Executive Summary

Dual-mode (GSM and Wi-Fi) services are becoming attractive to mobile operators and their subscribers because of three trends: a growing population of mobile subscribers, the prevalence of home broadband connections, and the availability of low-cost, home wireless access points that support GAN (previously known as Unlicensed Mobile Access, or UMA) technologies, such as Wi-Fi and Bluetooth. The research firm In-Stat forecasts that consumers will use more than 66 million dual-mode handsets by 2009, and Senza Fili Consulting predicts that the addressable market for dual-mode services will reach 55 million subscribers by 2010. With dual-mode services, subscribers make calls from outside the home as they would ordinarily, using the GSM radio network at the standard tariff rate. But inside the home, the call travels over the subscriber’s wireless broadband connection, so the operator can enjoy a similar economic structure as VoIP-over-broadband providers.

To offer GAN dual-mode services, mobile operators need handsets, network controllers, call control, the security to protect the mobile operator voice network from Internet-based threats, and wireless access points for their subscribers. Cisco Systems meets these requirements with a secure, scalable, flexible, highly secure multiservice IP architecture for GAN. Cisco GAN Enhanced Security Gateway architecture expands on the existing standards to provide a mobile operator with a complete, adaptive security solution that identifies and protects operator infrastructure.

The Cisco GAN Security Gateway solution is an integral part of the Cisco Mobile Exchange architecture. The Cisco Mobile Exchange is a standards-based framework that links the Radio Access Network (RAN) to IP networks and their value-added

---

1 In-Stat, “Wireless IP Phones Drive Future VoIP Markets,” August 2005
2 Senza Fili Consulting, “GAN and Beyond: Mobile Operators Benefit from Wi-Fi and Cellular Convergence,” January 2005
services. It comprises numerous different components (Figure 1), including IP gateways (Gateway GPRS Support Nodes [GGSNs], Packet Data Serving Nodes [PDSNs], Security Gateways, etc.), mobile services (application-layer charging, content filtering, service selection, policy control, etc.), load balancing, and network management services delivered on a range of Cisco platforms and application modules. Together, these components successfully address the many challenges that face mobile network operators as they seek profitability from their second-generation (2G), 2.5G, 3G, 4G, or GAN mobile packet infrastructures and their 802.11 public WLAN hotspots.

This paper explains the dual-mode service opportunity using GAN, the GAN standards, customer experience, operator security implications, and solution components of the Cisco GAN Enhanced Security Gateway. For a business solution description, refer to:

GAN Standardization—Accelerating Convergence Today
The Third-Generation Partnership Program (3GPP) GAN standards evolved from an UMA specification outlining and specifying the handling of secure connectivity, registration, and transmission of both GSM voice and General Packet Radio Service (GPRS) data signaling and bearer traffic over unlicensed RANs. Conceptually, GAN helps an end user both send and receive secure voice and data transmissions over the GSM or GPRS operator-controlled radio network, or a home Wi-Fi or Bluetooth private network using a single device maintaining a single phone number and identity. In addition, GAN allows the end user to transparently roam between a public GSM or GPRS network and private-home Wi-Fi or Bluetooth network without service interruption. The GAN specification seeks to extend the industry trend toward fixed mobile convergence (FMC), taking advantage of existing 2.5G mobile operator equipment, and provides a migration path toward an all-IP converged infrastructure.

Figure 2 shows how an end user with a dual-mode handset can extend coverage into an unlicensed private network. Regardless of the access type (licensed or unlicensed wireless), the end user is consistently able to access the core mobile network applications, allowing a mobile operator to extend the same network authentication and authorization mechanisms (subscriber identity module [SIM]-based authentication through home location register [HLR]) and services (IP multimedia subsystem [IMS], Push-to-Talk over Cellular [PTToC], Short Message Service [SMS], and multimedia messaging service [MMS]) into areas that are not covered by the operator’s radio network.
The GAN standardized functional architecture comprises five main components, depicted in Figure 3:

- **GAN controller (GANC)**—The GANC is integrated into existing 2.5G operator voice and data components through the standard 3GPP-defined network interfaces. For voice traffic, the GANC integrates directly into an operator Mobile Switching Center (MSC) through the A interface. For data traffic, the GANC integrates directly into an operator serving GPRS support node (SGSN) through the Gb interface. The GANC provides dual-mode handsets with alternative access to GSM voice and GPRS data services.

- **Security gateway**—As Figure 1 illustrates, the introduction of a GAN solution into an operator network raises numerous security implications and vulnerabilities inherent in an IP-based architecture. The security gateway provides two important security roles in the GAN: secure authentication (through Extensible Authentication Protocol–SIM [EAP-SIM] or EAP–Authentication and Key Agreement [EAP-AKA]) of mobile subscribers and termination of secure tunnels (through IP Security [IPSec] with Internet Key Exchange Version 2 [IKEv2]) from the handset.
• Authentication, authorization, and accounting (AAA) infrastructure—The AAA infrastructure interacts with numerous elements in the GAN architecture, including:

  – Security gateway: The AAA infrastructure interacts directly with the security gateway to validate mobile credentials during IPsec tunnel establishment. This includes the use of EAP mechanisms for SIM-based authentication using either EAP-SIM or EAP-AKA.

  – HLR: The AAA infrastructure includes a MAP Gateway function for communication to the operator HLR using the SS7 transport protocol. During authentication, the AAA infrastructure is responsible for converting RADIUS authentication messages from the security gateway into SS7 MAP Invoke messages to the HLR. This allows the existing HLR to verify a user on the GAN using the IMSI/triplets sequence that is standard for GSM/GPRS authentication.

    – GANC (optional): As an additional layer of authentication, the GANC may attempt to validate that the IMSI received during registration correlates to the IMSI received by the security gateway during IPsec tunnel establishment. This request can be sent to the security gateway itself or to the AAA infrastructure for validation. Additionally, the GANC is responsible for validating that the user is authorized to access via the unlicensed access point identified in the GAN Registration request as well as checking optional geographical restrictions, for example, using handset provided information.

• Dual-mode handset—The dual-mode handset allows an end user to connect to either a public GSM radio network or a private Wi-Fi or Bluetooth radio network and maintain the same service capabilities, including enhanced 911 (E911), SMS, GPRS, location services (LCS), MMS, Wireless Application Protocol (WAP), and IMS services. The dual-mode handset also contains an IPsec IKEv2 protocol stack for secure communications between the mobile subscriber and operator GANC.

• Standard Wi-Fi access point—A standard Wi-Fi access point (or hotspot) is used to provide Wi-Fi access to a dual-mode handset. This Wi-Fi access point may be enhanced with specific Quality of Service (QoS) and security mechanisms, such as rate-limiting for uplink traffic, Call Admission Control to limit the number of dual-mode handsets that may associate with it, 802.1x encryption, etc.

User Authentication—First Layer of Security
The 3GPP GAN standards specify the authentication and encryption methods for transmission of both voice and data across the public Internet. These standards take advantage of existing RFCs and drafts that already address the secure transport of information over IP. In order to ensure infrastructure security, these specifications address the following aspects:

• Shared air interface and backhaul security—In order to ensure secure transmission of voice and data across public shared media, 3GPP GAN standards specify IPsec with either 3DES or AES encryption to be used as a tunneling mechanism. In addition, 3GPP GAN standards rely on IKEv2 security association and key exchange. IKEv2 was chosen in order to reduce and optimize the tunnel establishment process, as is necessary when considering a roaming scenario where a new IPsec tunnel must be established before call handover can commence.

• Network Address Translation (NAT) traversal—The 3GPP GAN standards comply with RFC 3948, UDP Encapsulation of IPsec ESP Packets. This is a requirement in order to support the majority of home Wi-Fi users relying on either hardware firewalls or NAT routers to protect their devices. The NAT traversal standard also includes a keepalive mechanism sent outside the IPsec tunnel to ensure that the NAT entry is maintained.

• Triplet-based authentication—Although the 3GPP GAN standards introduce many new elements into a mobile operator network, user authentication is inherently based on the 3GPP-defined SIM-based authentication. In order to apply this authentication mechanism into the IP domain, 3GPP has extended RFC 3748 (Extensible Authentication Protocol for RADIUS) to include the use of GSM SIM-based authentication. This standard is known as EAP-SIM (RFC 4186). For a Universal SIM (USIM) in a Universal Mobile Telecommunications Service (UMTS) environment, a similar standard, EAP-AKA (RFC 4187), has been defined. IKEv2 specifies that EAP-SIM be used in conjunction with public key signature authentication. The following procedure (Figure 4) is used to communicate and authenticate based on SIM information.

  1. During the IPsec tunnel authentication phase, the mobile subscriber provides the security gateway with the subscriber’s IMSI. This IMSI is provided during the IKE_AUTH phase.
2. The security gateway embeds this information into a RADIUS access request message toward the AAA infrastructure. The access request includes an embedded EAP message specifying EAP-SIM and the subscriber’s IMSI.

3. The AAA infrastructure converts the RADIUS access request message into a SS7 MAP INVOKE message to the HLR requesting authentication information.

4. The HLR responds to this request including the triplets (response secret response [SRES], random number [RAND], and encryption key [Kc]) required for the SIM Challenge.

5. The AAA infrastructure calculates the MAC_RAND and responds to the security gateway with the triplets provided by the HLR.

6. The security gateway responds to the MS IKE_AUTH_REQ message by specifying a SIM Challenge and providing the triplets (SRES, RAND, and Kc).

7. Upon receiving this message, the mobile subscriber calculates the MAC_RAND value based on the triplets and sends a new IKE_AUTH_REQ message including the EAP response to the SIM Challenge.

8. The security gateway creates a new RADIUS access request message including the EAP response to the SIM Challenge (MAC_RAND) and sends it to the AAA infrastructure.

9. The AAA infrastructure compares the received MAC_RAND value to the previously calculated MAC_RAND. If these values match, the user is considered successfully authenticated and the AAA infrastructure sends an Access Accept message back to the security gateway.

10. The security gateway responds to the mobile subscriber with an EAP Success message.

11. Authentication proceeds are per the IKEv2 standard.
Infrastructure Security—A Gap in the Standards

Although using an IP-based GAN infrastructure allows an operator to quickly extend its services into unlicensed space, many security implications arise as the operator opens the network to the Internet. The GAN standards explicitly address the need for secure authentication and transport of traffic from the end user to the operator domain, but the standards do not address operator infrastructure security specifically. Although at a fundamental level the inclusion of secure IPSec tunnels with EAP-SIM authentication between the mobile subscriber and GANC prevents internal devices from being accessed by unauthenticated end users, both broad-based and targeted attacks against mobile operator infrastructure raise significant concerns and the need for a stronger solution than the standards specify.

One of the more prevalent IP-based attacks is denial of service (DoS), which is the result of either intentional maliciousness or misbehaving (virus-infected, for instance) devices; these attacks can affect numerous elements in the operator environment. Although firewalls and access control lists provide some defense, as viruses and hackers become more intelligent, preventing DoS attacks requires more than just static enforcement of security rules. Intelligent network devices that can detect and mitigate the effects of a DoS attack while still allowing legitimate traffic to pass are required. The components that may be affected by a DoS attack and the implications of such an attack can be summarized into two primary areas:

- **External devices**—External devices in a GAN solution include edge routers, security gateways, and external Domain Name System (DNS) servers that provide access to the operator network. A DoS attack against these elements can result in end users being unable to access the GAN network from private networks, and can originate from either a single Internet point or multiple Internet IP addresses (distributed DoS [DDoS]) simultaneously.

- **Internal devices**—Internal devices in a GAN solution include GANCs, AAA servers, media gateways, MSCs, and HLRs. These devices are accessible only by authenticated users who have established IPSec connections to the security gateway, and, as such, DoS attacks tend to be more targeted at a specific device, either intentionally or inadvertently. Because some of these devices, such as HLRs and MSCs, are used for GSM, GPRS, and GAN access, a DoS attack against these elements can result in end users being unable to access the operator services (voice or data) at all.

Cisco GAN Enhanced Security Gateway Components

Cisco GAN Enhanced Security Gateway architecture addresses both the 3GPP standards as well as enhanced security concerns, giving a mobile operator the option to deploy a complete, modular, and secure GAN solution. Figure 5 shows a logical view of the Cisco architecture, which comprises the following components:

**Figure 5. Cisco GAN Enhanced Security Gateway Logical View**
• **Outside tunnel threat defense system**—This system provides infrastructure security against broad-based attacks by unauthenticated users. These attacks, including DDoS attacks, can originate from one or many Internet IP addresses, and can be the result of widespread viruses or other malicious activity.

• **Cisco GAN Security Gateway**—The gateway component provides the 3GPP-specified functions of terminating IPSec tunnels and enabling authentication through the AAA infrastructure. In addition to standard functions, the Cisco GAN Security Gateway provides additional value-add functions, including:
  
  – **IKEv2 Security Association Call Admission Control (CAC):** CAC can be used to protect against overload conditions by restricting the total number of IPSec tunnels, restricting the total amount of resource utilization, and dropping calls when specified thresholds are exceeded.

  – **Hardware encryption:** GAN presents a unique traffic model as compared to other IPSec implementations because of the large number of simultaneous tunnels and low traffic rate and idleness per tunnel. In order to support incoming calls and manage roaming quickly and efficiently, handsets maintain an IPSec tunnel to the Security Gateway even when there is no voice or data call proceeding. Cisco GAN Security Gateway hardware supports hardware encryption to provide scalability.

  – **IKEv2 protocol-layer DoS protection:** Cisco GAN Security Gateway provides DoS protection against IKEv2-specific attacks. This allows the Security Gateway to not trigger continual authentication requests to the AAA infrastructure for a user whose access request has already been rejected.

  – **Quality of service:** QoS and bandwidth management features allow the GAN Security Gateway to deliver high transmission quality for time-sensitive applications such as voice and video. Each packet is tagged to identify the priority and time sensitivity of its payload, and traffic is sorted and routed based on its delivery priority. Cisco GAN Security Gateway supports classification based on Layer 2, Layer 3, or Layer 4 header information, allowing a mobile operator to create multiple service classes by marking the IP precedence.

  – **VRF support:** Virtual Routing and Forwarding (VRF) capability allows a mobile operator to “virtualize” the GAN Security Gateway architecture. VRF allows for the segregation of traffic into unique routing instances, with individual routing, security, and QoS policies per VRF. By virtualizing this architecture, the mobile operator can use the same infrastructure for multiple purposes, including an evolution to an IMS Tunnel Termination Gateway or IMS Packet Data Gateway.

• **Inside tunnel threat defense system**—This system provides infrastructure security against targeted attacks by authenticated users. These attacks, including DoS attacks, typically originate from one IP address, and can be the result of widespread viruses or other inadvertent activity (including misbehaving peer-to-peer protocols) by a subscriber.

• **AAA infrastructure**—The AAA subsystem provides the infrastructure that a mobile operator requires to perform EAP-SIM authentication. The Cisco AAA infrastructure for GAN solutions has two main components:

  – **Cisco Access Registrar:** Cisco Access Registrar provides the AAA function, which includes the authentication of the user, as well as proxying EAP messages to an external MAP gateway. For EAP-SIM authentication, Cisco Access Registrar can also compute the MAC_RAND based on the triplets provided by the HLR.

  – **MAP gateway:** The Cisco IP Transfer Point (ITP) provides the MAP gateway function into the SS7 network. The MAP gateway converts a RADIUS message into a MAP message, allowing a system running EAP-SIM, such as the Cisco GAN Security Gateway, to obtain authentication through the standard triplet-based challenge/response process inherent in a GPRS/UMTS network through the HLR.
Cisco Inside/Outside Tunnel Layered Threat Defense System

The unique Cisco Defense In Depth solution consists of numerous functions designed to provide additional network infrastructure security. The modular system allows operators to build an architecture that meets their specific security requirements. Figure 6 depicts the entire system provided by the Cisco GAN Enhanced Security Gateway.

The components in the Cisco GAN Enhanced Security Gateway architecture provide the mobile operator with static and dynamic detection of threats, as well as immediate adaptation of rules to block malicious traffic without affecting the flow of legitimate subscriber GSM voice and GPRS data traffic. The layered threat defense system is divided into five layers (Figure 7).

- **Outside tunnel threat defense system**—Protection for external DNS servers and security gateways from unauthenticated user attacks:
  
  **Layer 1**
  - *Cisco IOS® Firewall*: Cisco IOS Firewall provides static firewall rules to prevent access by unauthenticated users. In general, only IKE and ESP traffic related to IPSec should be reaching the Security Gateway. Cisco IOS Firewall protects the Cisco GAN Security Gateway from non-IPSec attacks.

  **Layer 2**
  - *Cisco Guard DDoS mitigation appliances and Cisco Traffic Anomaly Detectors*: Cisco Guard appliances provide DDoS detection and blocking. They work in conjunction with the Cisco Traffic Anomaly Detector module, which analyzes and determines traffic that is categorized as outside of “normal.” This module can provide either recommended actions or dynamically enforce rules for traffic that statistically or behaviorally deviates from normal.

  **Layer 3**
  - *Cisco GAN VPN Module*: The Cisco GAN VPN Module provides secure access control to the GAN by providing an authentication and authorization mechanism through IPSec, and EAP-SIM over RADIUS. For additional security functions, refer to the *Cisco GAN Enhanced Security Gateway Components* section.

- **Inside tunnel threat defense system**—Protection for internal DNS servers, GANCs, AAA infrastructure, HLRs, MSCs, and media gateways from authenticated user attacks:

  **Layer 4**
  - *Cisco PIX® Firewall*: The Cisco PIX Firewall provides hardware-assisted static firewall protection for targeted network attacks against operator infrastructure. This infrastructure includes operator AAA, HLR, MSC, and other IP-based 2.5G and 3G network equipment.

  **Layer 5**
  - *Cisco Intrusion Prevention System (IPS)*: The Cisco IPS provides internal protection against targeted network attacks. The location of the IPS in the Cisco GAN Enhanced Security Gateway provides for higher-layer traffic inspection and attack mitigation. Utilizing signature databases, the Cisco IPS inspects digital signatures associated with a protocol, and looks for malicious activity in that protocol.

- **Cisco Monitoring and Response System (MARS)**—The Cisco MARS platform ties the Defense In Depth solution together by analyzing information received from various network elements to provide a mobile operator with a complete security picture. By maintaining a correlation between inside IP address, outside IP address, and IMSI, the Cisco MARS can push dynamic security policies to multiple elements in the network.

**Why Cisco**

The Cisco GAN architecture provides several advantages for mobile operators that want to offer GAN dual-mode services.

**Carrier-Class Stability and Security**

The Cisco GAN Enhanced Security Gateway solution provides carrier-class stability and security—IPSec VPNs, DDoS-attack mitigation, firewall, intrusion detection and prevention, network monitoring, and attack correlation—in addition to the 3GPP GAN standard requirements. Other unique benefits of this solution include:

- **Proven platform**—The Cisco GAN security architecture employs components currently in use at some of the world’s largest DSL, cable, and mobile operators. Security services modules are deployed in Cisco 7600 Series routers, among the most widely deployed edge routers.

- **More effective DoS-attack detection**—The solution employs statistical DoS, which is more flexible and accurate than the IKEv2-based DoS mechanisms used in other vendors’ solutions.
Figure 6. Cisco GAN Enhanced Security Gateway Physical Architecture

Figure 7. Cisco GAN Enhanced Security Gateway Layered
• **Ability to support multiple services**—One physical security gateway can support multiple virtual gateways, one for each application. Therefore, mobile operators can capitalize on the same infrastructure investment to introduce additional fixed mobile convergence applications in the future.

• **High availability**—Advanced load-balancing techniques enable more efficient use of physical resources and higher service availability.

**Scalability**

GAN-based services require a highly scalable security solution, potentially creating a tunnel for every handset. For a large operator, this scenario can add up to millions or even tens of millions of sessions. Additionally, traffic models for these IPSec tunnels deviate significantly from normal functions of a VPN concentrator, because most established tunnels have low throughput and high idleness.

The Cisco GAN Enhanced Security Gateway solution provides scalability in multiple dimensions. For example, the Cisco IOS Software Server Load Balancing (SLB) solution helps enable the Cisco GAN Security Gateway blade and threat defense systems—both outside and inside the tunnel—to scale to provide more throughput. Cisco IOS SLB provides serverfarm-based load balancing, allowing for multiple hardware modules to be scaled both in chassis and cross chassis. Tunnel termination, in contrast, scales to accommodate more subscribers. As the subscriber base grows and usage patterns change, these dimensions do not always grow at the same rate. Therefore, the ability to scale each dimension separately helps the mobile operator support more subscribers and more minutes without unnecessary capital expense.

**Industry Leadership in Encryption and Authentication**

Cisco Systems® is an industry leader in several important respects:

- Among the largest network security vendors in the world, Cisco has shipped IPSec solutions for more than 10 years.
- Cisco is one of only nine vendors that participated in an Internet Computer Security Association (ICSA) evaluation for IPSec IKEv2 VPN technology since February 2005. ICSA Labs, an independent division of Cybertrust, sets standards for information security products and certifies more than 95 percent of the installed base of antivirus, firewall, IPSec, cryptography, and PC firewall products in the world today.
- Cisco is coauthor of the original EAP-SIM framework, used for subscriber authentication.

**Conclusion**

GAN dual-mode services to the home give mobile operators the opportunity for a significant competitive advantage by accelerating fixed-mobile substitution, increasing penetration, and reducing turnover. The Cisco GAN architecture provides an essential prerequisite for dual-mode services—protecting the mobile operator’s voice network from threats originating from the Internet. Because the security infrastructure that is used to offer dual-mode services can be reused for other services, including IMS, the investment in the Cisco GAN solution provides a competitive advantage for tomorrow’s services as well as today’s. For more information about the Cisco GAN architecture and the Cisco GAN solution, visit: http://www.cisco.com/go/mobile.