Round Trip Time and Hop Count for Geolocation in Request Tracing and Classification

CS4991 Capstone Report, 2024

Connor Wilson
Computer Science
The University of Virginia
School of Engineering and Applied Science
Charlottesville, Virginia USA
crw8eg@virginia.edu

ABSTRACT

Request classification and filtering in modern network security is becoming increasingly challenging as threat actors more and more commonly utilize IP spoofing techniques and proxies. One potential solution to aid in classifying such malicious requests utilizes Round Trip Time (RTT) to improve the Hop Count Filtering (HCF) technique. The value of HCF and RTT lies in their potential to estimate the region from which a signal is originating and classify requests based on that information. Combining this method with the existing methods utilizing RTT and HCF to detect IP spoofing enables network administrators and security personnel to rapidly classify and filter malicious requests based on Hop count (HC) and estimated area of origin. This technique has the potential to be employed in a network security context on an enterprise level by identifying regions from which a majority of requests are malicious and flagging or filtering traffic from said regions before it reaches the intended network or endpoint. Future work on this topic should examine patterns in packet travel, to include HC and RTT, consistent with advanced persistent threats (APT) living off the land or using home routers as a point of origin.

1. INTRODUCTION

In the realm of modern network security, the task of request classification and filtering faces escalating challenges posed by increasingly sophisticated threat actors employing IP spoofing techniques and proxies. The ubiquity of IP spoofing has propelled the significance of accurate request tracing and classification, necessitating innovative methodologies for identifying malicious traffic (Singh & Pandy, 2020). There are many existing approaches, the most notable of which are RTT and HCF for the purpose of this analysis.

RTT is a measure of latency, or the amount of time required to get a packet from one endpoint to another. Typically measured in milliseconds, an example of a round trip time is that of a UDP packet traveling from the East Coast to the West Coast and back in 40ms (Fei, et. al, 1998). For a TCP packet, this RTT will often be doubled for a trip from the East Coast to the West. RTT is specifically valuable not only as a measure of latency and therefore efficiency/speed but also as a statistic that can be used to identify network attacks in which traffic is being intercepted (Sengupta, et. al., 2022).

HCF is another networking statistic that measures the number of steps, or hops, taken by a packet on its way from one endpoint to another. This is important especially in the context of facing the IP spoofing challenge because an attacker cannot maintain a consistent HC while IP spoofing (Wang, et. al., 2007). Between HC and RTT, a network administrator has the tools to effectively

classify the validity of various incoming packets and draw other conclusions about other information such as the region of origin.

2. RELATED WORKS

Mukaddam & Elhajj (2012) provides much of the foundation for this report as the key technique used in the proposed regional classification of packets is that of HCF as improved by RTT. The proposed approach will differ from Mukaddam & Elhajj significantly in that it does not propose any changes to their method as a defense mechanism but, rather, applies it to the problem of best classifying the region of a packet's origin.

Wang, et. al. (2007) is one of the foundational works describing how HCF can be specifically applied to counter the threat posed by IP spoofing. This is a key factor to both establish and consider when evaluating any modern network attack, as such attacks are often using spoofed IP addresses or some other IP-obscuring mechanism. Application of this technique to classify packet region of origin, therefore, not only contributes to increased knowledge of the network environment and potential threat level but also to attribution.

Together, these reports provide the pillars upon which modern network defense techniques utilizing HCF and more recently a combination of HCF and RTT operate and translate directly to this report as the stepping stone from which a network administrator will be able to further classify incoming data.

3. PROPOSAL DESIGN

The proposed model uses existing or learned knowledge of global network topology in conjunction with HC and RTT to both identify IP spoofing and gain knowledge of true geographic region of origination.

3.1 Assumptions and Components

Two assumptions are made in order to control the complexity of the proposed model. First, RTT can vary based on several different factors including but not limited to time of day, day of week, and distance (Fei, et. al, 1998). While a significant increase in the volume of network traffic may affect RTT and HC, it is not to such a degree that one would not be able to associate geographic regions to HC and RTT with an acceptable degree of error. Second, while variation in the actual route of packets occurs frequently, there is a dominant path**,** and deviant paths will have minimal variation in HC from the dominant path. (Fei, et. al, 1998, Paxson, 1997). With those assumptions made, the proposed model would allow for the following inputs in order to estimate geographic location:

- RTT in milliseconds (ms). This is the backbone of the approach and would utilize measurements from a tool such as Traceroute.
- A database with average RTTs (in ms) to servers around the globe that receive the most traffic in each region. This could also be paired with a map to show the radius in miles or kilometers that a specific RTT limits possible IP addresses to.
- HC. Hop count, or the number of routers a packet visits on the path to its destination, can also be measured with Traceroute.
- Current router IP address. The current IP address of the router of the private network receiving traffic (should not be a spoofed address).
- Expected dominant paths. A database of dominant paths from to and from the servers that receive the highest traffic on a global scale. Traceroute dataset parsing libraries already exist for this purpose (Dan, et. al, 2021).
- IP address of the neighboring server on packet departure and return.

- The receiving network must have the capability to run Traceroute or a similar tool in order to measure HC and RTT.

3.2 Methodology

Upon detecting incoming packets from an unknown/untrusted source IP address attempting to enter the local network, a network administrator sees that the traffic in question is flagged as suspicious or abnormal. Upon deciding that IP spoofing may be involved and that it would be beneficial to know the geographic location the source IP most likely originated from, that administrator takes the following steps:

Using Traceroute or a similar tool, the network administrator determines the path to the source IP and makes note of the path, RTT, and hop count.

Reference the table generated by the HCF algorithm proposed in Wang, et. al (2007) to see if the HC from source to destination matches the expected or recorded HC. If it does not, IP spoofing is likely being used.

If the actual HC matches the recorded HC, it is still possible that IP spoofing is in effect and that there are multiple allowable HCs (due to multiple paths) for the packet in question. Analyze the RTT of the packet to see if it also matches the expected RTT value. If not, it is highly likely the sender is using some sort of IP spoofing technique (Mukaddam & Elhajj, 2012).

Based on the RTT database, estimate a distance from the receiving (friendly) network that the original packet likely came from. Use that distance to create a radius along which the packet likely originated.

Narrow the sending region by using knowledge of the receiving server's neighbors (from the prerequisite database of dominant paths) to eliminate as many routes as possible.

Explore possible route branches and use the RTT determined by the initial Traceroute to find the approximate region via elimination.

Choose to allow/disallow that traffic based on the findings and create a rule to disallow or flag traffic with similar characteristics or sources in the future.

3.3 Expected Limitations

This method is reliant on using Traceroute or a similar tool in order to classify incoming packets. As such, this methodology would require a not-insignificant amount of compute to perform regularly or in high volumes. As such, the recommended context for using this technique is when one's network is receiving traffic that is suspicious or abnormal. Those are broad categories that include but are not limited to traffic that comes from an unknown or suspicious IP, comes in high volume at an unexpected time, or comes together to create a suspicious or dangerous request or payload.

On a similar note, the memory required to perform this method on a large scale or at a high frequency could quickly become significant. In order to combat this, administrators could utilize either databases (recommended at scale) or data structures such as hash tables that are dynamically updated with relevant information throughout the process.

It is also important to acknowledge the error that will be inherent in these calculations. As this method is highly reliant on RTT, things like bandwidth, time of day or week and associated traffic, and queuing behavior of nodes on the path of a packet may create slightly varied results (Mukaddam & Elhajj, 2012, Fei, et. al, 1998). If a single measurement is to be taken, the reliability of such a measurement could be questioned. The proposed database of RTTs to highly trafficked global servers is one way to combat this, as it would ideally be an average determined from RTT samples taken under various conditions and different times.

Also notable are the findings of Mukaddam & Elhajj (2012) suggesting that

91% of IP addresses fall within 100ms of RTT. For this method, this finding effectively means the closer the receiver is to the sender, the higher confidence a network administrator can have in the results.

4. ANTICIPATED RESULTS

Based on similar research using Traceroute, without RTT and HC (Dan, et. al, 2021), it is highly likely that this method could produce results that are accurate to within 10km for senders who are not using IP spoofing techniques. The novelty of this method mainly lies in its capability to compare HC and RTT and determine from their overlap within a country or autonomous system if a sender is using IP spoofing and determine the region from which the original request came. In this case, the accuracy would likely be limited to a region covering several hundred square miles or kilometers rather than the more accurate result in the scenario without IP spoofing.

5. CONCLUSION

The proposed methodology uses a combination of existing HCF techniques and RTT and leverages knowledge of existing dominant paths in global routing with the intent of creating a filtering technique that can determine the true geographic origin of the traffic regardless of whether or not IP spoofing is being used. Were this methodology to be successfully implemented in an organization's network, it would permit network administrators to better understand, classify, and filter traffic entering their networks. This is specifically promising for the defense contexts as it would allow them to identify traffic incoming to sensitive networks from adversaries.

6. FUTURE WORK

The logical next step would be to implement and test the methodology proposed above. This could likely be achieved by combining the databases mentioned in Dan, et. al (2021) with the approach to HCF utilizing RTT as proposed in Mukaddam & Elhajj. Initially this would be most easily created in a sandbox environment simulating IP addresses from different geographic regions, some of which use IP spoofing techniques and some of which do not.

Successful implementation should be followed by using verified data to identify or improve the known profile of APTs using IP spoofing techniques or living off the land.

REFERENCES

Dan, O., Parikh, V., & Davison, B. D. (2021). IP Geolocation Using Traceroute Location Propagation and IP Range Location Interpolation. In *Companion Proceedings of the Web Conference 2021* (pp. 332-338).

Fei, A., Pei, G., Liu, R., & Zhang, L. (1998). Measurements on delay and hop-count of the internet. In *IEEE GLOBECOM'98-Internet Mini-Conference*.

Mukaddam, A., & Elhajj, I. H. (2012). Round trip time to improve hop count filtering. In *2012 Symposium on Broadband Networks and Fast Internet (RELABIRA)* (pp. 66-72). IEEE.

Paxson, V. (1997). End-to-end routing behavior in the Internet. *IEEE/ACM transactions on Networking*, *5*(5), 601-615.

Sengupta, S., Kim, H., & Rexford, J. (2022). Continuous in-network round-trip time monitoring. In *Proceedings of the ACM SIGCOMM 2022 Conference* (pp. 473-485).

Singh, V., & Pandey, S. K. (2020). Revisiting cloud security threats: IP spoofing. In *Soft*

Computing: Theories and Applications: Proceedings of SoCTA 2018 (pp. 225-

Wang, H., Jin, C., & Shin, K. G. (2007). Defense against spoofed IP traffic using hop-count filtering. *IEEE/ACM Transactions on networking*, *15*(1), 40-53.