Unified Heterogeneous Networking Design

Amandeep Singh, Gaston Ormazabal, Henning Schulzrinne
Columbia University
New York, USA
{aman, gso, hgs}@cs.columbia.edu

Yan Zou, Peter Thermos
Columbia University
New York, USA
{yz2437, pt81}@columbia.edu

Sateesh Addepalli
Cisco Systems, Inc.
San Jose, USA
sateeshk@cisco.com

ABSTRACT

The Internet was designed under the assumption that end-hosts are stationary and have one interface. Current mobile devices have multiple network interfaces, such as Wi-Fi, LTE, WiMAX, and possibly Ethernet. Such diverse network connectivity can be used to increase both reliability and performance by running applications over multiple links sequentially, for a seamless user experience, or in parallel, for bandwidth and performance enhancements. Users are also consuming Internet services from multiple locations and devices, such as smartphones, tablets, laptops, and IP-enabled TVs. The existing networking stack, however, offers almost no support for intelligently exploiting such network, location and device diversity.

Since, most Internet devices today are mobile, we propose a unified networking architecture that makes optimal use of a heterogeneous dynamic environment, both in terms of networks and user devices. The system core functionalities include mobility, multi-homing, multipath, and disruption tolerance. The system enables mobile nodes to make decisions about *how* and *when* to use each or a combination of networks, in a secure manner. With this new architecture, we envision a shift from current applications supporting a single network, location, and device at a time, to applications that can support multiple networks, multiple locations, and multiple devices.

Categories and Subject Descriptors

D.2.11 [**Software Engineering**]: Software Architectures; C.2.2 [**Computer Communication Networks**]: Network Protocols

General Terms

Design, Security, Management

Keywords

Heterogeneous access, Mobility, Policy management, Identity Management, Ubiquitous computing

1. INTRODUCTION

Packet-based communications systems were originally designed without considering the concept of mobility. Mobile Internet bandwidth consumption will soon surpass fixed line broadband [1]. Moreover, mobile users have developed very high expectations of their communication needs regardless of where they are located, how they are moving, what devices they are using, and which applications they are running. Yet current networks are generally not designed to handle such expectations of mobility, which call for seamless network transitions, and session mobility persistence, as the user moves about even while changing devices.

The architecture proposed in this work is based on a user-centric, unified network design, which makes better use of heterogeneous networking environments, while providing transparent control to access networks and devices [2]. Heterogeneous networks can be different network technologies or multiple administrative domains. The architecture introduces an enhanced networking stack comprised of additional control functions for each networking layer, managed by a policy-based decision-making system that provides a seamless user experience. The network control function provides terminal mobility, multi-homing and multipath functionalities. The link control function provides network discovery and handover optimizations functionalities. And the physical control function provides control of radio spectrum for optimal physical channel selection. Additionally, a control middleware module manages these control functions, and provides granular network access control, based upon user-defined policies, having various criteria, such as, cost, bandwidth, location and security, on a per-application basis. Moreover, the service providing network access is abstracted into two logical planes - infrastructure and service, by decoupling the physical infrastructure from the service-providing functionalities. The service plane provides mobility management, additional network information for optimal connectivity, service access control, and unified billing and accounting for all nodes. The infrastructure plane provides the actual physical network access.

In this work, the architecture design and a prototype implementation is described, including a proposed security framework. As the architecture supports a heterogeneous networking environment, the proposed design attempts to unify the various authentication schemes, resulting in reduced credential management complexity.

The remainder of the paper is structured as follows: Section 2 describes the design goals of the architecture. Section 3 describes mobility and the security challenges associated with it. Section 4 discusses an overview of the related work. Section 5 describes the architecture design. Section 6 describes secure heterogeneous access using the proposed design. Section 7 describes the implementation and the protocols used. The conclusions are presented in Section 8.

2. DESIGN GOALS

As all voice, multimedia, and data services are converging to IP, a generalized approach is required to understand mobility communication requirements impacting application users, developers, and network providers. The proposed architecture design goals include:

1) Seamless user experience: Users should not experience any degradation of service when a change in network point of attachment, network type, service provider or device occurs. Real-time applications, such as an active voice or video call, should not experience any disconnections or delays upon handover, either from one network service to another (e.g., LTE to Wi-Fi), or upon changing devices (e.g., from a mobile phone to a laptop).

2) User and mobile node independence: Users and their mobile nodes (MN) should be decoupled from physical network infrastructure. Networks are heterogeneous in terms of access technologies (Wi-Fi, LTE, WiMAX, satellite), service providers, performance and cost. Users can access Internet applications from many locations (home, office, outdoors), devices (PC, smartphone, tablet) and from more than one network at a time.

3) Policy management: Users should be able to control their network access. Application developers should be able to select either static (pre-configured) or dynamically configured network access policies. For example, a user may decide to download OS upgrades, or upload videos, when network bandwidth is plentiful or cheap. Conversely, voice or video services, or interactive games, may require low-latency mobile connections.

4) Multi-device support: MNs should support multi-device communication sessions. MNs should be able to discover neighboring nodes, and their corresponding shared services, such as, audio, video (camera, display), and storage, in a secure manner, to enable ubiquitous computing for a next-generation multimedia experience. For example, a user can transfer a video feed from a mobile device to a fixed TV, while maintaining security associations, and application context.

5) Secure communication: There should be a well-defined security model for connection (resource access), context (protocols), and content (data) abstractions without introducing additional latency. For example, cryptographic credentials can be reused to reduce redundant operations.

6) Disruption tolerance: Applications should continue to operate in the absence of network connectivity both short and long term, to provide resiliency and better user experience. For example, a file transfer to and from a cloud-based storage service can be controlled by user-defined policies that take into account network conditions and priority, delaying packet delivery until a suitable network becomes available.

7) Network intelligence: MNs should leverage network-based resources seamlessly, for making better network selection decisions, and storing data packets for future deliveries. These resources may provide additional computation and storage, helping MNs to reduce energy consumption. For example, a service can store a user's daily route geographic map of nearby networks, and based upon usage-pattern analysis, the service can delay data transfers, until either a high bandwidth, or a low cost network, becomes available.

8) Backward compatibility: Existing legacy applications should work without any modifications. The standard networking socket API should be adhered to, since it is the most programing interface.

The overall objective of the proposed architecture is to improve the quality and continuity of the mobile user experience. With these comprehensive goals, a shift from the traditional network-centric to a user-centric approach, in network architectures, is implied. In this architecture, users and applications are able to determine what kind of network connectivity they want to use in terms of availability, resource consumption, bandwidth, cost, and QoS.

3. SECURITY AND MOBILITY

Mobile communications are vulnerable to interception and analysis by third parties. Since the proposed architecture involves managing connections across heterogeneous networks, the most significant challenge is maintaining identity, confidentiality and integrity in a non-uniform security environment. For example, a device may transition from a secure LTE network to a public Wi-Fi access point that may not support adequate protection mechanisms, and consequently render communications vulnerable to attacks. The identification, categorization, and prioritization of applicable network threats helps define and enforce correspondingly uniform security controls to mitigate them.

3.1 Security Threats

General security threats can be commonly described under the following categories: unauthorized access; eavesdropping and traffic analysis; and service disruption and denial of service (DoS). Unauthorized access occurs when an attacker, using vulnerabilities in the communication protocol or user applications, gains access to restricted information (e.g., passwords, confidential files), and takes control of the system. Eavesdropping or network sniffing is a network attack consisting of capturing data packets transmitted by other nodes, and analyzing data content for access to sensitive information, such as session tokens, or other kind of confidential information. DoS attacks aim to disrupt communication services, affecting a single node, a collection of nodes or the entire network. These attacks take advantage of vulnerabilities in the protocol implementation, through protocol manipulation, or by flooding the application (or service) with bogus requests, such as reflection and amplification, resulting in blocking and replaying.

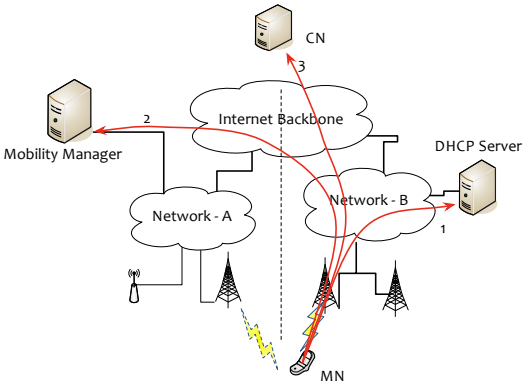

Figure 1: Global Mobility and Security

3.2 Mobility

The maintenance of transport layer connectivity is a significant requirement of mobility management, resulting in a location/identity split, where an invariant identifier is mapped to the changing locations (IP addresses). A MN experiences *local mobility* when moving within a single administrative domain, thus the network maintains the IP address unchanged across access technologies, because mobility is managed internally within the domain. *Global mobility* is defined when a MN moves across administrative domains, resulting in new IP address acquisition, in each domain. In order to maintain MN reachability, new location updates need to be sent to an *external* Mobility Manager (MM) service that can act as a location registrar, and a common rendezvous point, when both nodes are moving concurrently. In *opportunistic* mode, however, a MN can

send new location updates directly to a stationary Correspondent Node (CN), eliminating the need for MM support.

In local mobility, the MM is managed *internally* by a single provider, thus uniform and consistent security controls can be implemented to prevent attacks. Whereas in global mobility, the typically external MM can be subjected to the attacks described above, originating from networks that do not necessarily have common security guidelines. In a global mobility event, there are three channels vulnerable to these security attacks: network access for new IP address acquisition, new location update message forwarding to MM, and/or Route Optimization (RO) message forwarding to all active CNs, as shown in Fig. 1.

4. RELATED WORK

There have been several proposals that attempt to provide mobility, multi-homing, multipath, and flow management functionalities by introducing "shim sub-layer" based solutions, and new APIs to provide uniform abstractions to upper layers. Legacy host support has been typically provided via an external network proxy module, either natively, or by protocol extensions. The solutions can be categorized as either end-host or end-site, based upon where the enhancements in the networking stack take place. End-host solutions require changes to the *host's* networking stack, while usually not requiring any network infrastructure changes. End-site solutions require changes at the site's *exit routers* without any changes to the end-hosts. End-site solutions, in general, require additional protocol-specific network elements for routing support.

In the end-host category, at the network shim sub-layer, solutions providing mobility, multi-homing and multipath support were evaluated. The Multipath Transmission Control Protocol (MPTCP) [3] provides multipath support using the resource-pooling principle, and opportunistic terminal mobility. MPTCP supports existing applications without any changes. The Stream Control Transmission Protocol (SCTP) [4] provides multi-homing and multi-streaming functionality natively, while mobility support can be added using extensions [5]. Existing applications require modifications, and the limited firewall support makes SCTP deployment difficult. The Host Identity Protocol (HIP) [6] provides multi-homing, terminal and network mobility, and multipath support [7]. The location/identifier split required for mobility is implemented using the Host Identity Tag (HIT)[1], which acts as a constant identifier. In addition, it also provides native end-to-end security. The Mobile IPv6 (MIPv6) [8] protocol enables mobility support in IPv6 by maintaining a permanent IP address (home address), with optional security provided by IPsec. The multiple care-of-address registration extension [9] adds multi-homing support. For legacy hosts' support, a network-based mobility protocol called Proxy MIPv6 [10] has been also proposed. The End-to-end Connection Control Protocol (ECCP) [11] provides multipath and multi-homing support by abstracting the transport layer into data-delivery and connection control sub-layers. ECCP can be seen as a hybrid of HIP and MIPv6 protocols. At the link shim sub-layer, heterogeneous network discovery process optimization and handover-delay reduction solutions were evaluated. The IEEE 802.21 Media Independent Handoff (MIH) [12] framework introduces the MIH Function (MIHF) as a network access abstraction, along with interface-specific standard control functions called Service Access Points (SAP). The handover-delay can be further optimized using the MIH network information service (NIS), with additional network attributes, such as bandwidth, latency and cost. The Access Network Discovery and Selection Function (ANDSF) [13] is a 3rd Generation Partnership Project (3GPP) standard, with similar functionalities as the MIH framework. It defines LTE Evolved Packet Core connectivity for a non-3GPP (Wi-Fi, WiMAX) access interface. It enables seamless vertical handovers, and allows operators to provide MNs a list of preferred networks and corresponding access policies. Dynamic Spectrum Access (DSA) [14] physical shim sub-layer solutions were also evaluated, to make use of potentially additional location-based wireless spectrum (e.g., white spaces[2]). DSA solutions enable access to additional spectrum by using cognitive radio methods [14] opportunistically, or by querying a network based geospatial spectrum database [15].

In the end-site category, at the network shim sub-layer, the Network Prefix Translation (NPTv6/NAT66) [16] mechanism uses an address-rewriting procedure to support multi-homing natively, for IPv6 enabled hosts. At the link shim sub-layer, the Location Identification Separation Protocol (LISP) [17] provides multi-homing, using link layer encapsulation, and traffic engineering support, using flow priority and weight attributes. LISP requires additional network element support, namely Ingress and Egress Tunnel Routers for traffic encapsulation and routing, resulting in a separate mapping overlay network [18]. Furthermore, for mobility support, LISP requires Tunnel Router functionalities to be also implemented at the end-hosts, providing an exception to the expectation that end-site solutions do not generally require changes to end-hosts.

Based on our evaluation, the requirements to provide mobility, multi-homing, and multipath can be best fulfilled at the end-host, implemented in the network shim sub-layer. Also, the link shim sub-layer may enable common functions for network discovery and selection, and for reducing handover delay, in heterogeneous networking environments.

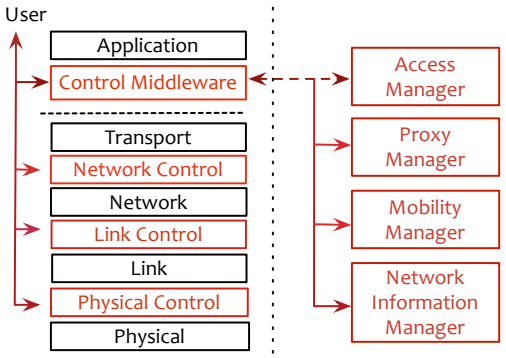

Figure 2: Architecture Design

5. ARCHITECTURE DESIGN

In the previous section, several proposals for improving the current Internet architecture at the transport, network, link, and physical layers of the IP protocol suite were discussed. Using similar design considerations, this architecture design proposes an enhanced networking stack that supports terminal mobility, multi-homing, multipath, and disruption tolerance, as core functionalities. Each networking layer is enhanced with a corresponding control function, managed by a policy-based control middleware (CM). The control functions transfer information to upper layers via event propagation mechanisms, while the CM acts as a decision-making engine by processing control events, as shown in Fig. 2. The CM decisions result in a dynamic change in the system default behavior, by supplementing or complementing an existing flow with possible parallel connections. The CM itself is independent of any

[1] 128-bit cryptographic hash of the public key.

[2] Frequencies allocated to a broadcasting service but not used locally.

transport, network, and link layer protocols. Additionally, a new network control event feedback API has been designed to provide network flow control to applications.

The proposed architecture is further divided into two independent logical systems: the networking stack at the end-host level, and a suite of network-based services that assist the networking stack with external support for mobility and network information. As such, the system operates in two modes: *assisted* and *opportunistic*. In assisted mode, the networking stack control functions leverage external mobility and network information entities, as shown in Fig. 2. In opportunistic mode, however, network discovery can be performed locally, and the new location updates can be sent directly to a stationary CN, rendering the additional entities required for mobility and network information unnecessary.

The architecture is an end-to-end solution that requires both communicating nodes to support identical networking stacks. Legacy systems backward compatibility support is proposed using a network-based proxy manager (PM) service that translates native packets (CM-aware) to static IP packets.

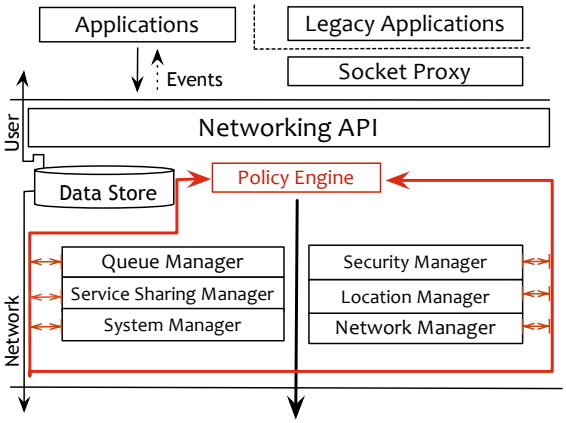

Figure 3: Control Middleware

5.1 Networking Stack

Each control function manages the corresponding networking layer protocol. They are designed to be independent of the underlying protocols, and any new present or future protocol can be integrated, without changing the upper layers. The Physical Control Function (PCF) provides DSA functionality to enable access to additional wireless spectrum, based upon location, network conditions, and application-specific requirements. The Link Control Function (LCF) decouples network and link layer functionality, and provides secure network access using interface specific authentication procedures. For example, in case of LTE, the LCF uses the Extensible Authentication Protocol – Authentication and Key Agreement (EAP-AKA') [19] mechanism. It also provides handover delay optimization capabilities, such as link status (up/down), link control, and monitoring. The Network Control Function (NCF) provides mobility, multi-homing, multipath, and adaptive flow management. The NCF provides to the transport layer a constant flow identifier that is mapped to the changing IP addresses, for each active flow, without causing any transport disconnections. The provision of NCF functionalities at the network shim sub-layer results in a system wide effect, allowing all applications to work seamlessly and uninterruptedly. The NCF provides optional end-to-end security with a symmetric shared secret, derived from an asymmetric cryptographic step, for the mutual authentication at the beginning of a session. This step requires a certificate-based public/private key-pair. The certificate can be either locally generated and self-signed, or issued by a Certificate Authority (CA), which is always preferable, to prevent potential man-in-the-middle attacks.

The CM is comprised of a policy engine (PE) that makes dynamic decisions, based upon control event inputs from various attribute managers, as shown in Fig. 3. The PE evaluates a state-vector of these current control events against pre-defined policies, resulting in a modification of system behavior. The state-vector defines a *context* at any given time, for example, location, time of day, and network type, cost, bandwidth, and latency. The CM's attribute managers include: The *network manager* (NM) that maintains and monitors all active network interfaces information. The NM also provides network information to the PE and executes the network handover decisions. The *security manager* (SM) maintains networks and devices access credentials, and end-to-end communications public/private key pairs. The *location manager* (LM) provides MN location information to the PE based upon GPS coordinates, or indoor positioning parameters, such as, Wi-Fi network identifier. The *service-sharing manager* (SSM) provides a centralized service registration function for local network discovery. The *system manager* maintains system parameters such as CPU, bandwidth and battery usage, to enable application-specific usage constraints, such as maximum bandwidth limit or battery utilization. And the *queue manager* maintains application specific queues to store data packets when there is no network connectivity, or a policy enforces no network usage, enabling application disruption tolerance support. The *data store* (DS) provides a structured key-value repository for each attribute manager respectively. Additionally, the *socket proxy* provides legacy application support by intercepting socket system calls.

5.1.1 Policy Management

Policies can be defined by the user, application or network. The policy control scope can be defined at three levels: global (system-wide), group (set of applications) or local (application-specific), which allows granular network access control for a given application. For example, a user can specify application updates delivery only when the location is home (Wi-Fi), or the network cost is free, and time is after midnight, with global scope. Conversely, a streaming application can be configured to always access one or higher bandwidth networks, irrespective of cost, but with only local scope. The policies are defined according to a semantically configurable policy-specification language [20], and stored locally in the data store. At policy evaluation time, the attribute managers provide their control event inputs to the PE for processing, and network control enforcement. The PE maintains a mapping of all active socket connections, corresponding to each application, and performs decisions based upon new control event states, e.g., new network, new location, new device, time-of-day, and system parameter(s) status. There are three possible policy evaluation outcomes: network interface addition, for a new connection (or update of an existing connection), packet storage in application specific queues (in the queue manager) for later delivery if no network interface is selected, or new connection blocking, if no suitable network interface is available. A CM-enabled MN can always connect to the best available network, based upon user policies, resulting in optimal seamless handovers without disrupting active applications, illustrating user-centric networking. For example, a MN may transitions from a LTE to a Wi-Fi network based on best bandwidth, latency, signal strength, and location attributes.

5.2 Network Services

The networking stack may be assisted with network-based external services to provide reliable mobility and additional network information support. The MM service, acting as an independent

centralized location registrar, assists the NCF to provide node reachability, while the node moves across domains, and acts as a rendezvous point, when both communicating nodes are moving concurrently. The Network Information Manager (NIM) service assists the LCF with additional network information, such as nearby networks presence, bandwidth, cost, latency, and spectrum information for optimal network discovery and selection. The NIM service helps reduce handover-delay and battery energy consumption. It may also use analytics to provide even more refined network selection capabilities [21]. The network-based PM service enables incremental deployment of the system by converting the enhanced networking stack packets into the traditional stack. These services provide common functionalities across heterogeneous administrative domains, organically defining a separate logical service plane.

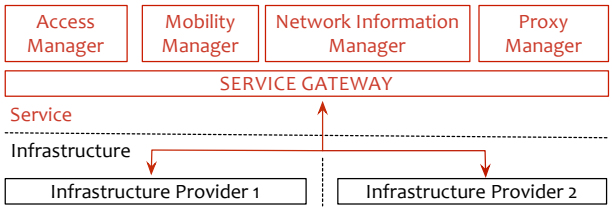

Figure 4: Service and Infrastructure Plane

The traditional service provider user management services such as identity, billing and accounting are generally replicated across service providers. These services can be abstracted into a single service management logical entity that can be also deployed in the common service plane described above. This results in service providers that are responsible for only technology-specific infrastructure access, while all the service functionality can be deployed in the service plane. This decoupling of the service functionality from the infrastructure access, as shown in Fig. 4, allows a service provider to offer services across multiple infrastructure providers, resulting in better management of heterogeneity and service scalability. Moreover, a single infrastructure provider can host multiple service providers, thus maximizing the infrastructure resource utilization, using for example, software defined network mechanisms [22, 23].

The new common service management entity, called the Access Manager (AM), may perform identity management and access control (authentication and authorization) for various administrative domains, comprised of heterogeneous network technologies. The AM provides a federated identity management[3] [24] mechanism, using the Extensible Authentication Protocol (EAP) framework [25]. Additionally, a Service Gateway (SG) function may provide an application layer firewall to further secure the interface between the service and the infrastructure plane, to prevent and mitigate the security threats discussed in Section 4. These service plane components are logical functions that can be deployed in a cloud environment. In this design, the service plane provides a *control-plane* abstraction, while the infrastructure plane provides a *data-plane* abstraction.

6. HETEROGENEOUS ACCESS

In this section, using the previously described architecture components, we illustrate how a MN gains secure network access, and how they can also be leveraged to enable secure device-to-device communication in local networks. Secure access is a two-step procedure involving discovery and trust establishment. The discovery step involves the infrastructure plane for location of the appropriate authentication service. The trust establishment step involves the service plane for credential verification (authentication) and assignment of corresponding authorization level (access control).

6.1 Network Access

When a MN moves across infrastructure domains, the network access step has to be performed for every domain. Heterogeneous network access typically requires management of multiple access credentials. For members of a service federation, using EAP-based mechanisms, however, network access methods can reuse the same pre-shared keys, as the service providers have prior agreements with the infrastructure providers[4]. For example, using the EAP-AKA' [19] mechanism, the same pre-shared key of LTE can be also used for Wi-Fi network authentication. For visited networks outside of the service federation (e.g., a home network), the network access methods may still use EAP-based mechanisms (e.g., EAP-TLS), while the credentials can be stored and managed in the SM. The CM, leveraging LCF, enables the logical separation of the network authentication process into service and infrastructure access replicating the EAP pre-shared secret mechanism abstractions [19].

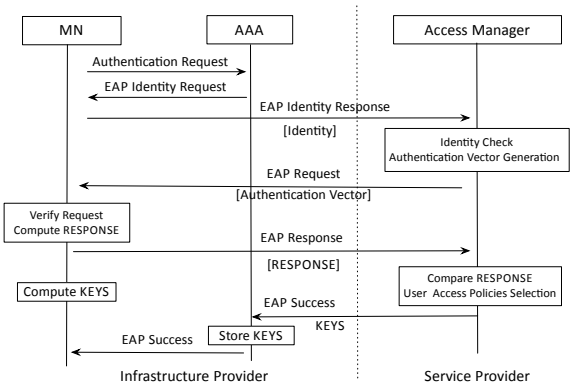

Figure 5: Service Access

6.1.1 Service Access

In the first step, the CM initiates an authentication request using a LCF interface-specific access function. The access function presents the MN's pseudo-identity[5] to the infrastructure provider, over a temporary access channel. The infrastructure provider's Authentication Authorization Accounting (AAA) service then maps the pseudo-identity to the corresponding MN's service provider. Upon service provider resolution, the AAA service initiates the EAP authentication procedure between the MN and its service provider. The LCF performs the authentication and authorization procedure with the service provider, using the symmetric pre-shared keys configured initially in the MN and AM respectively, as shown in Fig. 5. Upon successful authentication, the AM service returns a validated master session key, derived from the initial pre-shared keys, along with optional network access policies, to the infrastructure provider's AAA server, so that the MN can initiate the infrastructure access procedures.

[3] Federation of identity enables the portability of identity information across otherwise autonomous security domains permitting users of one domain to securely access data or systems of another domain seamlessly, and without the need for completely redundant user administration.

[4] The communication between the service and infrastructure providers is assumed to be secure, as there is an *a priori* trust relationship established between them.

[5] The pseudo-identity string is used to hide the MN's real identity from the infrastructure providers during service access, thus enabling additional privacy.

6.1.2 Infrastructure Access

The MN can initiate, using the validated master session key generated in the above procedure, the physical infrastructure technology-specific access procedures. A temporary security context is created between the MN and the network point of attachment using this key. After successful infrastructure authentication, the MN can initiate a DHCP request for an IP address.

Subsequently, in assisted mode, the MN needs to also establish secure channels for each control function, with the corresponding service plane components; specifically the LCF with the NIM, and the NCF with the MM respectively. Furthermore, the public/private key pair used in the asymmetric step, can be also used to generate shared secrets, for securing end-to-end communication channels using any of the well-known symmetric key-based security protocols (e.g., Encapsulating Security Payload (ESP) [26], TLS, or application specific security).

6.2 Device Access

The proposed enhanced networking stack can also be leveraged to enable secure device-to-device service sharing in local networks. It provides a general platform for application service sharing using the publish/subscribe mechanism. Similar to network access, device access can be abstracted into two logical planes – *interface* and *service*. The interface access abstraction provides technology specific authentication for a device, by providing device *discovery* and *trust establishment* between devices. The device identity can be locally generated, from stored credentials using link layer access technology (e.g. MAC address), or pre-assigned by a third-party. On the other hand, the service plane abstraction provides service registration, discovery, and authorization functionalities, to the locally running services, using the SSM in the CM as described in Section 5.1

The NM coordinates with the LCF to obtain a temporary device access channel, for performing authentication. Upon successful authentication, the PE can signal to all the locally registered running services the new device availability status, and may enforce application-specific device policies. Subsequently, applications can start using the newly discovered local network services. For example, users can transfer a video stream from a smartphone to an IP-enabled TV if the streaming services are compatible.

7. IMPLEMENTATION

The enhanced networking stack prototype was implemented using a Linux kernel (version 3.2.21). The LCF used the ODTONE open implementation of the IEEE 802.21 MIH framework [28], with the corresponding NIS service, which was deployed as the NIM entity. The NCF was implemented using a dual-stack configuration of the HIP [29], and MIPv6 [30] protocols and their corresponding MM support implementations. In the HIP case, the MM is called Rendezvous Server (RVS), while in the MIPv6 case it is called Home Agent (HA). The dual-stack configuration enables a gradual deployment for HIP, while providing backward compatibility with IPv6 hosts. The NCF provides end-to-end security using the ESP [26] protocol[6]. In the HIP case, a secure channel is established using the shared secret derived from the native Base Exchange (BEX)[7] mechanism, while in the MIPv6 case, the IPsec Internet Key Exchange (IKEv2) protocol is used.

[6] Security is natively implemented in HIP, while in MIPv6 it is optional.

[7] The BEX mechanism is a four-way handshake that takes place between two communicating nodes to establish a shared secret using Diffie-Hellman mechanism.

The CM including the PE and the attribute managers were implemented in user-space, using an event-driven architecture. The socket proxy was implemented using a SOCKSv5 open implementation [31]. The PM support was implemented using the SOCKSv5 relay mechanism.

In the current prototype, only the NM, SM, and LM attribute managers have been implemented. The LCF and NCF user-space protocol controllers were modified to communicate with the NM directly. The NM maintains and monitors network interfaces using the LCF user-space controller. A Trusted Access Service (TAS) table is maintained and dynamically updated by a service provider and/or user. The TAS table stores the following access service (network or device) information: service-identifier, authentication procedure, service type, credentials, cost, bandwidth, and access history. The SM maintains an Access Credential table with the following information: service-identifier, type, credential, name and additional information (location). The LM was implemented using a command line interface to masquerade as the input from a GPS-based device.

To initiate a new connection, the CM selects the NCF protocol based upon the CN's destination address (HIT, IPv6, or IPv4 address). As a default, and when both communicating nodes support HIP, a flow is always configured with HIP. Conversely, when the destination address is only IPv6 enabled, the flow is configured with the MIPv6 protocol. For legacy systems having only IPv4 support, the PM service is used for protocol translation.

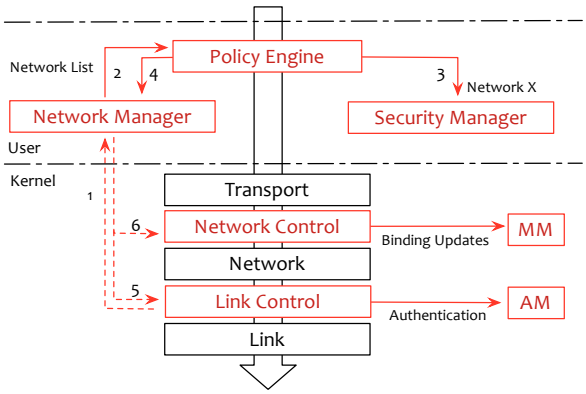

Figure 6: Control Middleware and Mobile Handover

7.1 Mobile Handover

A handover is a two-step process: link layer pre-authentication and acquisition of new IP address, and mobility control protocol signaling. As discussed in Section 3, a handover requires three entities to communicate securely, namely, the MN, CN and MM.

For optimal network connectivity, the NM is constantly monitoring the relevant network performance attributes, such as signal strength, latency, cost, and available bandwidth. If any of the attributes' values fall below a minimum, or exceeds a maximum threshold (such as minimum signal strength, or maximum cost), the NM triggers a network control event and forwards it to the PE, as shown in Fig. 6. The PE evaluates each active application's policies, and makes a candidate network selection decision for a handover (if any). Subsequently, the PE checks the network type, and requests the SM for any stored access credentials, of the candidate network. The PE signals the NM to perform pre-authentication and mobile handover. The NM commands the LCF to perform network authentication by following the same procedure described in Section 6.1. On successful authentication, the LCF signals the authentication result to the NM. The NM requests a new IP address, and subsequently, signals the respec-

tive NCF protocols to perform the new location update (IP address), to the corresponding MM. In the HIP case, the RVS and the CN are notified using the LOCATOR parameter in the HIP UPDATE message [5]. In the MIPv6 case, a BU message is sent to update the HA, and the Route Optimization[8] (RO) process is performed for all flows. In opportunistic mode, when there is no MM support, HIP may send location updates directly to all active CNs, and in the MIPv6 case, the RO process can also be performed directly, without HA intervention [27].

8. CONCLUSION

A unified, context-aware mobility-enabled architecture design has been presented, as an evolution of the traditional networking stack. This enhanced stack is complemented with a control middleware that abstracts networking complexity, and provides a policy-based decision making system. The policies take into account context information, such as location, cost, time and bandwidth, and override the system defaults, providing granular network access control, on a per-application basis. The system enables the provision of native mobility capabilities to perform seamless handovers, in a heterogeneous environment, and provides user and mobile node independence, from network and access technologies. The architecture also abstracts the physical service-providing network into separate planes, based upon infrastructure and service management criteria. This decoupling enables service providers to offer services across multiple infrastructure providers, resulting in better management of heterogeneity, and service scalability. Last, a generalized security framework for heterogeneous environments was proposed, using a federated identity management scheme, with the Extensible Authentication Protocol mechanism for network and device access. Asymmetric cryptography was used to secure mobility events, and we demonstrated that it could be reused for the end-to-end channels and applications.

9. FUTURE WORK

Policy definition languages will be evaluated and a formal specification proposed. The networking stack will be enhanced with the PCF implementation, including a location-based spectrum availability database. A system evaluation test-bed will be built for performance measurements. Additionally, the networking stack will be ported from the current Linux version to an Android based system.

10. ACKNOWLEDGEMENTS

This work was funded in part, by generous gifts from the Silicon Valley Community Foundation, and New York Center for Advance Technology – Telecommunications (CATT).

11. REFERENCES

[1] Cisco Visual Networking Index. White Paper. Cisco Systems Inc., Jun. 2013.

[2] A. Singh, G. Ormazabal, S. Addepalli, and H. Schulzrinne. *Heterogeneous Access: Survey and Design Considerations*. Technical Report. Columbia University, Oct. 2013.

[3] A. Ford, C. Raiciu, M. Handley, and O. Bonaventure. TCP Extensions for Multipath Operation with Multiple Addresses. IETF RFC 6824, Jan. 2013.

[4] R. Stewart. Stream Control Transmission Protocol. IETF RFC 4960, Sep. 2007.

[5] M. Riegel, and M. Tuexen. Mobile SCTP. IETF Internet Draft, Nov. 2007.

[6] P. Nikander, T. Henderson, C. Vogt, and J. Akko. End-Host Mobility and Multi-homing with Host Identity Protocol. IETF RFC 5206, Apr. 2008.

[7] T. Polishchuk and A. Gurtov. mHIP: TCP-Friendly Secure Multipath Transport. *In Proc. of 5th International Conference on Access Networks (ACCESSNETS)*, Nov. 2010.

[8] C. Perkins, D. Johnson, and J. Arkko. Mobility Support in IPv6. IETF RFC 6275, July 2011.

[9] R. Walikawa, V. Devarapalli, G. Tsirtsis, T Ernst, and K. Nagami. Multiple Care-of Addresses Registration. IETF RFC 5648, Oct. 2009.

[10] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil. Proxy Mobile IPv6. IETF RFC 5213, Aug. 2008.

[11] M. Arye, E. Nordstrom, R. Kiefer, J. Rexford, and M. J. Freedman. A Formally-Verified Migration Protocol For Mobile, Multi-Homed Hosts. *IEEE International Conference on Network Protocols (ICNP)*, 2012.

[12] IEEE Standard for Local and metropolitan area networks - Part 21: Media Independent Handover Services, http://www.ieee802.org/21/

[13] Access Network Discovery and Selection Function (ANDSF) Management Object (MO), http://www.3gpp.org/ftp/Specs/html- info/24312.htm

[14] I. F. Akyildiz, W. Lee, M. C. Vuran, and S. Mohanty. NeXt Generation Dynamic Spectrum Access / Cognitive Radio Wireless Networks: A Survey. *Computer Networks Journal*, Vol. 50, 2006, pp. 2127-2159.

[15] A. Mancuso, S. Probasco, and B. Patil. Protocol to Access White-Space (PAWS) Databases: Use Case and Requirements. IETF RFC 6953, May 2013.

[16] M. Wasserman and F. Baker. IPv6-to-IPv6 Network Prefix Translation. IETF RFC 6296, Jun. 2011.

[17] D. Farinacci, V. Fuller, D. Meyer, and D. Lewis. Locator / ID Separation Protocol (LISP). IETF RFC 6830, Jan. 2013.

[18] V. Fuller, D. Farinacci, D. Meyer, and D. Lewis. LISP Alternative Logical Topology (LISP+ALT). IETF RFC 6836, Jan. 2013.

[19] J. Arkko, V. Lehtovirta, and P. Eronen. Improved EAP for 3GPPP (AKA'). May 2009.

[20] L. Kagal. Rei: A Policy Language for the Me-Centric Project, HP Labs Technical Report, Sep. 2002.

[21] G. Yavas, et al. A data mining approach for location prediction in mobile environments. *Data & Knowledge Engineering*, *54*(2), 121-146.

[22] Open Networking Foundation, https://www.opennetworking.org

[23] GEYSERS - Generalized Architecture for Dynamic Infrastructure Services, http://www.geysers.eu

[24] D. W. Chadwick. Federated Identity Management, *Foundation of Security Analysis and Design*, 2009, Vol. 5705, pp. 96-120.

[25] B. Aboba, L. Blunk, J. Vollbrecht, J. Carlson, H. Levkowetz. Extensible Authentication Protocol (EAP). IETF RFC 3748, Jun 2004.

[26] S. Frankel and S. Krishnan. IP Security (IPsec) and Internet Key Exchange (IKE) Document Roadmap. IETF RFC 6071, Feb 2011.

[27] G. Hampel and T. Klein. Mobile IPv6 Route Optimization without Home Agent. IETF Internet Draft, Feb. 2011.

[28] IEEE MIH Implementation. ODTONE: One Dot Twenty One, http://hng.av.it.pt/projects/odtone

[29] HIP Implementation. OpenHIP, http://www.openhip.org/

[30] Mobile IPv6 Implementation. UMIP, http://umip.org/

[31] SOCKSv5 Proxy Server. Srelay, http://socks-relay.sourceforge.net/

[8] Enables nodes to communicate via a direct routing path, without the HA tunnel overhead.