Mobile Application Protocols

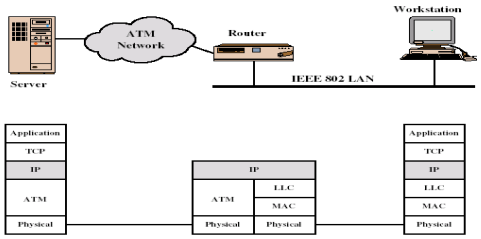
David Tipper
Associate Professor
Graduate Telecommunications and Networking Program
University of Pittsburgh

dtipper@mail.sis.pitt.edu

Slides 18

Mobile Protocols

- Consider existing Internet protocol stack
- Can we just reuse higher layer protocols and replace layer 2 and 1 with wireless layer?



Mobile Services

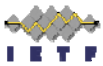
- Goal: enable computers to maintain Internet connectivity while moving from one Internet attachment point to another (wired or wireless)
- How goal accomplished depends on user needs
  - Nomadic use: Internet connection is terminated each time the user moves and a new connection is initiated when the user reconnects
    - For example, move laptop from work to home
    - New temporary IP address is assigned by DHCP
    - Note user is accessing services not providing them
  - Mobile use: wants to offer services from mobile node, user's point of attachment changes dynamically and want all connections automatically maintained despite the change
    - Change the IP-address?
      - adjust the host IP address depending on the current location
      - DNS updates take to long time
      - TCP connections break, security problems
    - Modify IP to support mobility - Mobile IP

Motivation for Mobile IP

- Note IP address is used for dual purposes
- Routing
  - based on IP destination address, network prefix (e.g. 129.13.42) determines physical subnet
  - change of physical subnet implies change of IP address to have a topological correct address or needs special entries in the routing tables
  - How would a sender know IP address changes?
    - Change of all routing table entries to forward packets to the right destination
    - Does not scale with the number of mobile hosts and frequent changes in the location, security problems
- End point identifier
  - Socket includes IP address
  - TCP connection can't survive change in IP address
  - Affects application performance

Mobile IP Goals (IETF RFC 3344)

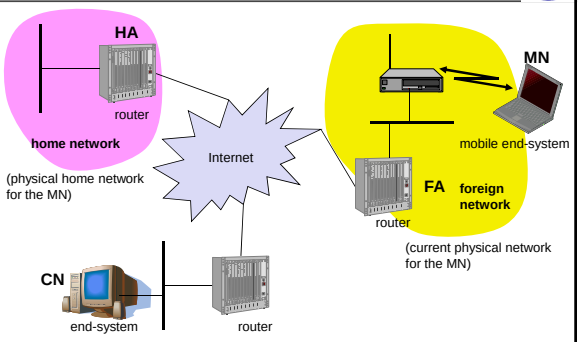
- Transparency
  - mobile end-systems *keep* their IP address
  - Invisible to higher layer protocols
  - continuation of communication after interruption of link possible
  - point of connection to the fixed network can be changed
- Compatibility
  - support of the same layer 2 protocols as IP
  - no changes to current end-systems and routers required
  - mobile end-systems can communicate with fixed systems
- Security
  - authentication of all registration messages
- Efficiency and scalability
  - Minimize additional messages to the mobile system required (connection typically via a low bandwidth link, conserve battery power)
  - world-wide support of a large number of mobile systems
- See http://www.ietf.org



Terminology

- Mobile Node (MN)
  - System (node) that can change the point of connection to the network without changing its IP address
- Correspondent Node (CN)
  - Communication partner (can be fixed or mobile)
- Home Network (HN)
  - Particular network where mobile node's *home IP address* resides
- Foreign Network (FN)
  - Network where mobile node is visiting
- Home Agent (HA)
  - System in the home network of the MN, typically a router, that manages IP layer mobility.
- Foreign Agent (FA)
  - System in the current foreign network of the MN, typically a router that manages the network mobility

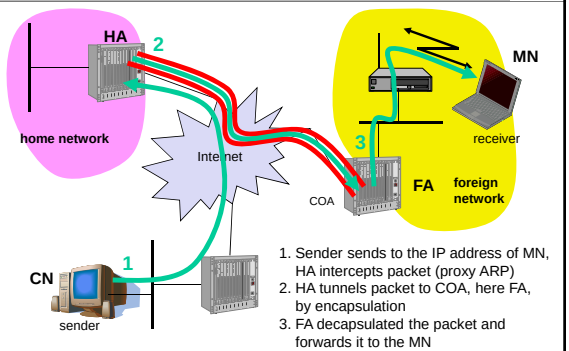
Example Scenario



Mobile IP Structure

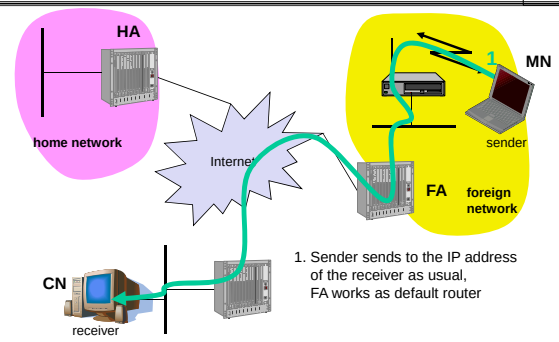
- Home address
  - Long term IP address assigned to MN in the Home Network
  - remains unchanged regardless MN location,
  - used by DNS to locate MN
- Care-of Address (COA)
  - IP address in the Foreign Network that is the reference pointer to the MN when it is visiting the FN
  - Usually IP address of Foreign Agent
  - Option for MN to act as it's own FA in which case it is a co-located COA
- How Does Mobile IP deliver the data??
  - Home Agent
    - registers the location of the MN, reroutes IP packets sent to the MNs home address to the COA using a encapsulation/tunneling procedure
  - Foreign Agent (FA)
    - forwards the tunneled packets to the MN within the FN

Data transfer to the mobile system

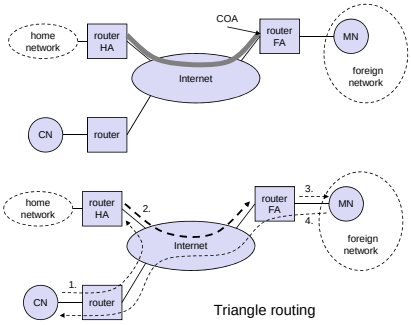


1. Sender sends to the IP address of MN, HA intercepts packet (proxy ARP)
2. HA tunnels packet to COA, here FA, by encapsulation
3. FA decapsulated the packet and forwards it to the MN

Data from the mobile system



Mobile IP Overview



Network Integration

- Mobile IP requires the following capabilities
  - Discovery :
    - MN uses discovery procedure to determine is it has changed networks and to identify prospective home and foreign agents
  - Registration:
    - mobile node uses an authenticated registration procedure to inform home agent of its care-of address
  - Tunneling
    - used to forward IP packets from home address to a care-of address

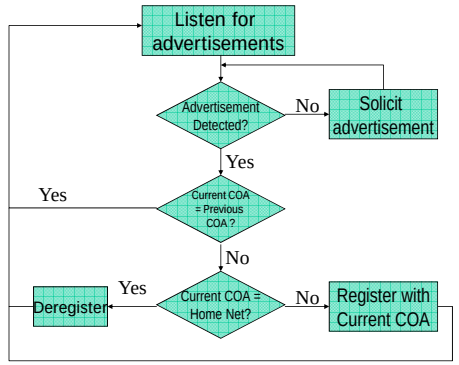
Discovery

- Mobile node is responsible for ongoing discovery process
  - HA and FA periodically send advertisement messages into their physical subnets
  - MN listens to these messages and detects, if it is in the home or a foreign network
    - Uses network prefix of agents IP address
  - MN reads a COA from the FA advertisement messages
- A mobility extension to ICMP is used for advertisement

Agent Advertisement

- Advertisement contains the relevant information
  - Is it a Home Agent or a Foreign Agent?
  - COA associated with the FA
  - Busy or not
  - Whether minimal encapsulation is permitted
  - Whether reverse tunneling is permitted (later)
  - Whether registration is mandatory
- The Agent Advertisement packet is a broadcast message on the subnet
- The same agent may act as both a HA and a FA
- If the MN gets an advertisement from its HA, it **must** deregister its COA's and enable a gratuitous ARP
- If a MN does not "hear" any advertisement, it must solicit an agent advertisement using ICMP

Discovery Search Flow Chart



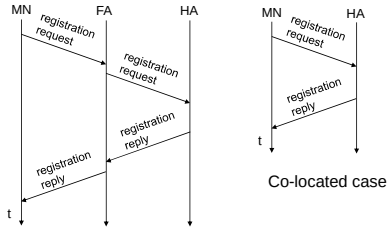
Co-Located Addresses

- If mobile node moves to a network that has no foreign agents, or all foreign agents are busy, it can act as its own foreign agent
- Mobile agent uses co-located care-of address
  - IP address obtained by mobile node associated with mobile node's current network interface
- Means to acquire co-located address:
  - Temporary IP address through an Internet service, such as DHCP
  - May be owned by the mobile node as a long-term address for use while visiting a given foreign network

Registration

- Purpose:
  - Inform the HA about the COA
  - FA can obtain approval from the HA to provide service to the MN
  - Authenticated to prevent malicious attacks
- Procedure
  - Mobile node sends registration request to foreign agent requesting forwarding service
  - Foreign agent relays request to home agent
  - Home agent accepts or denies request and sends registration reply to foreign agent
  - Foreign agent relays reply to mobile node
  - Note MN can act as co-located FA

Registration



Typical Registration Procedure

Co-located case

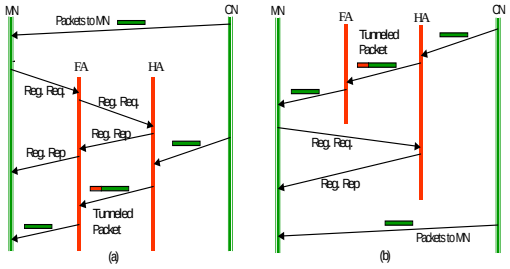
Registration

- UDP packets are used for registration
- A *nonce* called an identification field is used in the request and another in the reply to prevent replay attacks
- HA creates a ***binding*** between the MN's home address and the current COA
  - This binding has a *fixed lifetime*
  - MN should re-register before the expiration of the binding
- Registration reply indicates if the registration is successful or not
- Rejection is possible by either HA or FA due to
  - Insufficient resources, header compression not supported, HA unreachable, too many simultaneous bindings, failed authentication
- Upon a valid registration, the HA should create an entry for a mobile node that has:
  - Mobile node's care of address
  - Identification field
  - Remaining lifetime of registration

Registration

- Each Foreign Agent maintains a visitor list containing the following information:
  - Link layer address of the mobile node
  - Mobile node's home IP address
  - UDP registration request source port
  - HA IP address
  - Identification field
  - Registration lifetime
  - Remaining lifetime of pending or current registration
- Deregistration
  - Deregistration involves "registering" the home address with the HA
  - If multiple COAs are not explicitly requested, each new registration request wipes out the previous binding.

Registration Examples



MN moves to a foreign network

MN moves to Home Network

Packet Encapsulation by HA

- Forwarding packets is achieved by encapsulation (tunneling)
  - Virtual pipe between tunnel entry point (HA) and tunnel termination point (FA)
- The datagram from the CN is made the payload of ***another*** IP packet
- Three types of encapsulation are provided
  - IP in IP encapsulation
  - Minimal encapsulation (reduces overhead)
  - Generic routing encapsulation
    - Pre Mobile IP formulation

Encapsulation I

- Mandatory implementation (mandatory, RFC 2003)
- The outer header uses IP-in-IP as the protocol type
- The whole tunnel is equivalent to one hop from the original packet's point of view IP-in-IP-encapsulation tunnel between HA and COA

ver.	IHL	DS (TOS)	length				
IP identification			flags	fragment offset			
TTL		*IP-in-IP*	IP checksum				
IP address of HA							
Care-of address COA							
ver.	IHL	DS (TOS)	length				
IP identification			flags	fragment offset			
TTL		lay. 4 prot.	IP checksum				
IP address of CN							
IP address of MN							
TCP/UDP/ ... payload							

Encapsulation II

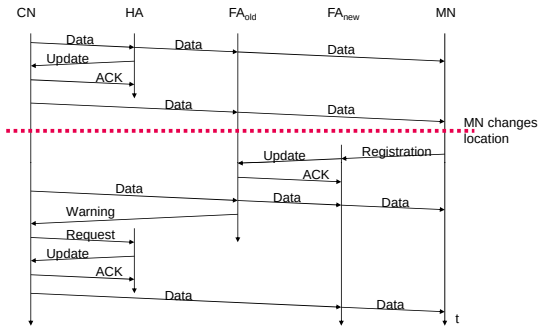
- Minimal encapsulation (optional)
  - avoids repetition of identical fields
  - e.g. TTL, IHL, version, DS (RFC 2474)
  - only applicable for unfragmented packets, no space left for fragment identification

ver.	IHL	DS (TOS)	length				
IP identification			flags	fragment offset			
TTL		*min. encap.*	IP checksum				
IP address of HA							
care-of address COA							
lay. 4 protoc.	S	reserved	IP checksum				
IP address of MN							
original sender IP address (if S=1)							
TCP/UDP/ ... payload							

Optimization of packet forwarding

- Triangular Routing
  - sender sends all packets via HA to MN
  - higher latency and network load
- “Solutions”
  - sender learns the current location of MN
  - direct tunneling to this location
  - HA informs a sender about the location of MN
  - big security problems!
- Change of FA
  - packets on-the-fly during the change can be lost
  - new FA informs old FA to avoid packet loss, old FA now forwards remaining packets to new FA
  - this information also enables the old FA to release resources for the MN

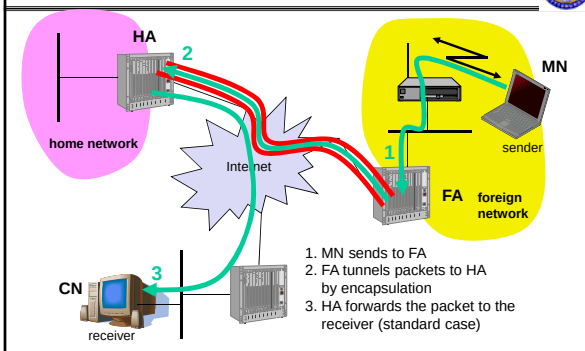
Change of Foreign Agent



Mobile IP with reverse tunneling

- Router accept often only “topological correct“ addresses (firewall!)
  - a packet from the MN encapsulated by the FA is now topological correct
  - furthermore multicast and TTL problems solved (TTL in the home network correct, but MN is to far away from the receiver)
- Reverse tunneling does not solve
  - the reverse tunnel can be abused to circumvent security mechanisms (tunnel hijacking)
  - optimization of data paths, i.e. packets will be forwarded through the tunnel via the HA to a sender (double triangular routing)
- The standard is backwards compatible
  - the extensions can be implemented easily and cooperate with current implementations without these extensions
  - Agent Advertisements can carry requests for reverse tunneling

Reverse tunneling (RFC 3024)



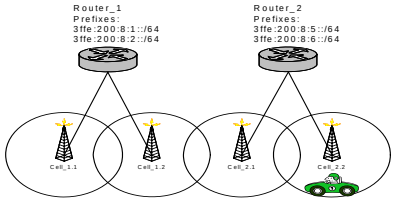
1. MN sends to FA
2. FA tunnels packets to HA by encapsulation
3. HA forwards the packet to the receiver (standard case)

Mobile IP and IPv6

- Mobile IP was developed for IPv4, but IPv6 simplifies the protocols
  - security is integrated and not an add-on, authentication of registration is included
  - COA can be assigned via auto-configuration (DHCPv6 is one method), every node has address autoconfiguration
  - no need for a separate FA, **all** routers perform router advertisement which can be used instead of the special agent advertisement; addresses are always co-located
  - MN can signal a sender directly the COA, sending via HA not needed in this case (automatic path optimization)
  - ``soft IP hand-over'', i.e. without packet loss, between two subnets is supported
    - MN sends the new COA to its old router
    - the old router encapsulates all incoming packets for the MN and forwards them to the new COA
    - authentication is always granted

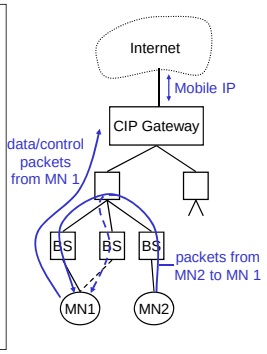
IP Micro-mobility support

- Micro-mobility support:
  - Efficient local handover inside a foreign domain without involving a home agent
  - Reduces control traffic on backbone
  - Especially needed in case of route optimization
- Mobile IP doesn't address this problem



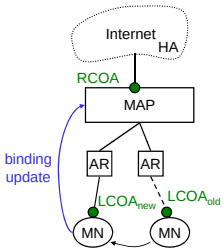
Cellular IP

- Operation:
  - CIP Nodes maintain routing entries (soft state) for MNs
  - Multiple entries possible
  - Routing entries updated based on packets sent by MN
- CIP Gateway:
  - Mobile IP tunnel endpoint
  - Initial registration processing
- Security provisions:
  - all CIP Nodes share network key
  - MN key: (net key, IP addr)
  - MN gets key upon registration



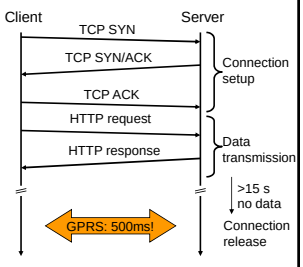
Hierarchical Mobile IPv6 (HMIPv6)

- Operation:
  - Network contains mobility anchor point (MAP)
    - mapping of regional COA (RCOA) to link COA (LCOA)
  - Upon handover, MN informs MAP only
    - gets new LCOA, keeps RCOA
  - HA is only contacted if MAP changes
- Security provisions:
  - no HMIP-specific security provisions
  - binding updates should be authenticated



Transport Layer

- HTTP (used by web services) typically uses TCP
  - Reliable transport between client and server required
- TCP
  - Steam oriented, not transaction oriented
  - Network friendly: time-out
    - → congestion
    - → slow down transmission
- Well known – TCP guesses quite often wrong in wireless and mobile networks
  - Packet loss due to transmission errors
  - Packet loss due to change of network
- Result
  - Severe performance degradation

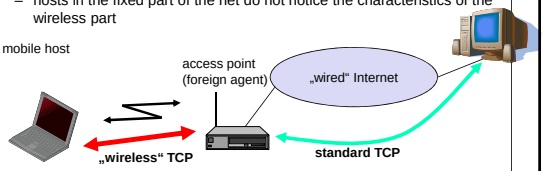


Mobility on TCP-mechanisms

- TCP assumes congestion if packets are dropped
  - typically wrong in wireless networks, here we often have packet loss due to *transmission errors*
  - furthermore, *mobility* itself can cause packet loss, if e.g. a mobile node roams from one access point to another while there are still packets in transit to the wrong access point and forwarding is not possible
- The performance of an unchanged TCP is poor
  - however, TCP cannot be changed fundamentally due to the large base of installation in the fixed network, TCP for mobility has to remain compatible
  - Learn to live with
    - Data rates: 64 kbit/s up, 115-384 kbit/s down; asymmetry: 3 -1000 times, periodic allocation/release of channels
    - High latency, high jitter, packet loss

Basic Method Indirect TCP

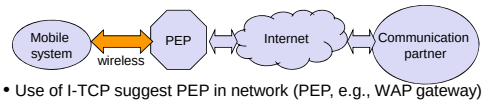
- Indirect TCP or I-TCP segments the connection
  - no changes to the TCP protocol for hosts connected to the wired Internet, millions of computers use (variants of) this protocol
  - optimized TCP protocol for mobile hosts
  - splitting of the TCP connection at, e.g., the foreign agent into 2 TCP connections, no real end-to-end connection any longer
  - hosts in the fixed part of the net do not notice the characteristics of the wireless part



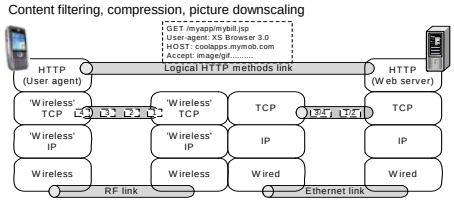
Indirect TCP II

- Fine tuning TCP on mobile link
  - Suggestions from use in 2.5G networks (i-mode, GPRS)
    - Large (initial) sending windows, large maximum transfer unit, selective acknowledgement, explicit congestion notification, time stamp, no header compression
    - WAP 2.0 ("TCP with wireless profile")
- Advantages
  - no changes in the fixed network necessary, no changes for the hosts (TCP protocol) necessary, all current optimizations to TCP still work
  - transmission errors on the wireless link do not propagate into the fixed network
  - simple to control, mobile TCP is used only for one hop between, e.g., a foreign agent and mobile host
- Disadvantages
  - loss of end-to-end semantics, an acknowledgement to a sender does now not any longer mean that a receiver really got a packet, foreign agents might crash
  - higher latency possible due to buffering of data within the foreign agent and forwarding to a new foreign agent
- Other TCP options have been proposed but not widely adopted

Performance Enhancing Proxy (PEP)



- Use of I-TCP suggest PEP in network (PEP, e.g., WAP gateway)
  - Transport layer
    - local retransmissions and acknowledgements
  - Application layer
    - Content filtering, compression, picture downscaling



WAP - Wireless Application Protocol

- Goals
  - deliver Internet content and enhanced services to mobile devices and users (mobile phones, PDAs)
  - independence from wireless network standards
  - open for everyone to participate, protocol specifications will be proposed to standardization bodies
  - applications should scale well beyond current transport media and device types and should also be applicable to future developments
- Platforms
  - 2G: GSM (900, 1800, 1900), CDMAONE IS-95,
  - 2.5G: GPRS, etc
  - 3G: UMTS, cdma2000 1x EV-DO

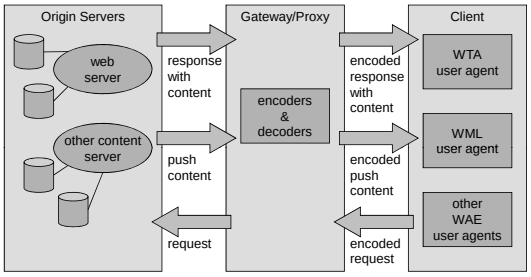
WAP - Wireless Application Protocol

- WAP was initiated as an open standard by an industry consortium
  - WAP Forum, co-founded by Ericsson, Motorola, Nokia, Unwired Planet, further information www.wapforum.org
- WAP based on tailoring existing standards to wireless mobile environment
  - (e.g., IP, XML, HTML, HTTP, etc.)
  - Add when necessary – especially security functions
  - Optimize for efficient use of device resources
  - Enable pesonalization and customization of device and content
- WAP Forum folded into open mobile alliance
  - **Open Mobile Alliance** www.openmobilealliance.org (Open Mobile Architecture + WAP Forum + SyncML + …)
  - Defined WAE

WAE - Wireless Application Environment

- Goals
  - network independent application environment for wireless mobile devices
  - integrated Internet/WWW programming model with high interoperability
- Requirements
  - device and network independent, international support
  - manufacturers can determine look-and-feel, user interface
  - considerations of slow links, limited memory, low computing power, small display, simple user interface (compared to desktop computers)
- Components
  - Architecture: application model, micro-browser, gateway/proxy, server
  - WML: XML-Syntax, based on card stacks, variables, ...
  - WMLScript: procedural, loops, conditions, ... (similar to JavaScript)
  - WTA: telephone services, such as call control, text messages, phone book, ... (accessible from WML/WMLScript)
  - Content formats: vCard, vCalendar, Wireless Bitmap, ...
  - Protocal Layers (WAP)

WAE logical model



WAP Proxy/Gateway

- WAP Architectural specification specifies the term ***WAP Proxy.***
- WAP utilizes proxy technology to optimize and enhance the connection between wireless domain and WWW. WAP proxy provides various functions including:
  - Protocol Gateway: Translates requests from a wireless protocol stack to the WWW protocols. Also performs DNS look up
  - Content Encoders and Decoders: Translate WAP content into a compact format due to slow underlying wireless link and vice versa
  - User Agent Profile Management: Enable personalization and customization of the device
  - Caching proxy: Improves perceived performance and network utilization by maintaining a cache of frequently accessed resources

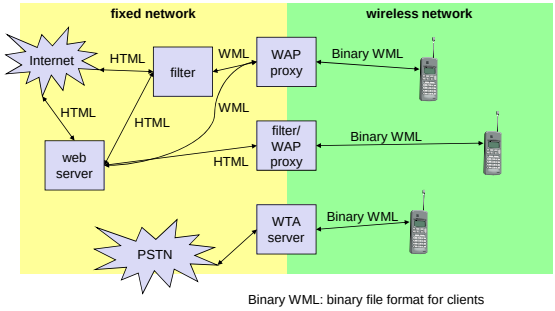
WAP Client

- Primarily includes wireless phones, PDAs, handheld PCs and pagers
- Beginning to support more memory, faster processing power and longer battery life
- Contains a user agent or a mini-browser that implements WAE specification and can execute any WAP compliant application.
- Available in thousands of different models and types. A WAP compliant application written once can reach and be executed on all of theses devices

Application Servers

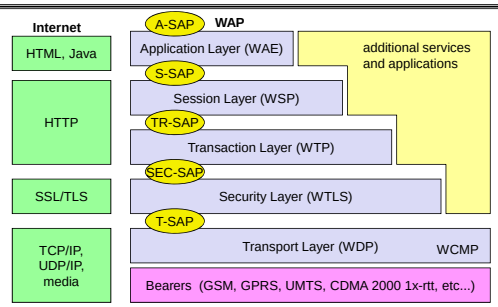
- Real power of WAP lies in the fact that it leverages existing Internet infrastructure to extend reach of applications to millions of users with wireless devices
- Application servers typically consist of three tiers:
  - Web Server; understands HTTP protocol and responds to HTTP requests from the clients. E.g. Apache, iPlanet, Microsoft IIS etc
  - Application Server; encodes elements like personalization, commerce, security and data persistence logic. E.g. iPlanet, WebLogic etc
  - Database Server; used for persistence storage of application data. E.g. Oracle, Sybase, Informix etc

Typical WAP network elements



Binary WML: binary file format for clients

WAP 1.x - Reference model and protocols



WAE comprises WML (Wireless Markup Language), WML Script, WTAI etc.

Wireless Application Environment (WAE)

- WAE specifies an application framework for wireless devices
- WAE elements:
  - WAE User agents – software that executes in the wireless device (e.g., microbrowser requirements)
  - Content generators – applications that produce standard content formats in response to requests from user agents in the mobile terminal (e.g., WML, WML Script)
  - Standard content encoding : defined to allow a WAE user agent to support color, audio, video, images, phone book records, animation etc. (e.g.,Wireless bitmap, V-card format)
  - Wireless telephony applications (WTAI) : collection of telephony-specific extensions for call and feature control mechanisms
  - Push: provides a general mechanism for the network to initiate the transmission of data to applications resident on WAP devices
  - Multimedia Messaging; Multimedia Message Service (MMS) provides for the transfer and processing of multimedia messages such as email and instant messages to WAP devices

Wireless Markup Language (WML)

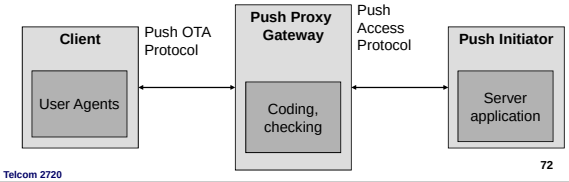
- WML follows deck and card metaphor
  - WML documents subdivided into cards, which specify one or more units of interaction
  - Cards are grouped into decks, a deck is similar to an HTML page, unit of content transmission
  - WML is based on XML
  - Support for navigation among cards and decks – includes provisions for event handling; used for navigation or executing scripts
  - presentation depends on device capabilities
- Features
  - text and images
  - user interaction
  - navigation
  - context management

Wireless Telephony Application (WTA)

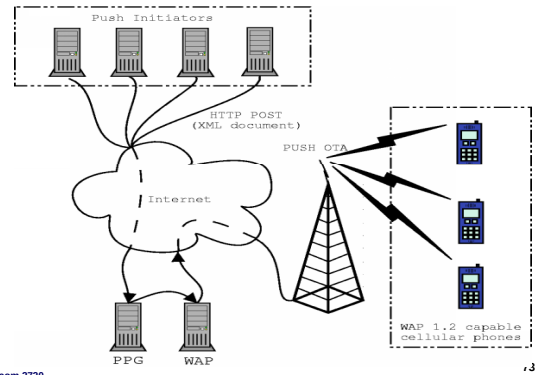
- Collection of telephony specific extensions
- Extension of basic WAE application model
  - access to telephony functions
    - any application on the client may access telephony functions (place/answer call, call forwarding, etc.)
  - content push
    - server can push content to the client
  - handling of network events
    - table indicating how to react on certain events from the network
- Example
  - calling a number (WML)
    `wtai://wp/mc;4126247400`
  - calling a number (WMLScript)
    `WTAPublic.makeCall("4126247400");`

WAP Push Architecture

- Normal client–server model is 'pull' technology (e.g., web browsing)
- In 'push' technology, there is no explicit request from the client before the server transmits its contents.
  - Beneficial for time and location based services. (e.g. traffic alerts of accident ahead on the highway, weather alerts, listing of nearby restaurants, etc)
- WAP Push Architecture
  - Push Access Protocol
    - Content transmission between server and PPG
  - Push OTA (Over The Air) Protocol
    - Simple, optimized, mapped onto WSP



WAP Push Architecture Example



Push Components

- Push Initiator (PI)
  - Responsible for generating the message to be pushed and passing it on to PPG.
  - Messages are all XML based
  - Responsible for authenticating itself with the PPG usually using X.509 based digital client certificates
- Push Proxy Gateway (PPG)
  - PI identification and authentication
  - Parsing of and error detection in push content
  - Translates client address provided by PI into a format understood by mobile network
  - Store the content if client is currently unavailable
  - Notify PI about final outcome of push submission
  - Protocol conversion

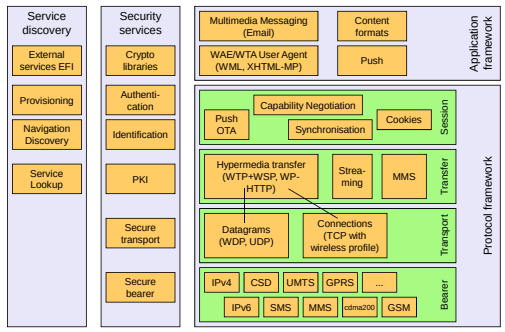
Push Protocols (PAP)

- Push Access Protocol (PAP)
  - XML based communication protocol by which a PI pushes content to mobile network addressing its PPG
  - Supports following operations:
    - Push Submission (PI to PPG)
    - Result Notification (PPG to PI)
    - Push Cancellation (PI to PPG)
    - Push Replacement (PI to PPG)
    - Status Query (PI to PPG)
    - Client Capabilities Query (PI to PPG)
- Push Over the Air (OTA)
  - Provides both connectionless (mandatory) and connection-oriented (optional) services
  - Connectionless service relies upon WSP
  - Connection-oriented service may be provided in conjunction with WSP (OTA-WSP) and HTTP (OTA-HTTP)

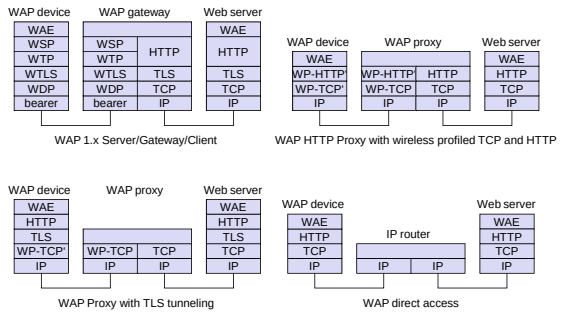
WAP 2.0

- New for developers
  - XHTML with "Mobile Profile" (XHTML-MP)
    - Sub/super set of XHTML (e.g., no frames, telephony support)
  - Wireless Profile HTTP (WP-HTTP)
  - Wireless Profile TCP (WP-TCP)
  - End-to-end encryption with TLS tunneling
    - Supports PKI
  - Data Synchronization with SyncML
  - Capability Negotiation
  - Multimedia messaging.
  - Interface to a storage device.
  - Support for plug-ins in the browser.
- New applications
  - Color graphics
  - Animation
  - Large file download
  - Location based/Smart services
  - Pop-up/context sensitive menus
- Goal: integration of WWW, Internet, WAP, i-mode

WAP 2.0 architecture



WAP 2.0 example protocol stacks



WAP 2.0 Benefits

- WAP has a complete framework for operators to develop and deploy applications
- Large base of users/potential users
- Large base of developers
- Backward compatible with WAP 1.0
- WAP 2.0 borrows from existing standards (XHTML, TCP, etc.)
- WAP is a Thin Client architecture