PRIVACY PROTECTION AND MOBILITY ENHANCEMENT IN INTERNET

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Ping Zhang

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

May 2019

Purdue University

Indianapolis, Indiana

www.manaraa.com

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Arjan Durresi, Chair

Department of Computer Science

Dr. Yao Liang

Department of Computer Science

Dr. Xukai Zou

Department of Computer Science

Dr. Rajeev R. Raje

Department of Computer Science

Approved by:

Dr. Shiaofen Fang

Head of the Graduate Program

www.manaraa.com

TABLE OF CONTENTS

www.manaraa.com

Page

www.manaraa.com

Page

www.manaraa.com

LIST OF FIGURES

www.manaraa.com

ABSTRACT

Zhang, Ping Ph.D., Purdue University, May 2019. Privacy Protection and Mobility Enhancement in Internet. Major Professor: Arjan Durresi.

The Internet has substantially embraced mobility since last decade. Cellular data network carries majority of Internet mobile access traffic and become the de facto solution of accessing Internet in mobile fashion, while many clean-slate Internet mobility solutions were proposed but none of them has been largely deployed. Internet mobile users increasingly concern more about their privacy as both researches and real-world incidents show leaking of communication and location privacy could lead to serious consequences. Just the communication itself between mobile user and their peer users or websites could leak considerable privacy of mobile user, such as location history, to other parties. Additionally, comparing to ordinary Internet access, connecting through cellular network yet provides equivalent connection stability or longevity.

In this research we proposed a novelty paradigm that leverages concurrent far-side proxies to maximize network location privacy protection and minimize interruption and performance penalty brought by mobility. To avoid the deployment feasibility hurdle we also investigated the root causes impeding popularity of existing Internet mobility proposals and proposed guidelines on how to create an economical feasible solution for this goal. Based on these findings we designed a mobility support system offered as a value-added service by mobility service providers and built on elastic infrastructure that leverages various cloud aided designs, to satisfy economic feasibility and explore the architectural trade-offs among service QoS, economic viability, security and privacy.

www.manaraa.com

1 INTRODUCTION

Internet access paradigm has changed dramatically since last decade. Beyond traditionally accessing Internet from a computer, now Internet equipped devices have been ubiquitous and greatly altered Internet ecology. Internet access and usage is becoming much more user centric, and rapidly shifting to full mobility. Smart phones, for instance, cover more population than traditional computer users, and are carried by people and connect to Internet almost anytime. IoT devices, as another recent example, grow rapidly and are being deployed faster in order of magnitude than any other previous prevailing Internet devices. These anytime, anywhere, and from anything types of Internet connectivity hatch innumerous Internet applications, that cover almost every aspect of everyday life as they unprecedentedly make information so available, close, and convenient to access. Either through long range wireless access network such as cellular data network, or shorter range access such as IEEE 802.11/Wi-Fi or Bluetooth, Internet applications' client side more and more resides on mobility-capable devices and connect through wireless networking. The upcoming 5G cellular network aims to replace traditional wired Internet last hops with cellular data network. The edge of Internet is becoming full mobile and wireless.

However, on the other hand accessing Internet in mobile fashion still relies on the Internet IP core to route traffic from source to destination, which is almost same as two decades ago. Cellular data network provides limited roaming support but still couldn't solve mobility issue completely due to identity and locator coupling. Semi-seamless Internet roaming access is achieved through workarounds by Push service or individual applications. Additionally and more critically, this increase of Internet mobility also increases privacy exposure. In particular network location and identity privacy are facing more challenges, since Internet endhost essentially becomes mobile endhost. Thus there are more data and characteristics exposed from the mobile style network

www.manaraa.com

access, but protection mechanism doesn't get improved equally, nor being addressed by new access technology. For example, when a mobile endhost connects to another endhost, connections are setup with its exposed public IP address, which is either its actual public Internet attach point, or a gateway close to its physical location. That means from its exposed IP address all its peers can deduce the approximate geolocation of it.

This is today's real-world Internet mobility support and accompanied privacy vulnerability. The upcoming 5G and included Mobile Edge Computing (MEC) would not change the basic routing pattern but could bring more privacy concern when it opens previous restricted access network to 3rd parties. We are motivated to find a solution to fill this left gap of network privacy protection and generic Internet mobility support.

1.1 Internet Mobility Proposals Failed in Real World

Mobility has been and still is one of the top requirements for current and proposed next generation Internet. Significant research efforts are dedicated to find appropriate solutions for mobility in the Internet, aiming to implement a "anywhere, anytime" Internet connectivity experience. [1–14].

The overloaded IP address is recognized as one major impediment for Internet mobility [6, 7], and there are several proposals to split the tight bound of identifiers (ID) and addresses (Locator) of communicating entities, so logical connection can be kept when address changed due to mobility [8–10,15]. Other proposals include routing via invariant intermediate point [11, 12], or migrating connections from old address to new ones [13, 14]. However, none of the proposed Internet mobility solutions have been largely deployed, and the only available method to access Internet in mobile manner is through cellular data network.

David Clark, one of the Internet Architects once said: "*Internet is about routing money; routing packets is a side-effect.*" The Internet experience clearly indicates that

www.manaraa.com

no solution will be used in the Internet if it is not economically viable, independently how technically sound the solution is. A long list of examples illustrates this "axiom." So, various QoS solutions, including Differentiated Services (DiffServ) and Integrated Services (IntServ), whereas considered "technically" scalable, after more than one decade of intense research, and implementation in almost all endpoints and routers, are not being used extensively, mostly because they are not economically viable in the Internet. On the other hand, applications bridging the mobility gap have been successful, such as Skype, WeChat, Messenger, etc., because users (directly or indirectly) pay for the QoS of their applications and the corresponding service provider generates revenues from such service.

We believe that several critical economic flaws have also made many "technically feasible" mobility solutions infeasible in real world. *First*, existing solutions, based on static intermediary forwarding, such as Mobile IP [11], HIP [9, 16] and similar ones, require modifications on access networks. But, such ubiquitous deployment of network changes does not offer enough economic incentives, especially for service providers. *Second*, existing solutions require that all Internet users pay the cost of deployment and operations of the given mobility support, even though a large portion of users might not be mobile. *Finally*, while technical collaborations among involved service providers are required, there is no mechanism to split the revenues among them. Therefore, there are not enough economic incentives for such mobility services.

1.2 Cellular Data Network Taking Over

Despite research community's enormous effort, today's majority Internet mobility support is done through cellular service providers: mobile device receives a private IP address that's routable within cellular service provider network, and Internet traffic will go through a nearby Internet Gateway to public Internet. Generally, the IP address issued by cellular network is allowed to roam across limited distance and time, until then a new IP will be assigned and may also accompanied with changing

www.manaraa.com

of gateway. Existing connections must be terminated then re-initiated by the mobile devices.

For the last decade accessing Internet from cellular network grows tremendously. Comparing to other options cellular network is the most available and cost effective one to access Internet in a mobile fashion, and practically dominates all popular applications and platforms. The accessing devices are not limited to only cellphones, but also to other personal electronics, vehicles, buildings, or just simply replacing wired connections especially for the upcoming 5G.

Cellular network, although it is the most popular and successful business network that provides proved mobility mechanisms, cannot solve Internet mobility all by itself (which makes Internet an "overlay" above cellular network) and in fact it does need help from Internet mobility support to ease the burden on its internal backbones and gateways. Directly copying from cellular data network will not benefit Internet mobility researches much as these two types of networks are based on opposite principles and ownership model and Internet also does not have such regular and well optimized network topology as cellular network.

Due to the nature of intermittent communication and non-routable address behind gateway, it's difficult to resolve mobile device's network location and initiate connection to it by peer host itself. To address this issue, major mobile OS vendors and application vendors implemented "Push Notification" to emulate an on-demand message pushing service, such as Apple Push Notification Service (APNS), Google Cloud Messaging(GCM), or Microsoft Push Notification Service(MPNS). Under the hood mobile devices keep live connections with Push service providers to receive real time message. When one mobile user wants to communicate with another user, either the message is delivered through Push service, or leverage Push service to bootstrap a direct connection between mobile device and peer node.

The general availability is another issue of cellular based Internet mobility. Besides it is not available to devices that are not cellular network equipped, Push notification system are centralized proprietary services that different systems are not compatible

www.manaraa.com

with others. For example, in order to send message through APNS, both sender and receiver must be able talk to APNS and having APNS client installed. Also, the Push sender must register with APNS in prior. A device with only MPNS and another device only has APNS won't be able to leverage Push service to communicate. To enable cross Push system communication, applications have to manage the identity mapping and communication channel translation themselves with extra external services. Push notification as an indirect communication mode, nevertheless, cannot solve privacy issue solely. Due to the architecture limitation it can only be used to send small piece of data, e.g. 4KB as current standard. If peers want to use high-bandwidth communication such as video stream, a direct connection not through Push service must be created separately. Additionally, Push service are usually OS/vendor bounded, and without any legacy support. Existing applications cannot benefit from Push service unless reconstructed. Usually it's not an easy task and expensive as communication model is different.

The upcoming 5G cellular network won't change much in these areas as the major improvements are in network speed and latency and aimed to replace landline Internet. On the other hand, one new official component Mobile Edge Computing (MEC) provides infrastructure for us to implement functionality to enhance. We will discuss our use of MEC in following chapters in detail.

1.3 Mobile Internet Threats Location Privacy

When a mobile host connects to its peers, connections are setup on its exposed public IP address, which is either its actual public Internet attach point, or a gateway close to its physical location. That means using its IP address all its peers can identify the approximate geolocation of it. Even worse, peers can not only track the trajectory of the mobile host for its IP changes, but also capture high fidelity movement timings. Unfortunately, any website can track their user's IP history and run all kinds of analysis and data mining to model user's behavior. Mobile applications step one

www.manaraa.com

level further that can accurately track a single user's movement and is able to form a precise network address timeline, even when application is not granted access to GPS location. Additionally, any mobile app providing direct communication exposes mobile node's location history not only to the mobile app vendor, but possibly to all other contacts using the same app. Those information can be further used to project its future location statistically [17–19]. On one hand this type of prediction can be useful for certain purpose [20], but for a privacy concerned user it's definitely not good news.

The tighter bound of people's identity to their mobile devices and applications started to raise lots of privacy concerns. Research communities and industry have responded: on the service side a number of technologies were proposed to anonymize identity information when aggregating statistic or providing location based service [21]; on the client side mobile users are promoted to give explicit permission to applications of using device's geographic location, contact list, storage, etc. in order to protect users' privacy [22]. However, network access is usually granted without explicit approval since most applications need Internet access to function. As result mobile users cannot easily protect themselves from application vendors who can continuously track a mobile user's network locations through the periodical communication between mobile devices and their servers. These network location, even not as precise as GPS location, still reveals mobile user's relative geo-location. With a history of network locations it is not difficult to profile and identify individual users, probe on their current and past where-about, and estimate the places where they would be [18, 23].

VPN has become a popular service as more and more Internet users start to concern about their privacy. Through either private VPN service or multiple relay networks like Tor, Internet users can hide where they are when communicate with peer hosts or websites and hide who they talk to from ISP. However, today's VPN service doesn't provide extra support of mobility, and adds performance overhead as traffic always go through a static relay end host. The overhead will increase when mobile

www.manaraa.com

host moves to different network while still using same relay point. Additionally, VPN services don't handle incoming connection well.

Mobile users can choose to use Virtual Private Network (VPN) to proxy their traffic for protecting their real network locations: VPN can tunnel traffic via VPN server that is away from mobile user, so remote peers could only see VPN server's network address instead of mobile node's. Usually VPN service providers offer a few location options for their customers to choose in prior. Once selected all traffic of a mobile node will go through the chosen VPN server. It becomes a dilemma to choose VPN server: too close it correlates mobile node's location; too far away connections will be diverted away from optimal route which then incurs latency overhead and other limitations. When mobile node moves, the changes of network attach point will interrupt connections. Performance will be degraded when node moves away from selected VPN server.

1.4 Dissertation Structure

The goal of this research is to propose a viable solution that can protect Internet mobile user's network privacy while enhancing generic mobility support. We started by identifying the privacy issue and mobility support gap in current Mobile Internet, then researched the root cause of why they are left unresolved, and at the end proposed theory, new paradigm, and solution to solve them in this dissertation. In Chapter 2 we will review notable previous researches on Internet mobility and privacy protection. Then we propose our theory and design of a mobility support and network privacy protection system in Chapter 3. Next, we present details of system design including algorithms and simulation results for validation in Chapter 4. At last we conclude in Chapter 5.

www.manaraa.com

2 BACKGROUND AND RELATED WORKS

In this chapter we will investigate the fundamentals of Internet mobility and privacy, summarize basic requirements, review a few most notable researches before we propose our theory and design in next chapter.

2.1 Requirement of Internet Mobility

The best way to recognize Internet mobility is to identify function requirements. In this section we will separate them into two tiers: the first tier that are fundamental and must be provided by any solutions, and second tier that are still critical but only start to receive more attention recently.

2.1.1 Basic requirement

Basic requirements are mostly functionality requirements, i.e. without satisfying all of them, Internet mobility support is not complete.

- Reachability. This is the most fundamental requirement of mobility support that a mobile Internet user is always reachable by other Internet users, which means the movement of an Internet user will not prevent the delivery of messages designated to it. This requirement does not imply that communicating peers would know the exact Point-of-Attachment address of the mobile node since that depends on the support mechanism, but it ensures Internet end hosts can always send data to a mobile Internet user and know whether the data are delivered, and vice versa.

- Continuity. The established communication should not be broken by the movement of the mobile Internet users. This requirement generally can be referred to

www.manaraa.com

handover, although the continuous communication does not necessarily mean continuous connection for example a TCP connection could be reestablished without breaking context at application level, or restricted on the same device such as that an Internet user can migrate his identity among multiple devices or uses them simultaneously.

- Ubiquity. The mobility support should be available anywhere when a generic Internet connection presents. That suggests the Internet mobility support should not bind to any specific link layer or physical layer networks, devices, or protocols, otherwise absence of any precondition would void the support.

- Security. Besides the security issues which inherit from static Internet and wireless network, a few new security problems are introduced accompanied with Internet mobility where one user may have arbitrary Point of Attachment (PoA) and one PoA would be reused by different identities at distinct time. Authentication on identity becomes necessary for both initial conversation setup and following movement or PoA change. Address authentication might be required as well in some special scenarios. Confidentiality and integrity are facing more challenges as well. The previous or later PoA occupier should not be able to deceive the communication, nor chances to perform any Man-In-The-Middle or relaying attack. Mobility support solutions should also prevent attacks aiming to "block" a legitimate user by maliciously updating false PoA, or other DoS attacks. Roaming across the boundary where different access policies are applied would produce more sophisticated regulation requirement.

- Scalability. This is always one fundamental requirement for any network architecture and protocols, though it is also hard to justify. Since the Internet mobility support is for the whole Internet, then at least any general solution should be able to support millions of users simultaneously and could be tailored to a specific scale when needed.

www.manaraa.com

2.1.2 Advanced requirement

Besides the basic requirements that every Internet mobility support solution must satisfy, several additional requirements may be required by specific applications, or

- Legacy application compatibility. Backward compatibility is necessary for general mobility support solutions, though it depends on how much benefit could the legacy applications gain from mobility support and how transparent the support mechanism is. Any solutions that need modification on existing applications are not considered legacy application compatible. This backward compatibility also could be extended to operating system, i.e. no need to upgrade or modify OS kernel.

- QoS: Latency. Generally, for wireless communication devices it is difficult to provide assurance for QoS. On the other hand, de facto we all know traditionally Internet architecture doesn't provide QoS also. So, when these two come together, we can image how difficult to support QoS.

  As QoS comes to Internet mobility support, latency turns into most important criteria. Latency are a highly application dependent criteria. It may refer to the time used to find peers(resolution), setup connections, reconfigure after moving(handover), or routing overhead. Latency of resolution and handover are more significant since they may be performed repeatedly, and extra routing overhead may apply if indirectly routing is employed. It is generally bounded by time out length of communication protocols. For real time applications this interval may not exceed several RTTs, and for others this interval may be limited to several seconds.

- Privacy. Privacy is not an original design goal of Internet, and to some extent implementation of privacy could impair performance. Nevertheless, for mobile Internet more privacy issues appear. One issue is the location privacy. Current Internet doesn't provide mechanism to protect location privacy, though its mix

www.manaraa.com

of identity and address helps a little. Several projects are proposed to hide IP during communication [24, 25], though high cost such as efficiency has to be paid. For mobile Internet, the location information of a specific user might become public. That means any peers having connection to this user might know where he was from his PoA address. Then the exposed information could be used to trace this user by mapping PoA address to geographic location and even disclose his real identity and life pattern. Another issue is that access network provider may have chance to touch personal information of users using its Internet access.

- Energy and computation resource constrain. Apparently, a large portion of devices used to access mobile Internet are hand handle or wireless devices. For those devices, energy and computation resource usually are restricted. Besides Physical and Data Link layer, designs of higher layer protocols and architecture could have implicit impact on resource consuming, especially for wireless devices. For example, reducing the amount of signaling would reduce the energy consumption.

- Device temporary offline. Device sleeping had been proved to be necessary and effective by cellular and wireless sensor network industry, to encounter energy constrain. In addition, wireless connection could be affected by various factors and the radio signal would not always cover all areas. Both need special treatment: all communication may be stopped for a while so state resume mechanism is needed; network may lose track of nodes or have duplicate/inconsistent records; a large amount of undelivered data may need to be cached and redelivered; actively node wakeup is needed in case of emergency.

- Accountability. Accountability is a dampened requirement of original Internet since at that time there was not much commercial factor in Internet. Now Internet is ran by many independent commercial companies and Internet users pay for their Internet access. Mobility support may need accountability if the

www.manaraa.com

payment needs to be shared. Another fact we want to note here is that a large portion of Internet nodes would always be static, and they would not like to pay for.

Requirements listed above are extensive but not comprehensive. Despite that, to satisfy these requirements is not an easy task either.

2.2 Internet Mobility Paradigms

In this section we will review several primary methods of supporting Internet mobility. They may not be completely parallel with each other to some extent, and some of them certainly can be combined to provide more comprehensive support.

2.2.1 Message Box

Message box is one of the oldest methods to delivery message indirectly between applications and is still a major one. Its inherited indirect pattern avoids the resolution of message receiver's address. Messages sent are stored at intermediate node, the message boxes, so wherever receivers move they can always check their message box using any computer having Internet access. No message will be lost due to receivers' roaming or off line, and neither the sender nor the box needs to know PoA address of the receiver. One typical example of message box is Email that is one of the oldest Internet killer applications. This active retrieval paradigm is also adopted by a few solutions, such as i3 [15] which is discussed in Sec.2.3.5.

On the down side, the message box method has a few inherited disadvantages for mobility support: 1)It is not real time communication. The latency of receiving message depends on how frequently the receiver check his message box intentionally. 2)The sender cannot know whether the receiver is online or offline, nor whether the message arrives in a timely manner. It is a practically one-way communication and the acknowledgment message must go through the same process. 3) It requires the mes-

www.manaraa.com

sage box having a considerable storage capacity. 4) Routing inefficiency. Typically, messages are routed through the sender, receiver's message box, and the receiver. In some solutions the message may be relayed via several intermediate servers, such as email.

2.2.2 State/session resume

Session layer is part of OSI reference model and originally designated to support session suspension and resume, which could implement part of handover functions. However, this mechanism was not needed much in the early stage of Internet and the TCP/IP model leaded to the obsolete of session layer. The function of session layer is then merged into application layer and handled by applications. The result of this is that applications use their own methods to implement session instead of a general one. "Cookie" used in web browser is an example of session abstraction implemented in application layer [26]. It is a piece of data which stores sets of states and is generated by server or webpage scripts but stored at client side by browser. Each time when the user visit websites the browser will send corresponding cookie to present a gloss that communication is not disturbed by disconnection. TESLA [27] is a proposal of providing a general solution aiming to help applications implement session without extra effort.

The state/session suspension and resume mechanisms are good complement to mobility support solutions in transport layer or below. On the mobile user side, state/session resume mechanism could maintain the application states and reconstruct connection which may be closed due to timeout at lower layers. Note that though the session layer in OSI reference model is independent and distinguished with other layers, the state/session resume mechanism could be used in protocols in any layer and used more than one time.

www.manaraa.com

2.2.3 ID/Locator split

One well known issue of supporting Internet mobility is the tight coupling of IP address and host identity, which is inherited from IP protocol. In static network routing a packet to an address practically deliveries that packet to the station that owns the address. In mobile environment, nodes typically change their location as well as PoA frequently while moving. The most intuitive solutions are whether let the node "carry" the IP address with it or give the node a "name" and bind connections on that name instead of IP address. The former one would greatly degrade IP routing efficiency except alternative mechanisms used (refer to Sec.2.2.5). The latter is to separate communication identity and routing locator by setting up another level of abstract name above IP address. This is the well-known and accepted "ID/Locator split" idea and is recognized as one radical cure to Internet mobility. Separating ID and locator also could benefit Internet for other means, such as mentioned in [28]. However, this behavior introduces challenges to legacy Internet protocols. TCP and other connection-based protocols will become invalid since connections are bounded on both sides' IP address. Furthermore, hosts need a way to discover the IP address of mobile nodes and follow its change. On the other hand, every scheme which employs ID/Locator split paradigm could claim its support of mobility in a degree. However, only introduce an ID system will not simply solve all mobility issue. A few open issues are left such as transport layer modification and management of ID/locator mapping, etc.

2.2.4 End-to-end connection reconfiguration/migration

AS we discussed above that current Internet protocols are IP address bound, then another intuitive idea of remedy is to reconfigure or migrate the existing connections after each move. For example, when a mobile user moves and receives a new IP address, it will send this new address to its communicating peers and then both sides could simultaneously reconfigure previous connections by binding them on new

www.manaraa.com

IP address. This is often referred as host mobility and is a well examined topic. SCTP [29] has capability of address reconfiguration [30], and another typical TCP solution is presented in [13]. Most of ID/Locator split solutions also adopts end-to-end connection reconfiguration for host mobility, such as HIP [31].

This type of solutions follows the Internet end-to-end principle that the change of endpoint addresses can be accomplished without a third party. It requires no changes to the IP forwarding infrastructure either, instead modifying transport protocols and applications at the end hosts. Nevertheless, some open issues are still left too. First, this mechanism may disturb the connection related states of legacy protocols and applications, such as states for flow control and congestion control.The synchronize and keep-alive mechanism may timeout due to no incoming signal for long time. Next, the modification on end host's protocol stack introduces backward compatibility difficulty. Unless both sides deploy the same modified protocol, they cannot talk to each other. Additionally, the lack of third party bring an interesting issue: "dual moving", in which both sides move simultaneously and send the update messages to the old address of the other. Therefore none of them would be able to receive the update.Finally, an external helper is always needed for initial lookup and connection setup, and authentication and privacy when needed.

2.2.5 Indirection routing

Current Internet architecture are based on end-to-end principle which emphasizes the intelligence and state information are kept at end points and the core of Internet is kept simple and stateless. Indirection routing solutions loosen the hold of principle a little and are proposed to support mobility more transparently. Generally, this type of solutions employs one or more fixed end points as intermediate points and do indirection routing via the intermediate points. Through this way, the address change of mobile nodes can be hidden from senders since senders would think that they are communicating with a static peer and always send to the fixed delegates. The most

www.manaraa.com

obvious problem of this type of solutions is the inefficiency of routing and the need of delegation. Modification on IP and transport protocols may also be needed at the mobile nodes. The scenario that both sides are mobile nodes would make this type of methods more complicated and less efficient.

2.2.6 Dumb terminal

The dumb terminal architecture has the longest history in computer networks and is the always standard of telecom industry including PSTN and cellular network. For a dumb terminal the mobility issue would become much simpler since all applications are running at the centralized server and most of states are maintained at server side as well. Therefore, temporary disconnection will not affect the applications, and handover can be simply implemented by reopening connection. Recently the popularity of "Cloud Computing" and powerful large data center bring the dumb terminal idea back again. With the help of data center, an individual user can store up to several GigaBytes data and even run complex applications such as 3D games on centralized servers. To enable dumb Internet mobile terminal, the Internet network infrastructure must become intelligent to do more work than it has been. However, to abandon the powerful end point computers and construct a tremendous rich-feature global network does not sound very scalable, and ironically cellular networks are simplifying their backbone and push more works to cellphones.

2.3 Notable Mobility Works and Researches

2.3.1 Mobile IP and Variants

Mobile IP(MIP) is a family of IETF standards primarily defined in [11, 32, 33]. It is among the most popular Internet mobility support solutions and has many variants and enhancements [2,8,34,35]. NEMO (NEtwork MObility) described in [36] is about the ubiquitous support of MIP. MIP uses static IP address as invariant identifier and

www.manaraa.com

tunnels packet between invariant Home Agent (HA) and local Foreign Agent (FA) where Mobile Host (MH) resides. MIP family solutions are compatible with legacy applications to some extents since IP addresses bound in sockets are static. However, MIP requires access network modification and collaborated charging that is only feasible in cellular and similar proprietary networks, plus the drawbacks of inefficient triangle routing and mix of IP address and identity.

Milind Buddhikot et al. propose a MIP compatible architecture MobileNAT [12], which locally translates between the invariant virtual IP address as ID and the dynamic one as locator. The key ideas in this architecture are: 1) use two IP addresses - an invariant virtual IP address for host identification at the application layer and an actual routable address at the network layer that changes due to mobility. 2) DHCP enhancements to distribute the two addresses. 3) a signaling element called Mobility Manager (MM) that uses Middlebox Communication (MIDCOM) framework to signal the changes in packet processing rules to the Network Address Translators (NATs) in the event of node mobility. This proposal does not require any modifications to the access networks and can seamlessly co-exist with the existing Mobile IP mechanisms and therefore can be used to provide seamless mobility across heterogeneous wireline and wireless networks. On the contrary, Proxy MIP [37] is another MIP variant that eliminates the requirement of modification on mobile node, by introducing a Mobile Proxy Agent at access network to delegate MIP functions on behalf of the mobile node to make mobility transparent and let the mobile node think it never leaves the home network.

2.3.2 DHARMA

DHARMA [38] is an overlay network improvement over MIP that provides session layer function to support constant connectivity while roaming or sleeping. DHARMA selects a location-optimized one from a set of distributed home agents to minimize routing overheads. Set management and optimization are done using the PlanetLab

www.manaraa.com

overlay network. DHARMA's session support facilitates transitions between home agents and improves intermittent connectivity. Cross-layer information sharing between the session layer and the overlay network are used to exploit multiple wireless links when available. DHARMA improves routing efficiency when do triangle or rectangle routing and is compatible with current and legacy application. However, a few issues exist such as HA has no knowledge of legacy application semantic thus may result in connection close from server side, and a well distributed overlay network must present, and each node must be stable for a long period and capable to share resource and bandwidth.

2.3.3 HIP

The Host Identity Protocol (HIP) [9,39] is an architecture that separates identifier and IP address by introducing Host Identity (HI), which is based on public keys and in the format of IPv6 address, to replace IP address for connection setup. HI is initially acquired by DNS lookup [40], and mobile node keeps updating peers and DNS record during move [31]. A HIP local daemon is responsible for replacing HI with corresponding IP address when packets are sent to IP network. Rendezvous server is defined for highly mobile nodes [41]. HIP is initially designed for end-to-end security and supports mobility by the benefit of ID/locator split and end-to-end locator update. It is one of most recent proposed solution and drawing much of attention. HIP presents to be a clean-slate solution and aims to stack up IP layer at end points. For current state it is not a complete solution for mobility support yet. For example, it lacks definition of an efficient mobility management system. The simple end-to-end locator update mechanism cannot deal with scenarios such as dual moving, backward compatibility, and location privacy.

www.manaraa.com

2.3.4 FARA

FARA [6]is an abstract high-level architecture model aimed to provide general guide line and a flexible framework for clean slate Internet architecture. The major concept of FARA are using decoupled communication entities from network forwarding mechanism, logic connection between entities called "associations", and new forwarding substrate called "Forwarding Directive" to form a flexible Internet architecture. Mobility is one of the major concerns of FARA, and it is primarily addressed by ID/locater split. FARA suggests using rendezvous point to setup initial connection to mobile node, or use directory service (DS) to lookup and keep track of mobile node. M-FARA is a conceptual implementation of FARA which targets mobility support. In M-FARA a "M-Agent (Mobile Agent)" is a static third-party rendezvous point used to update address information to support mobility.

2.3.5 i3

Internet Indirection Infrastructure (i3) [15]is an overlay network that offers an indirect passive routing model. Instead of pushing data actively from sender side, receivers express their interest of data in order to "pull" data from i3 network. The interest is marked by an identifier (including sender's address and port) of a specific type of packets, which is called a "trigger". i3 provides an alternative abstraction of Internet's end-to-end principle and emphasizes the motivation of receiver. This approach benefits routing schemes of multicast and anycast, especially for the case that sender does not have enough or exact information of receiver's identity or location. It also alleviates the difficulty of deployment by using overlay technique. Due to the indirect routing mechanism, i3 supports simple mobility under its architecture, though sophisticated function may not be able to be implemented due to lack of direct and responsive channel.

Robust Overlay Architecture for Mobility (ROAM) [42] is a proposal to provide seamless mobility for Internet hosts based on i3. ROAM takes advantage of end-host

www.manaraa.com

ability to control the placement of indirection points in i3 to provide more efficient routing and fast handover and preserve location privacy for mobile hosts. In addition, ROAM allows end hosts to move simultaneously, and is as robust as the underlying IP network to node failure. Hi3 [43] is a combination of HIP and i3. It inherits the architecture of HIP and use i3 as the mobility management system.

2.3.6 LISP

The Locator/Identifier Separation Protocol (LISP) [28] is another architecture based on the separation of identifier, called Endpoint Identifiers (EIDs), and address, called Routing Locators (RLOCs). EID-to-RLOC mapping are performed at RLOC router and routing are accomplished by tunneling between RLOC routers. LISP is a clean-slate architecture solution.

2.3.7 MobilityFirst

MobilityFirst [44] is another future Internet architecture that tries to address mobility as the first level foundation of further Internet. The authors argued that Content Centric Networking (CCN) was good for locating content in network but was not scalable for routing on Internet. Instead they employ an identity and locator separation mechanism, hybrid GUID and network address (HGN) routing, that uses a Global Unique Identifier (GUID) to identify content and a distributed service Global Name Resolution Service (GNRS) to map GUID to network address. The GUID was considered as the "narrow waist" of MobilityFirst architecture.

Neither MobileIP nor DNS was considered suitable to manage ID/locator mapping for MobilityFirst. Instead MobilityFirst translates human readable name to GUID by "name assignment" services, then registers in distributed database implemented by DMap [45] which is "a single overlay hop DHT". DMap hash GUID into network address such as IP and use router that "owns" that IP to store the actual GUID to mobile node's address mapping. To improve efficiency K random storage network

www.manaraa.com

addresses are selected that hopefully could place replica closer to originate of lookup request. In case of hashing result in unallocated IP DMap would choose a deputy AS router that has minimum "IP distance" for it. Although the authors claimed that DMap does not require storage of additional state information, it still need to be deployed to every router of all ASes. The simulation result showed lookup latency could be around 100ms for 26000 ASes using data gathered from DIMES.

For routing and forwarding every router will make decision of whether store packet, resolve GUID, reroute, etc. As result concept of End-to-End connection will not hold any more, and packet will change its destination network address on the fly.

The author also vision that cellular carrier could potentially build a distributed virtual private network on top of public Internet, and cellphones connected through carrier's E-UTRAN and EPC actually transmit traffic on public Internet infrastructure [46]. As a result, cellular carrier only needs to keep radio infrastructure and customer relationships.

2.3.8 Cellular IP and Columbia IP Micro-mobility Suite (CIMS)

Cellular IP [47] is a protocol that allows routing IP datagrams to a mobile host. It is intended to provide local mobility, hard and semi-soft fast handover support, and IP paging. Cellular IP uses mobile originated data packets to maintain reverse path routes. IP addresses is used to identify mobile hosts. Cellular IP semisoft handover exploits the notion that some mobile hosts can simultaneously receive packets from the new and old base stations during handover. Distinguishing idle and active mobile hosts reduces power consumption at the terminal side. The location of idle hosts is tracked only approximately by Cellular IP. Therefore, mobile hosts do not have to update their location after each handover. When packets need to be sent to an idle mobile host, the host is paged using a limited scope broadcast. A mobile host becomes active upon reception of a paging packet and starts updating its location until it moves to an idle state again. CIMS [48] is mobility support set including

www.manaraa.com

Cellular IP, Hawaii, and Hierarchical Mobile IP. The Hawaii supports Unicast Non-Forwarding (UNF) and Multiple Stream Forwarding (MSF) schemes.

2.4 Mobile Internet and Cellular Data Network

Cellular network is the most successful and largest mobile network all over the world. Because mobile Internet and cellular network share many similarities and comprehensive researches have been done on cellular network, researchers could learn a lot from it. Nevertheless, researches of mobile Internet are not intended to "reinvent the wheel", since these two networks are based on different purposes and design philosophy. The Internet mobility cannot be implemented by simply relying on cellular network either. Before discussing the difference, we will take a brief review on how cellular network deals with mobility and how it supports data communication. Though there are various cellular protocols sets available, they share similar architecture. Here we use GSM [49] and General Packet Radio Service(GPRS) [50] as the examples.

2.4.1 Cellular Mobility Management

In GSM a subscriber identity module (SIM) chip is used to represent a cellular user(cellular service subscriber). It stores certain parameters including: cellphone number, international mobile subscriber number(IMSI) , and other security and auxiliary information. A subscriber's identity is bound to the corresponding SIM chip. Each subscriber has a Home Location Register(HLR), which is a database server permanently storing all data of the SIM and detail of service parameter including billing information, and current location of the subscriber. When a mobile station(cellphone) connects to cellular network, local Gateway Mobile Services Switching Center(GMSC), which controls local cellular network, will request information from HLR according to the identifier reported by mobile station, and then stores in Visitor Location Register(VLR). The VLR ID is then updated at HLR by GMSC in order to

www.manaraa.com

paging the subscriber later. The mobile station(cellphone) is the most important part to achieve mobility management. When it moves, it will periodically check location area codes broadcast by base stations and employ a periodic location update procedure to update its location information at VLR and HLR. This process of roaming to another cellular is performed similar as the cellphone initial enters the network: user information is retrieved from HLR and roaming agreement is checked as well; then local network will decide whether accepting the roaming cellphone then updating HLR. When a call comes in, it will first reach the subscriber's HLR derived from IMSI, then HLR will return a temporary number provided by VLR, or an error code if no VLR is currently registered.

2.4.2 Cellular data network

Due to the circuit switching mode, in cellular network calls are relatively easy to manage and quality of service could be guaranteed since resources can be assigned in prior. However, this circuit switching mode doesn't suit the discrete data packets transmission well. Users do not want to pay the cost for idle connection which is charged the same even when the cellphone has no data packet in transmission. GPRS is a packet-oriented solution intends to support OSI model and IP protocol to provide burst data packet transmission on a shared TDMA channel. It works at IP layer and below to present a general IP interface to both ends. The cellular network assigns an IP address to the cellphone, either public or private according to cellular company's policy. A Serving GPRS Support Node (SGSN) connected by several base stations is responsible for diverting traffic to voice network or IP network. Then a Gateway GPRS Support Node (GGSN) which connects SGSN will act as an Internet router to forward the packets. The SGSN and GGSN works as intermediate proxy between mobile host and Internet. Since the roaming changes the attach point of the mobile station, in order to maintain a transparent continuous connection cellular network uses tunnel between original GGSN and local SGSN to maintain the exposed public

www.manaraa.com

IP address plus port. Therefore, the IP packets from public Internet are still received by the original GGSN, and then forwarded by the cellular network to the new local SGSN. The IP address assigned to the cellphone is also kept so sockets bound to it will not be affected as well as the applications. Through this way the mobility is handled inside the cellular network, with the tradeoff of increased routing overhead. From the description above we could infer that cellular data network has an ID/locator split naming framework and employs architecture very similar to MIP. For example, HA corresponds to HLR plus GGSN and FA corresponds to VLR plus SGSN. MIP is also very popular in cellular network and used to increase the flexibility and compatibility of cellular data network, such as tunneling between GGSNs in public Internet instead of purely within cellular network.

Though currently cellular data network is the largest network of accessing Internet mobility, it cannot solve all Internet Mobility issues. To support Internet mobility via cellular network will de facto make Internet an "overlay" network above cellular network. Providing major functions in the "underlay" cellular network is not efficient either for routing or protocols operation . For example, TCP is not able to know packet dropping is caused by network congestion or by radio interference. The retransmission mechanism of base station will make the scenario more complicate. The routing efficiency could be impaired too because all incoming and outgoing traffic must go through the original GGSN no matter where the cellphone moves, unless MIP is employed to stack up another layer of mobility support. When both endpoints are cellular subscribers the overhead will increase further. Additionally, accessing Internet from cellular network will tightly bind Internet users to specific cellular service providers, which differentiates the mobility support service to other common Internet services that can be accessed anywhere. In the case when a user having generic Internet access but is out of the coverage of his cellular service provider, Internet mobility is not available neither.

The experience of cellular network cannot be directly copied to Internet. Though cellular network backbone is on the trend of All IP Network(AIPN), cellular networks

www.manaraa.com

are completely proprietary, and the topology of cellular access network are highly corresponding to geographic topology. The proprietary network leads to cheap routing within own network and contrary when outside. It also makes collaboration billing feasible within own network. The correspondence to geography also makes handover and mobility management much easier, though "jump" of Internet mobile nodes (two access networks are geographically close but topologically far away) may show different mobility pattern compared to cellular network [51]. In addition, the rule of dumb terminal and intelligent network of cellular network opposites Internet's traditional end-to-end principle that makes the copy of cellular network more difficult.

2.4.3 5G and Mobile/Multi-access Edge Computing (MEC)

The fifth-generation cellular network (5G) [52, 53] is already on horizon and has started pilot deployment. Compared to current 4G cellular data network, 5G focuses on much faster speed and lower latency, but no notable changes to mobility management paradigm or IP backbone. On the other hand, one new component Mobile Edge Computing (or Multi-access Edge Computing, MEC) brings ubiquitous public computing that creates opportunity for systems that can benefits from computing at Internet edge.

Edge Computing(EC), the concept of moving data and computation to network edge rooted from Content Delivery Network (CDN) [54]. Recently it came back to draw more attention: a few paradigms, such as Fog Computing [55, 56], Mobile Edge Computing [57–59], Edge Cloud Computing, etc. are all further extending this idea. Compared to traditional Cloud Computing which aggregates computation and data at centralized data centers, Edge Computing utilizes the computation and storage capability of the edge. This architecture not only leverages the sparsely distributed but combined vast edge resources, but also pushes data and their processing closer to end users. After all, no matter how much higher bandwidth modern Internet provides, the speed of light still remains the same: the latency of sending a bit across countries

www.manaraa.com

didn't change much albeit the bandwidth is magnitudes higher. Additionally, there are always sensitive data that end users don't want to transfer to data centers, or at least must be partially processed and trimmed before sending to remote backend servers.

MEC is one of the most practical EC paradigms, mostly because of its already ubiquitous presence and capable network and computation resource, while at the same time cellular network is already powering the majority of mobile Internet users. Essentially, it's analogous of cloud computing that mobile radio service provider/cellular operators allow generic application to run on their ubiquitous Radio Access Network (RAN) network controllers and base stations. Compared to data center-based cloud service offering, this cloud is closer to end customers, having less latency and higher availability, and more importantly this architecture offers capability to application to access or manage cellular network traffic and configuration, thus easier and cheaper to implement a number of network services, and with less operation cost. The notion of Multi-access EC beyond Mobile EC wants to expand its ubiquity even further such as including Wi-Fi network into the picture. MEC as one of the most practical Edge Computing architectures, has great potentials of implementing(or enhancing) emerging network paradigms, such as 5G network, Fog computing, and Edge Cloud Computing. It roots in cellular networks, and could quickly become prevailing without long ramp-up years.

2.5 Notable Privacy Works and Researches

Privacy issue caused by mobility, especially location, has been well studied. Montjoye et al. [60] found that from a large set of anonymous movement data, using four data points of hourly data can identify 95% unique users. Given the identified movement pattern of this identified user, they can even construct a history of this user's locations from the anonymized data set. Ma et al. reached the same conclusion [23]. This clearly shows location information, even after anonymization, can greatly threat

www.manaraa.com

a mobile user's privacy when it can be collected by adversary. Cloud Computing and Big Data just make this exploit more available and accessible [61]. However, without proper protection, a mobile host cannot hide its location since it is network location will be exposed to any peer it communicates and the network location is approximation of its physical location.

The TOR network [25] is an overlay network designed to allow an Internet end-host to secretly communicate with peer end-hosts without exposing either content, source, or destination to traffic carrier. It achieves that by source creating a circuit step by step until reach destination and being the only one has that knowledge. Each redirection point in a circuit selected by source, called Onion Router, only knows its ancestor and descendent and owns a unique session key with source. Source uses several layers of encryption to achieve confidentiality and control of exit point (leak-pipe) so that no Onion Router can trace more than one step of the traffic either north or south.

Tor was designed and implemented as a non-prototype application can be used on Internet by real users, and quickly became one of the most popular utility to access Internet privately. It provides perfect forward secrecy and source-controlled path promises anonymity and can avoid filtering and traffic analysis from ISPs. It also enables a way to provide anonymous service. The most trade-off of Tor's privacy protection is increased latency, especially for application such as webpage browsing due to small file size and multiple concurrent TCP connections. Tor could be attacked on exit node if no end-to-end encryption and authentication deployed, or when majority of Onion Routers are controlled by single identity. Tor users also face dilemma between performance and path length.

Another popular research area is to protect against user profiling from Location Based Service (LBS) while can still use it [62–64]. Wernke et al. [65] surveyed different identify protection types and common mechanisms to protect and attack privacy. Usually a compromise between quality of service and privacy is optimized by manipulating location reporting frequency, precision, or both [66]. For example Shokri and

www.manaraa.com

Theodorakopoulos et al. [67] designed an approach to hide user's profile against adversary by solving it as Bayesian Stackelberg game. Primault et al. proposed mechanism to reduce profiling exposure, by hiding POI where user stops, and let user exchange their trajectories when meet [68]. On the other hand, researches also show obfuscated location data while can improve location privacy but cannot stop adversary to infer relative precise Point Of Interests (POI) [69]

Another focused area is to increase anonymity of collected user location data [21], limiting shared location information, or evaluating privacy exposure level before sharing location data [22]. On the other hand, there are researches pointed out that because human mobility trace is very unique [60], even completed anatomized data can still be used to extract patterns and identify individuals [17–19]. So as long as relative location and movement is collected, location privacy can be compromised to certain context.

Privacy is also a major concern of Internet Of Things (IoT) [70–72], as wearable devices constantly collect sensitive information and communicate with other Internet hosts, more data will be available to profile a user more precisely.

MobilityFirst [44] is a proposed new Internet architecture that emphasizes mobility support, comparing to existing Internet architecture. Privacy and communication security are one major challenge for MobilityFirst. Access control is proposed to apply to MobilityFirst so that only allowed network entities can contact a host or resolve its network locations. [73]

Shi et al. [74] summarized the characteristic of "new" Edge Computing pushed by blooming of Cloud Computing and popularity of Internet Of Things (IoT), and listed a few challenges and opportunities of this new area. Standards of MEC are being actively developed by cellular industry and standardization group [57]. It is considered as one building block of 5G network [75,76], and its capability of offloading computation from core network is especially emphasized [77].

www.manaraa.com

Roman et al. [78] studied and summarized common security threats and challenges of multiple paradigms used in mobile edge computing and mobile cloud computing, and pointed out needed improvement of security design.

2.6 Internet of Things

Internet of Things(IoT) has received substantial interest from academic research and industry in last a few years [72, 79]. With capability of ubiquitously and continuously collecting, processing, and responding to real world environment, it has vast potential that can become next greatest innovation after Internet. Personal electronics, office equipment, house appliances, public infrastructures, and ubiquitous sensors: in IoT everything can be connected to Internet and also interconnected. They communicate locally, regionally, and finally to Internet. Collections of these devices and their supporting backend processing system are combined to create intelligent system, such as smart home [80], smart infrastructure [81], or even smart city [82], etc. IoT is expected to greatly automate daily routines, free people from repeatedly intervention, improve efficiency and eventually create new life styles and business opportunities.

IoT devices collect data at different scale and granularity. Typically, even a small setup will collect massive amount of metrics and data, and upload large quantity of raw or processed data to remote service, in order to store and/or further process. For example, a smart home setup can upload house's temperature, humidity, and electric usage every a few minutes. Additionally, it could continuously upload status of garage, doors, and windows for security. For controlling the smart home may continuously monitor presence of other IoT devices to turn on or off of lights accordingly. The house's centralized system waits to receive control signals from remote service which relays input house owner sent from remote devices: such as increasing thermostats' temperature setting since they are coming back to home.

Most of data collected by IoT device need to be processed and aggregated before sending to remote backend service [80], since amount of raw data is huge that usually

www.manaraa.com

it is infeasible to send all back, and most of time it's not necessary either. Backend service can usually make good enough evaluation based on aggregated data. However, one challenge is that not every IoT device has enough power to aggregate and process all data, especially when distributed processing among heterogeneous devices is involved. The second challenge is latency. The pure data transmission latency, usually measured by Round Trip Time (RTT) and bandwidth between IoT device and remote server, is more perceptive and critical than traditional web applications, especially when control is involved in feedback loop. A completely remote controlled IoT device can hardly be managed in real time just because of the communication delay and potential network jitter. The third challenge is privacy. Collecting and uploading large amount of data inherently do not play well with privacy [83]. Additionally, although a lot of researches focused on IoT security [84–86], privacy protection receives in-proportional less attention [87]. IoT system user cannot fully rely on IoT system vendors to protect their privacy, because data sent from their IoT devices are stored in vendor's remote data storage which may subject to leak and abuse. Also, some research concluded that privacy invasion is more severe than researches anticipated [88]. Network location privacy, among one of many privacy concerns, needs further protection. Because no matter how obscure and anonymous the uploaded data is, the network location and traffic pattern will expose enough attacking vectors for adversaries.

2.7 Distributed SDN Control Plane

Software Defined Network (SDN) has been prevailing since last decade and gradually taking over of managing network programmatically and automatically, from small office network to cloud data center network. It employs standard and cheaper networking hardware and configures them globally and dynamically according to needs and usage, to achieve an agile and resilient network configuration while simplifies network management. Centralized control (either logically or physically) [89, 90] is the

www.manaraa.com

most adopted SDN control plane paradigm, not only because it is logically simpler but also because generally optimal configuration requires global view, state maintenance, and centralized control. The scalability and single-point-of-failure issues inherited from centralized model can be mitigated by implementing centralized logic control plane as distributed system [91, 92]. Particularly cloud service providers have complete control from edge to core inside their cloud data centers so that they can monitor and control every participant, including router/switch, gate ways, services, etc. For reliability and load distribution the web service interface and back end logic are distributed and replicated on multiple hosts. On the other hand, due to the real time characteristic, sometimes sub-optimal solution is acceptable or even preferred as long as it can satisfy the performance requirements.

For example, Hong et al. [90] developed a new approach to update WAN scale network configuration without causing temporary congestion or fluctuation between data centers. Their approach is to control switches by globally coordinating service sending rates and globally allocating paths. They developed an algorithm that is less computational complex for better scalability for computing large scale forwarding rule configuration, by computing sub-optimal rather than optimal solution. The overall controller is a logically centralized process and evaluate distributed collected inputs for every 5 minutes.

One notable idea of Hong et al.'s approach is to reserve "scratch" capacity (e.g. 10%) of each link so that when flipping configuration overflowed traffic can use them to avoid congestion. Hong et al. proved a congestion-free update can be produced, and they produced an algorithm can minimize "steps", which are deployment stages, of an update. One benefit enabled by this frequent configuration updates is that forwarding rules can be simplified to smaller size to ease the load further on configuration computation and deployment.

www.manaraa.com

3 THEORY AND DESIGN OF MOBILITY SUPPORT SYSTEM

We want to provide a solution to fill the last gap of mobility support and protecting location privacy, while also want to ensure our solution will not have the feasibility issue of previous solutions. First in Sec 3.1 we will identify how an Internet mobility solution can become successful in real world: that it must be a ubiquitous "add-on" and having only beneficial paying for it. Next in Sec 3.2 we propose an architecture, Mobility Support Service (MSS) that combines mobility support and proxy protection while satisfying the requirements: a system strategically creates a dynamic Proxy network for a mobile node to achieve best balance between privacy, performance, and cost. A centralized SDN controller manages relay servers in multiple cloud data centers, and proxies are dynamically allocated on demand to form a proxy network for each mobile node. All connections between mobile node and peer nodes are through proxies that are close to peer nodes. As result both the real network location and mobility characteristic is hidden completely from peer nodes, and mobile node can enjoy full legacy compatible seamless mobility support for any Internet applications. Additionally, we observed the opportunities of leveraging MEC to improve performance and lower operation cost. We further introduce the metrics we developed for measuring performance and privacy protections, and their use in MSS in Sec 3.3. The we discuss typical scenarios in Sec 3.4 and attack models in Sec 3.5.

3.1 Feasibility of Internet Mobility Related Proposal

Internet now is not a research tool any more, and the running of Internet involves lots of business factors and numerous individual organizations. A solution for Internet would not succeed any more if it overlooks the business feasibility, even if it is perfect in technique, and there is no exception for Internet mobility support. While several

www.manaraa.com

clean-slate Internet mobility support solutions exist [93], we support the idea that current Internet architecture would continue to survive for a decade or more [94] and it is further backed up by business factors discussed in following sections. We also suggest that current IP routing and forwarding infrastructure is capable to support Internet mobility.

3.1.1 Mobility support should be ubiquitous and optional

The nature of Internet mobility suggests that the mobility support should be accessible at any time and in most places when an ordinary Internet connection is available. Internet mobility support solutions relying on the modification on access networks, such as installing new infrastructures or agents, would be merely available at a small portion of Internet subnets and consequently lose ubiquity and continuity. In addition, endpoints of Internet vary from platforms and access techniques. Solutions optimized for specific access techniques would be not general enough. For example, handover mechanisms optimized for cellular network may be not applied to WLAN. Portable devices will not embrace solutions having heavy signaling overhead or requiring uninterrupted connection due to energy constrain. This requirement suggests that Internet mobility support should not bind to specific networks or depends on subnet modification. This requirement also weakens the incremental deployment idea for extending coverage area and inspiring incentive that may work in other fields.

Once Internet mobility support is deployed, the number scale of mobile users could be up to thousands of millions. This is a huge number that challenges the scalability of any proposal, but also implicates that the mobility support will only benefit a fraction of Internet endpoints, since a large portion of Internet endpoints will always be static stations. The cost of implementing, deploying, and running Internet mobility support should be paid only by the mobility users rather than every Internet user. Solutions that require change on Internet basic infrastructures such

www.manaraa.com

as DNS or subnet routers will practically bind mobile users and static users together and indistinguishably "charge" them the cost.

Implication of this characteristic shows Internet mobility support should be provided as an "add-on service" rather than a built-in/default feature of Internet, at least for current Internet architecture.

3.1.2 Business factor of Internet Mobility Support

Internet is a network owned and managed by various organizations all over the world with different intention and targeting customers, which means no one is capable to regulate all subnets of the Internet nor independently provide Internet access. This is a radical difference between Internet and cellular networks or like. One cellular service provider usually possesses of tremendous proprietary networks with full control and make internal change all by their will. In order to deliver calls and data cellular carries signs one-to-one peering agreements with a few other equivalent providers, or even provide services entirely by themselves for "in-network" communication. On the contrary ISPs need to collaborate to guarantee the connectivity of Internet by constructing a shared IP routing and forwarding infrastructure to connect every subnet. Generally, ISPs only sign peering agreements with adjacent or higher/lower level ISP/NSPs so that privilege can be implicitly inherited hierarchically. One ordinary IP packet delivered may go through several ISPs among which do not know or talk to each other.

On the other side, collaboration between ISP/NSPs is limited to the scope of basic IP routing and forwarding, which is the "thin-waist" of protocol stack and the only common functionality provided by all Internet infrastructures. First, consensus more than that can hardly be achieved among ISP/NSPs. For example, policy and classes of DiffServ are not honored the same way by different ISPs. Even not each bit of IP options is respected and treated the same way by all ISP/NSPs. Second, cost of collaborated management is exorbitant. To sign detailed peering contract

www.manaraa.com

with all potential ISPs is an infeasible job for any ISP. Techniques of implementing collaboration management at Internet scale are also proved not scalable for current Internet architecture, such as IntServ. In addition, collaborate accounting, which distributes the exact income for each participant, is always difficult to achieve and already expedited failure of several proposals including IntServ, DiffServ, and MIP.

Implementing Internet mobility support associates with cost. Experience shown that it was almost impossible to push all ISPs to upgrade their networks to accommodate a new service, besides above reasons. Business companies need enough incentive to deploy and operate new services. For Internet mobility support, customers who are willing to pay it need to know from whom they can purchase the support, and organizations willing to provide the support need to be able to collect the payment from mobility users to make profit. All these require clear definitions of roles and relations, to build feasible business models. When the profit is enough, even some not so scalable solutions could be feasible in business. While QoS failed in Internet but succeeded in cellular network, one of the reasons, besides the data pattern and network ownership, is cellular carriers can collect enough payment and users are willing to pay for it.

The ownership, cooperation, and incentive pattern suggest that Internet mobility support should be an independent service separated from network infrastructure, and is accessible from any place of Internet based on the most basic functions of IP routing and forwarding infrastructure, to avoid explicit collaboration as much as possible.

3.1.3 End-to-end host mobility

The end-to-end(E2E) mobility issue has be intensively researched [4, 8, 9, 13, 14, 27, 38, 42]. , which to some extent could be considered as a solved issue. The idea is similar: update addresses at peers after each change of address so existing connections could be reestablished or migrated. This idea can be implemented in different layers

www.manaraa.com

and in different format at endpoints, though it also partially adopted by network-based mobility support solution [33, 35].

E2E mobility solutions have relative lower deployment cost since most non-E2E solutions also need to modify endpoints, and they usually have less impact on rest static endpoints. We believe it is a necessary feature for all Internet mobility support solutions, and E2E handover can be performed totally at endpoints: lossless handover could be achieved by soft handover through protocols support multihoming [30] and state resume for hard handover.

However, E2E mobility support solutions solely cannot implement all functionalities of Internet mobility support. Besides potential conflicts with legacy applications and OS protocol stacks which can be solved by adding middle layer or agents, external help must exist for initial connection and dual movement that peers at both ends change addresses simultaneously. In addition, disconnections are always inevitable, varying from short period where physical channel is interfered or shielded, to long period where mobile devices are out of network coverage or during active sleep. When the handover is between different ISPs or a vertical handover such as between cellular networks and WLAN, a "jump" in network topology, i.e. a long topology distance handover, could happen. This "jump" might not be a rare instance for Internet mobility and it may impair solutions specifically optimized for adjacent subnet roaming. The various types of physical network interfaces and indoor/outdoor environment could limit the efficiency of solutions depending on handover prediction. E2E mobility solutions cannot achieve complete transparency at endpoints. Help from access network could improve performance and user experience, but this kind of help can only be optional, to avoid the dependence on specific access networks.

3.1.4 Security and privacy

Internet mobility brings more challenges to security and privacy than ordinary static Internet use for that mobility brings more vulnerabilities in these two aspects.

www.manaraa.com

Solutions that do not take them into account are not complete and have less chance to be adopted. Besides end-to-end encryption to ensure message confidentiality and integrity, Internet mobility support solution must authenticate communication peers and provide forward secrecy since the same address would be shared by different users at different time.

Similarly, privacy protection, especially for location privacy, is another required functionality Internet mobility support solution must provide. While several point solutions have been proposed [95–98], the mobility support system must integrate network location privacy protection from beginning of design. The system should also give its user capability to specify and control the trade-off for location privacy protection, as under different scenarios users can have different privacy choices, or even turning off network location privacy protection when user already sharing GPS location with their communication peers. Therefore, one essential capability is to have fine grain, per connection location privacy protection, instead of globally on/off choice.

3.2 Mobility Support Service Architecture

We have proposed the original Mobility Support Service(MSS) in [99, 100] and improvements in [101–104]. MSS is a system that provides mobility support and privacy protection to MSS customers/subscribers, while satisfying the economical and feasibility requirements we proposed in Sec 3.1.

MSS is provided as a distributed service over the Internet that it can serve customers virtually anywhere. It does not require any change on access networks, existing network infrastructure, legacy applications, and operating systems. Instead MSS is provided as a value-added service to customers who are willing to pay for enhanced mobility and privacy protection, based on single one-to-one contract between customer and service provider. Therefore, providers generate its own revenue and justify the business and investments.

www.manaraa.com

While keeping global service coverage, service provider can choose scaling between minimum presence to almost everywhere, and setup system on various underlay architectures, such on premise servers, on pubic cloud servers, on Mobile Edge Computing, or being hybrid involving multiple source of servers. The different choice of underlay architecture and server allocation will result in different regional performance and operation cost tradeoff, up to the favor of service provider. Single service provider can provide service to any customer, and at the same time customer could use multiple different provider if needs.

In this section we will describe MSS basic concepts, design principles, and theoretical evaluation. First, we give system overview in Sec 3.2.1 and then walk through details of major components. At the end of section we will illustrate how MSP can freely scale their footprint to suite their business and operation needs.

3.2.1 Parties and Entities

There are three different parties in system: Mobile Node, Peer Node, and Mobility Service Provider(MSP). Mobile Node (MN) is device the MSS subscriber used to access Internet and connect to peers. Peer Node (PN) is the peer side Internet end-host on the other end of connection, that it either receives connection from Mobile Node, or initiates connection to Mobile Node. Peer Nodes can be web server, ordinary Internet host, or another Mobile Node. MSP is the MSS service provider providing mobility support and location privacy protection service to Mobile Node. To receive MSS's privacy protection and mobility enhancement, Mobile Node user only needs to sign a single contract with an MSP and this MSP becomes the only one knowing Mobile Node's network whereabouts. MSPs are independent from infrastructure provider (such as ISP), and any MSP can have global presence, though their capability of where and how much they can allocate server will determine their services' performance. A MN could also use multiple MSPs at the same time for connecting different Peer Nodes, if desired.

www.manaraa.com

The key component to enhance mobility support and protect network location privacy is Proxy, which relay traffic between Mobile Node and Peer Node. The data link between Mobile Node and Peer Node are identity-based mobility friendly connection such as HIP [39], and data link between Proxy and Peer Node are ordinary TCP or UDP connection. Applications on Mobile Node still connect to Peer Node through the original protocol such as HTTPS and they are not aware of their traffic actually being tunneled by MSS agent on MN and relayed at Proxy then reaching Peer Node. Therefore, Mobile Node can move and change Internet attach point freely and existing connections applications depending on will not be interrupted. Network location privacy is also protected since the exposed network address is Proxy, not the actual attach point MN has.

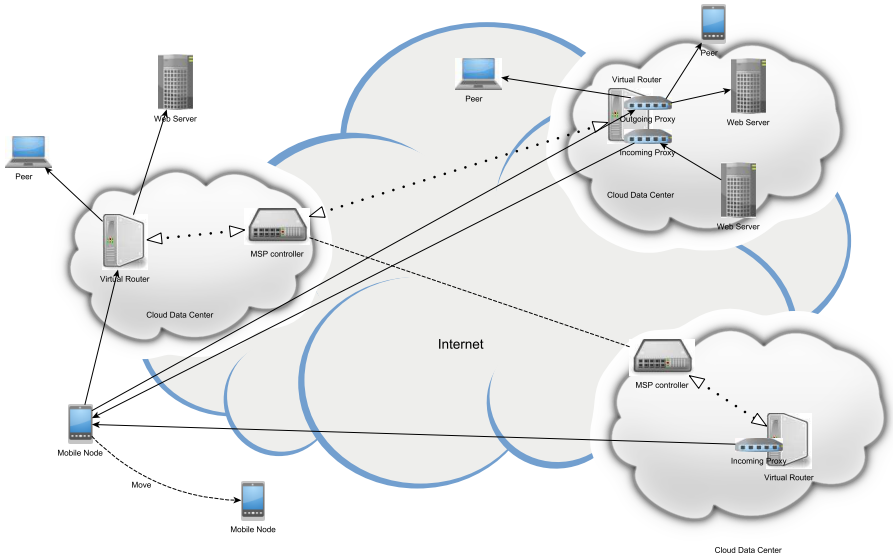

Figure 3.1.: MSS network architecture

To avoid the performance penalty of this VPN-like traffic relay, MSS employs a novelty paradigm: aggressively pushing Proxy close to Peer Node, and have multiple concurrent Proxies serving different Peer Nodes. When Proxy is very near to Peer Node no matter where Mobile node moves to, the route via Proxy is almost identical to optimal route. Having concurrent Proxies will make sure all routes Mobile Node has

www.manaraa.com

been near optimal so that it doesn't need to scarifies some Peer Nodes' performance for some other Peer Nodes, as shown in Fig 3.1.

This design doesn't require any change on the Peer Node side or Peer Node's network so it avoid the lacking incentive dilemma. It can also support legacy applications without modification further reduce the gap of adopting this new service. Additionally, it doesn't have performance penalty or difficulty of choosing relay location that traditional VPN has.

3.2.2 Mobile Node

Mobile Node is the mobile device hosting mobile user's identity and applications. It roams across different network and continuously communicates with its peers. During its movement, Mobile Node keeps changing its network attach point and exposes different public network address (such as public IP address) at times.

Mobile Node has MSS agent deployed, which handles incoming and outgoing traffic globally or per user specification. Connections between Mobile Node and Proxies are identity based, and all traffic are tunneled through these connections, as shown in Fig 3.2.

The reason MSS agent is designed this way is because it requires minimum modifications on operating system and legacy application thus has least endpoint deployment cost. Compared to the layer insertion between IP and transport [9, 10] this method would be more complex but provide more flexibility. For example, a multi-homing capable transport protocol can thus be used to implement lossless soft handover. End-to-end authentication and symmetric session key generation between Mobile Node and Proxy are performed for each address change or connection timeout. MSS also have capability to seamlessly upgrade service or deploy new services, because all these changes would merely need to upgrade the deployed MSS client.

When a Mobile Node wants to connect to a Peer Node, MSS agent on host will request a Proxy from MSP control plane. This Proxy will be allocated at a network

www.manaraa.com

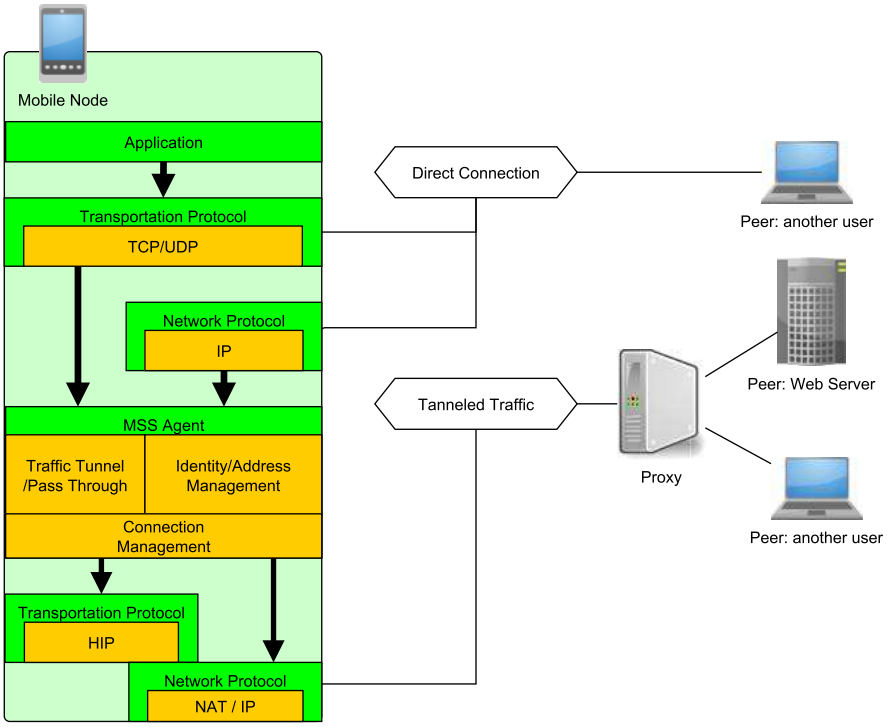

Figure 3.2.: Mobile Node

location as close as possible to Peer Node, and it connects directly to Peer Node. The traffic between Mobile Node and Peer Node is tunneled though Proxy using identity-based connection [105] between Mobile Node and Proxy. Proxy will be the exposed network address and will not change no matter where Mobile Node moves to. Therefore, Mobile Node's real network address and movement are completely hidden from Peer Node. When Peer Node is an ordinary Internet host and connection is bound on IP address, MSS provides additional mobility support that traditional network protocols can work transparently without interruption even if Mobile node is temporarily offline or changes network addresses. All these benefits are achieved without much network latency overhead because Mobile Node's movement will never deviate its route from optimal one much, including scenarios Mobile Node performing vertical handover which could change network location dramatically.

www.manaraa.com

3.2.3 Proxy

Proxy is a process that relays Mobile Node's traffic, and it runs on an MSS managed server, called "Virtual Router".

There are two types of proxies: outgoing Proxy and incoming Proxy. Outgoing Proxy is short lived and created on-demand when Mobile Node wants to create new connection to Peer Node. It is initialized by Mobile Node sending request to MSP controller. MSP controller will then create new Proxy that has minimum communicate cost to Peer Node, and the selection is also limited by current resource availability and customer's SLA. The Virtual Router, which is assigned to host this Proxy, then just chooses a random outgoing port and uses its own address to create connections to Peer Node. Then it instructs the chosen Proxy to update/initialize and take over the connection while waiting form tunnel opening request from Mobile Node. Example sequence is illustrated in Fig 3.3.

Incoming connection means Proxy must listen on a given port for incoming connection requests. Therefore, incoming Proxy are exclusive, since one specific listening port can be exposed for only one Mobile Node on a Virtual Router. Listing Proxy must be created in prior and are dynamically adjusted according to recent address queries, amortized management algorithm, and also historically statistic. Due to the resource scarceï¼Œlisting Proxy is more expensive than outgoing Proxy, and popular port (such as 80 or 443) are more expensive than non-popular ones. MSP will advertise those listening Proxy through regional/geographic DNS record so that Peer Node will resolve Mobile Node's DNS name to a nearby incoming Proxy. Example sequence is illustrated in Fig 3.4.

Since at any moment a Mobile Node can be behind a few Proxies, by nature MSS Mobile Node is considered multi-homing. Its Peer Nodes generally will only see one exposed network location of the Mobile Node, but it's also possible a Peer Node connects a Mobile Node through two or more different Proxies, especially when these connections are setup long time apart.

www.manaraa.com

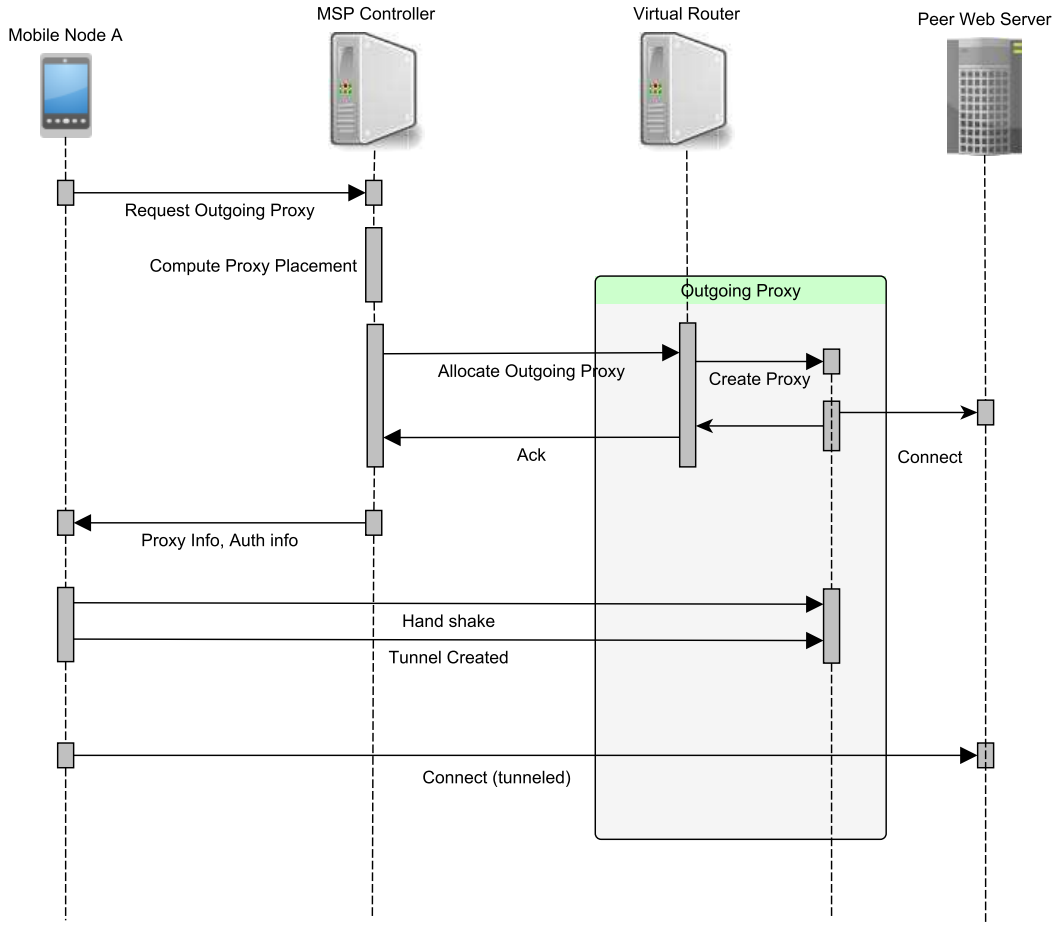

Figure 3.3.: Mobile Node Request Outgoing Proxy

A special scenario is both end-hosts are Mobile Nodes behind Proxies. Additionally, they may belong to different MSPs which additionally limits the data to optimize performance. When two Mobile Nodes belonging to same MSP, since MSP controller knows locations of both MSP will choose one "pivot" point between them to optimize for performance. If one Mobile Node knows the other end is also a Mobile Node, it may leverage that to detect how far away the other Mobile Node is away from it. To mitigate that MSP controller must set a lower bound of route path length, to avoid choosing a pivot point too close to a Mobile Node. When Mobile Nodes belong to different MSPs, both only exposed Proxy to the other side, and the traffic will go

www.manaraa.com

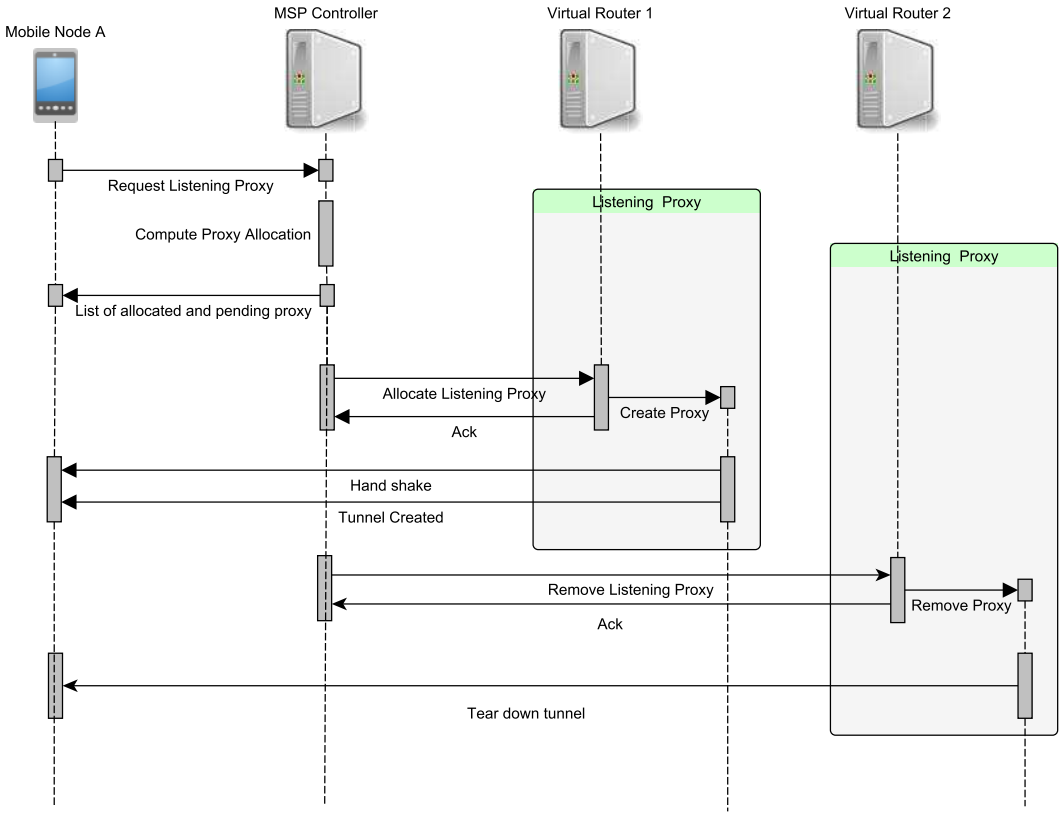

Figure 3.4.: Mobile Node Request Listening Proxy

through two Proxies. It might result in an inefficient trombone routing, unless MSPs can cooperate to share some location knowledge.

3.2.4 Mobility Service Provider (MSP)

MSP manages a fleet of servers, called "Virtual Routers", that are dynamically allocated and released from public cloud service provider's data centers. Each Virtual Router can host multiple Proxies that relay different Mobile Nodes' traffic, up to server's resource limit. Proxy also functions as profile server that serves Mobile Node's profile or DNS name lookups. For each Mobile Node, its Proxies together create an overlay network to propagate control signals. At any moment Mobile Node has a

www.manaraa.com

master Proxy which is created or designated at a location close to Mobile Node. This Proxy may only be used for relaying traffic with lower privacy setting or close by Peer Node, but its major tasks are to manage the Proxy Overlay Network and delegate communication between Mobile Node and SDN Controller. When Mobile Node moves away from master Proxy, a new master Proxy will be created to take over the task.

A mobile user signs a single agreement with an MSP for leveraging the service to protect his/her privacy. Powered by cloud even a single MSP could offer reasonable location and performance coverage for most of places. On the other hand, different MSPs will have their own strategy, strength, and goals. For mobile users want to maximize privacy or cost/performance, they could choose to sign up for multiple MSPs and use accordingly for different connections.

MSP manages its Virtual Router fleet similarly at larger scope with same strategy. It removes unoccupied Virtual Router or creates new ones to maintain a healthy load ratio and global presence. Public cloud enables this architecture that Virtual Routers can be created/removed in almost all major areas around the world, in the manner of on-demand that Virtual Router can be allocated or removed within minutes dynamically.

Virtual Router

Virtual Routers are the real servers that host Proxies. It is called "virtual" because it is not physical server or routers, but instead applications/processes running on virtual host, such as cloud host. That grants the capability of dynamically allocating and removing capacity when needed. On the other hand, provisioning Virtual Router is still time consuming (e.g. in terms of minutes rather than seconds), so adjustment of Virtual Router allocation needs to plan ahead, depending on current overall system load, and prediction from historical statistic. Proxy and profile servers are applications running on Virtual Router. They are launched on demand: MSP

www.manaraa.com

controller create/remove them when needs, since creation and update can complete immediately when requested, as shown in Fig 3.5

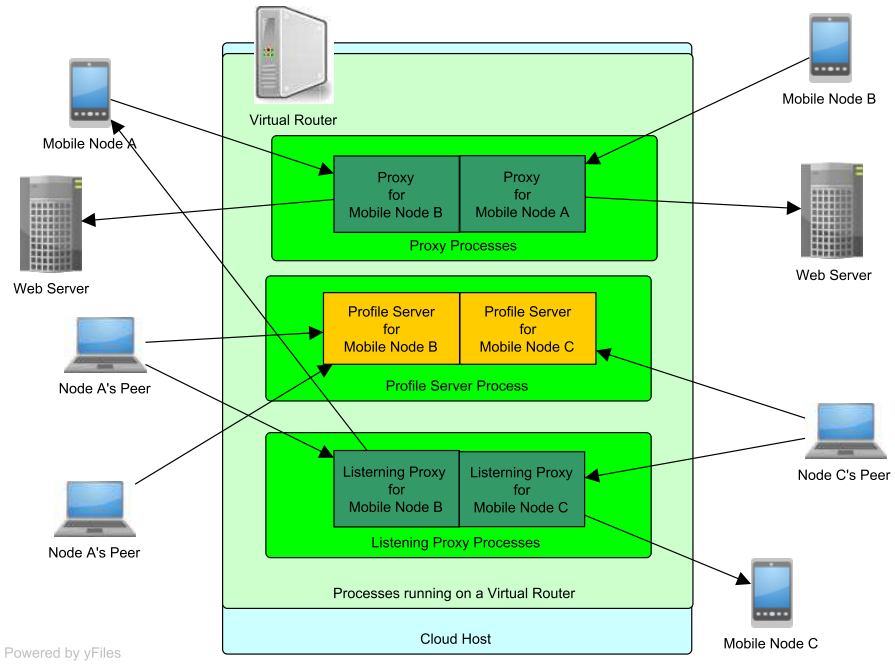

Figure 3.5.: Virtual Router

Multiple Proxies and Profile Servers shares a Virtual Router, as long as they don't have port conflict and the combined computation and bandwidth requirement doesn't exceed Virtual Router's capacity. These Proxy and Profile Server applications run independently so they can belong to the same Mobile Node or different Mobile Nodes.

MSP Endpoint Server are front end of MSP controller, and rendezvous point of bootstrapping Mobile Host and Peer Host.

MSP has a logically centralized SDN-like controller which controls the allocation of Virtual Routers, Proxies, Profile Servers, and network topology for each Mobile Node's Profile Server overlay network.

Controller maintains a distribution graph of all Proxies and Profile Servers of every Mobile Node, and receives periodical report collected by Virtual Router. Therefore, the total message is bound by number of Virtual Routers, although one message from

www.manaraa.com

a Virtual Router may contains multiple records regarding the state of Proxies and Profile Servers running on it.

When Mobile Node wants to setup a new outgoing connection, it can either reuse its current connected Proxies, or can request a new Proxy from SDN controller. Proxy may also reject request for new connection if it's overloaded or pending for removal, and in this case, it will forward the request to SDN controller while telling Mobile Node to wait for further instruction. Controller will examine location of Mobile Node and its Peer Node, and assign a proxy according to availability and Mobile Node's SLA.

MSP controller periodically checks Virtual Routers' load and dynamically launch or remove virtual host to balance load. For each individual Mobile Node, SDN also periodically updates its view of Mobile Node's listening Proxies and Profile Servers, and adjusts them according to recent statistic and pattern learnt from historical data.

3.2.5 MSP Infrastructure

There are two levels of fleet management operated by MSP: Proxy management and Virtual Router management. Each Mobile Node's Proxy fleet forms an overlay control plane to monitor and aggregate metrics, and delegate signaling between Mobile Node and MSP control plane. In addition to on-demand Proxy allocation, MSP also periodically adjusts Proxy distribution to comply to user's Service Level Agreement(SLA). On a higher and broader level, MSP continuously monitors resource utilization of its Virtual Routers, and dynamically allocate or remove capacity to maintain a healthy availability. It employs similar philosophy as public Cloud service provider that even individual customer usage can be volatile, the aggregated usage of many users is relative stable and predictable. Additionally, the system benefit greatly from on-demand resource allocation of public Cloud. These two are the keys to drive down operation cost.

www.manaraa.com

The deployment of MSP server is resilient that it can shrink or extent in terms of MSP's will, and although distributed servers are deployed, MSS is still completely apart from access networks. On one hand, MSP servers provide information access via web services. Thus, an Internet user can always access desired information via an ordinary Internet access. On the other hand, while server distribution and profile replication devote to facilitate mobility management, it can also improve availability of MSP and overall performance. The payment of subscribers for mobility management and extra services would generate income for MSPs.

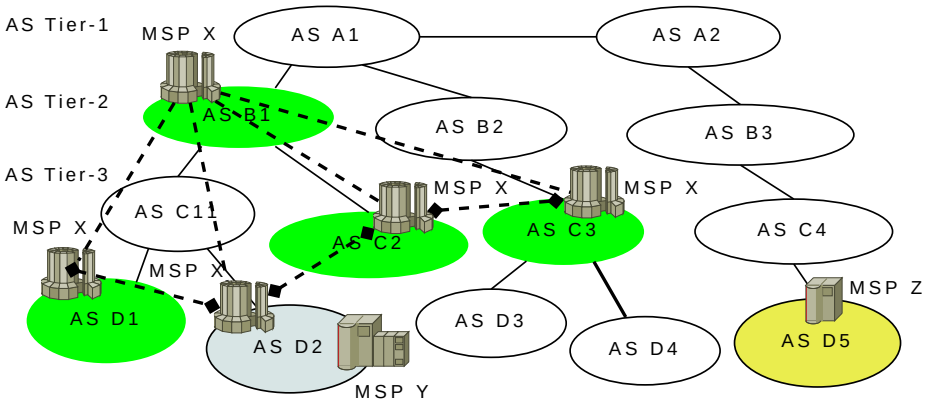

Figure 3.6.: Examples of MSP coexistence and cooperation

For the minimum an MSP may have only one server such as MSP Y in Fig.3.6, where AS D2 is covered both by MSP X and Y. All Y's subscribers and peers can still use its service from all other places of Internet, but only have the minimum overhead when they are in AS D2. MSP Y prefers to serve a small group of people around a certain area, such as a college or an enterprise. Hence MSP Y practically degrades to become a rendezvous server with additional features such as PKI and relaying for a group of users. Please note MSP Y can deploy more servers in AS D2 if it finds it necessary. MSP Z is another MSP having servers deployed in AS D5. Suppose MSP Y and Z belongs to nearby universities and they want to share each other's services for convenience. Hence, they could have a "peering" agreement to serve subscribers of each other in their covered network.

www.manaraa.com

Different MSPs can provide services to the same region simultaneously, and subscribers will have the freedom of selection.

MSP Scaling

For one of the most extreme configurations, MSP can have only server which hosts MSP control plane and also acts as Virtual Router hosting all Proxies as shown in Fig 3.7. It can still serve Mobile Node at anywhere of Internet, although it degrades performance into a static VPN server since all traffic will be proxied at this server. The mobility support and privacy protection function still work without interruption, though the performance apparently won't be good, and privacy protection level will be limited too. On the other hand, this configuration has the lowest operation cost, and everyone has constant Internet connection can run their own MSP.

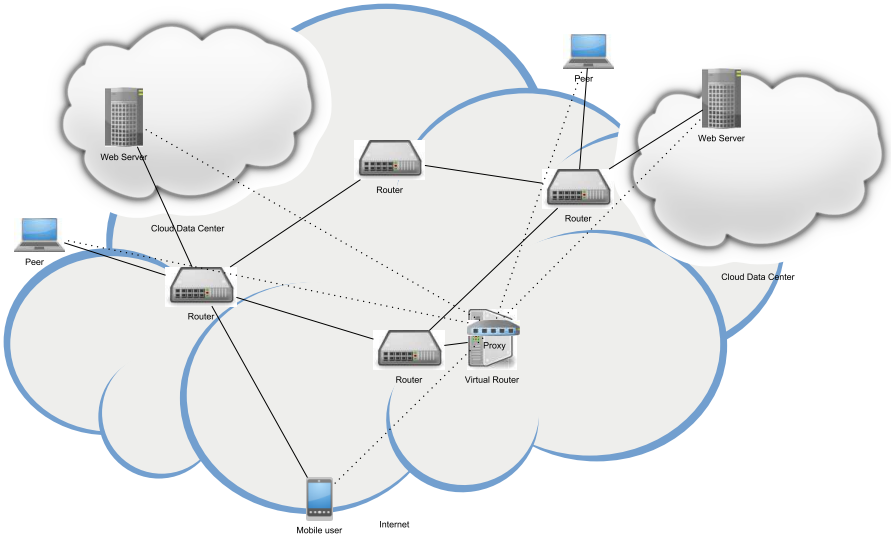

Figure 3.7.: Examples of MSP having only one server at an edge location

Another side of extreme config is MSP has ubiquitous servers everywhere, covering every AS even subnet as shown in Fig 3.8. This is not realistic, but it will have best performance since there won't be any performance penalty and every network router are mobility aware.

www.manaraa.com

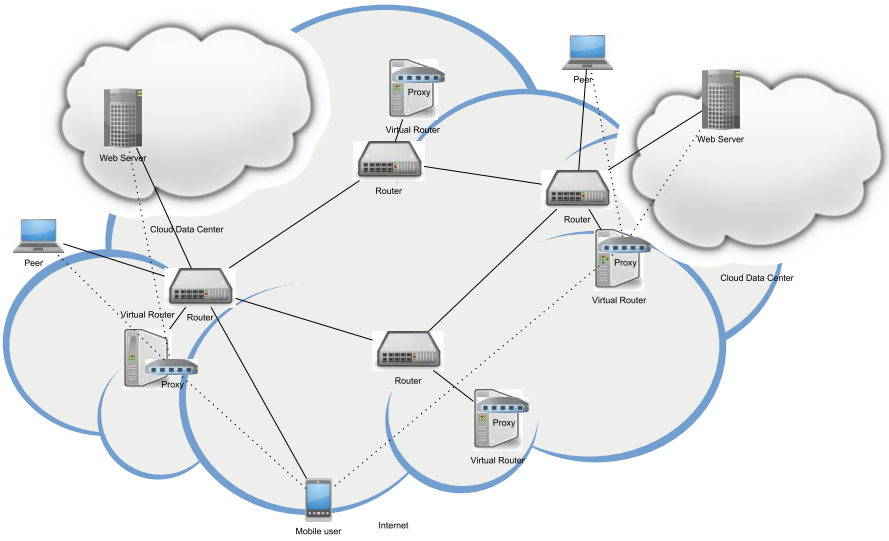

Figure 3.8.: Examples of MSP having servers at every subnet

Cloud computing services already are being used to enhance various Internet functionalities. For example, Amazon Route 53, built on Amazon Cloud Computing infrastructure, implements a DNS service. In this paper we extend the design of MSS by using cloud computing platforms. Furthermore, we show that cloud-based architectures offer interesting tradeoffs among performance, security, privacy and economic viability. Cloud computing not only can improve the technical performance of our MSS, but most importantly can make much more economically attractive the MSS to service providers. In particular cloud computing, with its elastic nature, lowers the initial investment cost to start the business of MSS, and then keeps the cost scalable to the service demand. Therefore, the use of cloud computing in MSS will help innovation, by keeping lower its initial and ongoing costs, as shown in Fig 3.9

3.3 Metric Definition

We propose three major metrics to quantify mobile node's network location privacy, performance overhead, and corresponding operation cost. They are main targets

www.manaraa.com

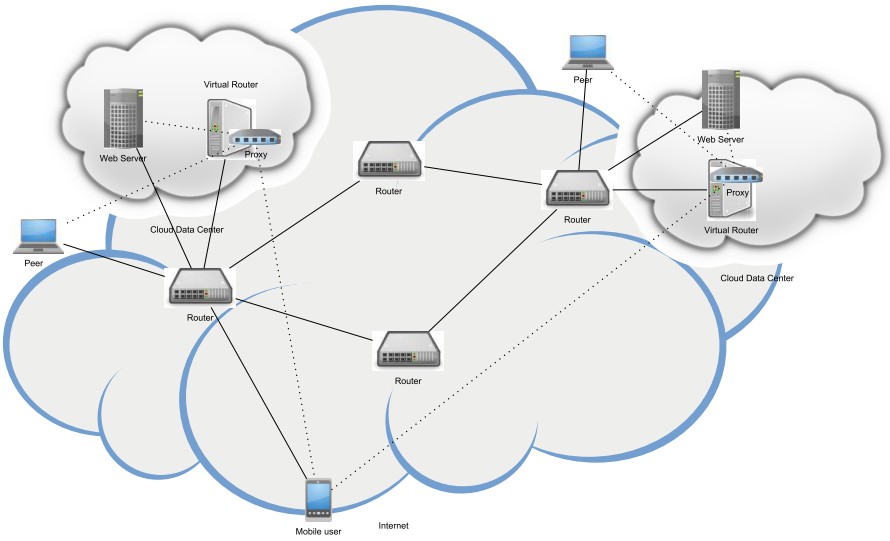

Figure 3.9.: Examples of MSP having servers only in public cloud data centers

MSS system optimize for. To simplify multiple target optimization, we condensed their definition and further added an aggregated metric definition, **score**.

In following discussion, we use m, n, and p to represent Mobile Node, Peer Node, and Proxy Server; use M, N, and P represent corresponding set. Particularly p_l and p_o represents Listening Proxy and Outgoing Proxy.

3.3.1 Location Privacy

To protect location privacy from communication peers, traditional LBS (Location Based Service) privacy metrics such as k-anonymity [21] doesn't apply as the mobile node is the single endpoint of connections to each peer, i.e. there is no "hiding" from other identities. Fundamentally location privacy in communication can be quantified by coherence of the Mobile Node's actual location and location observed from Peer Node side. To quantify privacy protection for network location in an end-to-end communication scenario, we propose two metrics to measuring privacy: **distance** and **timing**. In the case when Peer Node believes the exposed location is real location, the uncertainty is 100%. However, since we cannot quantify how much Peer Node believe

www.manaraa.com

Mobile Node's exposed location, we will always assume Peer Node knows Mobile Node is behind a Proxy, and the proxy is on a strategy point that won't introduce unreasonable latency penalty, i.e. at some point along the route between Mobile node and Peer Node. (Our metrics can also quantify privacy and performance of artificially away Proxy as well, i.e. Proxy is selected far away from Mobile Node and Peer Node, to create an illusion of being away for Peer Node, in cost of performance penalty.)

Distance

Distance, λ, is defined as how different the exposed network address is in term of relative geographical distance, which is derived from the mapping of network address to registered geographical locations. MSS only manages network communication so location is determined by network address, which can be mapped to approximate geographical locations. Other location information such as GPS coordinates are not directly exposed by communication channel nor needed for general communication, so they are out of scope of this research.

Distance stands for two different types of measurement in MSS: network distance and geographical distance. Network distance can be measured by network hops of end-to-end connection, or hops of network segments such as Autonomous System (AS). Geographical distance is measured by the distance of corresponding geographic locations of exposed network address and actual network address. This metric bears similarity to distance error described in [106, 107].

For example, a Mobile Node in New York City with IP address 128.59.a.b talks with a Peer Node in Los Angeles with IP address 128.97.x.y, via a Proxy Server in Indiana with IP address 129.79.m.n. The network distance is the hop distance between 128.59.a.b and 129.79.m.n, and the geographic distance is about 700 miles between New York City and Indianapolis.

Network distance is harder to quantify and compare. Network address, such as IP address, are usually not uniformly distributed. It is already a hard job to estimate

www.manaraa.com

hop distances given two arbitrary IP addresses. Also, the IP address or AS exposed may tell which organization owns the address, but this information generally not reveal much personal information.

Geographic distance mapped from IP address is easier to quantify and relative reliable. Even though there are cases that IP address incorrectly mapped to wrong geographical locations (mostly depending on IP database), in reality this doesn't impair privacy. In our research we will assume all IP addresses can be correctly mapped so our evaluation can rely on geographical distance for comparison.

In general, the larger the distance the better privacy. We use function $dis(x, y)$ to represent the approximate geographical location between network attach point x and y. Then for a given combination of Mobile Node (m), Peer Node (n), and Proxy Server (p), location privacy $\lambda m, n, p$ can be evaluated as:

$$\lambda m, n, p = \frac{dis(m, p)}{dis(m, n)}$$

as shown in Figure 3.10. On the other hand, the performance overhead ϕ is quantified as:

$$\phi m, n, p = distance(m, p) + distance(p, n) - dis(m, n)$$

λ can equal to 0, between 0 and 1, equal to 1, or greater than 1. We use Figure 3.11 as reference to explain that:

- when no Proxy is leveraged, which means Mobile Node's location is directly exposed, then $dis(m, p) = 0$ and $\lambda = 0$.
- when Proxy is located somewhere between Mobile Node and Peer Node, $0 < \lambda < 1$, such as Scenario A and B in Figure 3.11. When $dis(m, p)$ of Scenario A and B are same, they have same distance privacy metric λ. On the other hand, overhead ϕ of Scenario A is lower.
- when Proxy is located next to Peer Node, $\lambda = 1$. In this case overhead ϕ is minimized to 0, and λ is 1.

www.manaraa.com

- when Proxy is located far away from between Mobile Node and Peer Node, $\lambda > 1$, such as Scenario C in Figure 3.11. This is a valid, but not very reasonable case that distance privacy is high and the overhead is also much higher.

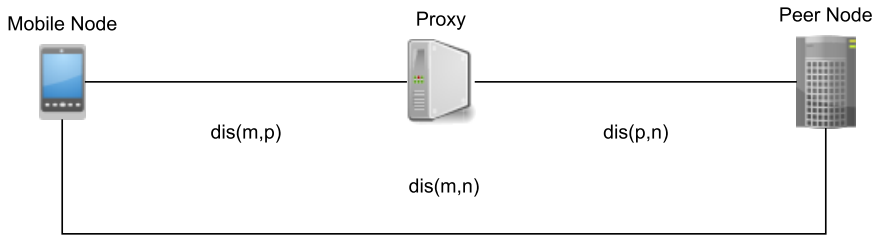

Figure 3.10.: Distance Legend

The relation between distance λ and overhead ϕ is further illustrated Figure 3.12. Assuming all Proxies Mobile Node connects has same distance $dis(m,p)$, which forms a circle around Mobile Node. $dis(m,p) = dis(p,n)$ means Peer Node happens to be on the circle as well. According to our equation all Proxies have same λ. On the other hand, the Proxy which locates at the same location of Peer Node has minimum ϕ, 0.

Distance metric can be evaluated for each connection/reconnection between Proxy and Peer Node. Then for Mobile Node (m), its overall location privacy at a specific time can be quantified as:

$$\lambda_m = min_{m \in M, p \in P}(\frac{distance(m,p)}{distance(m,n)})$$

Timing

Timing, δ , measures how correlated inferring mobile host's movement (i.e. changing network attach point) versus real movements mobile host has. When no protection mechanism is applied, adversary can know exactly when mobile host moves from one location to another. Timing is measured by two types of correlations: number of real

www.manaraa.com

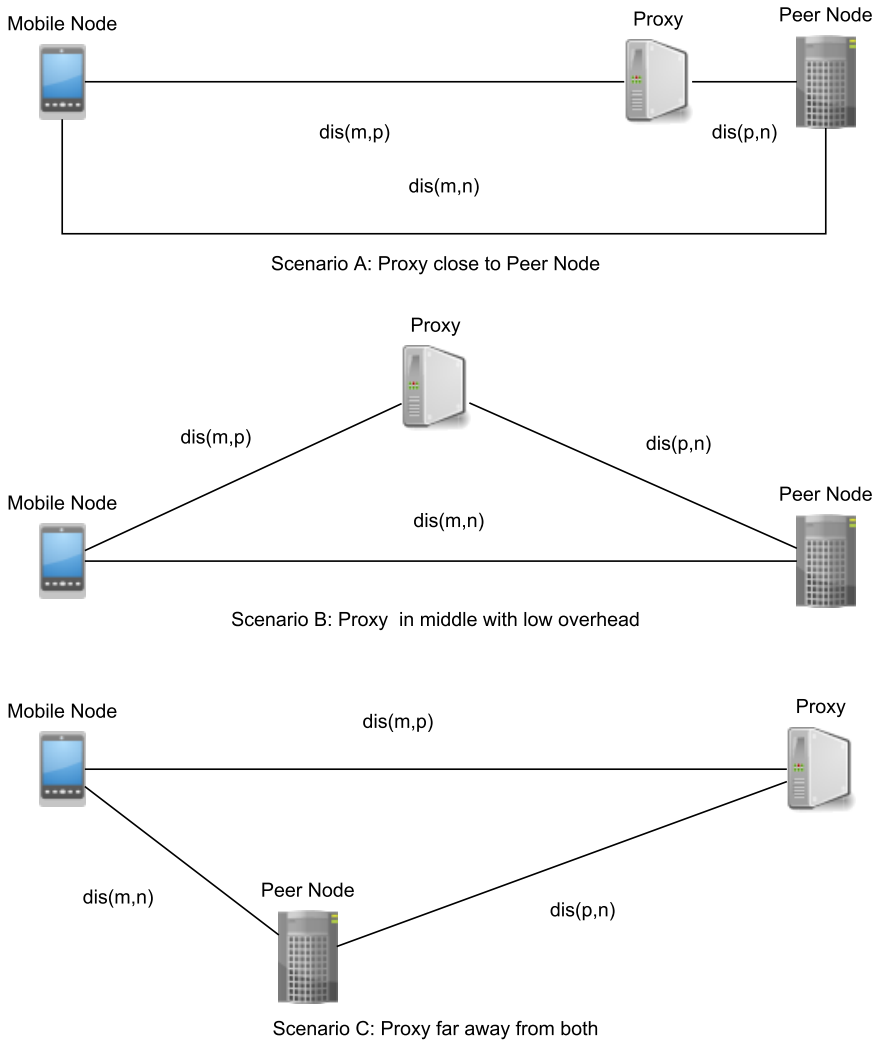

Figure 3.11.: Distance Scenarios

network address changes versus exposed network address changes during whole communication, and difference of timestamp between real address changes and exposed address changes. The larger the correlation, the better privacy. Timing privacy must be evaluated for a period of time: during a time range, we assume Mobile Node moves i times and Proxy changes j times. Function $\delta(x, y)$ is used to represent the timestamp difference between event x and y. Then for a Mobile Node m went through a series of network address change events E^m while its Peer Node n observed another

www.manaraa.com

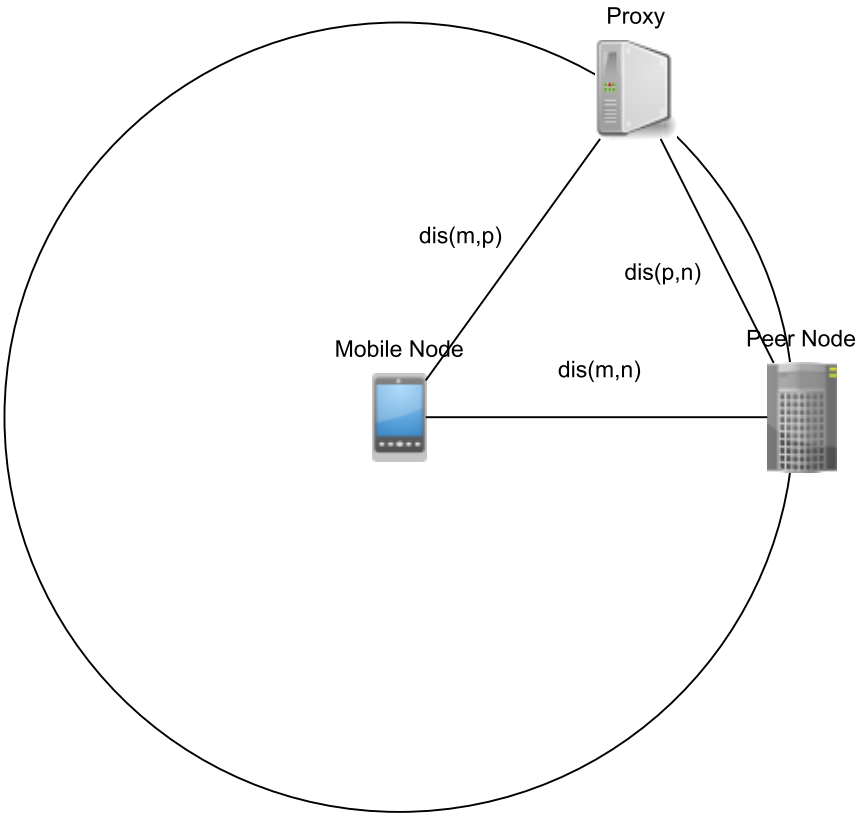

Figure 3.12.: Distance Radius

series of change of Proxy Server events E^p. Then for a matched Mobile Node move E_i^m and Proxy Server change E_j^p, the timing metric is evaluated as the time difference of these two events:

$$\delta_{E_i^m, E_j^p} = \left| t_i^m - t_j^p \right|$$

where t is event's timestamp.

When evaluating timing privacy, we always use best match of Mobile Node move events versus Proxy Server change events, to assume Peer Node has best knowledge to leverage that correlation. In another word, we assume the worst case for Mobile Node that each of its Proxy server change will be associated with its most recent

www.manaraa.com

move. Given that assumption and our δ equation, the overall timing privacy can be evaluated as the sum of :

$$\Delta_{m,n,p} = \frac{\sum_{i \in E^m, j \in E^p} |time(x,y)|}{j}$$

The higher value of $\Delta_{m,n,p}$, the better timing privacy. The lower bound of $\Delta_{m,n,p}$ is 0 that for each Mobile Node move, Peer Node can detect its Proxy Server change at exactly the same time, i.e. equivalent to no Proxy server. The upper bound is ∞ that when Proxy Server does not change at all, the exposed address becomes completely static.

3.3.2 Cost

Cost, χ, is total operation cost of running all proxies and control plane controllers, plus traffic/bandwidth cost if applies.

$$C = \sum cost_{proxy} + cost_{controller} + cost_{traffic}$$

Generally, the controller is relatively fixed fleet, and much smaller compared to proxy fleet in all three setups. When implemented distributed it can actually be hosted on existing Proxies. Therefore, we generally consider it as a constant. Traffic can be either absorbed into Proxy cost in form of flat rate pricing or can be separated charged to Mobile Node. Since the total traffic does not affect Proxy selection as long as bandwidth requirement satisfies, we remove it from MSS Cost metric.

Single proxy cost is determined by its location, cloud service provider and cloud host type, and bandwidth Proxy node provides. To unify different virtual host type with different capacities, we define a Unit Stream Capacity (USC) as 1MB/S. Therefore, we use virtual host's hourly rate, a popular cloud service charging unit, as dividend and number of USC it can support as divisor, to define a Proxy's operation cost:

$$\chi_{proxy} = \frac{hourlyrate}{\frac{ProxyStreamCapacity}{UnitStreamCapacity}}$$

www.manaraa.com

In each setup, MSS controller will use a Cost Database to lookup predicted running cost for each candidate Proxy. Apparently keeping the database data precise and up-to-date is important and MSP and MEC will have advantage here.

As result for a Mobile Node its overall operation cost is sum of all hour charge of Proxy Servers and bandwidth cost, which MSS optimize the former.

$$\chi_{MN} = \chi_{proxy} * hour_{proxy} + \sum \chi_{traffic}$$

3.3.3 Performance

Communication latency and bandwidth are two most perceptible measurement of mobile user's experience. Since generally proxy can always provide enough bandwidth for a Mobile Node, MSS doesn't consider bandwidth as a variance for performance optimization, but rather an SLA requirement for MSS control plane to choose Proxy for Mobile Node. Round Trip Time (RTT) of traffic between Mobile Node and its Peer Node directly and via Proxy are the major performance difference MSS optimizes for. It is defined as proxy overhead, the difference between direct connect RTT and via proxy RTT, that smaller means better performance.

$$\mu = RTT(m,p) + RTT(p,n) - RTT(m,n)$$

3.3.4 Score

Score, S, is one aggregated metric unifying performance and privacy metric, based on 5 inputs: Mobile Node location, Peer Node location, Proxy candidate location, acceptable minimum performance, and acceptable minimum privacy. The value range of Score is between 100 and -100, while any value below or equal 0 means the candidate does not satisfy minimum requirement.

$$S = \frac{max_\mu - \mu}{max_\mu + \mu} * k_\mu + \frac{\lambda m, n, p - min_\lambda}{\lambda m, n, p + min_\lambda} * k_\lambda$$

k is weighing constant that the sum of all k equals to 100%.

www.manaraa.com

Score can be further combined with Cost as:

$$S = \frac{max_\mu - \mu}{max_\mu + \mu} * k_\mu + \frac{\lambda m, n, p - min_\lambda}{\lambda m, n, p + min_\lambda} * k_\lambda - \chi_{proxy} * k_\chi$$

3.4 Connection Scenarios

In this section we will discuss typical scenarios that how a mobile node connects to its peer nodes through MSS for privacy protection and additional mobility support. Fig 3.13 shows a simplified illustration of one cellular provider with two Radio Access Networks (RAN) at different geographic locations connecting Internet through their gateways, while two RANs are physically isolated but also directly connected through private link. It is an abstraction of a typical cellular provider that has many RANs vastly deployed covering almost everywhere, either independently or through collaboration. Usually these RANs are also privately connected so direct traffic between RANs are routed via private links for better security, cost, and performance.

A user carries a cellphone deployed with MSS agent and connects public Internet through cellular provider's data network. Initially it connects to Base Station A1 in RAN A. It communicates with a few Internet peer nodes: two web servers and another peer user. Cellular data network will assign an internally routable IP address to cellphone for routing traffic within cellular data network. For traffic going out, without using MSS or VPN this traffic will exit from Internet gateway selected by RAN A controllers, generally the one close to its current Base Station such as Gateway G1 in figure. The public IP address of this gateway will be the exposed network location of the cellphone, which can be approximately mapped to its geo-location. When a pre-selected VPN server is utilized to proxy all traffic, network packets will still exit from Gateway G1 but then redirect at VPN server. The exposed IP address will be the one from VPN server, which does not correlate with cellphone's real location. The overall latency overhead of proxying is determined by the distance between gateway and VPN server and the distances between VPN server and peer nodes. When MSS with MEC is utilized, multiple proxy points are selected that all

www.manaraa.com

of them are allocated closing to traffic target. For example: Gateway G1 will be selected for Web Server X, Gateway G3 will be selected for Peer User, and a Virtual Router within same public Cloud data center of Web Server Y will be selected for it. The exposed public IP address are different for each peer, and not correlates with cellphone's real location either. Although Gateway G1 is selected, it is just because it is close to Web Server X, regardless of cellphone's location (it will be further illustrated in following scenarios when user moves). The overall latency overhead is minimum to none since all proxies locate on or near optimal routes. Connections between proxies and peer nodes are ordinary TCP/IP connections that bound to IP addresses, while the connections between cellphone (via MSS agent) and proxies are identity based and resilient to soft or hard handover [105]. Both VPN and MSS will incur some extra operation cost.

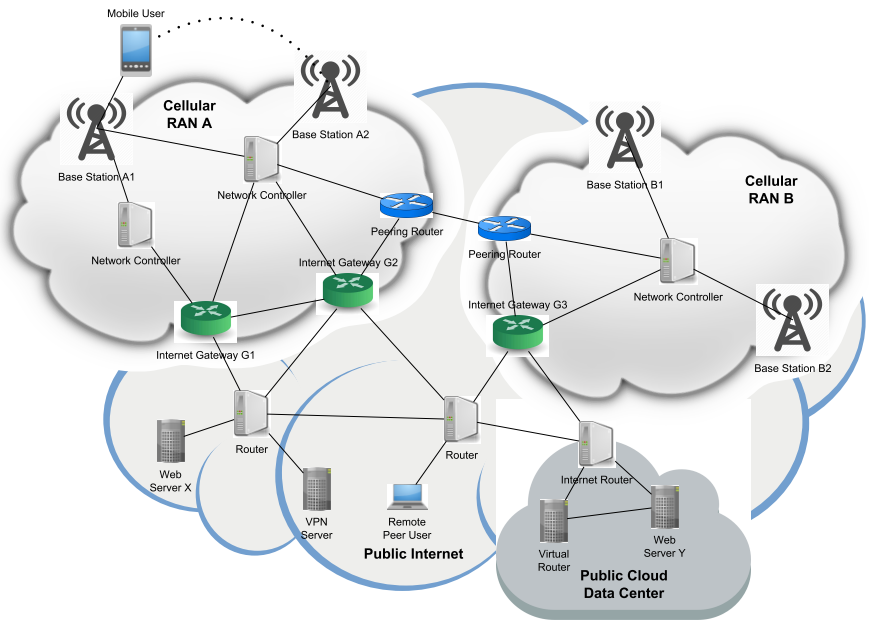

Figure 3.13.: Mobile node roams

When user moves a distance the cellphone handover connection to Base Station A2, which still belongs to same RAN A. It may receive a new internal IP address or

www.manaraa.com

changing Internet gateway, though cellular data network may try to minimize impact to existing connections by keep current Internet gateway unchanged, with protocols such as Proxy Mobile IP [108]. Therefore, the original exposed IP address will remain while latency and operation cost will increase a bit. For VPN nothing changes, except it also suffers the increased network latency from cellular network. If cellular network decides to change Internet gateway, VPN still protects cellphone's real network address, but ongoing connections may be interrupted, and the latency change depends on whether new gateway is closer or further from VPN server. MSS with MEC will instruct cellular provider always assign most efficient internal IP address and Internet Gateway, without need of PMIP. Proxies in MEC will handover ongoing connections to the new internal IP address, while proxies in public Cloud will handover connection to a new gateway closet to cellphone. The exposed public IP address remain unchanged and ongoing connections are preserved. Still the overall latency overhead is minimum to none since all proxies are still on or near optimal routes.

When user moves longer distance eventually it will move to RAN B and connects to Base Station B1, and cellular network has to change its Internet gateway to G3. All IP based connections will be disrupted, and without protection remote peer nodes will be able to detect cellphone's movement and new location by its new exposed public IP address. For VPN all connections between VPN server and previous gateway will be changed to new gateway G3. It still protects cellphone's real network address, and the new latency overhead is determined by location of gateway G3. For MSS with MEC new optimal routes between cellphone and proxies will replace old ones. Privacy protection and latency overhead will remain at the same level as before, regardless of where user moves to.

In everyday use there is another typical scenario: vertical handovers between cellular data network and Wi-Fi. It is analogous to above discussed scenarios that exposed network location changes. Without protection it could pose higher privacy risk since public IP address from Wi-Fi network usually to be more specific to location. VPN will provide same protection while latency overhead is still determined by the

www.manaraa.com

distance triangle. MSS with MEC once again provides best privacy protection with minimum latency overhead. When peer node in inside cellular data network or in same MEC, MSS with MEC will have optimal performance since traffic are proxied at network switch connecting peer node. VPN server on the other hand will have the worst performance among three because traffic will have to traverse in and out of cellular data network.

3.4.1 When both ends employ MSS

When both ends employ MSS, MSS can still protect both of their network privacy, but could trade off with higher latency overhead. Depending on whether they use same MSS Service Provider(MSP) there could be two different scenarios:

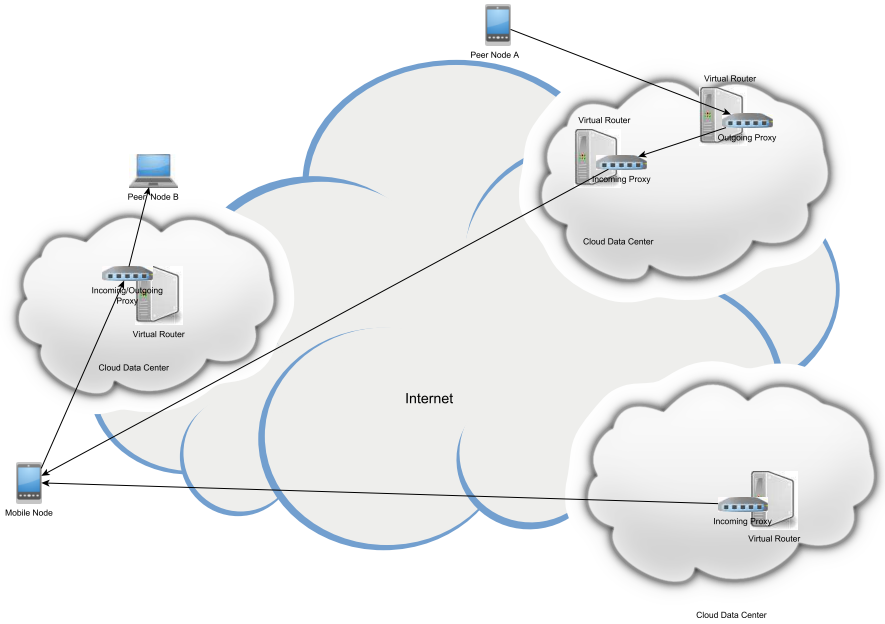

Figure 3.14.: Both ends employ MSS

- Peer Node uses a different MSP. This is the generic scenario. Mobile Node has a few listening Proxy allocated at different places. When Peer Node A want to connect Mobile Node (or the other way around by switching their roles) as

www.manaraa.com

shown in Fig 3.14, it can only resolve to Mobile Node 's current listening proxies' network addresses. Due to Mobile Node's de facto multi-homing, Peer Node A could receive one location that is close to where it initiates the lookup request, or even receiving multiple locations. In this case Peer Node A will have to choose one (or a few) deemed best for it, without guarantee of low performance overhead. Unfortunately, this is an inevitable tradeoff of maximizing privacy protection.

- Same MSP optimization mode. When Peer Node B uses same MSP and both agree to connect on identity, the connection initiator (Mobile Node in Fig 3.14 will ask MSP to allocate an outgoing Proxy to Peer Node B's identity, rather than an IP address. MSP in this case will allocate a single Proxy functioning both as incoming and outgoing for Peer Node B and Mobile Node, respectively. This Proxy will be selected around the middle point of the optimal route as long as privacy SLA allows. It functions like a "pivot" point that doesn't need to change frequently since the optimal route will not depart from it much while privacy is well protected. Additionally, when needed, such as performance overhead deteriorates, new Proxy can be created and MSP will ask both ends to migrate their connections since both connections are mobility capable. This mode can mitigate the performance issue in above different MSP scenario while still protect network privacy at same level.

3.5 Privacy Attack Models

We assume Alice, the attack target, carrying a mobile device with her all the time so that the network/geolocation of her Mobile Node is an approximate of Alice's geolocation. The adversary Bob wants to know Alice's current geolocation and location history so he can take advantage of that. The more precise location history Bob knows about Alice, the more sophisticated attack he will be able to craft. In following sections we list 4 major distinct attack models that Bob can leverage to

www.manaraa.com

attack on Alice's privacy. Note that different attack models could be combined in certain circumstances to further enhance attack effect as describe in attack scenarios.

3.5.1 Direct Connection Attack

By accomplishing this type of attack Bob can successfully directly connect to Alice's device, and even maintain a connection to it. Attack succeeds when connection can be setup successfully so that Bob can acquire current location of Alice. For protocols only bound to network location, such as TCP, adversary might need to perform further communication to confirm Alice's device identity. Identity bound protocols, such HIP, may give Bob enough information for identify verification with just connection attempt. There is one precondition of this attack that Bob has to know Alice's network location prior to connect.

3.5.2 Location Registry Attack

By accomplishing this type of attack Bob can indirectly acquire Alice's network location from a registration service, such as DNS, without direct interaction with Mobile Node. A successful attack will reveal one temporary contact point (not necessarily real network location of Mobile Node such as when Proxy is leveraged), and give Bob chance to further verifying by attempting direct connection. There is one precondition of this attack that Bob must know Alice's network identity in prior.

3.5.3 Historical Location Attack

By accomplishing this type of attack Bob can collect a sequential list of where Alice has been, which can be used to profile Alice or aid other type of attacks. Sequence here is important as more precise the location sequence, the better resolution of profiling adversary can achieve. However, a location list with completely wrong sequence may still be useful to Bob to some extent. The preconditions of this attack are that: 1)

www.manaraa.com

Bob must know Alice's network identity in prior; 2) Bob is able to retrieve a subset list of Alice's location history.

3.5.4 Location Change Timing Attack

By accomplishing this type of attack Bob knows Alice's device handover time, i.e. when Alice moves from one location to next location. This attack by itself does not reveal privacy that much, but when it's combined with other attack models Bob can dramatically increase profiling precision and multiply privacy attack damage.

3.6 Privacy Attack Scenarios

3.6.1 Adversary directly connects to Mobile Node

The simplest yet most impactful attack on Alice's privacy is that Bob can keep a live connection directly to Alice's Mobile Node device. Therefore, Bob will be able to know exactly Alice's network attachment location which can be mapped to geolocation. Also, Bob will know when Alice's address changes. Having that Bob not only know the real time location of Alice, but also can create a history timeline of Alice's movement. This is combination of Attack Model 1, 3, and 4.

When Bob knows Alice's real-world identity, with Alice's real time location and historical location information he can launch all kinds of sophisticated attack or even threatening Alice's physical world safety. Without knowing Alice's real-world identity Bob can still easily profile Alice by knowing her unique location history. Note that Bob does not need to be a friend of Alice to able to trace her. Bob can be a website Alice is used to visit, or just a script embedded in an advertisement.

3.6.2 Adversary resolve Mobile Node's address via a Location Service

Based on Scenario 1, assume Alice enhances her security by deploying a local firewall on her Mobile Node to refuse connection from Bob. This would to some

www.manaraa.com

extend prevent Bob to acquire real time location information of Alice. However, there could be some public location service, such as DNS, that can be used to resolve Alice's identity to her location in order for Alice to be connected. Bob can keep sending location resolution requests to this service to collect Alice's location history. This is combination of Attack Model 2 and 3.

Compared to Attack Scenario 1 Bob's tracking capability is limited: first Bob won't be able to get deterministic real time location of Alice since he cannot directly connect to her; second since the location registration is always lag behind, and sometimes protected by throttling mechanism, Bob will not perceive precise timing or even complete location history of Alice. In this case it only accomplishes Attack Model 2. When the registration service has access control and Bob is not whitelisted to resolve Alice's address, he will not be able to track Alice. However, maintaining a whitelist is difficult and expensive, as modern Internet host usually have tens or hundreds of open connections to web servers and other hosts at any moment. On the other hand, if Bob is allowed to connect to Alice or allowed to resolve Alice's location, Alice's exposure is no different than Attack Scenario 1.

3.6.3 Adversary connects through Proxy moving along with Mobile Node

Alice can protect her location privacy while keep connectivity by sending/receiving traffic through a Proxy. In this case Bob can communicate with Alice at any time, but only the Proxy location is exposed to Bob. Bob will only observe Proxy's location history, and under most circumstances Bob will not be able to detect whether Alice is behind a Proxy. This is combination of Attack Model, 3, and 4.

A typical example is cellular data network. When Alice uses cellular network to access Internet, usually Alice's Mobile Host will be assigned a private network address that is routable within carrier's network, and Alice will have to route her traffic through her cellular carrier's Internet gateway for Internet access. For Bob he will only see Internet gateway's network address as Alice's exposed network address.

www.manaraa.com

In this case the carrier's Internet gateway becomes a de facto proxy. When Alice roams away new private network address will be assigned. When this new private address associates with another Internet gateway which are usually close to the cell Alice is in, Bob will observe connection interruption and location change.

www.manaraa.com

4 DETAIL DESIGN AND VALIDATION

We designed system to implement architecture we proposed in Chapter 3. In this chapter we will describe system design and show how it can protect privacy while provide mobility support efficiently. The final system design evolved from our previous work that is based on all on-premise servers forming a continuous overlay network and aims to only enhance mobility support [99], to designs based on public cloud servers composing distributed system protecting network location privacy in addition to mobility support [100,102], and next to integrating MEC into MSP Virtual Router fleet [103,104]. In this chapter we will only present the complete final version.

4.1 Incorporating MEC

The incorporation of MEC provides MSS capability to greatly expand geolocation presence and thus reduce performance overhead. It also provides opportunities for MSS to reduce system operation cost. In fact with MEC MSS elevates its network paradigm from edge of network to network core.

Public Cloud service is the most crucial enabler for MSS: with it MSP can allocate Proxies on-demand all over the world, to maintain an optimal operation state satisfying both its customer requirement and financial sustainability. However, there are also limitations of utilizing public Cloud host to proxy traffic. First, public Clouds are not truly ubiquitous that their data centers are only available near selective major cities. For example, until now there are barely more than ten of metropolitan areas in the US has major public Cloud data centers nearby, and less than 100 edge locations national wide combined. In other countries the density is even less. Second, hosting Proxy in public Cloud may have advantage connecting to web sites also hosted in public Cloud, but for Peer Nodes outside Cloud, the Proxy is still at another

www.manaraa.com

edge of network. Traffic going through multiple ISP and boundary creates difficulty to maintain consistency and keep SLA. Third but not the last, Virtual Routers are mostly used for network I/O, which leaves large amount of computation and storage capability that MSP purchase but may not able to fully consume.

MEC provides answers to above three issues. MEC's physical presence is far beyond public Cloud data centers that cellular providers already deployed their networks and routers to almost everywhere. With MEC MSS can even find "next-door" proxy for every Peer Node. These RANs are also generally owned by same provider and connected with globally managed private links, which can provide better QoS than public Internet. Additionally, those network controllers and routers which implement MEC are originally specialized of handling network traffic. MSS can leverage MEC to augment performance and reduce operation cost, as illustrated in Fig 4.1. To incorporate MEC, MSS designs are changed in following areas:

1. Proxy application. Proxy is changed that can directly run as MEC application where VM is abstracted out from platform interface. Otherwise Virtual Router will still be allocated as ordinary VM on mobile edge host. Both their functionality and role remain the same.

2. Specialized Proxy. When MEC allows access to RAN especially managing customize routing rules, Proxy is replaced by routing table entries and/or packet repeater. In this case Virtual Router will reduce to an application listening to MSP signals and accordingly maintaining routing rules and policies of the network device it runs on. This mode reduces operation cost, and can also improve performance due to less traffic processing.

3. Network representation. In the simple mode where MSP does not have deep integration with cellular network, MSP just treats MEC Edge Clouds as equivalent of public Cloud data centers with less capacity. When MSP can have deeper integration, such as routing traffic within cellular network, MSP's network graph distinguishes cellular network and public Internet where two network connects

www.manaraa.com

through multiple Internet gateways. In this mode any connection cross two network boundary must go through one selected Internet gateway, and this gateway implements Proxy if privacy metric meets requirement. Otherwise proxy will have to allocated in a public Cloud data centers and gateway will be chosen at the place with minimum cost only.

4. Proxy allocation algorithm. MSP gives priority to MEC when allocating Proxy, but will retreat to public Cloud when region's MEC is not sufficient either in availability or performance. So a Mobile Node's Proxies could consist of MEC routing rules, MEC application, and public Cloud hosted ones.

5. On Mobile Node agent remains mostly the same, except it can now distinguish connection based on private or public IP address and manage accordingly.

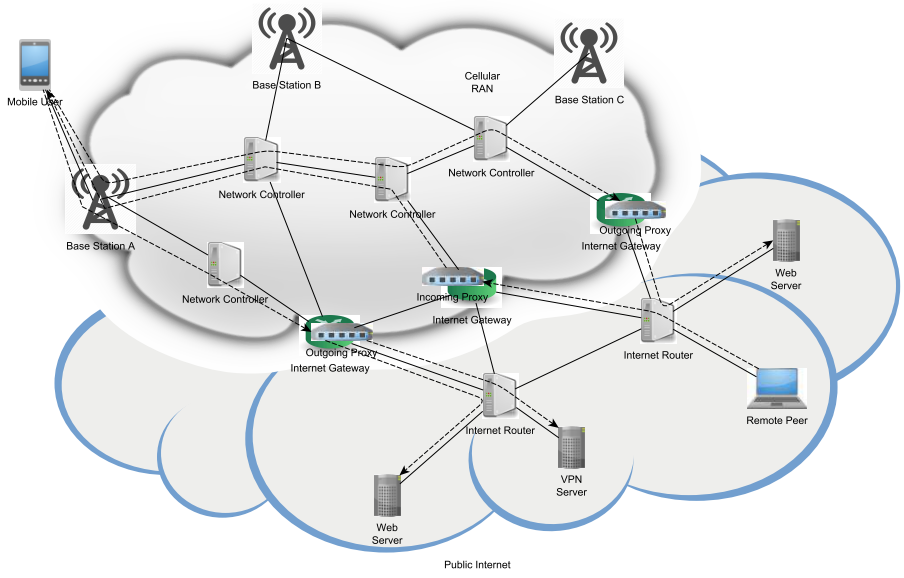

Figure 4.1.: Examples of MSP leveraging MEC

When MSS can have deeper integration with MEC (or cellular provider implements MSS) it gains more benefits besides the operation cost saving. For example, when MSP use MEC Edge Cloud as ordinary public Cloud, traffic from Mobile Node

www.manaraa.com

will first exit RAN into public Internet, go into another RAN to reach Proxy, exit second RAN, and finally reach Peer Node. As comparison a deeper integrated MSP will directly route traffic within cellular data network and exit at an Internet gateway very close to Peer Node. On the other hand, in fact cellular provider has inherent advantages becoming an MSP. For example, mobility Management is already one core functionality of cellular network. MSS's functionality such as tracking Mobile Node's location, looking up Mobile Node, replicating mobility information across network, etc. can be easily adapt from cellular network's similar functions. It also benefits MSS users since they don't even need to share their real network location to anyone else. Cellular provider becomes the only one knowing their network location (and they already know at the first place), a single one-to-one contract for mobile users is all they need to protect their network privacy completely.

Moreover, with deeper integration MSS can provide more options of SLAs for privacy protection and quality of service. For example:

- Different portion of routing traffic within cellular network and outside in public Internet implies different operation cost and QoS. A user could pay more to route traffic as much as possible within cellular network for not only minimizing public Internet exposure, but also more consistent performance.

- Cellular providers can allocate proxy from a larger group of device candidates, not limited to MEC Edge Cloud open for public. For example, those core/backbone switches are good candidate as they resides in the center of cellular network topology therefore having minimum performance penalty.

- Number of proxies, especially for multi-homing users. Since the cost of running proxy is greatly reduced and generally cellular provider have more distributed resources available, user can be given wider range of choice to find the sweat spot of cost versus number of listening proxies.

www.manaraa.com

4.2 Security Examine

In this section we briefly examine other security implications of MSS with MEC, besides the location privacy protection. We assume Alice, the attack target, carrying a mobile device with her all the time. Bob is an adversary tries to compromise Alice's communication via any possible means.

- **Anonymity**. In MSS with MEC every communication is bound to identity, such as cellular number, email address, or HIP's identity. All end-to-end communication is protected by encryption so no one other than end host can infer traffic content. However, there could be one special case that Proxy is instructed by Mobile Node to terminate encryption channel between it and Peer node for any reason, then Proxy itself will have access to communication traffic temporarily.

- **Unlinkability**.This is one of the major defenses MSS provides. All Alice's communications are distributed through different Proxies, which is dynamically assigned and also shared by many other mobile users. By observing one or a few Alice's Proxies will not be able to detect Alice's involvement with specific remote peer, nor whether Proxies are used by Alice.

- **Undetectability**. This is not main goal of MSS design. Bob could resolve to Alice's Proxies if Alice decides to advertise her identity and waiting for incoming connection request.

- **Repudiation**. It is up to Mobile Node's choice of end-to-end protocol when communicate with Peer Node. Between Mobile Node and Proxy repudiation is not critical and the end-to-end authentication and encryption are used to prove identity and sign communication. Mobile Node can choose to log or audit the packets it sends and receive from Proxy.

- **confidentiality**. When Mobile Node employs public/private key based end-to-end encryption protocol to Peer Node, no one other than two ends including

www.manaraa.com

Proxy, can infer the communication content. The communication can be further protected by employ similar end-to-end encryption between Mobile Node and Proxy when tunneling traffic so that no one else can infer what Peer Node is instructed to connect to.

4.2.1 Protect Against Attacks

MSS system enables ubiquitous mobility support while protects Mobile Node's privacy. For the four attack models we described:

- Defend Direct Connection Attack: all connections are indirect and through Proxy. So Bob will never be able to infer Alice's network address while being able to talk with her.
- Defend Location Registry Attack: all location registry only points to Alice's Proxy locations. Even if when Bob can acquire multiple Proxies' locations, he cannot infer Alice's location since these Proxies are setup near to Alice's Peer rather than Alice.
- Defend Historical Location Attack: Bob cannot acquire a list of Alice's real network location history either through communication with Alice or through registry.
- Defend Location Change Timing Attack: Bob cannot directly detect when Alice changes network location as the connection between him and Proxy is always unchanged.

4.3 Allocation Algorithms

There are two types of management need allocation algorithm:

- Mobile Node's overlay network management. For each Mobile Node, its overlay network summarizes all related metrics: 1) Cost, sum of Proxies(online and

www.manaraa.com

standby) determines overlay network's running cost; 2) load ratio of computation and network resource for each Proxy; 3) Performance, communicate latency and bandwidth usage for each Peer Connection; 4) Privacy, exposed distance and timing metric for each Peer Connection; 5) other attributes of Mobile Node, such as SLA, payment, authentication info, access control, etc. Having all these data, the overlay network makes decision on when and how to adjust Proxy, and which Proxy to take outgoing traffic or becoming Listening Proxy, so that SLA can be meet while running cost is minimized.

- MSP's region/data center management. MSP can optimize each region/data center individually as there is not much correlation of fleet management across region/data center(when there is, a third type of optimization for MSP to globally reduce running cost or improve performance across region/data centers could be added, such as by altering baseline parameter of a region, but we will limit our scope and not discuss it in this paper.). Within a region/data center, MSP needs to keep a healthy load ratio by dynamically add or remove Virtual Router, and by selecting proper Virtual Router to host new Proxy, while keeping running cost low. Running cost could be capped and in this situation the optimization algorithm needs to gracefully degrade by balancing load evenly.

4.3.1 Zone

For better scaling the MSP fleets are divided into multiple level of Zones, somewhat analogous to BGP Autonomous System but with hierarchy levels. Zone hierarchy forms a multi-root tree with attributes:

- A leaf Zone is a Zone does not contain other Zones.
- Leaf Zone is the minimum unit that MSP can allocate Virtual Router.
- For a given Internet address, there is one Leaf Zone and multiple non-leaf Zone containing it.

www.manaraa.com

- A Zone can be contained in one higher hierarchy level Zone, which is considered as its "Parent" Zone.
- A Zone can only overlap with its parent or ancestors.

Zone boundaries are generally decided according to geographic location as well as network topology, and the main use of Zone is for divide and conquer and global optimization to paralleled localized problem. Therefore, MSP can apply special zoning criteria to better match its interest.

4.3.2 Mobile Node Proxy Allocation Algorithms

Outgoing Proxy

When Mobile Node (MN) initiates connection to peer, it will submit a request to MSP for new outgoing Proxy. Given a MN already has a few outgoing Proxies, MSP can either create new Proxy at a Virtual Router (VR) already hosting this MN's proxies, or choose a VR not currently hosting MN's Proxy. For cloud and MEC based system the difference is trivial, as hosting multiple Proxies for a MN at the same VR only saves the update signaling when MN acquires a new network address. Therefore we simply algorithms to not take "reuse" VR into computation, and requesting new outgoing Proxy is simplified as independent operation.

To creating a new outgoing Proxy, there are three main constrains: privacy level, performance, and cost. Their definition and calculation are described in Sec 3.3, and the goal is to find the Proxy location having highest privacy and performance and lowest cost. This multi-target optimization is implemented by converting and merging numeric metric into a combined "score" metric, as describe in Sec 3.3, and MSP will select the Proxy with highest score. MN can also specify particular privacy or performance constrain, so the Proxy candidates will be filtered first to remove Proxy either not satisfying Privacy or Performance requirements. Similarly MSP could also filter based on cost. If after filtering no candidate is left, MSP will setup

www.manaraa.com

down filtering requirement per customer's SLA and negotiate with MN, or eventually fails the Proxy request if there is no location satisfying the requirement.

The allocation can also be constrained by Virtual Router's availability, i.e. if there is enough computation and bandwidth available at selected place to host this new outgoing Proxy. However this algorithm will always receive current fleet availability as input, which is managed by MSP using algorithm presented in Sec 4.3.3.

To scale the algorithm, the available Proxy locations are divided into hierarchy of zones, with the view as of multi-root tree. Algorithm will starting from top level of hierarchy which has small number of large coverage zones. After each run at one hierarchy level, only top 3 zones with highest scores will be selected and fed into next level search, until reach bottom level and find Proxy. The computation of Proxy scores in a selected Zone happen in the Zone SDN controller, and the summarize happens at the MSP server serving proxy creation request.

One example algorithm implementation is listed as in Algorithm 1. It uses a priority queue to store candidate , run a Breath-First search from top level zones.

Listening/Incoming Proxy

Mobile Node needs to pre-allocate Listening/Incoming Proxy if it wants to receive connections from its Peer Nodes. Since Peer Nodes' locations are unknown when setting up Listening Proxy, the score evaluation will based on prediction either made by customer manually (e.g. list potential Peer Node subnets or appoint locations where MN wants to have Listening Proxies) or computed by MSP using historical statistic.

Different than Outgoing Proxy which is allocated on-demand per connection, Listening Proxy allocation runs periodically for each MN, and each run generates the full distribution of all Listening Proxies. For a MN wants to listen on multiple ports, each port runs its own allocation and have independent Listening Proxy fleet.

www.manaraa.com

Algorithm 1 Algorithm of Allocating Outgoing Proxy

Require: MN location loc_{MN}, Peer Node location loc_{PN}, SLA,

Create Priority Queue *zones* with size limit 3, sort based on score

Create Priority Queue *candidates* with size limit 3, sort based on score

for all z in top level Zones **do**

$s_z \leftarrow$ score(z, loc_{MN}, loc_{PN}, SLA)

zones.push(z, s_z)

end for

while *zones* is not empty **do**

$z \leftarrow$ *zones*.pop()

if z is Virtual Router **then**

candidates.push(z, s_z)

else

$newZones \leftarrow$ requestCandidates($z, loc_{MN}, loc_{PN}, SLA$)

for all nz in $newZones$ **do**

$s_{nz} \leftarrow$ score(nz, loc_{MN}, loc_{PN}, SLA)

zones.push(nz, s_{nz})

end for

end if

end while

if *candidates* is not empty **then**

Fine sorting *candidates* based on extra requirement

Return top of *candidates*

else

Failed the request

end if

www.manaraa.com

The cost of setting up and tearing down Listening Proxy are considered as trivial for simplicity in our research, so as result each allocation will not take existing Listening Proxy distribution into account. For existing Listening Proxies having active connections but excluded from next iteration fleet, they will keep running but will not accept new connections, and terminate once current connections close.

Since Listening Proxy are expensive due to port is exclusive resource, allocation algorithm requires the maximum number of Listening Proxy, along with a list of expected addresses or Zones. MN's location is also required to evaluate and satisfy privacy SLA if specified. The first stage of algorithm is to merge expected Zones until number of Zones is less or equal to allowed Listening Proxies. The second stage of algorithm goes the opposite direction to search for the best location hosting Listening Proxy in each selected Zones. The example algorithm implementation is listed as in Algorithm 2.

4.3.3 MSP Server Fleet Allocation Algorithms

Each MSP needs to manage two different fleets: 1) control plane fleet that acts as controller managing system states and as webservice serving requests from customers; 2) data plane fleet that is composed of Virtual Routers carrying Proxies. Apparently, the latter fleet is much larger scale and dynamic, and contributes majority of MSP's operation cost. Additionally, since Virtual Routers are just process running on MSP' servers, the same servers can be used to run control plane controllers. Therefore the allocation algorithm will only use data plane usage (historical and future prediction) as input for the allocation algorithm and ignores control plane usage.

The fleet allocation algorithm needs to determine where to strategically put on-demand servers hosting Virtual Routers, so that all potential Proxy request can be accommodated until next run, while during this period the load on each server can maintain in healthy range. Since the allocated servers and resided Virtual Routers will serve all MNs, the algorithm will only optimize against the expected Peer Node.

www.manaraa.com

Algorithm 2 Algorithm of Allocating Listening/Incoming Proxy

Require: Port number p, allowed Proxy count max_{lp}, MN location loc_{MN}, List of locations loc_{PN}^{list}, SLA,

Create empty set $zones$

for all loc_{PN} in loc_{PN}^{list} **do**

$zones \leftarrow$ the hosting Zone of loc_{PN}

end for

while size($zones$) $> max_{lp}$ **do**

Find two Zones z_m and z_n in $zones$ having minimum distance

$zones.remove(z_m)$

$zones.remove(z_n)$

$z_x \leftarrow$ lowest common ancestor of z_m and z_n

$zones.add(z_x)$

end while

for all $zone$ in $zones$ **do**

Create empty set p

for all children Zone $zone_c$ in $zone$ **do**

$p_c \leftarrow \frac{expectedPeeraddress/Zoneinzone_c}{expectedPeeraddress/Zoneinzone}$

$p.add(p_c)$

end for

if largest p_c in p is greater than threshold **then**

repeat search in $zone_c$

else

allocate proxy in $zone_c$

end if

end for

www.manaraa.com

Apparently MSPs that only employ on-premise servers will not need this allocation adjustment as all servers are static. MSP run allocation algorithm periodically to ensure it maintains a healthy presence and server load.

first step of algorithm is to summarize count of potential Proxies in each Zone. Then for MSP employing hybrid fleet, i.e. having on-premise servers, those servers will be first employed to serve nearby Zones' Proxy requests until reaching their capacity limit, or depleting nearby Proxy requests (where nearby is defined by a network distance threshold). However either way these on-premise servers will not be considered again in the second step of computing server allocated from public Cloud Service data centers and Mobile Edge Computing gateways/switches. The reason is that this optimization is a NP-complete problem for a general graph since it needs to consider all possible solutions and replica layout are highly dependent(i.e. the setup of a replica at a specific location is based on the existence and position of all other replicas) [109], so optimization is usually conducted on reduced topology, i.e. a tree abstracted from general graph, and this optimization problem can be reduced to a $p-median$ problem in Discrete Location Theory, which can be solved by using existing dynamic programming (DP) solutions. We present an optimization algorithm based our previous published in [99, 100] which is based on the off-line optimal algorithm Tian and Cox described in [110]. A pseudo root Zone is added to the multi-room Zone trees.

Example algorithm is listed in Algorithm 3. It takes $onPremZones$ as Zones hosting on-premise servers; $onDemandZones$ as Zones that can launch on-demand servers; loc_{PN}^{list} as list of Peer Node, number of connections to/from them, and capacity requirement of each connection; $budget$ as total budget MSP can spend on on-demand server, and $serverCost(zone)$ as cost and $serverCap(zone)$ as capacity of on-demand server in corresponding Zone. $\beta(i, j)$ is the element to measure the cost to Proxy serving Peer Node from node j at node i; $\alpha(i, T_j)$ is the optimal cost of serving subtree T_j, including Peer Node connections originate from T_j when served by a

www.manaraa.com

Proxy at node i; and $\alpha(T_j)$ is the optimal cost of serving subtree T_j when Proxies are all within.

Algorithm 3 MSP Server Fleet Allocation Algorithm

Require: $onPremZones$, $onDemandZones$, loc_{PN}^{list}

create empty set r

for all $onPremZone$ in $onPremZones$ **do**

remove PN with distance and capacity threshold from loc_{PN}^{list}

$r.add(onPremZone)$

end for

for all $zone$ in $onDemandZones$ **do**

Sum up PN requests for $zone$ loc_{PN}^{list}

end for

Calculate β matrix for each pair of zones.

Compute α of each leaf zone.

Starting from bottom zones until root, do:

for all j As a non-leaf zone **do**

for all i As each other zone **do**

Compute $\alpha(i, T_j)$

end for

Compute $\alpha(T_j)$

end for

$r.addAll(serversofpseudorootZone))$

4.4 Protection for IoT Devices

For IoT devices that are directly connect Internet through cellular data network and capable to deploy MSS agent, they can leverage MSS to protect their network privacy just like ordinary Mobile Node. However, for IoT devices that are resource

www.manaraa.com

constrained or not capable to deploy MSS agent, MSS has a special protection model for them: adding one additional Master Proxy which is allocated close to IoT devices and this Master Proxy will function as IoT devices' MSS agent, as shown in Fig 4.2. IoT devices just need to configure passing all their traffic through this Master Proxy, for example set as their network gateway. Master Proxy will function just like an MSS agent that setup connections to Peer Node through outgoing and incoming Proxies. When IoT devices moves, Master Proxy will move along with it to keep a minimum distance in order to minimize performance overhead.

Particularly IoT devices within same subnet or behind same public Internet attach point can share same Master Proxy. For example, for a smart home, cameras, smart power switches, garage controls, etc. all can direct all their traffic to this nearby Master Proxy. For cellular network with MEC, this master Proxy could provide another layer of protection that other local MEC application will not be able to probe the remote sides of connections from IoT devices.

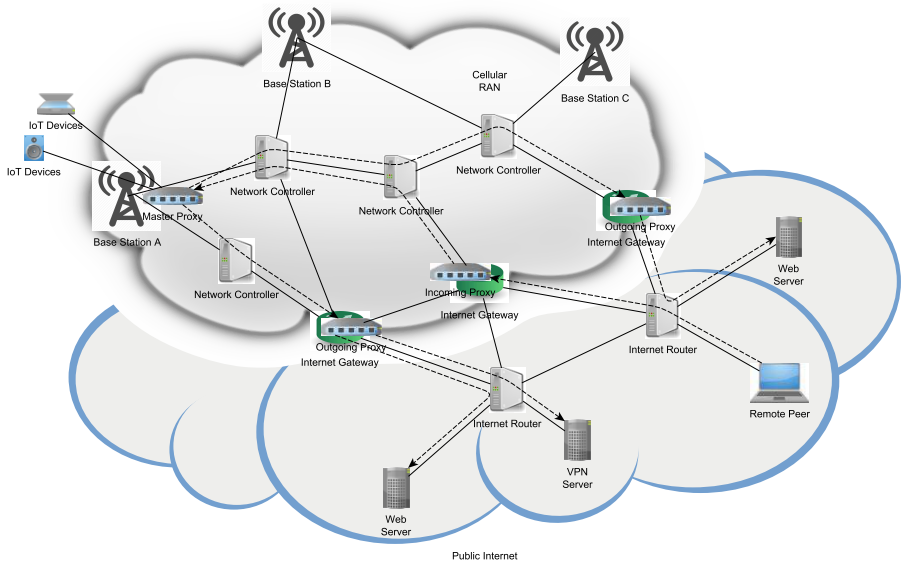

Figure 4.2.: IoT Devices leverage a Master Proxy

www.manaraa.com

4.5 Simulation

To validate our design and algorithm we setup simulation trying to mimic real world topology and traffic pattern. Simulations were conducted on a topology generated by Inet3 [111] topology generation tool. With it we created 5000 nodes on a 2D plain with dimension 10000 of 10000 to emulate a simplified view of Internet. Each node represents a public Internet subnet that is publicly routable and reachable. Their coordinates represent their geographic location, while hops between two nodes represent the network distance. Due to the mobility nature, every event in simulation has a timestamp and all events are merged onto same timeline during one simulation iteration. The Maximum event timestamp is 2000.

During this period, Mobile Node perform a random walk in 2D plane and connects to the geographic nearest node as its Internet attaching point. This simulates an Internet mobile user roams across different networks and uses nearby gateway as its exposed Internet address. Each Mobile Node randomly moves 20 times at random timestamp. For each Mobile Node we create a pool of 20 Peer Nodes, and their locations are randomly selected from the 5000 nodes. These are Peer Nodes either initiating connections to Mobile Node, or receiving connection coming from Mobile Node. For this simulation we do not count connection time but only use location attributes of one connection to compute metrics such as performance or privacy. Each Peer Node can have 1 to 10 connections during the simulation period, and timestamp is also uniformly random generated. On the other hand Mobile Nodes randomly nominate 3 locations for hosting listening proxies to receive incoming connection. This design is trying to simulate one ordinary Internet mobile use's average day access pattern. We generate traces of 100 Mobile Nodes and compare average of Mobile Node's metrics, to counter the potential skew caused by accidental non-uniform randomness.

We run simulation to compare MSS versus other models. The comparison is based on view of individual customer's privacy and performance metrics when one customer moves while communicates with its peers. Models include:

www.manaraa.com

1. Typical VPN user. We simulate by selecting one static Proxy Server selected around the middle of the optimal router between Mobile Node and its first Peer node. It does give favor to first Peer Node but statistically it does not make difference since we have 19 other Peer nodes for simulation.

2. Typical cell data mobile user. We simulate by keeping a proxy server that follows Mobile Node move (randomly picked within 2 hops of Mobile Node's attaching point, and re-selects when Mobile Node's access point moving away from attaching point). This proxy server simulates the Internet Gateway which relays a cellular data network user's IP packets.

3. MSS based on public Cloud only. Among the 5000 nodes we randomly select 10 "seed" nodes that are not too close to each other to represent public Cloud service provider's data centers. We randomly "grow" the data center footprint by tainting nodes directly connect to it in 2 to 3 hoops, to simulate public cloud data center covering large fleets of servers and owning large amount of IP addresses.

4. MSS based on MEC. Among the 5000 nodes we exclude leaf nodes (those have only one connection to neighbors) and core nodes (those have highest number of connections to neighbors), and all the rest are candidate of MEC nodes. This simulate the ubiquitous presence of cellular data network and potential coverage of MEC.

In real world network distance (either hop, latency, or subnet distance) is harder to quantify and does not reflect geographical distance well, because network address, such as IP, are usually not uniformly distributed or segmented. Less network hops don't always mean closer geographic distance. On the other hand geographic distance mapped from IP address is easier to quantify and relative reliable. Even though there are cases that IP address incorrectly mapped to wrong geographical locations (mostly depending on IP database), in reality this doesn't impair privacy. Therefore, in our

www.manaraa.com

simulation we will assume all IP addresses can be correctly mapped so our evaluation can rely on geographical distance of network attaching point for comparison.

Performance overhead is measured as the extra hops compared to optimal route, applying our Performance equation. Fig 4.3 shows Cumulative Distribution Function (CDF) of performance overhead of four competitors. "VPN" apparently has largest performance penalty as all traffics go through a static point no matter when Mobile Node moves to. On the other hand "Cellular Data Network" has the lowest performance overhead since the data path is almost always near optimal as the Internet Gateway is always nearby. However, note that this low overhead accompanied with multiple dysconnectivity of existing connections as each change of Internet Gateway will disturb ongoing traffic and connection. "MSS" in two configurations show great performance overhead that are very close to "Cellular Data Network" without sacrificing neither connectivity nor privacy.

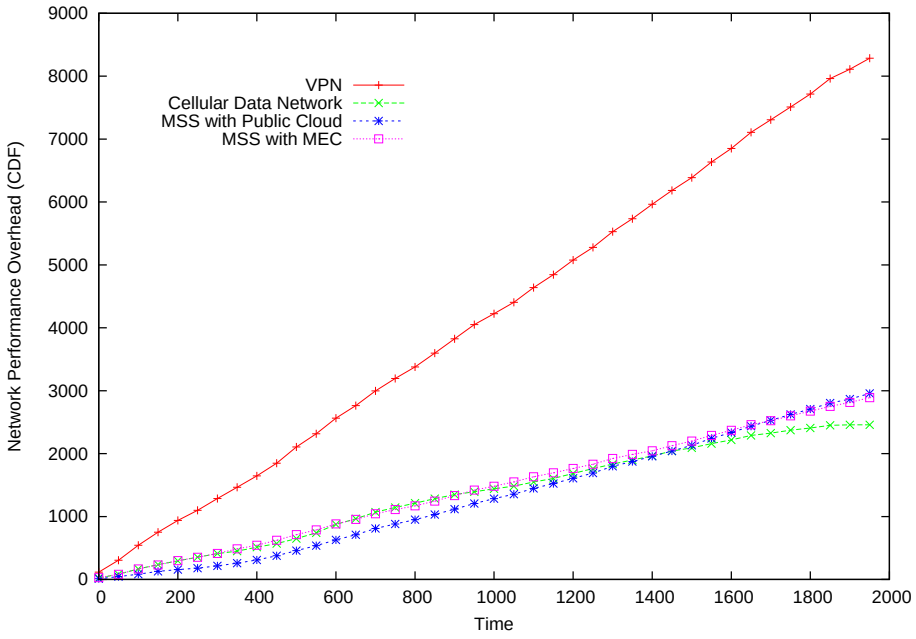

Figure 4.3.: Performance Overhead Comparison

Fig 4.4 presents a point-in-time performance overhead among four. It is clear that VPN always has high performance overhead cost. MEC based MSS generally

www.manaraa.com

has lower performance overhead comparing to Public Cloud based MSS. "Cellular Data Network" performance overhead decreased in last one third time range mostly due to Mobile Node not moving much during that period.

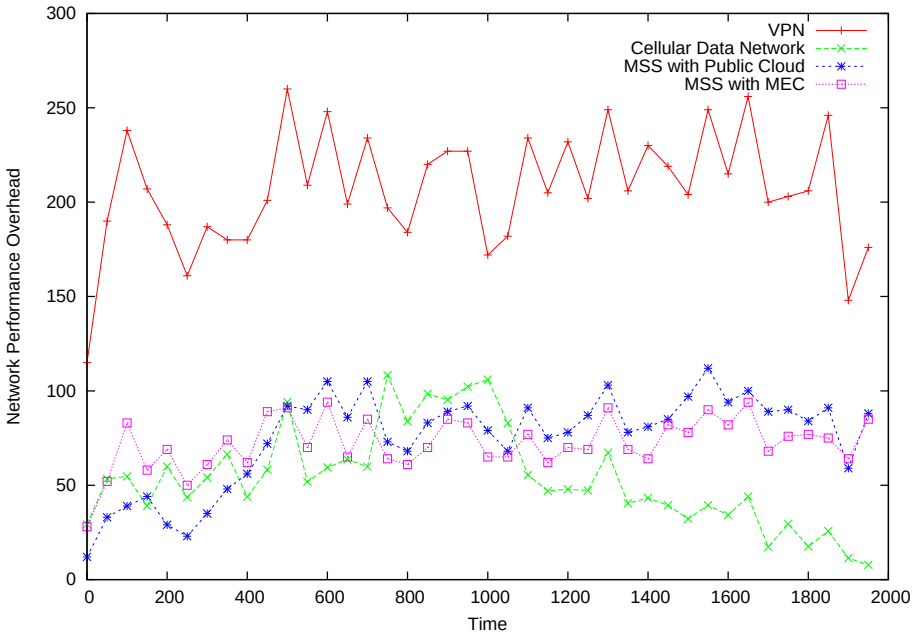

Figure 4.4.: Performance Overhead Per Sampling Point Comparison

Location privacy is measured based on our location privacy evaluation equation. Fig 4.5 displays the average location privacy metrics. VPN unsurprising dominate as it only exposed the VPN address no matter where the Mobile Node or Peer Node, and in general Mobile Node would select a VPN server far from its actually location. MSS in two configurations perform good, and MEC based one beat public cloud based one. Cellular Data Network generally doesn't protect Mobile Node's location privacy much, unless in rural large cell areas where an Internet Gateway actually covers large area.

The overall signaling cost, including outgoing (setting up connection to a Peer Node), lookup (incoming connection from Peer Node), and update (update all Proxies when Mobile Node changing attaching point) are illustrated in Fig 4.7

www.manaraa.com

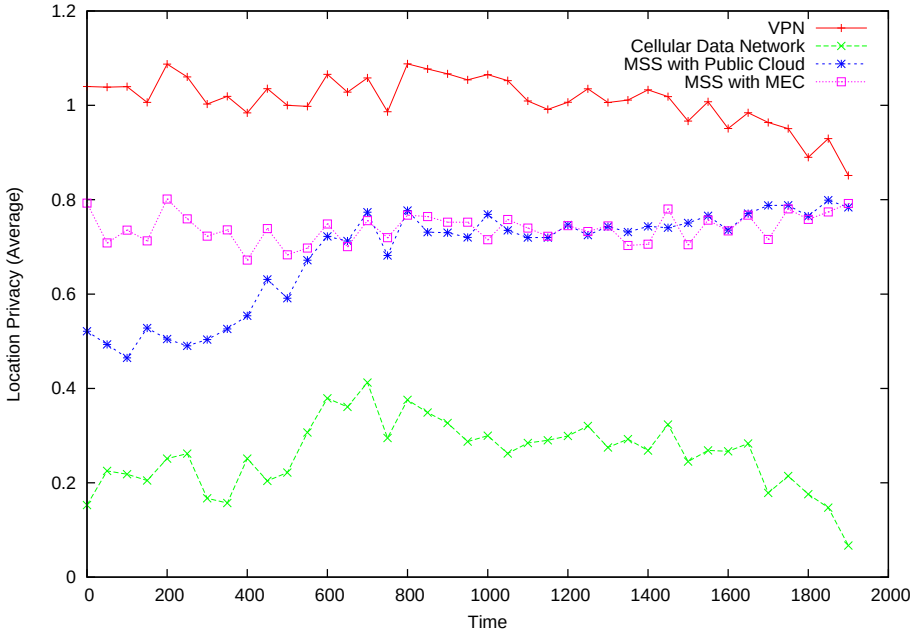

Figure 4.5.: Location Privacy Comparison (Average)

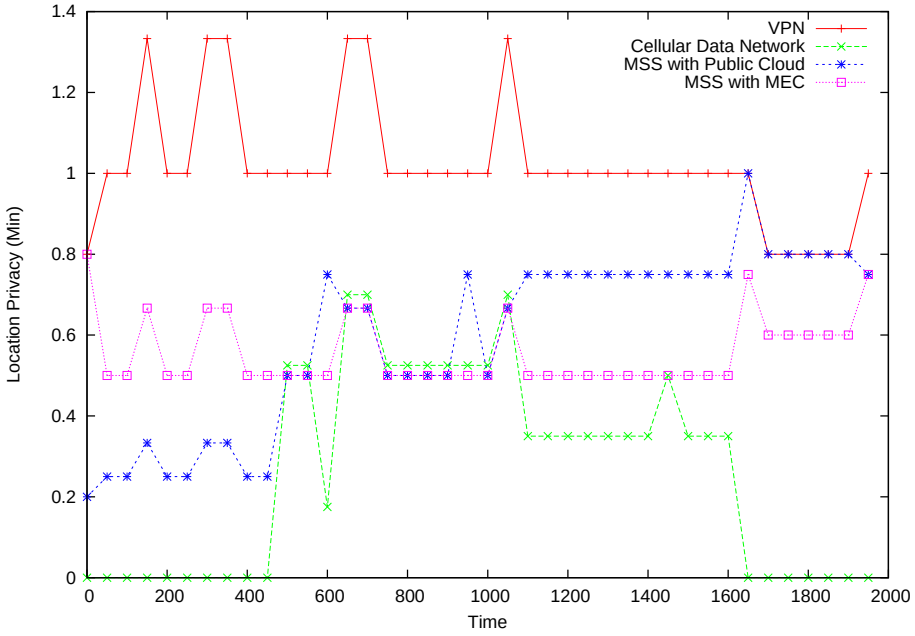

Figure 4.6.: Location Privacy Comparison (Min)

www.manaraa.com

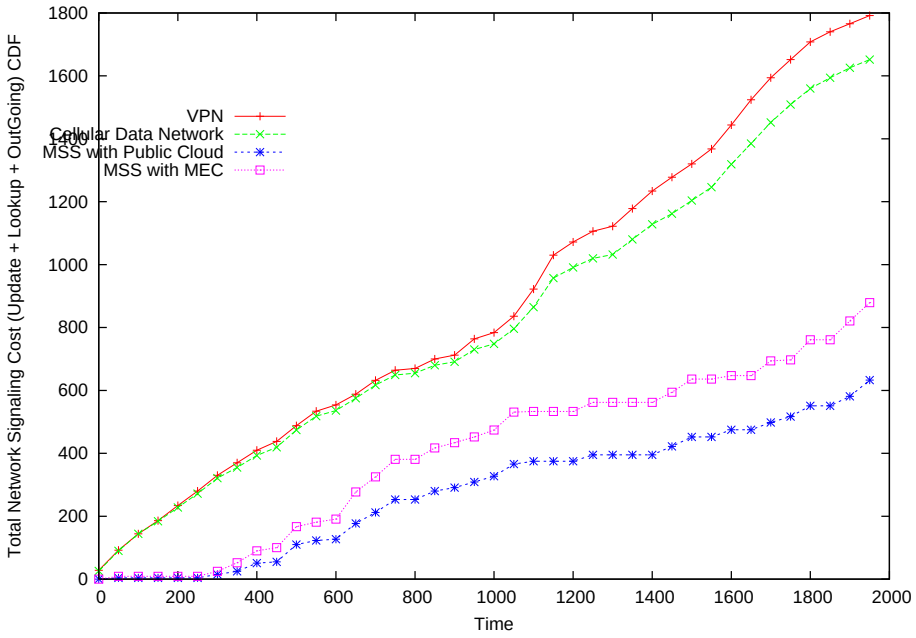

Figure 4.7.: Signaling Cost Comparison

We further experimented with different maximum allowed Proxies, i.e. maximum 1, 4, 16, and 32 Proxies for each Mobile Node in simulation. Fig 4.8 shows that when allowing more Proxies privacy metric improved. On the other hand the percentage of improvement starts to become less for our simulation setup when allowed maximum Proxies is more than 4, and became less significant when allowed Proxies increased to 16. Our further analysis shows the sweet spot is around 8, and the reason is that 8 Proxies are about large enough to cover the 20 Peer Nodes we selected for each Mobile Node.

www.manaraa.com

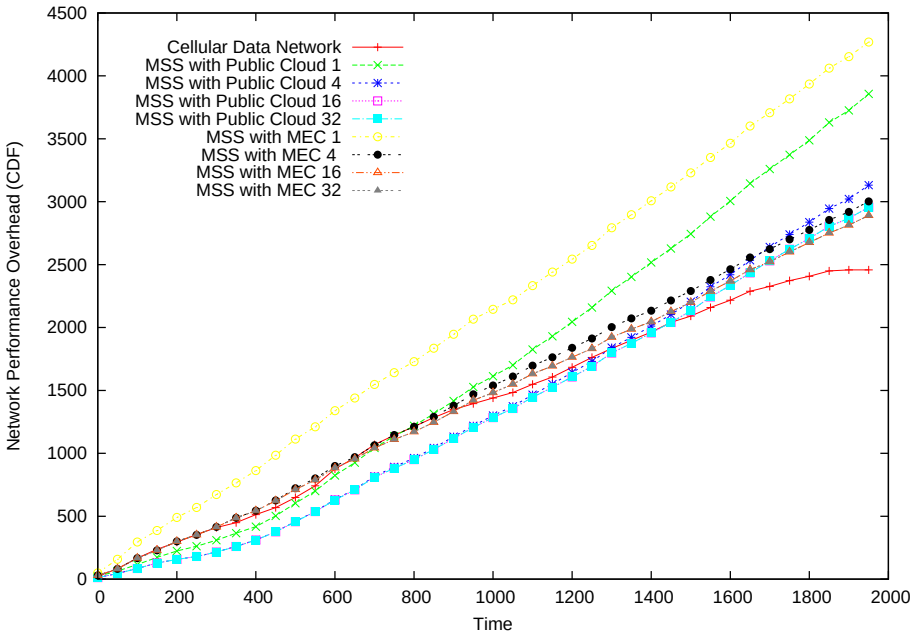

Figure 4.8.: Scaling Performance Overhead Comparison

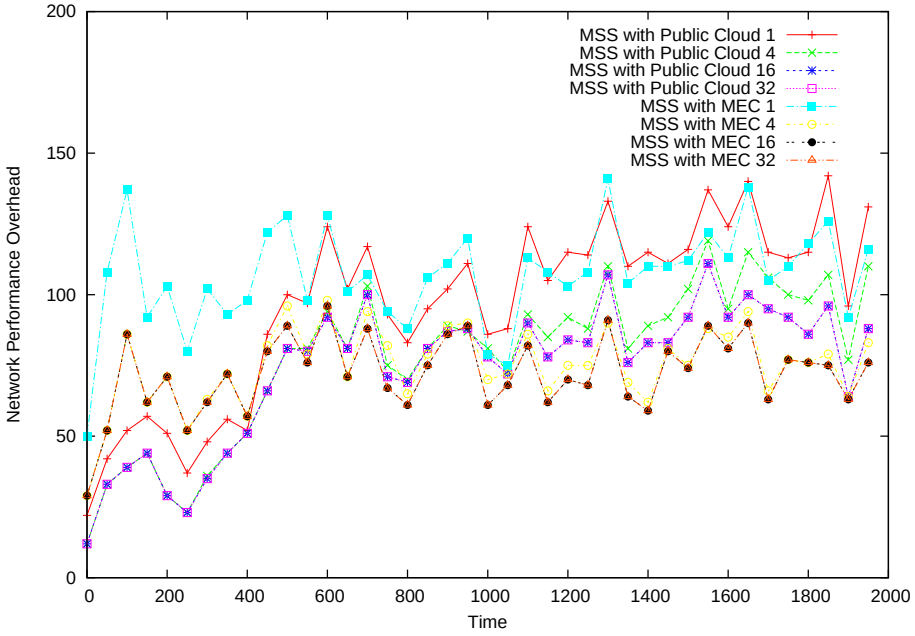

Figure 4.9.: Scaling Performance Overhead Per Sampling Point Comparison

www.manaraa.com

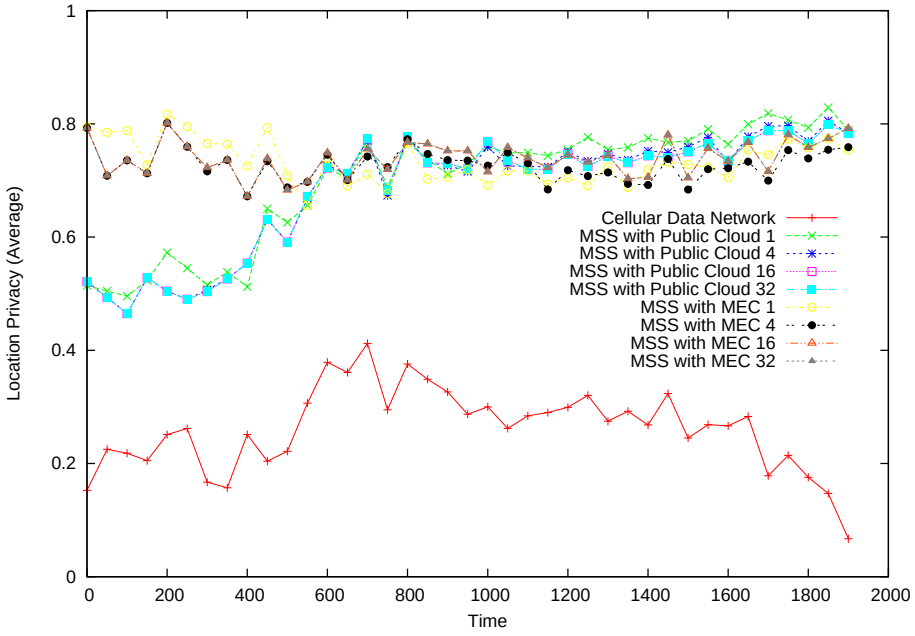

Figure 4.10.: Scaling Location Privacy Comparison (Average)

www.manaraa.com

5 CONCLUSION

This research originated from recognition of cellular base Internet mobility exposing privacy vulnerability and lacking generic Internet mobility support. Since lots of researches had been done to implement fundamentally mobility friendly or oriented Internet but none of them succeeded in real world, investigation of what Internet mobility requires and why these proposals are not real world feasible was conducted, and concluded with principals on how to create an economic feasible Internet mobility support solution. Based on the novel proxy paradigm we proposed, Mobility Support Service (MSS) architecture was introduced to enhance network location privacy protection and mobility support in Internet. Details of prototype implementation is discussed, and simulations were conducted for validation.

The major contributions and novelty of this research include: we reviewed the state-of-the-art Internet mobility support proposals and solutions, and found a few hurdling problems overlooked by previous researches; we studied the economic viability of protocols and system designed for Internet, especially mobility support system; we examined the role of network location privacy protection in Internet mobility support, identified most important factors, and proposed metrics to quantify and evaluate them; we proposed a novelty proxying paradigm that aggressively push proxy close to remote peer, and have multiple proxies simultaneously for single end host. This paradigm is the key for minimizing interruption and performance penalty brought by mobility, and maximizing network location privacy protection; we proposed an Internet mobility support system that can support a generic Internet mobility and protect network location privacy, and this framework can accommodate/complement some important existing protocols and solutions; we designed algorithms for managing proxy allocation against criteria including multiple performance metrics and overhead.

www.manaraa.com

This research could be further extended in a few areas. Statistical model is the most intuitive and usually most precise tool to predict request patterns. For example, system can track history of a subscriber's repeated outgoing and incoming connection peers along with time ranges, so that a generalized individual model can be created to adapt to this particular subscriber's usage. The system this information to improve user's performance by pre-allocating Proxies at expected zones, while lower operation cost by consolidating Proxy. Subscribers' everyday activities provide enough data set for Machine Learning to discover and improve prediction model, and other statistics tools such as Bayesian network can also be used to promote prediction successful rate. On the broader level, the overall usage pattern of Virtual Routers which serve all subscribers is an even better Machine Learning target, since the overall usage will be more regular and easy to predict.

www.manaraa.com

REFERENCES

www.manaraa.com

REFERENCES

[1] Thomas R. Henderson. Host mobility for ip networks: a comparison. *IEEE Network*, 17:18–26, November 2003.

[2] Eranga Perera, Vijay Sivaraman, and Aruna Seneviratne. Survey on network mobility support. *SIGMOBILE Mob. Comput. Commun. Rev.*, 8(2):7–19, 2004.

[3] Wesley M. Eddy. At what layer does mobility belong? *IEEE Communications Magazine*, 42:155–159, October 2004.

[4] Deguang Le, Xiaoming Fu, and Dieter Hogrefe. A review of mobility support paradigms for the internet. In *Communications Surveys and Tutorials, IEEE*, volume 8, pages 38–51, August 2006.

[5] Mohammed Atiquzzaman and Abu S. Reaz. Survey and classification of transport layer mobility management schemes. In *Personal, Indoor and Mobile Radio Communications, 2005. PIMRC 2005. IEEE 16th International Symposium on*, volume 4, pages 2109–2115. IEEE, IEEE, September 2005.

[6] David Clark, Robert Braden, Aaron Falk, and Venkata Pingali. Fara: reorganizing the addressing architecture. In *FDNA '03: Proceedings of the ACM SIGCOMM workshop on Future directions in network architecture*, pages 313–321, New York, NY, USA, 2003. ACM.

[7] David D. Clark, Karen Sollins, John Wroclawski, and Ted Faber. Addressing reality: an architectural response to real-world demands on the evolving internet. *SIGCOMM Comput. Commun. Rev.*, 33(4):247–257, 2003.

[8] Elin Wedlund and Henning Schulzrinne. Mobility support using sip. In *WOWMOM '99: Proceedings of the 2nd ACM international workshop on Wireless mobile multimedia*, pages 76–82, New York, NY, USA, 1999. ACM.

[9] R. Moskowitz, P. Nikander, P. Jokela, and T. Henderson. Host Identity Protocol. RFC 5201 (Experimental), April 2008.

[10] D. Farinacci, V. Fuller, D. Meyer, and D. Lewis. Locator/id separation protocol (lisp) draft-ietf-lisp-03.txt, July 2009.

[11] C. Perkins. IP Mobility Support for IPv4. RFC 3344 (Proposed Standard), August 2002. Updated by RFC 4721.

[12] Milind Buddhikot, Adiseshu Hari, Kundan Singh, and Scott Miller. Mobilenat: a new technique for mobility across heterogeneous address spaces. *Mob. Netw. Appl.*, 10(3):289–302, 2005.

www.manaraa.com

[13] Alex C. Snoeren and Hari Balakrishnan. An end-to-end approach to host mobility. In *MobiCom '00: Proceedings of the 6th annual international conference on Mobile computing and networking*, pages 155–166, New York, NY, USA, 2000. ACM.

[14] Ilknur Aydin. Cellular sctp: A transport-layer approach to internet mobility. In *Computer Communications and Networks, 2003. ICCCN*, pages 285–290, 2003.

[15] Ion Stoica, Daniel Adkins, Shelley Zhuang, Scott Shenker, and Sonesh Surana. Internet indirection infrastructure. In *SIGCOMM '02: Proceedings of the 2002 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 73–86, New York, NY, USA, 2002. ACM.

[16] J. Laganier, T. Koponen, and L. Eggert. Host Identity Protocol (HIP) Registration Extension. RFC 5203 (Experimental), April 2008.

[17] C. Y. T. Ma, D. K. Y. Yau, N. K. Yip, and N. S. V. Rao. Privacy vulnerability of published anonymous mobility traces. *IEEE/ACM Transactions on Networking*, 21(3):720–733, June 2013.

[18] M. Lin, H. Cao, V. Zheng, K. C. Chang, and S. Krishnaswamy. Mobile user verification/identification using statistical mobility profile. In *2015 International Conference on Big Data and Smart Computing (BIGCOMP)*, pages 15–18, Feb 2015.

[19] G. Tsoukaneri, G. Theodorakopoulos, H. Leather, and M. K. Marina. On the inference of user paths from anonymized mobility data. In *2016 IEEE European Symposium on Security and Privacy (EuroS P)*, pages 199–213, March 2016.

[20] Vaibhav Kulkarni, Arielle Moro, and Benoît Garbinato. A mobility prediction system leveraging realtime location data streams: Poster. In *Proceedings of the 22Nd Annual International Conference on Mobile Computing and Networking*, MobiCom '16, pages 430–432, New York, NY, USA, 2016. ACM.

[21] Buğra Gedik and Ling Liu. Protecting location privacy with personalized k-anonymity: Architecture and algorithms. *IEEE Transactions on Mobile Computing*, 7(1):1–18, January 2008.

[22] Ioannis Boutsis and Vana Kalogeraki. Location privacy for crowdsourcing applications. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '16, pages 694–705, New York, NY, USA, 2016. ACM.

[23] C. Y. T. Ma, D. K. Y. Yau, N. K. Yip, and N. S. V. Rao. Privacy vulnerability of published anonymous mobility traces. *IEEE/ACM Transactions on Networking*, 21(3):720–733, June 2013.

[24] D. Chaum. Untraceable electronic mail, return addresses and digital pseudonyms. In *Communications of the ACM*, volume 4(2), February 1981.

[25] Roger Dingledine, Nick Mathewson, and Paul Syverson. Tor: The second-generation onion router. In *Proceedings of the 13th USENIX Security Symposium*, 2004.

www.manaraa.com

[26] D. Kristol and L. Montulli. HTTP State Management Mechanism. RFC 2965 (Proposed Standard), October 2000.

[27] Jon Salz, Alex C. Snoeren, and Hari Balakrishnan. Tesla: a transparent, extensible session-layer architecture for end-to-end network services. In *USITS'03: Proceedings of the 4th conference on USENIX Symposium on Internet Technologies and Systems*, pages 16–16, Berkeley, CA, USA, 2003. USENIX Association.

[28] Bruno Quoitin, Luigi Iannone, Cédric de Launois, and Olivier Bonaventure. Evaluating the benefits of the locator/identifier separation. In *MobiArch '07: Proceedings of first ACM/IEEE international workshop on Mobility in the evolving internet architecture*, pages 1–6, New York, NY, USA, 2007. ACM.

[29] R. Stewart. Stream Control Transmission Protocol. RFC 4960 (Proposed Standard), September 2007.

[30] R. Stewart, Q. Xie, M. Tuexen, S. Maruyama, and M. Kozuka. Stream Control Transmission Protocol (SCTP) Dynamic Address Reconfiguration. RFC 5061 (Proposed Standard), September 2007.

[31] P. Nikander, T. Henderson, C. Vogt, and J. Arkko. End-Host Mobility and Multihoming with the Host Identity Protocol. RFC 5206 (Experimental), April 2008.

[32] C. Perkins, P. Calhoun, and J. Bharatia. Mobile IPv4 Challenge/Response Extensions (Revised). RFC 4721 (Proposed Standard), January 2007.

[33] D. Johnson, C. Perkins, and J. Arkko. Mobility Support in IPv6. RFC 3775 (Proposed Standard), June 2004.

[34] Stuart Cheshire and Mary Baker. *Internet mobility 4x4*, pages 388–400. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 1999.

[35] Charles E. Perkins and David B. Johnson. Mobility support in ipv6. In *MobiCom '96: Proceedings of the 2nd annual international conference on Mobile computing and networking*, pages 27–37, New York, NY, USA, 1996. ACM.

[36] V. Devarapalli, R. Wakikawa, A. Petrescu, and P. Thubert. Network Mobility (NEMO) Basic Support Protocol. RFC 3963 (Proposed Standard), January 2005.

[37] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil. Proxy Mobile IPv6. RFC 5213 (Proposed Standard), August 2008.

[38] Yun Mao, Bjï¿½ï¿½orn Knutsson, Honghui Lu, and Jonathan M. Smit. Dharma: Distributed home agent for robust mobile access. In *Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies 2005 (INFOCOM 2005)*, volume 2, pages 1196–1206, 2005.

[39] R. Moskowitz and P. Nikander. Host Identity Protocol (HIP) Architecture. RFC 4423 (Informational), May 2006.

[40] P. Nikander and J. Laganier. Host Identity Protocol (HIP) Domain Name System (DNS) Extensions. RFC 5205 (Experimental), April 2008.

www.manaraa.com

[41] J. Laganier and L. Eggert. Host Identity Protocol (HIP) Rendezvous Extension. RFC 5204 (Experimental), April 2008.

[42] Shelley Zhuang, Kevin Lai, Ion Stoica, Randy Katz, and Scott Shenker. Host mobility using an internet indirection infrastructure. *Wirel. Netw.*, 11(6):741–756, 2005.

[43] Andrei Gurtov, Dmitry Korzun, Andrey Lukyanenko, and Pekka Nikander. Hi3: An efficient and secure networking architecture for mobile hosts. *Comput. Commun.*, 31(10):2457–2467, 2008.

[44] Dipankar Raychaudhuri, Kiran Nagaraja, and Arun Venkataramani. Mobilityfirst: A robust and trustworthy mobility-centric architecture for the future internet. *SIGMOBILE Mob. Comput. Commun. Rev.*, 16(3):2–13, December 2012.

[45] Tam Vu, A. Baid, Y. Zhang, T.D. Nguyen, J. Fukuyama, R.P. Martin, and D. Raychaudhuri. Dmap: A shared hosting scheme for dynamic identifier to locator mappings in the global internet. In *Distributed Computing Systems (ICDCS), 2012 IEEE 32nd International Conference on*, pages 698–707, June 2012.

[46] R.D. Yates and W. Lehr. Mobilityfirst, lte and the evolution of mobile networks. In *Dynamic Spectrum Access Networks (DYSPAN), 2012 IEEE International Symposium on*, pages 180–188, Oct 2012.

[47] András G. Valkó. Cellular ip: a new approach to internet host mobility. *SIGCOMM Comput. Commun. Rev.*, 29(1):50–65, 1999.

[48] Andrew T. Campbell, Javier Gomez, Sanghyo Kim, Chieh yih Wan, Zoltan R. Turanyi, and G. Valko. Ip micro-mobility protocols. *IEEE Wireless Communications*, 9:45–54, 2002.

[49] Siegmund M. Redl, Matthias K. Weber, and Malcolm W. Oliphant. *An Introduction to GSM*. Artech House Publishers, 1995.

[50] Uyless Black. *Mobile and Wireless Networks*. Advance Communications Technology. Prentice Hall, second edition edition, 1999.

[51] Emir Halepovic and Carey Williamson. Characterizing and modeling user mobility in a cellular data network. In *PE-WASUN05*. ACM, Oct 2005.

[52] C. Wang, F. Haider, X. Gao, X. You, Y. Yang, D. Yuan, H. M. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir. Cellular architecture and key technologies for 5g wireless communication networks. *IEEE Communications Magazine*, 52(2):122–130, February 2014.

[53] Afif Osseiran, Jose F. Monserrat, and Patrick Marsch. *5G Mobile and Wireless Communications Technology*. Cambridge University Press, New York, NY, USA, 1st edition, 2016.

[54] H. H. Pang and K. L. Tan. Authenticating query results in edge computing. In *Proceedings. 20th International Conference on Data Engineering*, pages 560–571, March 2004.

www.manaraa.com

[55] Shanhe Yi, Cheng Li, and Qun Li. A survey of fog computing: Concepts, applications and issues. In *Proceedings of the 2015 Workshop on Mobile Big Data*, Mobidata '15, pages 37–42, New York, NY, USA, 2015. ACM.

[56] A. V. Dastjerdi and R. Buyya. Fog computing: Helping the internet of things realize its potential. *Computer*, 49(8):112–116, Aug 2016.

[57] P. Mach and Z. Becvar. Mobile edge computing: A survey on architecture and computation offloading. *IEEE Communications Surveys Tutorials*, 19(3):1628–1656, thirdquarter 2017.

[58] Michael Till Beck, Martin Werner, Sebastian Feld, and Thomas Schimper. Mobile edge computing: A taxonomy, 2015.

[59] A. Ahmed and E. Ahmed. A survey on mobile edge computing. In *2016 10th International Conference on Intelligent Systems and Control (ISCO)*, pages 1–8, Jan 2016.

[60] Yves-Alexandre de Montjoye, CÃ©sar A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports*, 3(1376), 2013.

[61] Rongxing Lu, Hui Zhu, Ximeng Liu, J. K. Liu, and Jun Shao. Toward efficient and privacy-preserving computing in big data era. *IEEE Network*, 28(4):46–50, July 2014.

[62] K. G. Shin, X. Ju, Z. Chen, and X. Hu. Privacy protection for users of location-based services. *IEEE Wireless Communications*, 19(1):30–39, February 2012.

[63] Maria Luisa Damiani. *European Data Protection: Coming of Age*, chapter Privacy Enhancing Techniques for the Protection of Mobility Patterns in LBS: Research Issues and Trends, pages 223–239. Springer Netherlands, Dordrecht, 2013.

[64] Nikos Pelekis and Yannis Theodoridis. *Mobility Data Management and Exploration*, chapter Privacy-Aware Mobility Data Exploration, pages 169–185. Springer New York, New York, NY, 2014.

[65] Marius Wernke, Pavel Skvortsov, Frank Dürr, and Kurt Rothermel. A classification of location privacy attacks and approaches. *Personal Ubiquitous Comput.*, 18(1):163–175, January 2014.

[66] B. Niu, X. Zhu, W. Li, H. Li, Y. Wang, and Z. Lu. A personalized two-tier cloaking scheme for privacy-aware location-based services. In *Computing, Networking and Communications (ICNC), 2015 International Conference on*, pages 94–98, Feb 2015.

[67] Reza Shokri, George Theodorakopoulos, Carmela Troncoso, Jean-Pierre Hubaux, and Jean-Yves Le Boudec. Protecting location privacy: Optimal strategy against localization attacks. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, CCS '12, pages 617–627, New York, NY, USA, 2012. ACM.

[68] V. Primault, S. B. Mokhtar, and L. Brunie. Privacy-preserving publication of mobility data with high utility. In *Distributed Computing Systems (ICDCS), 2015 IEEE 35th International Conference on*, pages 802–803, June 2015.

www.manaraa.com

[69] Vincent Primault, Sonia Ben Mokhtar, Cédric Lauradoux, and Lionel Brunie. Differentially private location privacy in practice. *CoRR*, abs/1410.7744, 2014.

[70] S. Sicaria, A. Rizzardia, L.A. Griecob, and A. Coen-Porisinia. Security, privacy and trust in internet of things: The road ahead. *Computer Networks*, 76, 2015.

[71] M. Abomhara and G. M. KÃ¸ien. Security and privacy in the internet of things: Current status and open issues. In *Privacy and Security in Mobile Systems (PRISMS), 2014 International Conference on*, pages 1–8, May 2014.

[72] G. A. Akpakwu, B. J. Silva, G. P. Hancke, and A. M. Abu-Mahfouz. A survey on 5g networks for the internet of things: Communication technologies and challenges. *IEEE Access*, 6:3619–3647, 2018.

[73] Xiruo Liu. *Integrating security and privacy protection into a mobility-centric internet architecture*. PhD thesis, Rutgers, The State University of New Jersey, May 2016.

[74] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu. Edge computing: Vision and challenges. *IEEE Internet of Things Journal*, 3(5):637–646, Oct 2016.

[75] T. X. Tran, A. Hajisami, P. Pandey, and D. Pompili. Collaborative mobile edge computing in 5g networks: New paradigms, scenarios, and challenges. *IEEE Communications Magazine*, 55(4):54–61, April 2017.

[76] S. Nunna, A. Kousaridas, M. Ibrahim, M. Dillinger, C. Thuemmler, H. Feussner, and A. Schneider. Enabling real-time context-aware collaboration through 5g and mobile edge computing. In *2015 12th International Conference on Information Technology - New Generations*, pages 601–605, April 2015.

[77] Tuyen X. Tran and Dario Pompili. Joint task offloading and resource allocation for multi-server mobile-edge computing networks. *CoRR*, abs/1705.00704, 2017.

[78] Rodrigo Roman, Javier Lopez, and Masahiro Mambo. Mobile edge computing, fog et al.: A survey and analysis of security threats and challenges. *Future Generation Computer Systems*, 78(Part 2):680 – 698, 2018.

[79] Jayavardhana Gubbi, Rajkumar Buyya, Slaven Marusic, and Marimuthu Palaniswami. Internet of things (iot): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7):1645 – 1660, 2013.

[80] S. D. T. Kelly, N. K. Suryadevara, and S. C. Mukhopadhyay. Towards the implementation of iot for environmental condition monitoring in homes. *IEEE Sensors Journal*, 13(10):3846–3853, Oct 2013.

[81] F. Tao, Y. Zuo, L. D. Xu, and L. Zhang. Iot-based intelligent perception and access of manufacturing resource toward cloud manufacturing. *IEEE Transactions on Industrial Informatics*, 10(2):1547–1557, May 2014.

[82] Luis Sanchez, Luis Munoz, Jose Antonio Galache, Pablo Sotres, Juan R. Santana, Veronica Gutierrez, Rajiv Ramdhany, Alex Gluhak, Srdjan Krco, Evangelos Theodoridis, and Dennis Pfisterer. Smartsantander: Iot experimentation over a smart city testbed. *Computer Networks*, 61(Supplement C):217 – 238, 2014. Special issue on Future Internet Testbeds - Part I.

www.manaraa.com

[83] A. McEwen and H. Cassimally. *Designing the Internet of Things*. Wiley, 2013.

[84] K. Zhao and L. Ge. A survey on the internet of things security. In *2013 Ninth International Conference on Computational Intelligence and Security*, pages 663–667, Dec 2013.

[85] Z. K. Zhang, M. C. Y. Cho, C. W. Wang, C. W. Hsu, C. K. Chen, and S. Shieh. Iot security: Ongoing challenges and research opportunities. In *2014 IEEE 7th International Conference on Service-Oriented Computing and Applications*, pages 230–234, Nov 2014.

[86] M. Sain, Y. J. Kang, and H. J. Lee. Survey on security in internet of things: State of the art and challenges. In *2017 19th International Conference on Advanced Communication Technology (ICACT)*, pages 699–704, Feb 2017.

[87] P.N. Howard. *Pax Technica: How the Internet of Things May Set Us Free Or Lock Us Up*. Yale University Press, 2015.

[88] Noura Aleisa and Karen Renaud. Privacy of the internet of things: A systematic literature review. In *Hawaii International Conference on System Sciences*, 01 2017.

[89] Sushant Jain, Alok Kumar, Subhasree Mandal, Joon Ong, Leon Poutievski, Arjun Singh, Subbaiah Venkata, Jim Wanderer, Junlan Zhou, Min Zhu, Jon Zolla, Urs Hölzle, Stephen Stuart, and Amin Vahdat. B4: Experience with a globally-deployed software defined wan. In *Proceedings of the ACM SIGCOMM 2013 Conference on SIGCOMM*, SIGCOMM '13, pages 3–14, New York, NY, USA, 2013. ACM.

[90] Chi-Yao Hong, Srikanth Kandula, Ratul Mahajan, Ming Zhang, Vijay Gill, Mohan Nanduri, and Roger Wattenhofer. Achieving high utilization with software-driven wan. In *Proceedings of the ACM SIGCOMM 2013 Conference on SIGCOMM*, SIGCOMM '13, pages 15–26, New York, NY, USA, 2013. ACM.

[91] Teemu Koponen, Martin Casado, Natasha Gude, Jeremy Stribling, Leon Poutievski, Min Zhu, Rajiv Ramanathan, Yuichiro Iwata, Hiroaki Inoue, Takayuki Hama, and Scott Shenker. Onix: A distributed control platform for large-scale production networks. In *Proceedings of the 9th USENIX Conference on Operating Systems Design and Implementation*, OSDI'10, pages 351–364, Berkeley, CA, USA, 2010. USENIX Association.

[92] Pankaj Berde, Matteo Gerola, Jonathan Hart, Yuta Higuchi, Masayoshi Kobayashi, Toshio Koide, Bob Lantz, Brian O'Connor, Pavlin Radoslavov, William Snow, and Guru Parulkar. Onos: Towards an open, distributed sdn os. In *Proceedings of the Third Workshop on Hot Topics in Software Defined Networking*, HotSDN '14, pages 1–6, New York, NY, USA, 2014. ACM.

[93] Anja Feldmann. Internet clean-slate design: what and why? *SIGCOMM Comput. Commun. Rev.*, 37(3):59–64, 2007.

[94] Constantine Dovrolis. What would darwin think about clean-slate architectures? *SIGCOMM Comput. Commun. Rev.*, 38(1):29–34, 2008.

www.manaraa.com

[95] Tao Jiang, Helen J. Wang, and Yih-Chun Hu. Preserving location privacy in wireless lans. In *MobiSys '07: Proceedings of the 5th international conference on Mobile systems, applications and services*, pages 246–257, New York, NY, USA, 2007. ACM.

[96] Reza Shokri, Julien Freudiger, Murtuza Jadliwala, and Jean-Pierre Hubaux. A distortion-based metric for location privacy. In *WPES '09: Proceedings of the 8th ACM workshop on Privacy in the electronic society*, pages 21–30, New York, NY, USA, 2009. ACM.

[97] Sebastian Kay Belle, Marcel Waldvogel, and Oliver Haase. Pathforge: faithful anonymization of movement data. In *MobiHeld '09: Proceedings of the 1st ACM workshop on Networking, systems, and applications for mobile handhelds*, pages 63–64, New York, NY, USA, 2009. ACM.

[98] Claudio Agostino Ardagna, Marco Cremonini, and Gabriele Gianini. Landscape-aware location-privacy protection in location-based services. *J. Syst. Archit.*, 55(4):243–254, 2009.

[99] Ping Zhang, Arjan Durresi, and Raj Jain. Economically viable support for internet mobility. In *Proceedings of IEEE International Conference on Communications, ICC 2011, Kyoto, Japan, 5-9 June, 2011*, pages 1–6, 2011.

[100] Ping Zhang, Arjan Durresi, and Raj Jain. Cloud aided internet mobility. In *Proceedings of IEEE International Conference on Communications, ICC 2013, Budapest, Hungary, June 9-13, 2013*, pages 3688–3693, 2013.

[101] Ping Zhang, Arjan Durresi, and Leonard Barolli. Policy-based mobility in heterogeneous networks. *J. Ambient Intelligence and Humanized Computing*, 4(3):331–338, 2013.

[102] Ping Zhang and Arjan Durresi. Cloud aided internet mobility for privacy protection. In *IEEE International Conference on Communications, ICC 2017, Paris, France, May 21-25, 2017*, pages 1–6, 2017.

[103] Ping Zhang, Mimoza Durresi, and Arjan Durresi. Enhanced internet mobility and privacy using public cloud. *Mobile Information Systems*, 2017:4725858:1–4725858:11, 2017.

[104] Ping Zhang, Mimoza Durresi, and Arjan Durresi. Mobile privacy protection enhanced with multi-access edge computing. In *32nd IEEE International Conference on Advanced Information Networking and Applications, AINA 2018, Krakow, Poland, May 16-18, 2018*, pages 724–731, 2018.

[105] J. Laganier, T. Koponen, and L. Eggert. Host Identity Protocol (HIP) Registration Extension. RFC 5203 (Experimental), April 2008.

[106] R. Shokri, G. Theodorakopoulos, J. Y. Le Boudec, and J. P. Hubaux. Quantifying location privacy. In *Security and Privacy (SP), 2011 IEEE Symposium on*, pages 247–262, May 2011.

[107] Miguel E. Andrés, Nicolás E. Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. Geo-indistinguishability: Differential privacy for location-based systems. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer; Communications Security*, CCS '13, pages 901–914, New York, NY, USA, 2013. ACM.

www.manaraa.com

[108] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil. Proxy mobile ipv6. RFC 5213, RFC Editor, August 2008.

[109] Ouri Wolfson, Sushil Jajodia, and Yixiu Huang. An adaptive data replication algorithm. *ACM Transactions on Database Systems*, 22:255–314, 1997.

[110] Karen Q. Tian and Donald C. Cox. *Mobility Management In Wireless Network: Data replication strategies and applications*. Kluwer Academic Publishers, Boston, December 2004.

[111] Jared Winick and Sugih Jamin. Inet-3.0: Internet topology generator. Technical report, Department of Electrical Engineering and Computer Science, University of Michigan, 2002.

www.manaraa.com

VITA

www.manaraa.com

VITA

Ping Zhang is a PhD student in the Department of Computer & Information Science at Indiana University Purdue University in Indianapolis. His research interests include networking, distributed system, and operating system. He has conducted researches in network routing, Internet mobility, privacy protection, and distributed trust management, and published papers in several conferences and journals. Ping received his B.S. from Beijing University of Technology in Computer Science in 2003, and worked as developer at AWS from 2011 building and helping shape the rising of Cloud Computing. He currently works at OCI to build the next generation of public Cloud. In his spare time he enjoys hiking, cooking, and reading.

www.manaraa.com