CISCO

Building the Mobile Internet

Pervasive, ubiquitous computing technologies
and protocols that are shaping the future of our
mobile experience

Mark Grayson
Kevin Shatzkamer
Klaas Wierenga

ciscopress.com

Building the Mobile Internet

Mark Grayson, Kevin Shatzkamer, Klaas Wierenga

Published by:
Cisco Press
800 East 96th Street
Indianapolis, IN 46240 USA

Printed in the United States of America

First Printing February 2011

The Library of Congress Cataloging-in-Publication data is on file.

ISBN-13: 978-1-58714-243-7

ISBN-10: 1-58714-243-0

Warning and Disclaimer

This book is designed to provide information on how to enable the mobile Internet. Every effort has been made to make this book as complete and as accurate as possible, but no warranty or fitness is implied.

The information is provided on an “as is” basis. The authors, Cisco Press, and Cisco Systems, Inc. shall have neither liability nor responsibility to any person or entity with respect to any loss or damages arising from the information contained in this book or from the use of the discs or programs that may accompany it.

The opinions expressed in this book belong to the authors and are not necessarily those of Cisco Systems, Inc.

Trademark Acknowledgments

All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Cisco Press or Cisco Systems, Inc., cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

Corporate and Government Sales

The publisher offers excellent discounts on this book when ordered in quantity for bulk purchases or special sales, which may include electronic versions and/or custom covers and content particular to your business, training goals, marketing focus, and branding interests. For more information, please contact: **U.S. Corporate and Government Sales** 1-800-382-3419 corpsales@pearsontechgroup.com

For sales outside the United States, please contact: **International Sales** international@pearsoned.com

Feedback Information

At Cisco Press, our goal is to create in-depth technical books of the highest quality and value. Each book is crafted with care and precision, undergoing rigorous development that involves the unique expertise of members from the professional technical community.

Readers' feedback is a natural continuation of this process. If you have any comments regarding how we could improve the quality of this book, or otherwise alter it to better suit your needs, you can contact us through email at feedback@ciscopress.com. Please make sure to include the book title and ISBN in your message.

We greatly appreciate your assistance.

Trademark Acknowledgments

All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Cisco Press or Cisco Systems, Inc. cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

Publisher: Paul Boger

Associate Publisher: Dave Dusthimer

Cisco Representative: Erik Ullanderson

Cisco Press Program Manager: Anand Sundaram

Executive Editor: Mary Beth Ray

Managing Editor: Sandra Schroeder

Senior Development Editor: Christopher Cleveland

Project Editor: Mandie Frank

Copy Editor: John Edwards

Technical Editor: Scott Brim

Editorial Assistant: Vanessa Evans

Designer: Sandra Schroeder

Composition: Tricia Bronkella

Indexer: Cheryl Lenser

Proofreader: Sarah Kearns

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV
Amsterdam, The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices.**

About the Authors

Mark Grayson is a distinguished consulting engineer at Cisco Systems with responsibility for leading Cisco's mobile architecture strategy. He has over 20 years of experience in the wireless industry, ranging from the development of military systems, the definition of satellite communication architectures, and the evolution of traditional cellular systems to the creation of the latest small-cell solutions. He holds a first class honors degree in electronics and communications engineering from the University of Birmingham (England) together with a Ph.D. in radio communications. Mark has been granted over 50 patents in the area of mobile communications and is the coauthor of *IP Design for Mobile Networks* (Cisco Press).

You can contact Mark Grayson at mgrayson@cisco.com.

Kevin Shatzkamer is a distinguished systems architect at Cisco Systems with responsibility for long-term strategy and architectural evolution of mobile wireless networks. He has worked at Cisco and in the mobile wireless industry for over 10 years, focusing on various technologies that include 3G and LTE networks, packet gateways, network-based services and security, video distribution, quality of service, and end-to-end design theory. Kevin holds four issued patents and has 16 pending patents related to all areas of work. Kevin holds a Bachelor of Engineering degree from the University of Florida and a Master of Business Administration from Indiana University.

Kevin Shatzkamer is a regular speaker at various trade shows and industry forums and has previously published *IP Design for Mobile Networks*, a Cisco Press book that discusses the technologies and requirements shaping the future of the mobile Internet, from RAN to services. Kevin's current area of focus is the end-to-end digital media value chain for mobility, working with both content providers and service providers to create unique mobile media service offerings.

You can contact Kevin Shatzkamer at kshatzka@cisco.com.

Klaas Wierenga is a senior consulting engineer in the office of the CTO at Cisco. His 15-plus years of experience include the planning, analysis, and design of numerous solutions for enterprises, municipalities, hospitals, and universities in the fields of mobility, security, and identity worldwide. Klaas is the original creator of the worldwide eduroam service for federated network access in academia and cocreator of the federated identity solution that forms the basis of the Dutch government's e-Identity portfolio. He is the author of numerous publications and has presented many times on wireless networking, security, and identity topics. Klaas is active within 3GPP, in the group responsible for the security architecture of future mobile networks. He serves as chairman of the Abfab Working Group in the IETF, which deals with federated access for non-web applications, as well as of the Task Force on Mobility and Network Middleware of TERENA, the European Association for Research and Education Networks. Klaas holds a master's degree in computer science from the University of Groningen (The Netherlands).

You can contact Klaas Wierenga at klaas@cisco.com.

About the Technical Reviewer

Scott Brim is a Senior Consulting Engineer in the office of the CTO at Cisco. He received a BA magna cum laude from Harvard University and has been active in developing communications technology since 1978. He has been at Cisco since 2000. Previous to that, he was research staff at Cornell University for 18 years and Director of Technology Strategy at Newbridge Networks for 3 years. Technically, he has spent 12 years on Internet routing, 5 years on Internet QoS, and 4 years on mobile services. His particular interest has always been making different technologies or technology layers interwork better. He is currently focused on the future mobile Internet infrastructure and how proposed fundamental changes to Internet architecture in routing, addressing, mobility and identity can create a robust, flexible, beneficial synergy.

He has also been active in a number of standards bodies, most recently the IETF, ITU-T, and GSMA.

Dedications

I dedicate this book to my parents, Anne and Bryan, for their ever-present encouragement and support. I would like to thank my wife Sharon and two sons, Charlie and Harry, not the least for their inspiration for Chapter 7; I'm sure it won't be long before your networked lives of iPods, iPads, and PCs become fully mobilized Internet experiences. Finally, I would also like to thank the many friends, coworkers, and mentors who, over the last 20-odd years, have helped me achieve so much.

—Mark Grayson

I dedicate this second book to my wife and family, who, having experienced the time and commitment to authoring a book during the first round, allowed me to write a second one. As I explained the context of this second book, my young children assured me that our content reviewers would catch that I did not reference SpongeBob SquarePants, who, much to my surprise, is the founder of the mobile Internet. Alas, you will find no reference to SpongeBob in this book, partially because we did not have adequate time to receive all relevant copyright information, and partially because we have sought to provide an alternative theory into the development of the mobile Internet. To my children—may your reality always consist of Santa Claus, the Easter Bunny, and SpongeBob SquarePants.

—Kevin Shatzkamer

I dedicate this book to my wife Licia, who has been very supportive and patient whenever I deviated from my regular pleasant and optimistic self—ahem ;-)—in trying to meet the deadlines of this book while doing my day job. To my parents for making me explain the things I was working on in a nongeek way. And to my former colleagues at SURFnet, current colleagues at Cisco, and all the others I have worked with in the past years for shaping my understanding of the topics at hand and providing the often-so-necessary critique. In particular, I would like to thank the participants in the Task Force on Mobility and Network Middleware of TERENA and the members of the Mobile Internet project team at Cisco, without what I learned in the many discussions, fights, meals, and beers I have had with you, I could not have written this book.

—Klaas Wierenga

Acknowledgments

We'd like to thank the Pearson production team for their time and effort in creating this book, for patience during the delays resulting from our jobs, and for providing valuable and insightful feedback during the entire process. Specifically:

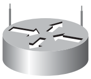
- Thanks to Mary Beth Ray for getting this book contracted and managing the process from beginning to end. We understand that we are not always the easiest to work with, and your involvement has made the authoring process a bit less painful.
- Thanks to Christopher Cleveland, Mandie Frank, and John Edwards for their fantastic editing of the book.
- Thanks to the many others at Pearson who were part of developing and producing this book. Sometimes it is those who remain faceless and nameless that tend to do the majority of the work, and we recognize that.
- Thanks to Moray Rumney at Agilent for giving his permission to use the chart in Figure 1-9 and for his comprehensive analysis of the radio frequency challenges in today's cellular systems.
- Thanks to Morgan Stanley Research for its permission to use the chart in Figure 1-1.
- Thanks to SURFnet for making the diagrams in Figures 3-3 and 3-9 available under a Creative Commons license.
- Thanks to TERENA for its permission to use the chart in Figure 3-7.

We'd like to thank Tom Carpenter for providing technical feedback on the many topics that this book covers. Also, thanks to all the technical reviewers, especially Scott Brim, who took the time to read our gibberish and turn it into gold.

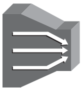
Contents at a Glance

Contents

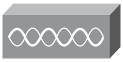
Icons Used in This Book



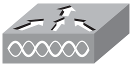
Wireless Residential Gateway



DSLAM

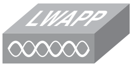


WiFi Access Point



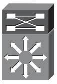
WLAN Controller

Wireless Transport



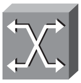
Lightweight Single Radio Access Point

Cisco ASA



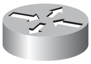
Route Switch Processor

Switch

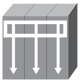


ATM Switch

Bridge



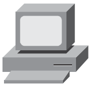
Router



Policy Server

SIP Proxy Server

Web Server



PC

Laptop

WiFi Enabled Tablet

Cellular Smartphone

Cell Phone

IP Phone

Wireless Connection

Introduction

This book examines the different techniques for building mobility into the Internet. The breadth of approaches currently in operation should cause us all to pose the question as to whether, in the future, a single utopian mobility solution can be defined that accommodates all scenarios, or whether solving "mobility" requires a decomposition of the "mobility problem space" into a number of distinct use cases.

The tremendous success of mobile broadband–based services based on cellular architectures where mobility has been effectively performed at the data link layer has shown how that approach is perfectly acceptable for providing wide-area mobility to use cases involving a single device with a single interface.

Should such data link layer techniques be enhanced to address alternative use cases? This is an important question to answer, because we confidently predict that the mobility use cases will broaden from today's homogeneous, cellular-only view of the world:

- Devices will become more heterogeneous from an access perspective. Wi-Fi dual-mode capabilities will become widely integrated into the next generation of cellular devices.
- Users will increasingly have access to more than a single cellphone for accessing the mobile Internet.
- As the majority of users who access the Internet become mobile, applications will increasingly look to become "mobile-aware," tailoring their operation to address specific limitations of mobile access, including being able to accommodate switches in access technologies and rapid fluctuations in available bandwidth.

This book takes a look at mobility from a broad perspective of use cases and examines how mobility solutions are in fact pervasive across all layers of the protocol stack. The book provides details of how mobility functionality has been added to these layers and describes use cases that demonstrate the different approaches to building the mobile Internet.

Who Should Read This Book

This book is intended to increase the reader's understanding of how mobility can be supported in IP networking.

The book assumes at least a basic understanding of standard networking technologies, including the Internet Protocol itself. Many concepts are introduced to give the reader exposure to the key mobility functionality that can coexist across different protocol layers. The book does not give recommendations on which of these technologies should be deployed for supporting mobility use cases, nor does it provide a transition plan for existing network operators for adding mobile functionality. Each network operator is expected to evaluate his or her mobility user case(s) that must be supported and make decisions based on his or her own criteria on which technique(s) to adopt for mobilizing the Internet.

This book is written for many levels of technical expertise, including network design engineers and network planning engineers looking to design and implement mobile network migrations toward an all-IP future, networking consultants interested in understanding the technology trends that affect mobile operators, students preparing for a career in IP networking that is increasingly being impacted by mobile technologies, and chief technology officers (CTO) seeking a further understanding of the convergence of IP and mobile technologies.

How This Book Is Organized

Although this book can be read from cover to cover, it is designed to be flexible and allows you to easily move between chapters and sections of chapters to learn just the information that you need.

This book covers the following topics:

- **Chapter 1, "Introduction to 'Mobility'":** This chapter defines the mobility market in terms of device proliferation, consumption trends, and radio-specific challenges in scaling for massive adoption of the mobile Internet.
- **Chapter 2, "Internet 'Sessions'":** This chapter explains the protocols and layers that make up the Internet architecture, as well as the fundamental problem with that architecture in supporting mobility.
- **Chapter 3, "Nomadicity":** This chapter describes how users and devices are authenticated for using the network and its applications, in particular those that are not operated by the operator that the user has a subscription with.
- **Chapter 4, "Data Link Layer Mobility":** This chapter explains the benefits of solving mobility at the data link layer. Contrasting approaches for delivering local- and wide-area wireless mobility are introduced and used with Wi-Fi and cellular technologies.
- **Chapter 5, "Network Layer Mobility":** This chapter provides an overview of a number of network layer solutions for delivering seamless mobility and session continuity.
- **Chapter 6, "Transport/Session Layer Mobility":** This chapter describes the advantages of integrating mobility functionality into the transport/session layer. The required mobile modifications to existing transport/session layer protocols are introduced.
- **Chapter 7, "Application Mobility":** This chapter describes how the application layer can be enhanced with additional mobility functionality, allowing advanced mobility use cases to be supported, including the ability to move media sessions between different devices.

- **Chapter 8, "Locator-Identifier Separation":** This chapter provides an overview of the approaches for redesigning the Internet architecture to allow better mobility, as well as a discussion of the pros and cons of some typical examples of those approaches.

This page intentionally left blank

Chapter 3

Nomadicity

Rather than focusing on keeping sessions alive, nomadicity is about being able to use the Internet and its services, regardless of location and time. The biggest challenge in gaining ubiquitous access is to be able to use networks and services that are not controlled by the operator that the user has a subscription with. This chapter explains the key concepts that make it possible for users and devices to gain access to IP networks and IP-based applications that are offered by others than their own operator. Nomadic or roaming use of the Internet refers to a usage pattern in which network connectivity is not available (or used) on a permanent basis, but rather intermittently and opportunistically. In other words, no session persistency at the transport layer is assumed. Therefore, this chapter does not cover Layer 2 or Layer 3 mobility, which are part of subsequent chapters. In particular, this chapter will not discuss roaming within the network of a cellular operator because that is based on Layer 2 roaming.

The basis of all Internet communications is, obviously, getting access to the Internet in the first place. In a local environment, getting access might be as simple as plugging an unshielded twisted-pair (UTP) cable into a wall outlet. As organizations grow bigger, in particular when wireless technologies are deployed or when users need to access the network from outside their own organization, security based on the ability to enter a particular building is no longer sufficient. There is a need to control access to the local network and the Internet in a scalable and efficient way.

A similar reasoning holds for application access. Users need to be able to access their networked applications, regardless of where they are. An increasingly popular phenomenon is that of offering applications "in the cloud," meaning that the application is hosted or offered by a third party somewhere on the Internet.

These are examples of the need for authentication, authorization, and accounting (AAA) mechanisms to control which persons or devices can gain access to the network and what they are allowed to do on that network. This chapter explains those mechanisms as well as the associated opportunities and challenges that come with that ability, in particular in a roaming situation.

Authentication and Authorization

A central concept for access to networks and applications is that of the digital identity—the digital representation of users or devices. The digital identity is usually associated with a unique identifier (such as a number or a name).

Authentication establishes the link between actual persons or devices and their digital representation. In other words, by successfully authenticating yourself, you prove to the network or the application that you are who you claim to be. After successful authentication, the network or the application then decides, based on policies that the operator or owner of the application has defined, what resources you get access to—the authorization.

As you can imagine, operators that often have millions of subscribers need to have sophisticated systems to keep track of all these subscribers and to provide adequate mechanisms for provisioning and deprovisioning, billing, authentication, and other services that are available to the subscribers. The servers that perform these tasks are generally referred to as AAA servers (pronounced *triple A servers*). As you will see in the sections that follow, in different domains, different types of electronic identifiers are used, which results in interoperability challenges. For example, the identifier that is used to gain access to a Long Term Evolution (LTE) network cannot be used just like that to gain access to a Wi-Fi network.

Authentication and Authorization in LTE

There are many types of cellular networks in use today. Standardization takes place in the Third Generation Partnership Project (3GPP)[1] and 3GPP2[2] (focusing mainly on the North American market). Instead of describing all the generations (1G, 2G, 2.5G, 3G, and 4G) and all the standards in those generations (CDMA, CDMA2000, EV-DO, HSDPA, GSM, UMTS, and many more) that all come with slightly different authentication methods and various roaming capabilities, this section provides a description of the LTE system for two main reasons:

- LTE is the technology that gains support from most mobile operators as the technology of choice for their future networks.
- LTE is the cellular technology that provides the most comprehensive system for roaming with other cellular but also noncellular network technologies.

Strictly speaking, LTE is only the radio access network technology. The core network architecture goes by the name System Architecture Evolution (SAE) and defines the Evolved Packet Core (EPC), the fixed part of a mobile operator network. But what is commonly referred to as LTE encompasses both the radio and the fixed network. This chapter will follow that convention.

Figure 3-1 shows the various components in an LTE network; these are defined in the list that follows. Chapter 4, “Data Link Layer Mobility,” describes the EPC and its associated mobility protocol in more detail.

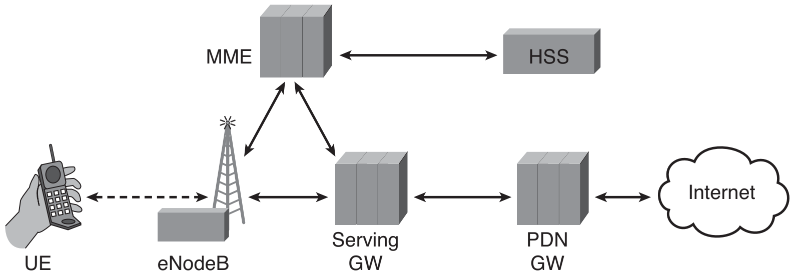


Figure 3-1 *LTE Architecture*

- The user equipment (UE) is the mobile device.
- The eNodeB is the *access point* to which the terminal connects through the wireless network and that is connected to the EPC.
- The Mobile Management Entity (MME) is the central control component in the EPC, and it is responsible for authenticating the user (by interfacing with the HSS—see the later bullet), assigning temporary identifiers to the terminals, roaming authorization, and lawful intercept.
- The Serving Gateway (Serving GW) routes packets to and from other 3GPP networks (General Packet Radio Service [GPRS], Universal Mobile Telecommunications System [UMTS]) and is a transient mobility anchor for the UE in those networks.
- The Packet Data Network Gateway (PDN GW) performs the routing to and from non-3GPP networks (like Wi-Fi, Code Division Multiple Access [CDMA] 1X, Evolution-Data-Optimized [EVDO], and WiMAX) and is the permanent mobility anchor for the UE roaming with those networks—in other words, the IP point of attachment.
- The Home Subscriber Server (HSS) contains the database with all subscriber data and is used to perform authentication and authorization as well as to provide user location.

The 3GPP specifications[3] define a number of identifiers to be used in cellular networks, the most important of which are those that identify, respectively, a user, a user subscription, and a device.

The International Mobile Subscriber Identity (IMSI) identifies users. The IMSI conforms to the ITU E.212 numbering standard and is usually 15 digits long (but can be shorter) and consists of a country code, a network operator code, and a mobile subscriber identity. The IMSI is stored in the SIM card and is used as the index key for subscriber data in the HSS, a database containing the data of all subscribers and the services they are entitled to. For privacy reasons, the IMSI is sent as little as possible over the network. Instead, after successful authentication, a temporary identifier, the Temporary Mobile Subscriber Identity (TMSI), is used.

An identifier called the Mobile Subscriber ISDN Number (MSISDN) is the phone number that corresponds with the SIM card in a mobile phone of a user. An MSISDN conforms to the ITU E.164 numbering standard and contains 15 digits that identify the country code, the network operator, and the subscriber.

Finally, the International Mobile Equipment Identity (IMEI) identifies the mobile device itself (not the SIM card inside).

Using the Authentication and Key Agreement (AKA) protocol defined in RFC 3310[4], a user authenticates to 3G and 4G networks and vice versa. The AKA procedure is a challenge-response mechanism based on a shared key that is stored on the SIM card of the terminal and in the Authentication Center (AuC) that is part of the HSS (in LTE) or HLR (in 3G). This shared key is used as input to algorithms to calculate other keys that are used for integrity (IK) and confidentiality (CK) protection of the data and for calculating the response to the challenge sent in the AKA.

Figure 3-2 and the list that follows show how the AKA procedure works in LTE. (Incidentally, UMTS networks also use the AKA.)

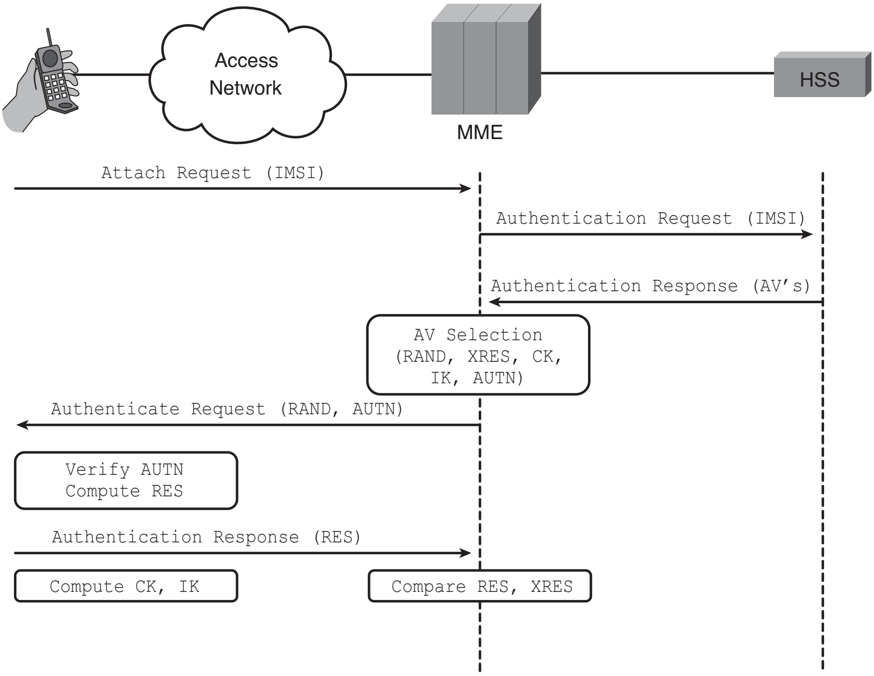

Access Network
MME
HSS
Attach Request (IMSI)
Authentication Request (IMSI)
Authentication Response (AV's)
AV Selection (RAND, XRES, CK, IK, AUTN)
Authenticate Request (RAND, AUTN)
Verify AUTN Compute RES
Authentication Response (RES)
Compute CK, IK
Compare RES, XRES

Figure 3-2 *AKA Authentication*

1. A shared secret (Ki) is defined beforehand and stored in both the SIM card and the Authentication Center (part of the HSS).

2. The terminal sends an Attach Request to the MME containing the IMSI or TMSI of the user.
3. The MME requests authentication information from the HSS.
4. The Authentication Center function in the HSS takes a random challenge (RAND), uses the shared key Ki that is associated with the IMSI to calculate the expected response (XRES) to that challenge, as well as CK and IK, and sends an authentication vector (AV) containing RAND, XRES, CK, and IK as well as the authentication token (AUTN) used by the SIM for authenticating the network to the UE.
5. The MME then sends an authentication request to the terminal containing the RAND and AUTN. The SIM authenticates the network by verifying the AUTN and calculates the CK and IK, as well as the response (RES) to the challenge RAND using the same algorithms the HSS used.
6. The RES is sent to the MME and compared with the XRES. If the RES and XRES match, the terminal gets access.

Authentication and Authorization in Wi-Fi Networks

As described in the sections that follow, authentication for Wi-Fi networks typically comes in two flavors—captive portals and IEEE 802.1X.

Captive Portals

With the captive portal approach, the device gets access to the local wireless IP network only. Whenever the user requests a web page outside the local network, the captive portal captures that request (hence the name) and instead shows a login page in which the user enters his or her username and password or credit card details. The user credentials are verified in some kind of user database, and upon successful verification, the user then gets access to the Internet. User identifiers take the form of a username.

802.1X and EAP

The IEEE 802.1X standard defines a framework for access control to a local-area network by encapsulating Extensible Authentication Protocol (EAP) messages. Wireless security standards such as WPA (Wi-Fi Protected Access) and WPA2 use 802.1X and EAP.

Figure 3-3 illustrates an 802.1X authentication. 802.1X defines three entities:

- The supplicant is a piece of code that runs on the user device.
- The authenticator is the device that gives the device network access; in Wi-Fi, this is the access point.
- The authentication server (typically a RADIUS server) verifies the user credentials in some sort of user database and informs the authenticator of the outcome.

The user identifier usually takes the form of a Network Access Identifier[5] (NAI), an identifier of the form *username@realm*, where the realm stands for the administrative domain to which the user belongs.

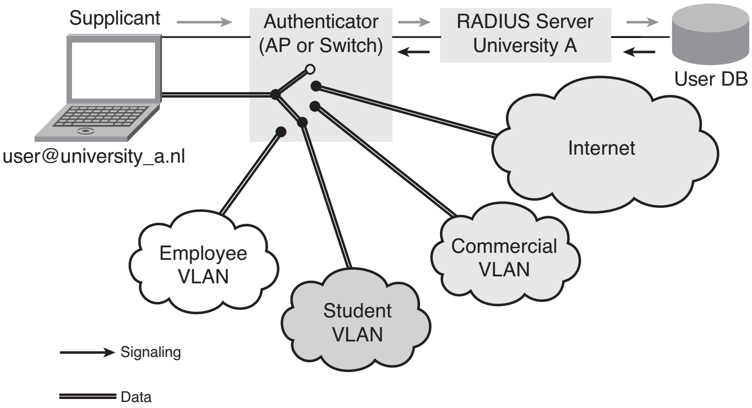


Figure 3-3 *802.1X Authentication Example of Student Gaining Access to the Campus Cetwork (Courtesy of SURFnet)*

User credentials are transported to the authentication server by using the EAP[6], a generic framework for forwarding encapsulated authentication data. EAP allows many types of credentials to be used, including username/password combinations, X.509 certificates, and others.

Figure 3-4 shows the EAP architecture. Between the supplicant and the authenticator, the EAP messages are encapsulated in Ethernet frames (EAP over LAN). Between the authenticator and the authentication server, EAP is usually encapsulated in RADIUS (or alternatively Diameter).

EAP methods define how authentication data should be encapsulated into EAP messages. Many different EAP methods exist. A number of EAP methods support confidentiality of user credentials in transit between the supplicant and authentication server. This means that neither the authenticator nor other network elements in the path between supplicant and authentication server can eavesdrop on the user credentials. In Wi-Fi networks, EAP-TTLS,[7] PEAP,[8] and EAP-FAST[9] are mainly used. All of these protect the user credentials against eavesdropping and allow mutual authentication of supplicant and authentication server. For roaming between cellular networks and Wi-Fi networks, EAP-AKA can be used, as discussed in the section "Non-3GPP Access," later in this chapter.

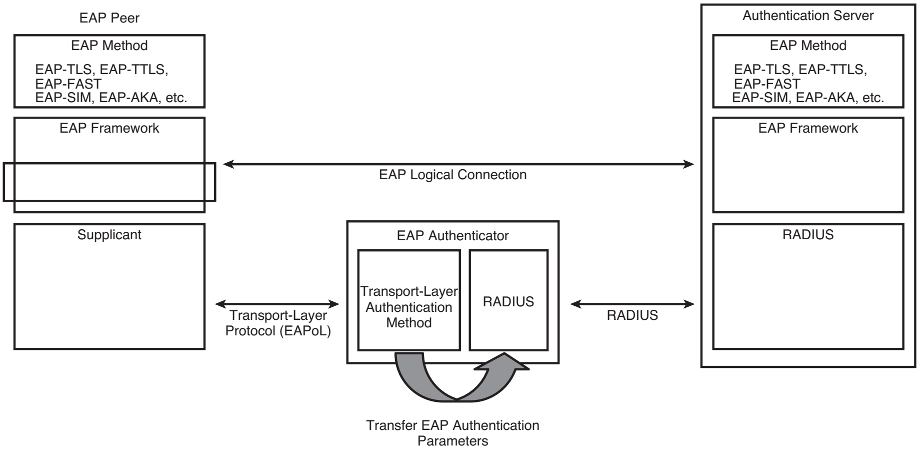


Figure 3-4 *EAP Message Communication*

Authentication and Authorization for Internet Applications

Authentication for network access is relatively difficult because there is no IP connectivity yet, so special protocols like 802.1X need to be used to transport user credentials to the authentication server. But as you saw in Chapter 2, “page 13,” after you have IP connectivity, the sky is the limit. Therefore, it is hard to say anything in general about authentication for networked applications. Many different protocols exist, such as Kerberos, NT LAN Manager (NTLM), HTTP Basic Authentication, and so on. Also, every possible authentication method, ranging from username/password combinations and one-time passwords to smartcard authentication, exists and is in use. For web-based applications, the most common one is still username/password over (hopefully) a Secure Socket Layer (SSL) connection.

Federated Identity

When you cross the border into another country or when you enter a shop that offers Wi-Fi, you really don’t want to have to sign up for a contract each time to connect to a network, not to mention the burden of remembering all the different usernames and passwords.

Here is where *federated identity* comes in. In the federated model, a user has a contract with only one (or a few) operators—the “home operator” that establishes the identity of the user. That one identity is then used to gain access to networks or applications managed by other operators. To make this work, the operators of the different networks need to establish a roaming or federation agreement. Such an agreement specifies under what conditions a visited network accepts an authentication statement (“this is a valid user”) from the home operator, how the authentication credentials and accounting data are exchanged, and what financial arrangement is in place for visiting users.

So, in this model, as shown in Figure 3-5, the home operator (in identity lingo called the Identity Provider [IdP]) acts as a trusted third party for the serving operator, called the Service Provider (SP) or Relying Party (RP). There is no direct contractual or trust relationship between the user and the visited network, only between the user and the IdP and between the IdP and the RP. Because there is a trust relationship between the user and IdP and between the IdP and RP, the RP "trusts" the user.

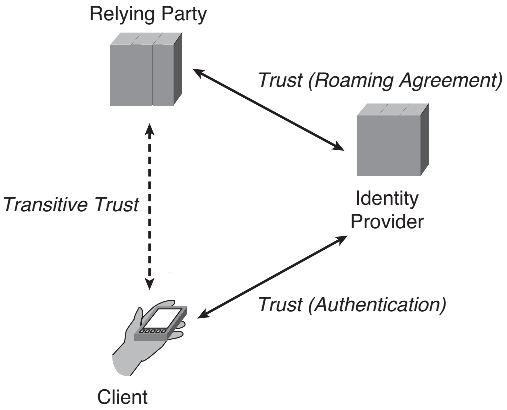


Figure 3-5 *Federated Identity*

The RADIUS[10] protocol allows forwarding of authentication requests to another RADIUS server; this is why RADIUS is widely used for network roaming. The authentication requests here are forwarded by the RP (usually also a RADIUS server) to the RADIUS server of the home operator (the IdP), and the outcome of the authentication is sent back. By combining RADIUS with EAP, the confidentiality of the user credentials can be preserved.

For access to web-based applications, a number of different protocols are used, such as Security Assertion Markup Language (SAML[11]), OpenID[12], or Open Authentication (OAuth)[13].

When users roam between networks using the same technology (that is, from one Long Term Evolution (LTE) network to another LTE or similar network), this is called *horizontal roaming*. Roaming between different types of networks, such as roaming from an LTE network to a Wi-Fi network, is called *vertical roaming*.

3GPP standardizes access to non-3GPP networks but places the LTE core network firmly in control; authentication is always performed in the LTE EPC. IEEE (the standardization body for the Wi-Fi standards) in its 802.21 standard[14] addresses vertical roaming with a more equal role for the different access technologies but does not address federated network authentication.

Federated Access in LTE

3GPP distinguishes two types of federated access:

- **3GPP access[15]:** Describes horizontal roaming.
- **Non-3GPP access[16]:** Describes vertical roaming.

In both cases, the home network needs to establish a roaming agreement beforehand (and the user's subscription should allow roaming access).

3GPP Access

3GPP access is access to an LTE network of another operator or to a UMTS or GPRS network. UMTS networks (and GPRS networks that support interworking with LTE) support AKA authentication.

Based on the IMSI (that contains a country and an operator code), the MME can ask the home HSS of the user to verify the user rather than the HSS in the serving network. The home HSS must check whether the subscription agreement with the user allows roaming, but apart from that, the authentication process is the same as for the nonroaming case.

Non-3GPP Access

Examples of non-3GPP access are CDMA-2000, WiMAX, and Wi-Fi. These networks don't use the same authentication methods, and the elements in these serving networks don't understand how to deal with an AKA authentication. So, rather than involving a network element in the serving network directly in the authentication flow with the home network, EAP is used instead. Here EAP provides the necessary abstraction from the actual authentication using AKA. For this purpose, EAP-AKA and its more secure successor EAP-AKA' (EAP-AKA Prime) have been created. The EAP identity contains the IMSI or a pseudonymous identifier that was established in a prior authentication to locate the authentication server for the user.

Figure 3-6 shows how AKA authentication can be encapsulated in EAP.

Although you can use the IP connectivity of the serving network, the typical use is to tunnel all traffic back to the home network using IPsec.

Federated Access to Wi-Fi Networks

Originally, Wi-Fi was intended to be used as a local-area network technology, typically covering an area with a radius of some 30–50 meters. Nowadays, sometimes hundreds to thousands of Wi-Fi access points together form *hotspots* that provide coverage to complete campuses or even cities. Still, the majority of the hotspot operators provide access to an area with a limited geographical scope (unlike the nationwide coverage that cellular operators provide). To provide coverage beyond the geographical region, hotspot operators need to collaborate so that subscribers of one operator can gain access to the network of another operator.

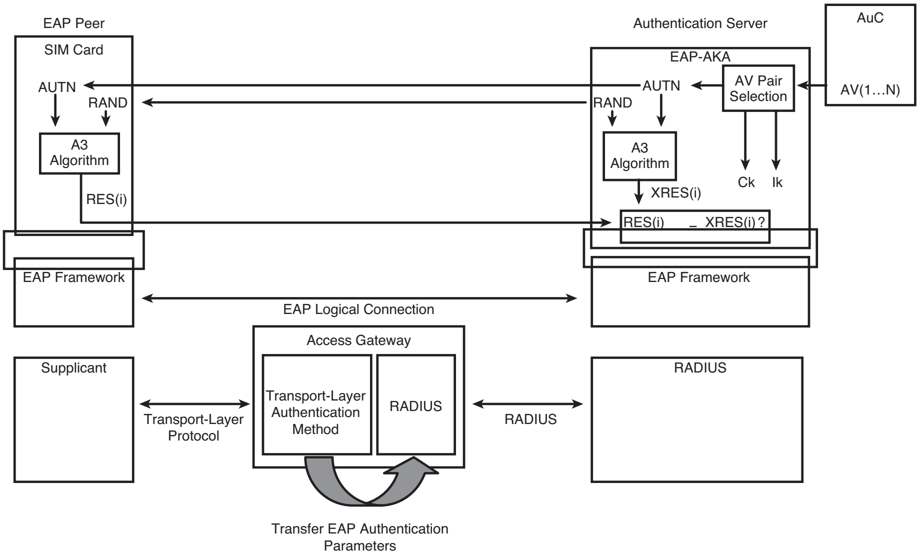


Figure 3-6 *EAP-AKA*

Roaming to Other Wi-Fi Networks

The main challenges in roaming access for Wi-Fi networks are setting up the roaming agreements and verifying the user credentials at the home network.

Because, unlike cellular networks, Wi-Fi hotspots are by virtue of the local-area character of the technology relatively small, setting up roaming agreements with a large number of Wi-Fi operators is hard to scale. To solve the scaling problem, three different types of organizational models emerge:

- The first model mimics the cellular model. A large operator acquires or leases a large number of hotspots and unifies the authentication across these hotspots. AT&T hotspots and T-Mobile hotspots are examples of this arrangement.
- In the second model, a third party acts as a broker for a large number of hotspot operators. The users have a contract with the broker and authenticate and pay for access to the broker. The broker in turn pays the hotspot operator. Examples of this arrangement are Boingo and iPass.
- The last model has individual hotspot operators join forces and agree on roaming conditions and credential verification methods. Examples of the latter are FON and the Wi-Fi roaming infrastructures that many schools worldwide participate in—eduroam. (This is further explained later in the section "Example of Wi-Fi Roaming: eduroam.")

Verification of the credentials of the users at the home network requires transporting the credentials to the home network and sending the outcome of the authentication back to the visited network. The dominant transport protocol for transporting the credentials is RADIUS.

The main advantage of the captive portal method for Wi-Fi is that it only requires a web browser on the user device. This is also why it is the most commonly used access method at public hotspots. The main downside is that because the Wi-Fi link is unprotected, simple MAC spoofing can be used by an attacker to piggyback an authenticated user's connection. Additionally, the user credentials are visible to every hotspot operator (they have to be entered in the web page that the captive portal shows) and can be observed by every RADIUS server in the path to the home RADIUS server. When 802.1X is used, the combination of 802.1X, EAP, and RADIUS allows user credential privacy. This means that users don't have to worry about giving their password to potentially thousands of hotspot operators, let alone rogue hotspot operators. An added benefit of using 802.1X is that all user traffic is encrypted over the Wi-Fi radio link, allowing the operator to be sure that every packet sent into the network originated from an authentic Wi-Fi user.

The added security features of 802.1X and better support in the most common operating systems have resulted in a slow but steady increase in use, especially in corporate environments.

802.11u

Two issues that are particularly important for Wi-Fi access are the fact that most Wi-Fi hotspots are relatively small and that there are thousands of them. In a densely populated area, a user easily often "sees" 30 or 40 different Wi-Fi networks, without knowing which of those will have a roaming agreement with the home operator and, if so, under what conditions.

This is the problem space that the upcoming IEEE 802.11u[17] standard addresses. Hotspots that are 802.11u enabled can broadcast information about the roaming consortia they belong to and under what conditions they can be used.

Example of Wi-Fi Roaming: eduroam

An example of a Wi-Fi roaming service is eduroam[18]. This service is limited to educational institutions. However, its technical setup and broad uptake (more than 500 universities in some 50 countries with over 10 million users) warrant attention.

eduroam started out in the Netherlands in 2003 and gained fast popularity in most European countries and later in Australia, Japan, Hong Kong, and Canada. Lately, U.S. schools are joining eduroam and Internet2 is supporting the initiative.

Figure 3-7 shows the European national research and education networks that participate in eduroam. (For an up-to-date overview of all participating institutions in Europe and elsewhere, refer to the eduroam website.[19])

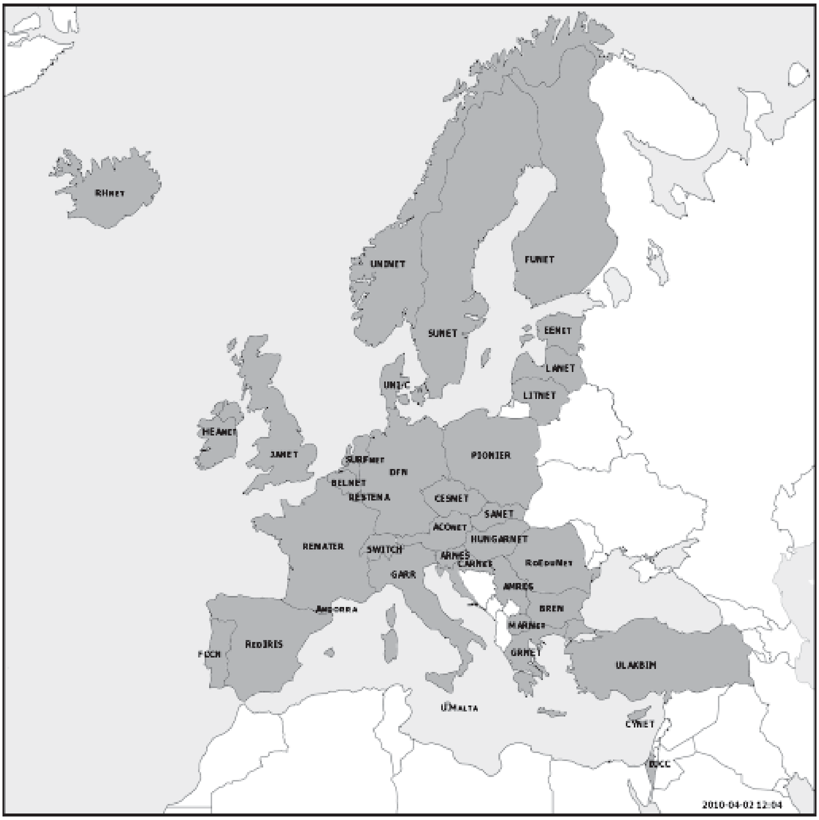


Figure 3-7 *European National Research and Education Networks Participating in eduroam as of May 2010 (courtesy of TERENA)*

eduroam consists of a few basic elements, described in the following paragraphs.

A RADIUS hierarchy is set up consisting of a set of institutional (redundant, for failover purposes), national, and continental RADIUS servers. All institutional RADIUS servers connect to the national servers in their country. All national servers connect to the top-level servers for their continent, and the continental servers (Europe, America, and Asia-Pacific) connect to each other.

Figure 3-8 shows the RADIUS hierarchy that constitutes eduroam. The top-level servers that are fully meshed know which top-level servers serve what national domains. The national servers are connected to all institutional servers in their country and to the top-level servers in their continent. The institutional servers are connected to their national servers.

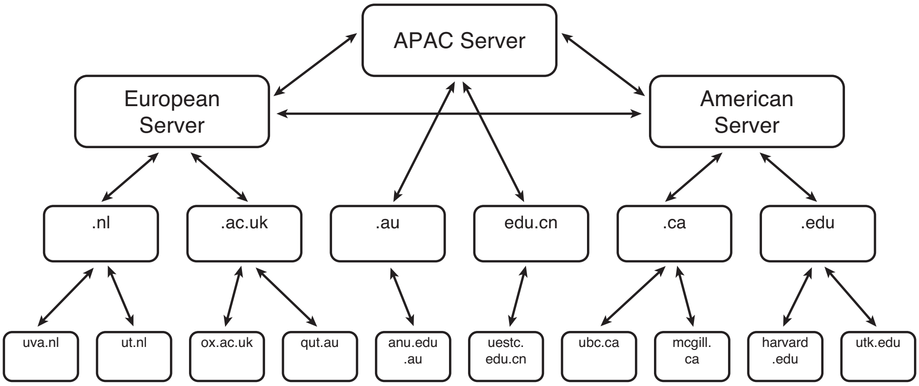


Figure 3-8 *The eduroam Hierarchy*

802.1X is used for secure access to the institutional Wi-Fi networks.

EAP is used to protect user credentials. EAP identities are of the form anonymous@*domain-name-of-institution* or, instead of anonymous, a pseudonymous identifier. Users' authentication requests are forwarded through the RADIUS hierarchy based on the domain name of the institution to which the user belongs.

In other words, the home institution authenticates the user and the serving institution authorizes the user for access. The home institution of the user can decide which authentication method and what EAP method to use.

Figure 3-9 shows a typical eduroam authentication, which is described in further detail in the list that follows.

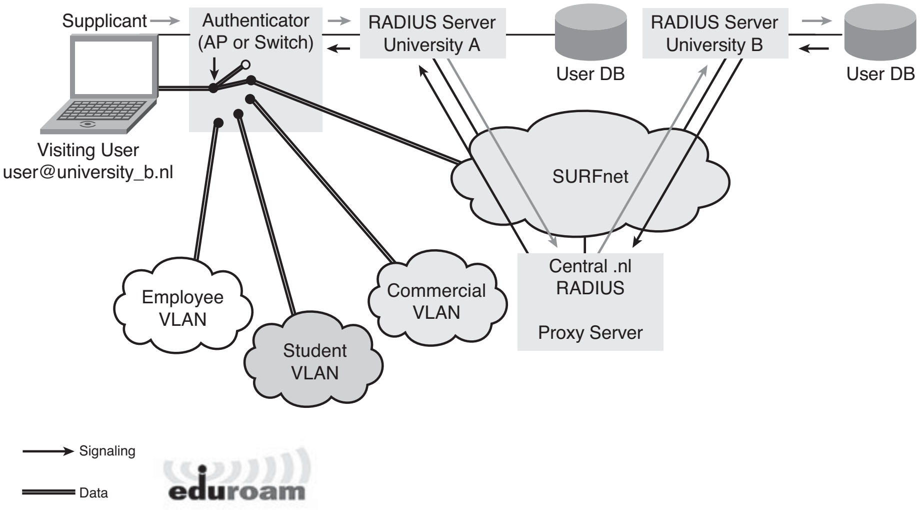

Supplicant
Authenticator (AP or Switch)
RADIUS Server University A
User DB
RADIUS Server University B
User DB
Visiting User user@university_b.nl
SURFnet
Central .nl RADIUS Proxy Server
Employee VLAN
Student VLAN
Commercial VLAN
Signaling
Data
eduroam

Figure 3-9 *The eduroam Basic Operation (Courtesy of SURFnet)*

1. A user from University B in the Netherlands tries to gain access to the network at University A, also in the Netherlands.
2. The authenticator asks the user (or rather the supplicant) to authenticate.
3. The user sends the authentication credentials encapsulated in EAP with an EAP identity of anonymous@university_b.nl to the authenticator.
4. The authenticator at University A forwards the EAP message to the RADIUS server of University A.
5. The University A RADIUS server observes that the EAP identity does not belong to University A and forwards the EAP message to the national RADIUS server for the Netherlands operated by SURFnet, the Dutch research and education network.
6. The SURFnet RADIUS server for the .nl domain sees that the EAP identity belongs to University B and forwards the EAP message to the University B RADIUS server. (If the EAP identity were not for the .nl domain, the EAP message would be forwarded to the European top-level server.)
7. The University B RADIUS server deencapsulates the EAP message and verifies the credentials.
8. University B sends the result of the authentication back along the same route.
9. The RADIUS server at University A instructs the authenticator to allow access to the user (and possibly to assign the user to a specific VLAN for guests).

Federated Access to Applications with SAML

When you assume that more and more applications will be offered “in the cloud,” it is imperative that scalable mechanisms exist for federated identity. The most widespread systems for federated identity to (mainly) web-based application make use of the Security Assertion Markup Language (SAML) protocol suite. SAML is an XML-based markup language for transporting authorization assertions between IdPs and RPs.

Figure 3-10 shows a typical SAML (version 2.0) flow, which is further described in the list that follows:

1. The user uses his browser to try to access a resource under control by the RP.
2. The RP issues an authentication request to the browser (plus a redirect to the IdP).
3. The browser sends the authentication request to the IdP and asks for an authentication statement.
4. The user (if not already authenticated) authenticates at the IdP.
5. The IdP issues an authentication statement to the browser stating that the user is successfully authenticated (plus a redirect back to the RP).
6. The browser presents the authentication statement to the RP.
7. The RP gives the user access to the resource (assuming that the user satisfies the RP’s policies and a roaming arrangement exists between the IdP and RP).

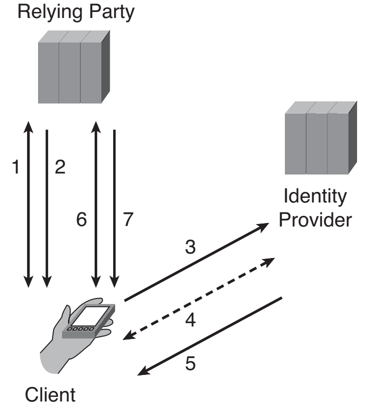


Figure 3-10 *SAML 2.0 Authentication Flow*

Location Information and Context Awareness

One particular characteristic of the mobile Internet is obviously that users are mobile—that is, not bound to a particular location. The location of users can be used for a number of location-based services, ranging from finding the nearest pizza parlor through turn-by-turn driving directions to finding a person that makes an emergency call.

Many modern devices contain a Global Positioning System (GPS) chip for providing location information. GPS is, however, not very accurate indoors or in the presence of buildings. Therefore, in many networks, additional ways of positioning a mobile terminal are used.

Location information is, however, just one example of the broader issue of context awareness. Knowledge about the network, the device that accesses the network, the user, and the applications can all be leveraged to offer the user a tailor-made user experience. But that is beyond the scope of this book.

The following paragraphs explain how location is determined in LTE and Wi-Fi networks.

Location Information in LTE

Nowadays many mobile devices contain a GPS receiver. This allows the applications to acquire the location of the device and offer services based on that location. In addition to that, many devices are able to use so-called assisted GPS (A-GPS) to overcome the inaccurate indoor positioning of GPS. With A-GPS, the network supplies location information (based on the position in relation to cell towers) or information about the geo-orbital position of the GPS satellites. This allows devices to acquire fast and reliable positioning information in all circumstances, even indoors. This is of particular importance in emergency situations (like 911 or 112 calls in, respectively, the United States and Europe).

Location Information for Wi-Fi Networks

Unlike in cellular systems, where there are relatively few radio towers of which the position is well known, for Wi-Fi networks, the location of all access points is not always well known and might change. This makes it comparatively harder to use the location of the access points to reliably determine the position of the mobile equipment. If the location of the access points is stable, however, the location of the access point can be used to determine the location of a mobile terminal accurate to typically about 50 meters, even better if triangulation is used between multiple access points.

An example of a Wi-Fi positioning system is the Cisco Wireless Position Appliance[20] that is part of the Cisco Context-Aware Mobility solution and that can be used to track assets and users.

Privacy and Security

Privacy concerns develop when user data is spread across many locations. Personal Identifiable Information (PII), such as street addresses, IP addresses, first and last names, and login credentials, can be traced back to an individual or a small group of individuals. Privacy regulations often dictate the amount of PII data that can be exchanged.

At the same time, users need to be properly authenticated when they are trying to access another network, and users often want to share their location to get location-based services.

Law enforcement requires the ability to track crime suspects and monitor their transactions and conversations. For that purpose, operators need to be able to redirect and monitor traffic of particular users without their knowledge. These Lawful Intercept (LI) requirements complicate roaming agreements, traffic offload, and other route optimization functions, because the easiest way to comply with these requirements is to direct all traffic through a central location, where it can be monitored.

A useful concept in federated access is that of a “pseudonym,” an identity that is unique for a specific user and often for a specific access network but that can only be linked to an individual user by the home network operator. The extent to which pseudonymity can be used varies from one access technology to another and from implementation to implementation.

From a security point of view, a benefit of the federated model is the fact that the sensitive user data is not distributed over many systems, but concentrated in the IdP.

Another benefit of having a centralized authentication server is that it is possible to introduce stronger authentication means (like smartcards) without the need to change all applications to support this type of authentication.

Apart from authentication and authorization data, the user traffic and the control traffic between the various elements in the network often need to be protected against eavesdropping and tampering. For this purpose, a wide variety of cryptographic means are used.

Privacy and Security in LTE

In LTE (unlike UMTS), a great deal of effort has gone into making sure that compromising the security of one network element will not imply compromising the security of the system as a whole. As an example of that, a complex system for the generation of cryptographic keys has been developed that is being used for securing the communication between the various other network elements. In particular, all keys inside a visited network are derived from a "master" key, which is specific for that serving network. This means that if the security in the serving network is breached, this will not have any implications for the home network and the integrity of the user credentials.

Traffic between the serving network and the home network is protected using IPsec.

In the initial AKA authentication, the IMSI is sent to the serving network, but after that, a temporary identifier is used. This means that the serving network is still capable of observing the IMSI, but at least the casual eavesdropper is unable to monitor the point of attachment of that particular IMSI.

Privacy and Security in Wi-Fi Networks

For captive portals, it is by the nature of that technology very difficult to provide location privacy and credential protection, the users submit their credentials after all at that specific location and to the captive portal that is used by the hotspot operator. These problematic security properties are worsened by the fact that users are in a way "trained" to submit their username and password or other authentication credential to every web page that remotely looks like a plausible hotspot page, instead of sharing their credentials only with their home network operator. Furthermore, unlike with 802.1X, typically all users get IP access to the local LAN that the hotspot is connected to, even before authentication. So, it is relatively easy to eavesdrop on the wireless traffic.

Using 802.1X in combination with EAP in contrast, it is possible to use pseudonymous identifiers for the users (identifiers such as anonymous@*homeprovider* or pseudonym12345@*homeprovider*), and in addition to that, 802.1X sets up a secure association between mobile equipment and access point, thereby protecting the user traffic against eavesdroppers on the wireless network.

Privacy and Security in SAML

SAML-based identity federations have been designed with user privacy and confidentiality in mind. Users are redirected to their own IdP to perform authentication so that the user credentials don't have to be shared with the RP. Instead of using the actual user identity for interacting with the RP, it is possible to use a pseudonymous identifier that is unique for the user and on a per-RP basis (a so-called *targeted identity*).

From a privacy aspect, there is one concern that has to do with the nature of SAML-based federations. The SAML model is geared toward an enterprise-centric model. That is to say, the IdP is always a party in a transaction, and therefore the IdP has a good insight in all the transactions that a user performs. In answer to this concern, there has been a lot of interest in what is called *user-centric identity*. In this model, the user uses an IdP for initial identity proofing and goes on wielding that proof of identity without having to involve the IdP in every transaction. So the Identity Provider does not need to know what services the user accesses. Examples of user-centric identity approaches are OpenID, OAuth, and Infocard.

DynDNS

So far, this chapter has concentrated on the user gaining access to the network or application. There is, however, another important issue to consider—how to find the mobile equipment if the other side initiates the communication. As you will see in future chapters, there are a number of solutions that provide a stable *anchoring point* that can be used to find the current point of attachment or to direct all traffic to, but they require changes in the protocol stack. The standard way of informing "the Internet" where a certain host resides is by using the Domain Name System (DNS). DNS, after all, contains mappings from host names to IP addresses. So, if the entry in DNS is updated every time a host changes its point of attachment (and thus IP address), DNS information can be used to find the target IP address of a connection.

This is precisely what Dynamic DNS (DynDNS) is—a DNS server that is optimized for frequent updates of the mapping information. A number of implementations of DynDNS exist, often provided for free. For this to work, a DNS client that updates the current name to IP address mapping every time the host changes IP address is required.

Because of the distributed nature of DNS (it takes some time before DNS resolvers become aware of a change), DynDNS is not practical when DNS changes occur very frequently, in the order of magnitude of seconds.

Another point of concern is that rogue DNS updates can be used to redirect traffic. Unless the DNS updates can be authenticated—for example by using DNS Secure (DNSSEC)—this is a security problem. DNSSEC, however, is not yet widely in use.

Summary

Nomadic use refers to a usage pattern of the Internet with intermittent access. Key to a user experience of anytime, anywhere access to the Internet and its applications are scalable, secure, and seamless access methods. Federated identity plays an important role in providing such a user experience in the presence of multiple-access networks and applications offered by different operators. Nomadic use requires some form of context awareness to provide services that are tailored to location, access method, and device.

Endnotes

1. The 3D Generation Partnership Project, http://www.3gpp.org.
2. The 3D Generation Partnership Project 2, http://www.3gpp2.org.
3. TR 21.905, "Vocabulary for 3GPP Specifications," http://www.3gpp.org/ftp/Specs/html-info/21905.htm.
4. RFC 3310, "HTTP Digest Authentication Using AKA," A. Niemi, J. Arkko, and V. Torvinen, http://www.ietf.org/rfc/rfc791.txt, September 2002.
5. RFC 4282, "The Network Access Identifier," B. Aboba, M. Beadless, J. Arkko, and P. Eronen, http://www.ietf.org/rfc/rfc4282.txt, December 2005.
6. RFC 5247, "Extensible Authentication Protocol Key Management Framework," B. Aboba, D. Simon, and P. Eronen, http://www.ietf.org/rfc/rfc4282.txt, August 2008.
7. RFC 5281, "Extensible Authentication Protocol Tunneled Transport Layer Security Authenticated Protocol Version 0," P. Funk and S. Blake-Wilson, http://www.ietf.org/rfc/rfc5281.txt, August 2008.
8. "Protected Extensible Authentication Protocol," http://en.wikipedia.org/wiki/Protected_Extensible_Authentication_Protocol.
9. RFC 4851, "The Flexible Authentication via Secure Tunneling Extensible Authentication Protocol Method," N. Cam-Winget, D. McGrew, J. Salowey, and H. Zhou, http://www.ietf.org/rfc/rfc4851.txt, May 2007.
10. RFC 2138, "Remote Authentication Dial In User Service," C. Rigney, A. Rubens, W. Simpson, and S. Willens, http://www.ietf.org/rfc/rfc2138.txt, April 1997.
11. "SAML V2.0 Executive Overview," P. Madsen, et al., http://www.oasis-open.org/committees/download.php/13525/sstc-saml-exec-overview-2.0-cd-01-2col.pdf, April 2005.
12. OpenID, http://openid.net.
13. OAuth, http://oauth.net.
14. IEEE 802.21, http://www.ieee802.org/21.
15. TS 33.401, "3GPP System Architecture Evolution: Security Architecture," http://www.3gpp.org/ftp/Specs/html-info/33401.htm.
16. TS 33.402, "3GPP System Architecture Evolution: Security Aspects of non-3GPP Accesses," http://www.3gpp.org/ftp/Specs/html-info/33402.htm.
17. IEEE 802.11u, http://en.wikipedia.org/wiki/IEEE_802.11u.

18. K. Wierenga and L. Florio. "eduroam: Past, Present and Future." *Computational Methods in Science and Technology*, Vol. 11, No. 2: February 2005.

19. eduroam website, http://www.eduroam.org.

20. "Cisco Location Solution Overview," https://www.cisco.com/en/US/solutions/collateral/ns340/ns394/ns348/ns753/net_brochure0900aecd8064fe9d_ps6386_Products_Brochure.html.

Index