Geotagging IP Packets for Location-Aware Software-Defined Networking in the Presence of Virtual Network Functions

[Vision Paper]

Tamraparni Dasu
AT&T Labs-Research
tamr@research.att.com

Yaron Kanza
AT&T Labs-Research
kanza@research.att.com

Divesh Srivastava
AT&T Labs-Research
divesh@research.att.com

ABSTRACT

A substantial portion of the global telecommunication is based on the Internet Protocol (IP), where IP packets are routed from a source host to a destination host via a communication network. While there is some loose connection between IP addresses and geospatial locations, associating packets to geographic coordinates merely according to IP addresses is hard, and often infeasible in real-time, given the rapidity and prodigious volume of packet traffic via routers and switches. This obstructs using geospatial information about the origin, destination, or route of IP packets or flows.

In this paper we introduce a vision of adding geotags to IP packets, to enhance the capabilities of communication networks and of location-based services. We explain how the augmentation can be done flexibly and effectively using two new networking technologies: *(1) software defined networking* (SDN)—a new architecture that facilitates the ability to control network flows, and *(2) network function virtualization* (NFV) which allows deploying virtual network services, and chaining such services to one another. We describe new applications that can be built using the enrichment of packets with spatial or temporal properties, including applications related to network security, geofencing and operations support systems. We discuss challenges and research directions in this domain.

KEYWORDS

Geotagging, IP, SDN, software defined networking, NFV, spatio-temporal analysis, geofencing, network security, IoT

ACM Reference format:
Tamraparni Dasu, Yaron Kanza, and Divesh Srivastava. 2016. Geotagging IP Packets for Location-Aware Software-Defined Networking in the Presence of Virtual Network Functions. In *Proceedings of ACM Conference, Washington, DC, USA, July 2017 (Conference'17),* 4 pages.
DOI: 10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The Internet is one of the foundations of the Information Age and the Digital Revolution. It connects billions of computers and devices, and facilitates the exchange of data between machines—both stationary and mobile. With the advent of the Internet of Things (IoT), the role of the Internet in connecting devices to one another is expected to grow dramatically.

The Internet Protocol (IP) is the underlying protocol for identifying hosts on the Internet. Every host on the internet has a unique IP address. The addresses are used to route IP packets from a source host to a destination host, via various intermediate nodes (routers, switches, middleboxes, etc.) Each packet is associated with a source IP and a destination IP. However, the association of an IP address to a geographic location is loose. Switches and routers, which route packets to their destination, are oblivious of the geographic source location or destination location of the packet. The spatio-temporal information, however, can be useful in various applications. Some of these applications will be described in this paper.

The IP addresses in the packets are used effectively for routing packets to their destination. But extracting the source or destination geographic location of a packet in real time is unpractical, due to the rapid flow of packets via the nodes of the network. There is no simple function that translates IP addresses to geospatial locations, so the translation is based on table lookups, and is not immediate. Furthermore, source IP addresses can easily be spoofed and manipulated, so, often applications are unable to rely on them.

To cope with this, we present a novel approach of geotagging IP packets. The main idea is to add to IP packets a geotag that contains the location and the time. A packet may be tagged just at the location where it is originated, or may accumulate tags as it visits nodes on the route to the destination. The geotags could then be used by various applications, e.g., to improve network management, strengthen network security or support location-based services.

Adding geotags to IP packets is not a simple task. The headers of packets are rigidly structured and leave no space for extra information. The payload (body) of the packet may not have enough space either. Furthermore, there is a need to control the addition of geotags so that geotags will be added only when necessary, without impeding traffic flow unnecessarily. To provide such a flexibility, we suggest to use two new network paradigms: SDN and NFV. They provide flexibility and enhanced control over packet routing.

In this paper, we describe the vision of using SDN and NFV to geotag IP packets. Section 2 provides some background on SDN and NFV. Section 3 illustrates how to add geotags to IP packets. We describe potential applications of the approach in Section 4. Finally, in Section 5, we conclude.

2 BACKGROUND

Software defined networking (SDN) is a new networking paradigm that decouples the control plane from the data plane in communication networks [5, 6]. In SDN, a controller that directs the entire flow in the network has access to the routers and switches of the network. The controller can change the network flow by modifying routing instructions in the routing tables of the routers and switches, e.g., using the OpenFlow communication interface [8], or some other interface that is defined using the P4 language [1].

Conference'17, Washington, DC, USA
2016. 978-x-xxxx-xxxx-x/YY/MM…$15.00
DOI: 10.1145/nnnnnnn.nnnnnnn

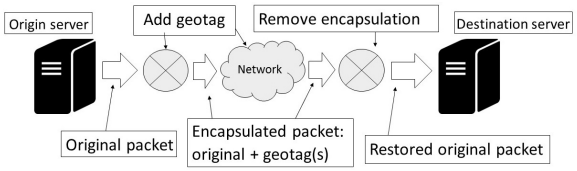

Figure 1: Geotagging an IP packet.

It can add different types of action instructions to the routing tables. When an arriving packet matches an action rule, the action is performed on it. This includes actions such as forwarding the packet to some node, sending a copy of the packet to the controller, increasing a counter or dropping the packet. The controller can also add an action that changes a packet header, which allows adding information to the packet. The controller can instruct routers and switches to add spatio-temporal information to packets. This could facilitate the implementation of services that rely on the location of the user and the time of the action, e.g., location-based services.

In non-SDN networks, protocols are hardcoded in the hardware, and networking rules are computed in a decentralized fashion. Thus, changes in networking protocols are cumbersome, expensive, difficult to implement and hard to deploy. SDN adds flexibility to the network. It supports rapid changes in protocols, without incurring a large cost. Our suggestion is to utilize SDN for associating network flows with spatial and temporal information. This has the potential of introducing a new set of capabilities, and will allow implementing *location-aware SDN*, where spatio-temporal information is combined with packet routing.

The temporal and spatial applications could be virtual network functions (VNFs), by using network function virtualization [3]. VNF provides a virtualization of tasks that otherwise are carried out on dedicated propriety hardware. VNFs allow executing them as services on commodity hardware. Such functions could be chained to other VNFs, such as intrusion detection functions, load balancer, firewall, etc. The chaining will lead to a better intrusion detection, a more effective load balancing, and so on. Furthermore, it may be desirable to add spatial and temporal information to flows that go via a VNF, e.g., for troubleshooting, since virtualization may decouple the VNF from the physical infrastructure on which the virtual machines are deployed.

3 GEOTAGGING

Spatio-temporal tagging of IP packets is the process of adding at a network node (e.g., a switch, router or middlebox) a tag specifying the geographic location of the node and the time of the addition, similarly to geotagging of social-media or multimedia content [7]. When the packet travels via the node, the geotag is added to the packet. Nodes that later inspect the packet can know where it was, geographically, and at what time. Since there are applications in which the authenticity of the tag is a concern, the node should also add a certificate attesting to the genuineness of the network node that added the tag. The geotag may be removed before reaching the destination node, see Fig. 1, or at the destination node.

Spatio-temporal geotagging requires allocating space in the packet for the added data. In IPv4 and IPv6, packet headers have

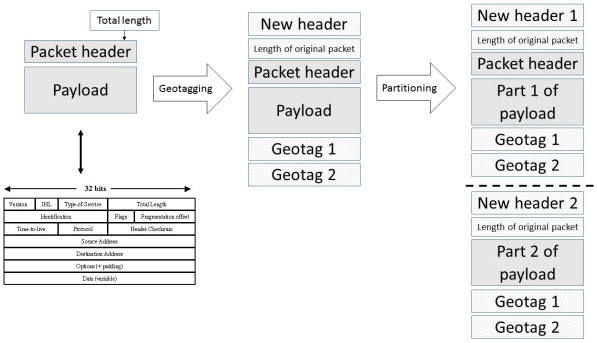

Figure 2: Encapsulating an IP packet.

a pre-defined strict structure, so the data cannot be added to the packet header. Instead, we can use *packet encapsulation* to wrap the existing packet and concatenate to it the added data (geotag and authentication information). Future IP protocols may consider the need to allocate space for metadata, so, in the future the geotag may be added to the header, in the space allocated for metadata. One may need to add geotags at different nodes along the packet route, i.e., several geotags may be added to a packet.

The encapsulation and the unwrapping can be executed by different nodes via which the packet travels, based on instructions of the SDN controller. In the encapsulation, the packet and the geotag are concatenated, to create a new payload. A new header is created, taking the destination of the original packet into account. Note that by using encapsulation and creating a new header, the existing packet remains untouched, and it is being transported without being modified at all at any time. That is, the geotagging is part of the Network Layer (layer 3 of the OSI model) and does not affect the Transport Layer (layer 4 of the OSI model).

To distinguish between the payload and the geotag, the payload of the new (encapsulating) packet may be as follows. The first two bytes indicate the length of the original packet. Then, there is the entire original packet. After that comes the geotag, or several geotags. Note that when a node adds a geotag to an already encapsulated packet, the size of the payload in the packet header must be changed accordingly. MPLS (Multi-Protocol Label Switching), or VLAN tagging can be used to mark encapsulated packets, to distinguish them from non-encapsulated packets [9].

When a packet is too big for encapsulation, it is partitioned. The first part of the packet is inserted into the first encapsulating packet, and the second part is inserted into the second encapsulating packet, both parts contain the relevant geotags. See Fig. 2.

There are different ways to represent the location and the time in a geotag. These representations may be distinguished by the precision of the time and the location. For instance, we may represent locations at a granularity of several square kilometers rather than at the granularity of meters, to reduce the amount of bits needed for representing the location information. Typically, a location is represented as a pair of latitude and longitude coordinates. However, many switches are expected to have the same location (e.g., if they are located at the same data center). So, it is possible to use hash codes for efficient representation of the locations. When using

hashing, the hash code of the location will be stored in the packet, instead of the coordinates. The translation from a hash code to a location or from a location to a hash code can be carried out as a VNF, and can be deployed on a virtual machine. Hence, there is no need to add physical machines to implement the geotagging.

For authentication, the network nodes could sign the geotag. When a geotag g is added to a packet p, there is a need to associate them to each other. This can be done as follows. First, a cryptographic hash h (e.g., SHA256) is applied to a concatenation of the packet and the geotag, yielding $h(p|g)$. This can be signed by the node with the private key, e.g., using RSA, yielding $sign(h(p|g), K_{private})$, where $K_{private}$ is the private key. Note that the private key may be shared by all the nodes (but will be unknown to any user). The cryptosystem may also be such that each node would have its own pair of private and public keys. In such a case, the geotag should include the identifier of the node, so that a verifier of the authenticity of the tag would know which public key to use.

4 APPLICATIONS

We describe now potential applications that could use geotagging of IP packets.

4.1 Location Verification

With the advent of smartphones, location-based services (LBSs) have become ubiquitous. In some LBSs, the location of the user is needed for improving the service (e.g., users from different countries may see the information in different languages). In other cases, the service relies on knowing the location of the user (e.g., a search of type "find the nearest ...", or a system that only provides service to users that are located in a specific area). In some LBSs, users may have an incentive to report a fake location, e.g., in Foursquare users can check-in at a place and can gain a "mayorship"[1] status, which may provide benefits. In such services, it is desired to prevent users from reporting a fake location [2].

There are two direct ways by which services gain the location of the user. One is by asking the user app to find and send back the location, typically using GPS. The other is by inferring the user location from the source IP of the network connection.

GPS locations are quite accurate (the typical error is of several meters), but they have two drawbacks. One is that GPS is not always available. GPS reception can be limited inside buildings, and often the GPS receiver is disabled to save the battery power of the device. Another limitation is that GPS can be spoofed,[2] and users can easily cause their device to report a fake location or rely on such [12]. Extracting a location from the source IP address of the TCP/IP connection is problematic for two main reasons. First it is inaccurate, in comparison to GPS locations. Second, it may not reflect the location of the user, e.g., when changed by the NAT, the server would see the modified IP address and not the one of the user device.

As an alternative method, we suggest a *verified source location approach*. In this approach, the access point, where the user device accesses the network, would add to the first packet of the flow, or to all the packets, a geotag, which may include *(1)* its location, and *(2)* the reception strength (for mobile devices). Authentication information, such as signing the data using a private key, can be added to indicate the authenticity of the access point. The geotag can either be removed by the destination host, which could then use the information, or by the last intermediate node before the destination host, and the destination host could get the location information from that intermediate node.

This has several advantages over relying on IP in a naive way. First, in the case of mobile devices, by including the reception strength, the location of the user device could be estimated more accurately by the service. Second, the location of the user could be carried on to the destination even when changed by intermediate nodes like the NAT, because it is part of the payload. Third, by cryptographically signing the location, servers could rely on the provided information, and the system would be less vulnerable to spoofing. This approach could facilitate the implementation of location proofs [10] and location corroborations [4].

4.2 Time Verification

Time verification is similar to location verification. Some services may depend on the time at the location from which the service is requested. For example, a conference management system may let users submit a paper till 11:59 PM at the local time of the submitting user. A request for tender may work the same way. As with GPS, it is easy to change the time zone of the client device. But if the access point (e.g., the WiFi access point or the cellular antenna) would add a cryptographically signed time information to packets, servers could use that and rely on the reported time.

4.3 Detect/Monitor Movement of Devices

The Internet of Things (IoT) is expected to revolutionize many areas, including transportation and domestic services, having cars talking to one another, domestic appliances reporting their status to various servers, and different mobile devices, such as drones or autonomous vacuum cleaners, remotely controlled. In the IoT, the items ("things") would talk to one another via the network. In the era of IoT, it would be essential to reliably monitor the location of items, even approximately. The ability to add the location of the access point to packets sent by entities on the IoT would provide a reliable way to track their approximate location, even for items that are not equipped with a GPS receiver. For example, a detection of a movement of a lawn mower in the middle of the night would send an alert to the owner. A company like Amazon may want to have the capability to track its drones, e.g., as a backup in a case of no GPS reception, or if the device get stolen.

4.4 Geofencing and Geoblocking

For security or privacy reasons, it may be needed to restrict routing of particular flows to bounded areas. Geographically-constrained routing and geofencing are limitations on the routing of packets based of geographic constraints. For example, some US agency may want all the IP packets of its transmissions to go merely via routers and switches that are located on American soil. In such a case, the restriction would be encoded in the packet header, and the SDN controller would instruct the routers not to forward packets with a

[1] https://support.foursquare.com/hc/en-us/articles/201065220-Mayorships
[2] https://play.google.com/store/apps/details?id=com.lexa.fakegps

restricting predicate to routers, switches and servers that do not satisfy the predicate.

Geoblocking can be used for protecting copyrights and transmission rights over the Web, e.g., a TV network that wants all its broadcast to only be viewed from within the USA. Relying on the source IP of the viewer does not prevent foreign users from watching the content using a proxy with an American IP address. But if the packets of the broadcast are tagged with a predicate that does not permit forwarding them to routers, switches and servers outside the US, bypassing the restriction would be much harder.

4.5 Depicting the Geography of Flows

Often, it is required to understand geographic effects on the network. For example, how do a storm in New Jersey, a freeze in Nebraska or a flood in Louisiana affect the network traffic? When packets that travel from the west coast to the east coast can be routed via routers in Dallas, Chicago or Kansas City, network operators should be able to examine whether there is a geographical influence on the latency or the bandwidth. This may show the effect of the geography or the weather on network traffic, and would help to plan the network better. A holiday or some other event could have a local effect on the network (e.g., greater usage during the holiday, a smaller number of network engineers to cope with service issues due a local event, etc.) This can be used by operations support systems.

Gathered spatio-temporal information about packets could be used for analyzing geospatial *behavior patterns*, for improving the network. For example, what is the level of network activity in Manhattan at different hours of the day, on different days of the week? What is the geographic distribution of the servers/devices the hosts and devices in Manhattan interact with? Are the people of Chicago connected more to services and people in the east coast or in the west coast? Such information could help to configure the network in the short term, e.g., deploying VNFs to cope with a large local demand for network services, or on the long term, e.g., deploying more routers and communication lines in places where they are needed.

To examine geographical influence on network traffic, the SDN controller can instruct the routers and the switches to geotag the IP packets they route, and to send mirrored tagged packets to the controller. This will allow collecting information about the geographic route of packets. To prevent congestion or causing too much load on the routers, the information can be added to only a small percentage of the packets yielding a sample that would still provide statistically useful geospatial information. By doing so, it would be possible to collect information about paths of packets, and analyze geographic effects on the traffic flow. The lack of flexibility makes it difficult to do so in a traditional network.

4.6 IP Traceback

Tracing anonymous packet flooding attacks in the Internet back towards their source can improve the resilience of the network [11]. This can be facilitated by geotagging. By adding to packets their provenance, i.e., their origin and travel history, the transparency of the system would increase, providing more information to protect the system from attacks or from misconduct. The SDN controller would instruct routers and switches to geotag packets that are somewhat suspicious, but are not suspicious enough to be dropped, so that in a case of a revealed attack, it may be easier to trace it back or conduct a postmortem analysis to understand better where the packets were initially noticed by trusted network nodes, and when that happened. This may help improving the system's capabilities of defending itself.

5 CONCLUSION

Geotagging IP packets, by adding location and time to packets, has many important applications. It can improve the way routing is being done, strengthen the security of the network, support geofencing, and so on. In this paper, we depict a way by which geotagging of IP packets can be done flexibly and effectively by using the capabilities of an SDN controller to add action rules to routing tables of switches and routers, in real time. The functionality of adding tags or processing them can be supported by VNFs, which can be deployed on commodity hardware, ad hoc, even for a short time. Thus, adding geotags can be adjustable and inexpensive.

Geotagged IP packets provide an opportunity to boost existing services. They give rise to new research directions and raise important questions. This includes finding new ways to use geotagged packets, improving existing location-based services, and making the geotagging process more efficient. Effectively using the geotags in routers, switches and middleboxes, and combining geolocation with routing protocols are challenging open research questions, for both the SIGSPATIAL community and the networking community.

REFERENCES

[1] Pat Bosshart, Dan Daly, Glen Gibb, Martin Izzard, Nick McKeown, Jennifer Rexford, Cole Schlesinger, Dan Talayco, Amin Vahdat, George Varghese, et al. 2014. P4: Programming protocol-independent packet processors. *ACM SIGCOMM Computer Communication Review* 44, 3 (2014), 87–95.

[2] Bogdan Carbunar, Radu Sion, Rahul Potharaju, and Moussa Ehsan. 2012. The shy mayor: Private badges in geosocial networks. In *International Conference on Applied Cryptography and Network Security*. Springer, 436–454.

[3] Kaustubh Joshi and Theophilus Benson. 2016. Network Function Virtualization. *IEEE Internet Computing* 20, 6 (2016), 7–9.

[4] Yaron Kanza. 2016. Location Corroborations by Mobile Devices Without Traces. In *Proc. of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM.

[5] Keith Kirkpatrick. 2013. Software-defined networking. *Commun. ACM* 56, 9 (2013), 16–19.

[6] Diego Kreutz, Fernando MV Ramos, Paulo Esteves Verissimo, Christian Esteve Rothenberg, Siamak Azodolmolky, and Steve Uhlig. 2015. Software-defined networking: A comprehensive survey. *Proc. IEEE* 103, 1 (2015), 14–76.

[7] Jiebo Luo, Dhiraj Joshi, Jie Yu, and Andrew Gallagher. 2011. Geotagging in multimedia and computer vision—a survey. *Multimedia Tools and Applications* 51, 1 (2011), 187–211.

[8] Nick McKeown, Tom Anderson, Hari Balakrishnan, Guru Parulkar, Larry Peterson, Jennifer Rexford, Scott Shenker, and Jonathan Turner. 2008. OpenFlow: enabling innovation in campus networks. *ACM SIGCOMM Computer Communication Review* 38, 2 (2008), 69–74.

[9] Joshua Reich, Christopher Monsanto, Nate Foster, Jennifer Rexford, and David Walker. 2013. Modular SDN programming with Pyretic. *Technical Reprot of USENIX* (2013).

[10] Stefan Saroiu and Alec Wolman. 2009. Enabling New Mobile Applications with Location Proofs. In *Proc. of the 10th Workshop on Mobile Computing Systems and Applications*. ACM.

[11] Stefan Savage, David Wetherall, Anna Karlin, and Tom Anderson. 2000. Practical Network Support for IP Traceback. *SIGCOMM Comput. Commun. Rev.* 30, 4 (Aug. 2000), 295–306.

[12] Nils Ole Tippenhauer, Christina Pöpper, Kasper Bonne Rasmussen, and Srdjan Capkun. 2011. On the Requirements for Successful GPS Spoofing Attacks. In *Proceedings of the 18th ACM Conference on Computer and Communications Security (CCS '11)*. ACM, New York, NY, USA, 75–86.