(2) All-IP Network Architecture for 4G Mobile Communications

Ravi Jain, Muhammad Mukarram Bin Tariq, James Kempf and Toshiro Kawahara

DoCoMo USA Labs has been conducting research on the next-generation 4G mobile network architecture. This article presents the rationale and overview of our vision of the 4G mobile network architecture, as well as a snapshot of our recent research accomplishments and the direction our work is taking.

1. Introduction

NTT DoCoMo has demonstrated worldwide leadership in mobile communications by being the first carrier to deploy a Third-Generation (3G) mobile communication service, Freedom Of Mobile multimedia Access (FOMA). The incipient success of FOMA indicates the widespread appeal of the ability to communicate anywhere, anytime, seamlessly, with high-speed multimedia and Internet access and a wide variety of services.

We view the shift to Fourth-Generation (4G) fundamentally not in terms of air interface technology or protocols but in terms of the innovative services and applications that will be made available to users. This orientation led us to several design choices in conceiving the next-generation architecture.

This article will briefly summarize the architecture and rationale of the next-generation 4G mobile network and highlight some of the key directions we have been pursuing to implement the 4G system in the areas of protocols, security and cryptography, programmability and support for value-added services. Finally we conclude with a brief summary and directions for further work.

2. Rationale and Key Features

Our orientation toward applications led us to several design decisions. The first design decision is that the 4G architecture is based on supporting the Internet Protocol (IP), particularly IPv6, as a fundamental construct in all parts of the system. IP is attractive not because of technical superiority or decreased cost, but because of the widespread proliferation of personnel, tools and support for development of applications, ease of integrating the telecommunications network with Internet applications, and the ability to build end-to-end applications even if some of core network facilities are lacking or delayed. The second design decision is that the architecture is defined by a layered family of Application Programming Interface (API)s, some public and some private, but all designed to facilitate access to the network resources in a secure, useful and billable manner. The third guiding decision, dictated by the migration of intelligence from the core to the edges of the system in both IP-based and public-switched networks, is that the 4G architecture will consist of intelligent terminals, an intelligent Radio Access Network (RAN), a dumb core, and intelligent control and overlay networks.

If 4G will be defined by the availability of innovative applications, where will those applications come from? It is unlikely that any single company can develop a stream of "killer apps" rapidly. This means that the 4G network must be a programmable IP-based network in some sense [1] so that third-party service providers can contribute new services.

We believe this requires a second waist [1] analogous to the IP waist [2], but at the interface between applications and the middleware of the 4G system to hide the heterogeneity of the protocol stack and software layers between the application and the IP waist. This second waist must offer a high level abstraction of essential support services, such as Authentication, Authorization and Accounting (AAA), Quality of Service (QoS), and location information, which most intelligent end services will rely on. Current efforts towards Web Services, including Web Service Definition Language (WSDL), Simple Object Access Protocol (SOAP) and UDDI (or their variants and successors) are steps in this direction.

Figure 1 shows a set of layered APIs, which we believe are required for 4G networks. In fact, the set of APIs can be viewed as an application-oriented abstract specification of the architecture. Access to every layer of the API is strictly controlled to ensure resources to be made available in a secure and billable manner.

APIs enable access to network resources in a well-defined manner, but this raises the question of where resources can most efficiently be located in the network. We observe that while the

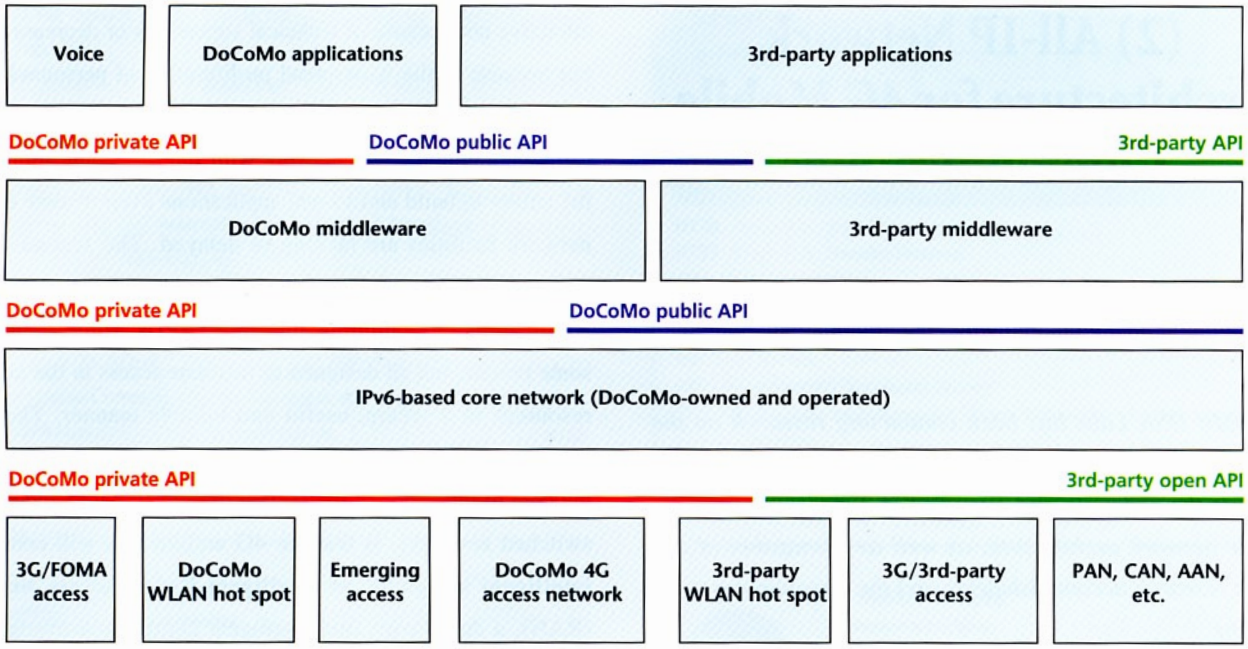

Figure 1 4G architecture as defined by layered APIs

telecommunications core has become less intelligent, the opposite is the case for terminals, the RAN, the control layer, and emerging overlay networks. We believe that the 4G architecture must be consistent with these historical trends, with smart terminals, smart RAN, a dumb core, a smart control layer, and smart overlays.

3. Architecture Overview

The design rationale presented above leads naturally to a 4G architecture such as depicted schematically in **Figure 2**. It consists of four main abstract layers: overlay, high-level control, IP core network, and access.

The IP core network has relatively little intelligence. Thus most core network functions, such as routing, are handled by existing and evolving Internet protocols. The high-level control layer focuses on functions that can be made available to applications and overlay network elements, such as AAA, and policy management. Below the core is a collection of access networks that serve different market sectors and needs.

Finally, the overlay layer provides higher-layer functionality and support services for applications, such as Application Layer Multicast (ALM), location services, and content distribution. This overlay can in fact be split into two tiers, with functions that are relatively close to the Core (such as ALM) in the lowertier and functions such as location services in the higher tier.

In **Figure 3** we show the functional aspects of the 4G architecture in greater detail. The four horizontal abstract layers discussed above are further sub-divided and some of the functions in each specified. The functions are grouped into vertical collections we call "facets" that contain key capabilities that span all or several layers (Security, Transport, etc.). Note that separate parallel planes deal with OA&M and user equipment; both have a similar layering and facet arrangement.

The lower layer (Layers 1 to 2.5) access network provides physical and MAC level connectivity, access control and local mobility, and QoS-aware switching capability. This layer is topped with an IP-based access network, which provides IP connectivity along with access control, integrated QoS management, address assignment, and inter-subnet handover capability with the Fast Mobile Internet Protocol (FMIP). The core network layer consists of a pure IP core. The overlay layer is divided into two tiers of support services. Tier 1 support services are mostly related to the transport functionality of the network whereas Tier 2 support services provide functionality for end-services.

4. Research Initiatives Towards 4G

Our research has focused on developing new protocols and algorithms supporting key functions needed to implement the 4G architecture. In the following sections, we briefly describe some of the technical solutions we have developed, selecting topics that reflect our architectural design decisions.

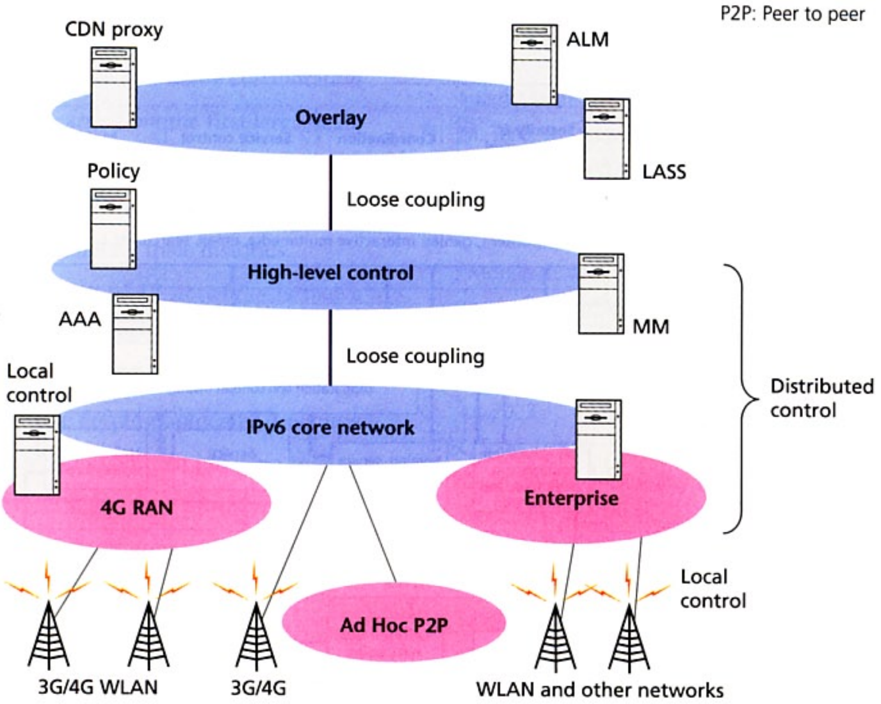

Figure 2 Schematic of 4G architecture

4.1 Mobility Management in 4G (FMIP)

Mobility management is a central concern for any mobile network design. In Fig.3, several layers of the architecture deal with mobility. At the IP layer, while the basic functions of mobility management will be embodied in the Mobile IP protocol for IPv6, we believe extensions are required to provide better performance and local control.

FMIP is used for streamlining handoffs at the IP layer in the IP access network. FMIP handoff restricts the handover signaling to the access network, thus requiring no global coordination. This is in accordance with the goal of distributed local control for access networks in the overall architecture, as presented earlier.

FMIPv6 consists of two parts:

1) Exchange of a Proxy Router Solicitation and Proxy Router Advertisement prior to handover to discover subnets and access routers in the vicinity,
2) After handover, a Fast Binding Update message sent to the previous access router by the mobile node to initiate the tunnel.

Completion of the FMIP protocol by the Internet Engineering Task Force (IETF) is near; for details see [3]. The key point to note is that fast handoff is achieved by FMIP without any centralized control.

4.2 Security and Cryptography

Our current research in the area of security focuses on cryptographic algorithms that are designed to allow DoCoMo to provide security services to a large client base in diverse environments. Among our contributions are three techniques designed to simplify management of DoCoMo's Public Key Infrastructure (PKI): Hierarchical Identity-Based Cryptography (HIBC), certificateless Public-Key Cryptography (PKC), and aggregate signatures. We have also developed techniques to enable secure content distribution: an exceptionally efficient micropayment scheme that uses the concept of microcredits, and a stream authentication scheme that allows an intermediate proxy to transcode the stream dynamically without breaking end-to-end security. This is covered in greater detail in the article “Cryptography and Security Technology” of this special issue. In addition, we have conducted research on access network authentication (in particular the Protocol for carrying Authentication for Network Access (PANA) [4]), Mobile Firewalls, AAA for ad hoc networks, and other AAA and security mechanisms [5]-[6].

4.3 Programmability

Our current research towards programmability focuses on both the overlay layer and the control layer.

For overlay functions we have introduced techniques to

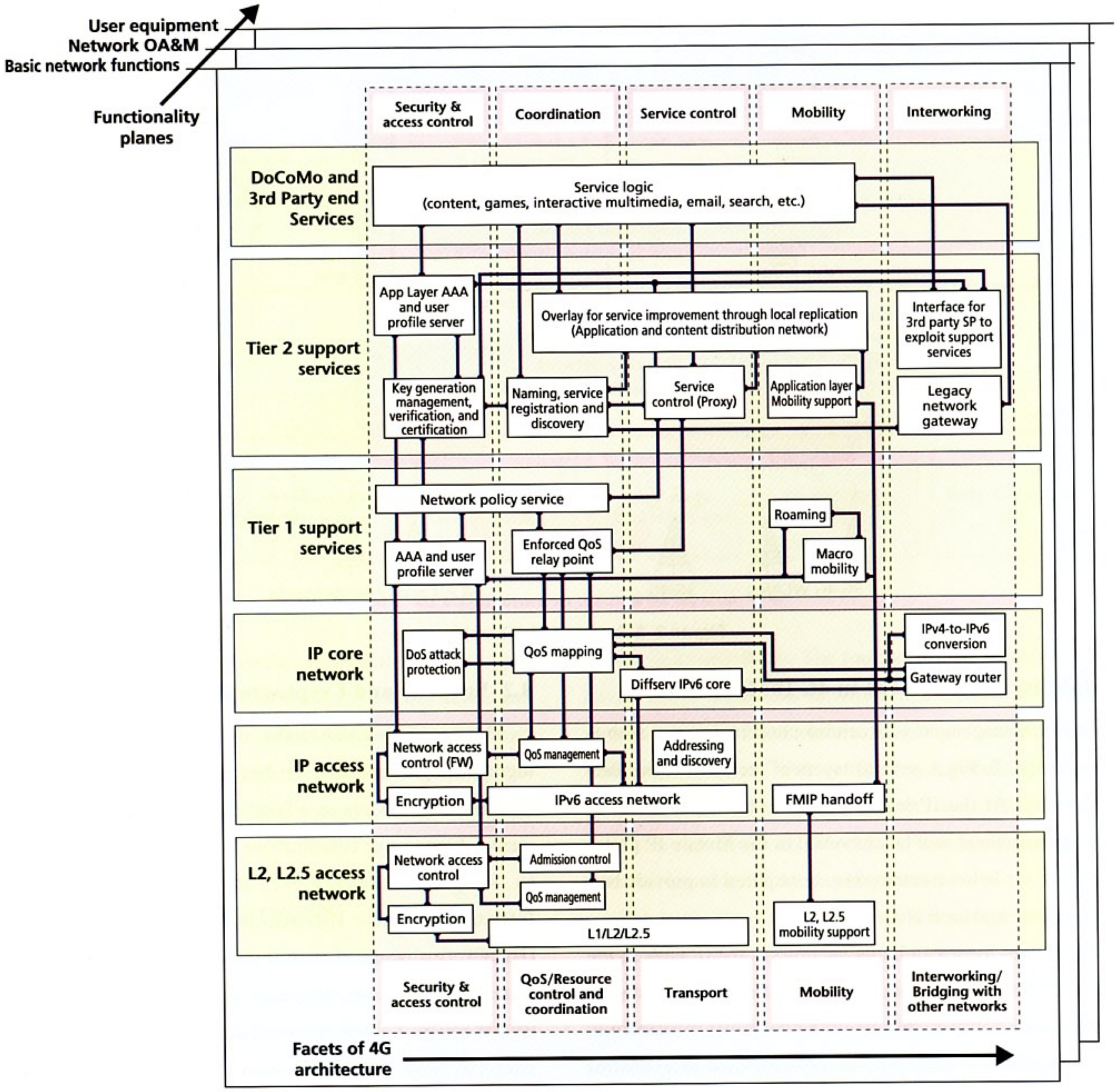
User equipment
Network OA&M
Basic network functions
Functionality planes
Security & access control
Coordination
Service control
Mobility
Interworking
DoCoMo and 3rd Party end Services
Service logic (content, games, interactive multimedia, email, search, etc.)
Tier 2 support services
App Layer AAA and user profile server
Overlay for service improvement through local replication (Application and content distribution network)
Interface for 3rd party SP to exploit support services
Key generation management, verification, and certification
Naming, service registration and discovery
Service control (Proxy)
Application layer Mobility support
Legacy network gateway
Tier 1 support services
Network policy service
Roaming
AAA and user profile server
Enforced QoS relay point
Macro mobility
IP core network
DoS attack protection
QoS mapping
IPv4-to-IPv6 conversion
Diffserv IPv6 core
Gateway router
IP access network
Network access control (FW)
QoS management
Addressing and discovery
Encryption
IPv6 access network
FMIP handoff
L2, L2.5 access network
Network access control
Admission control
QoS management
Encryption
L1/L2/L2.5
L2, L2.5 mobility support
Security & access control
QoS/Resource control and coordination
Transport
Mobility
Interworking/ Bridging with other networks
Facets of 4G architecture

Figure 3 Layers and facets of the All-IP 4G architecture

allow Content Distribution Networks (CDNs) to serve mobile users better. Yet we were also aware that in most cases CDNs offer little programmability to content providers or developers of advanced applications. Thus we have also been defining an API for CDN functions, which is important for making the 4G overlay network more attractive to content providers.

For control layer programmability we have been defining an API for the Mobile IP protocol. This API allows the applications running on the mobile nodes to learn about host mobility, and allows the applications to control certain management parameters.

4.4 Support for Value-Added Services

Our research on network support for value-added services has focused on the important category of leveraging and managing the user's location information, focusing on location estimation, prediction and privacy.

Our work on location estimation does not deal with traditional areas of radio signal estimation but with higher-layer issues, in particular how information from multiple location sources (e.g., for dual-mode terminals on Wireless Local Area Networks (WLANs) and FOMA) can be integrated to obtain better estimates. We have developed and experimentally evaluated new algorithms that can improve the accuracy of commercially available systems.

For location prediction we have focused on predicting the next cell that a mobile user will visit. In a research collaboration with Dartmouth College we have carried out the first large-scale experimental evaluation of classical prediction algorithms using a campus WLAN installation, and achieved a median prediction accuracy of about 72% for users with long trace histories.

One privacy-related vulnerability of IP networks is that a user's approximate geographical location can be inferred from an IP address subnet. To address this problem, we developed location privacy techniques that use cryptographically generated IP addresses to hide the user's location without inducing tunneling or sub-optimal routing overhead. Our prototype router implementation indicates our techniques have minimal processing overhead and virtually no overhead in terms of packet size.

5. Conclusion

DoCoMo USA Labs has been conducting cutting-edge research to bring our vision of the all-IP 4G architecture to fruition. We believe that an all-IP architecture is the best choice for developing commercially viable services, and have now developed the key enabling technologies to support this 4G architecture including mobility management, security and cryptography, network programmability, and value-added services. Other articles in this special issue examine these technologies in greater depth and detail.

Our research continues to develop techniques that flesh out further key areas of this architecture, including advanced mobility management, a comprehensive AAA system, and cryptographic techniques. These initiatives will enable DoCoMo to realize new business models and roles (e.g., a trusted broker status), seize upon new business opportunities in emerging markets and service areas, and evolve strategies for migrating from 3G to 4G.

References

[1] R. Jain: "4G Services, Architectures and Networks; Speculation and Challenges," Keynote address, International Conf. on Mobile Data Management (MDM), Jan. 2003.

[2] S. Deering: "Watching the Waist of the Protocol Hourglass," IAB Meeting, 51st IETF, London, UK, Aug. 2001.

[3] R. Koodli: "Fast Handovers for Mobile IPv6," draft-ietf-mobileip-fast-mipv6-07.txt, Internet Draft, work in progress, 2003.

[4] D. Forsberg, et al: "Protocol for Carrying Authentication for Network Access (PANA)," IETF Internet Draft. Work in progress. Jun. 2003.

[5] J. Kempf, P. C. Hwang and S. Okazaki: "CertBU: Certificate-based Techniques for Securing Mobile IPv6 Binding Updates," Proceedings Internetworking 2003, San Jose, CA., Jun. 2003.

[6] J. Arkko, J. Kempf, et al: "Secure Neighbor Discovery (SEND)," IETF Internet Draft. Work in progress. Jun. 2003.

[7] M. Tariq, R. Jain and T. Kawahara: "Mobility aware server selection for streaming multimedia content distribution networks," Proc. International Web Caching Workshop (IWCW), Sep. 2003.

[8] A. Yokote, A. Yegin, M. Tariq, G. Fu, C. Williams and A. Takeshita: "Mobile IP API" Pro. IEEE Mobile and Wireless Communications Networks (MWCN), Sep. 2002.

[9] Y. Gwon, T. Kawahara and R. Jain: "Robust Indoor Location Estimation and Tracking of Stationary and Mobile Users," submitted for publication Jul. 2003.

[10] L. Song, D. Kotz, R. Jain and X. He: "Evaluating Location Predictors with Extensive Wi-Fi Mobility Data," Poster presentation, Proc. ACM MobiCom, Sep. 2003.

[11] J. Trostle, H Matsuoka, M. Tariq, J. Kempf, R. Jain and T. Kawahara: "Cryptographically Protected Prefixes for Location Privacy in IPv6 networks," submitted for publication. Sep. 2003.

Abbreviations

AAA: Authentication, Authorization and Accounting
AAN: Automobile Area Network
All-IP: All Internet Protocol
ALM: Application Layer Multicast
API: Application Programming Interface
CAN: Campus Area Network
CDN: Content Distribution Network
FMIP: Fast Mobile Internet Protocol
FOMA: Freedom Of Mobile multimedia Access
HIBC: Hierarchical Identity-Based Cryptography
IEFT: Internet Engineering Task Force
IP: Internet Protocol
IPv6: IP version6
LAN: Local Area Network
LASS: Location-Aware Service Server
MAC: Medium Access Control
MM: Mobility Management
OA & M: Operations, Administration and Management
PAN: Personal Area Network
PANA: Protocol for carrying Authentication for Network Access
PKC: Public-Key Cryptography
PKI: Public Key Infrastructure
P2P: Peer to Peer
QoS: Quality of Service
RAN: Radio Access Network
SOAP: Simple Object Access Protocol
UDDI: the Universal Description, Discovery, and Integration
WLAN: Wireless Local Area Network
WSDL: Web Service Definition Language