11-24. [Online]. Available: <https://developers.google.com/maps/documentation/geocoding/overview>
- [27] G. Chen, W. Meng, and J. Copeland, "Revisiting mobile advertising threats with madlife," in *The World Wide Web Conference*, 2019, pp. 207–217.
- [28] T. Liu, H. Wang, L. Li, X. Luo, F. Dong, Y. Guo, L. Wang, T. Bissyandé, and J. Klein, "Maddroid: Characterizing and detecting devious ad contents for android apps," in *Proceedings of The Web Conference 2020*, 2020, pp. 1715–1726.