Celtic Project MEVICO
Tutorial on Customer Edge Swiching

Nicklas Beijar, Aalto University

Project-all f2f meeting
Espoo, 22 November 2012

For MEVICO internal use

Agenda

- Background and Basic Concepts
- Customer Edge Traversal Protocol (CETP)
- Private Realm Gateway (PRGW)
- Customer Edge Switching in EPC
- Protocol Compatibility and Application Layer Gateways
- Prototype
- Conclusions

Tutorial on Customer Edge Switching

Background

Background

- Address shortage
  - Increasing number of UEs
  - UE continuously connected to the Internet
- It is difficult to replace IPv4

→ NATs will be deployed to enable growth

- NATs/firewalls deployed for security reasons
  - Reduce unwanted traffic (attacks, port scans, spam, …)
  - Prevent costs to users due to unwanted traffic

NATs introduce new problems

- NATs prevent inbound connections
  - Conversational applications, distributed applications (Skype, P2P), servers, games
  - Mobile networks replacing fixed networks

→ NAT traversal

NAT traversal is not a solution

- NAT traversal mechanisms includes network layer functionality into each application
  - Several types of NATs → complexity
- NAT traversal creates security risks
  - Utilize “weaknesses” of NATs in an uncontrolled way
- NAT traversal causes additional traffic
  - Each application provide their own NAT traversal
  - Drains battery of mobile device, just for waiting for an inbound connection
  - Relaying of traffic trough TURN relays
- NAT traversal causes additional delay
  - ICE setup can take seconds

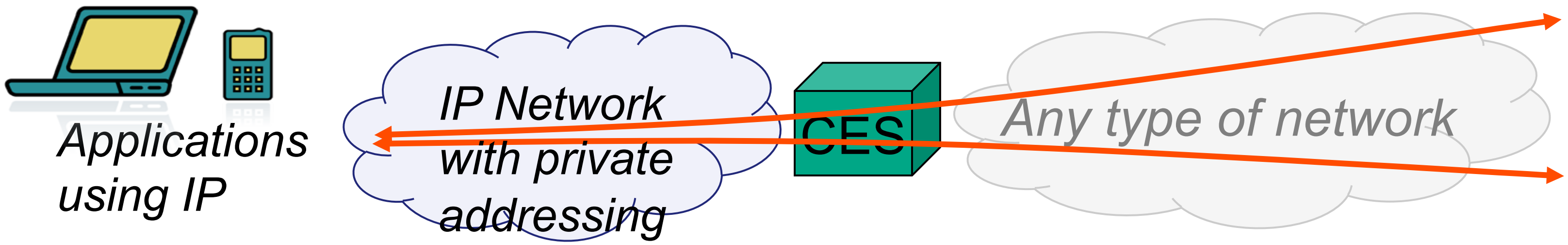
Tutorial on Customer Edge Switching

Customer Edge Switching - basic concepts

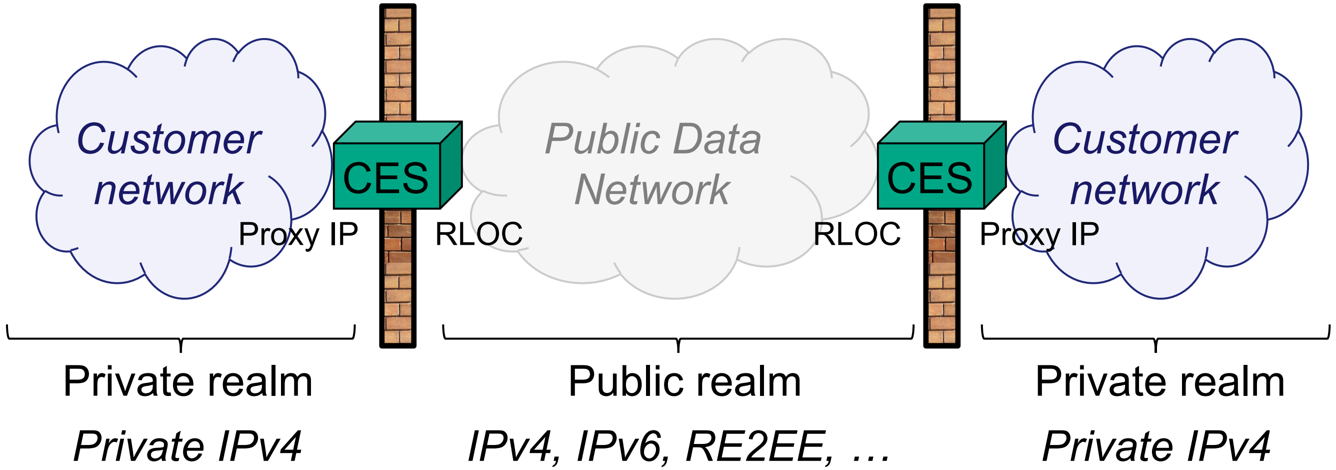
For MEVICO internal use

Goals of CES

- CES is a development of the NAT/firewall concept
- Enable inbound traffic in a secure way (based on policies)
- Avoid NAT traversal mechanisms
- Reduce unwanted traffic in the packet core
- Separate operator network from public network
- Improve scalability (similar to LISP)
- Simplify deployment of new technologies
- No changes to hosts, applications, IP stack



Customer Edge Switching



- Separates the operator network from the public network
- Separates the name from the routing address
- Each network can use different routing and different transport

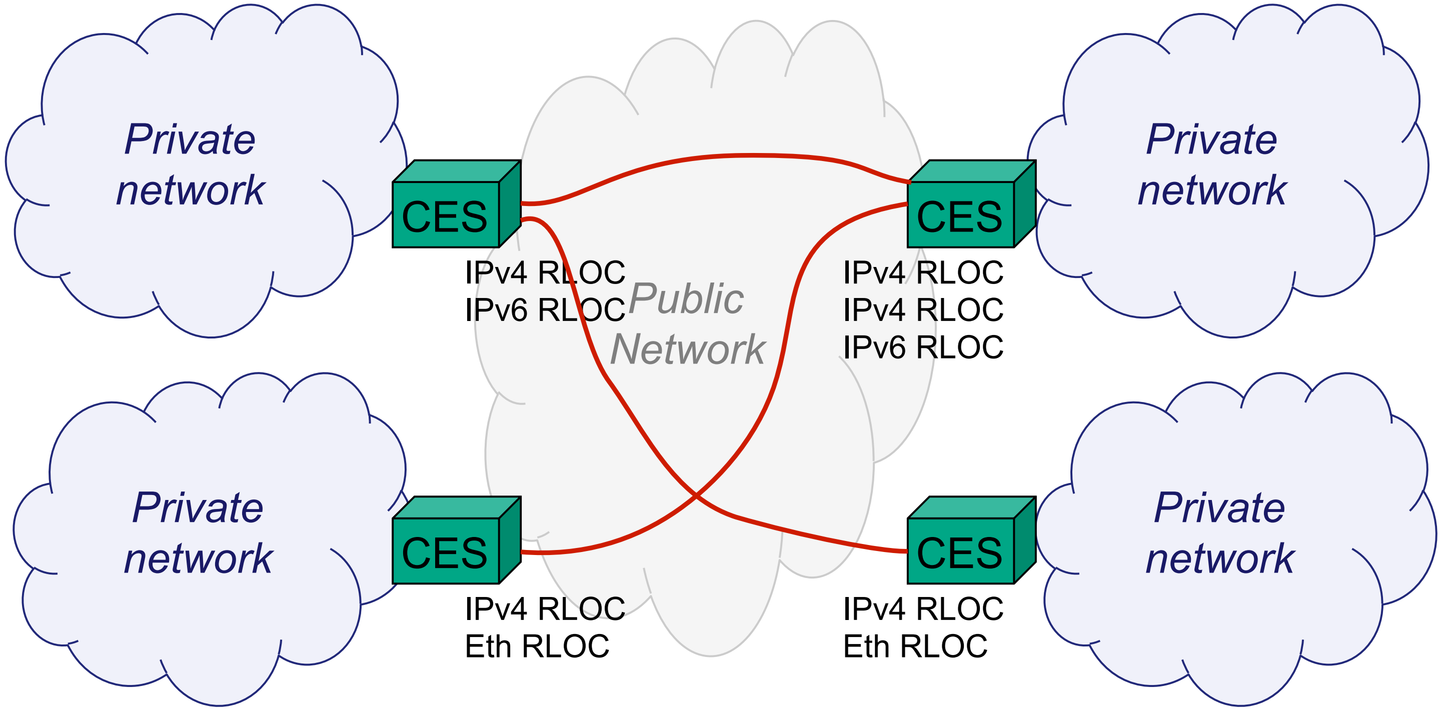
Identifier/locator split

- Names (FQDN) are used by applications and visible to users
- Addresses (IPv4, IPv6, other addresses) are used for routing within a realm
- Identifiers are used by policies to identify users
  - Simplest case: hash of FQDN
  - Anonymous use: random value
  - Identified user: e.g. mobile operator assured IDs

CES is a component of a Internet Trust Framework

- CES integrates functions of a firewall
- In contrast to ordinary firewalls, the CES of the sender and receivers communicate information for security mechanisms
- In addition to CES
  - Reputation System
  - Deep Packet Inspection
  - Policy Management
  - Identity Management

Tunneling of data through public network



Destination URLs

Examples: id RLOC type RLOC

- `dest:1234?eth=12:34:56:78:90:AB`
- `dest:1234?ip=123.45.67.89`
- `dest:1234?ipv6=1234:56::7890:ABCD`

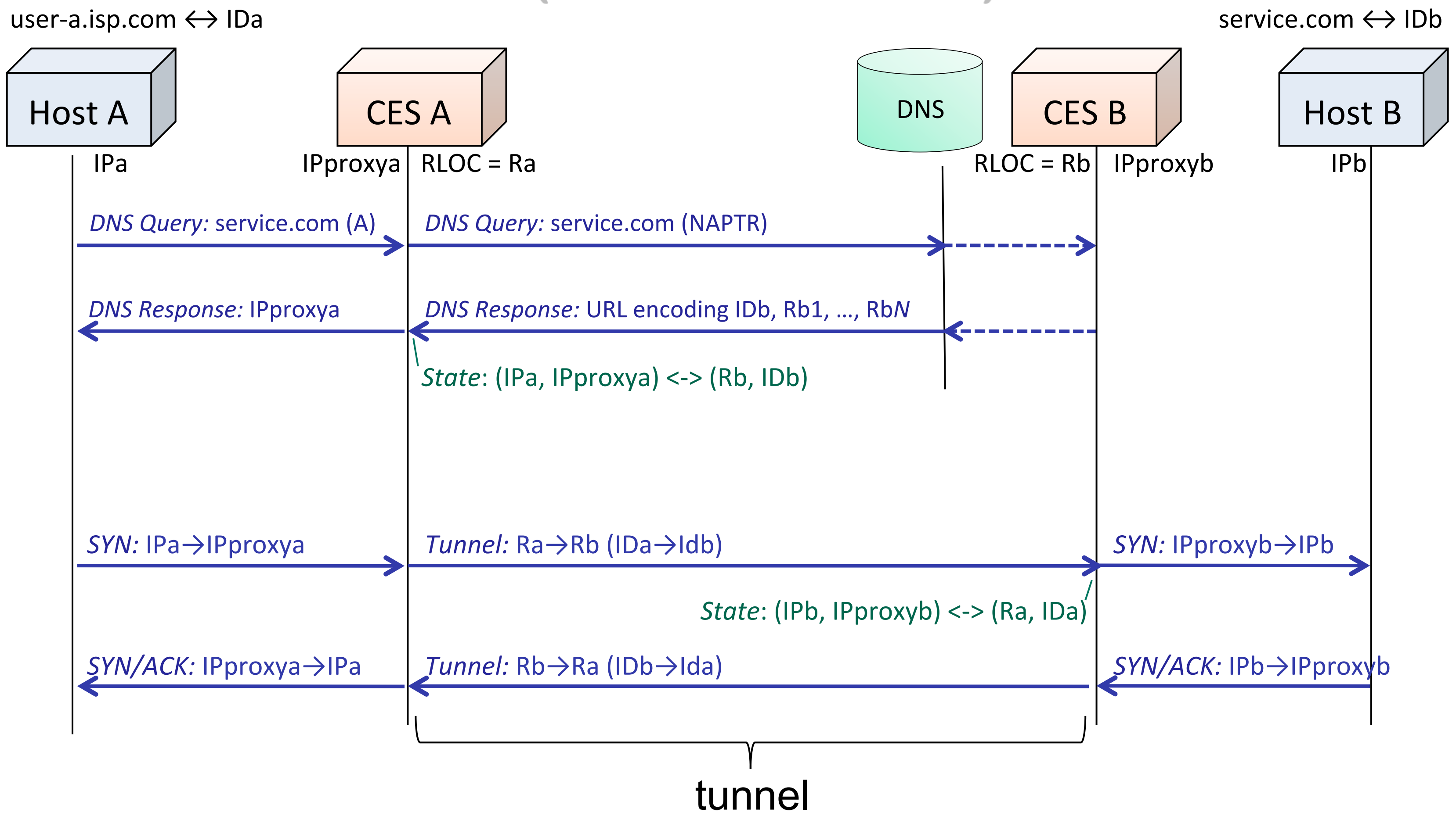
Combined example:

- `dest:1234?eth=12:34:56:78:90:AB&ip=123.45.67.89&ipv6=1234:56::7890:ABCD`

In a NAPTR record:

```
IN NAPTR   100 10 "U" "ID+idprotocol"
     "!^(.*)$!dest:1234?eth=12:34:56:78:90:AB!" .
```

Basic connection setup example (without CETP)

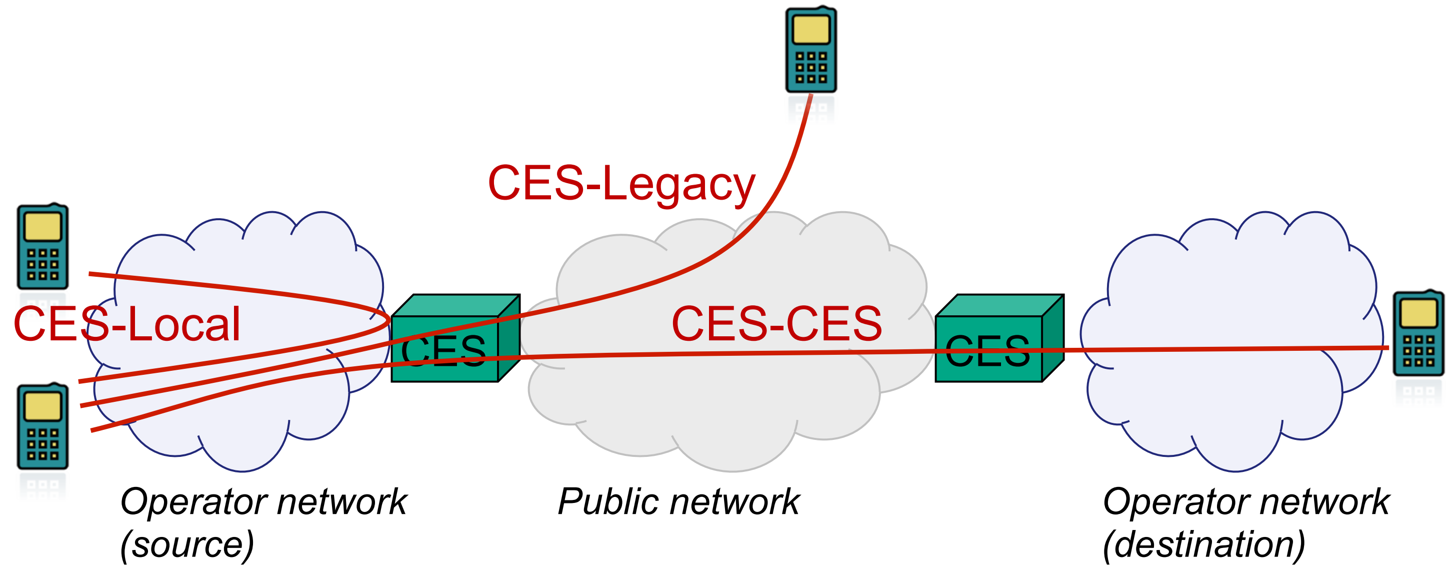


Scenarios

CES-CES: Both users are behind a different CES

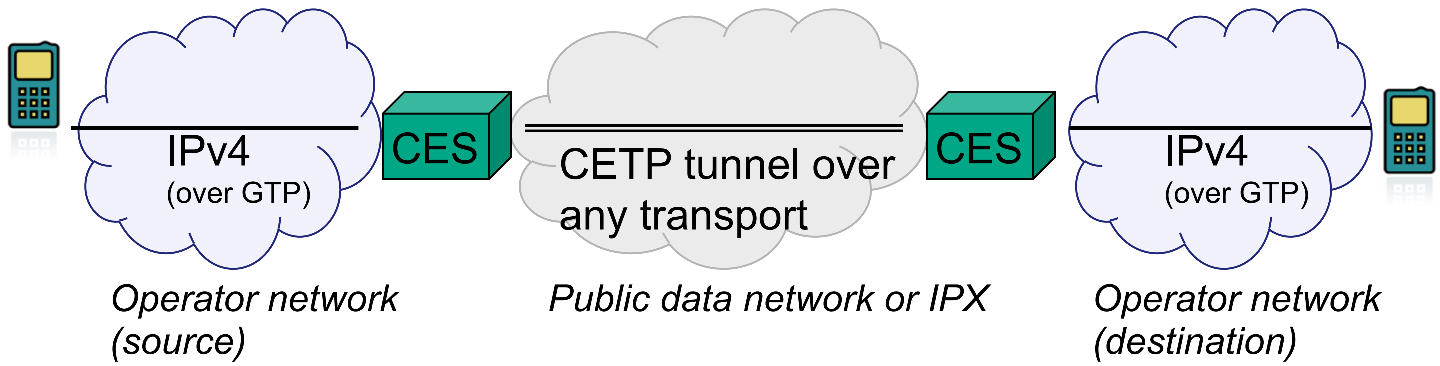
CES-Legacy: One of the users is behind a CES

CES-Local: Both users are behind the same CES

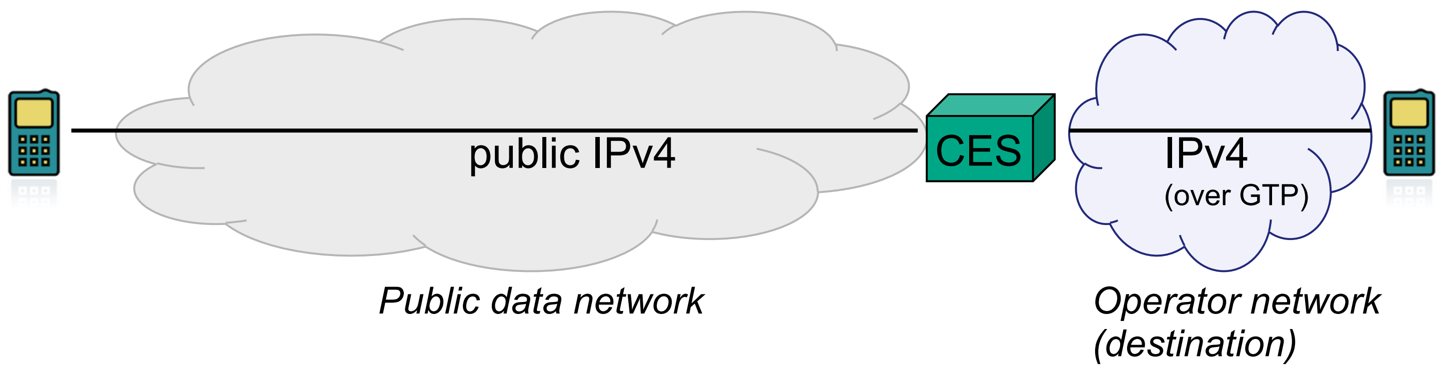


Scenarios

CES-to-CES scenario: Both users are behind a CES



CES-legacy scenario: One user is behind a CES



Tutorial on Customer Edge Switching

Customer Edge Traversal Protocol (CETP)

For MEVICO internal use

Customer Edge Traversal Protocol (CEPT)

- Control signaling between CES devices Control plane
  - Signaling for security methods
  - Robustness and multihoming
  - Control of connection state
  - Negotiation of used ID types
- Tunneling with header compressions Data plane
  - Transports the source and destination IDs
- TLV encoding → Extensible
- Can be transported on top of IPv4, IPv6, Ethernet, ...

Security related methods

- Policy control for accepting traffic
- Return routability checks
- Postponing connection state creation
- Validity checks for IDs
- Negotiation of ID types
- (Limited) attack reporting
- Signatures
- Revocation of invalid IDs

Packet structure

Header	Control TLVs	Payload TLV

- Header
  - Source and destination IDs (type, length, value)
  - Flags and lengths
- Control TLVs
  - For control signaling (Queries, Responses, Acks)
- Payload TLV
  - For tunneling
  - Compressed IP packet or full Ethernet frame

Possible ID types

- Random ID generated by CES based on its own algorithm
- Local (corporate) network certified ID
  - corporate network has its own CA
- Mobile operator assured ID
  - used in ”closed” networks, like in IMS
- User certificate obtained from Mobile Operator
- FQDN
- Temporary ID allocated by a visited network
- Internet of Things objects have their own ID schemas

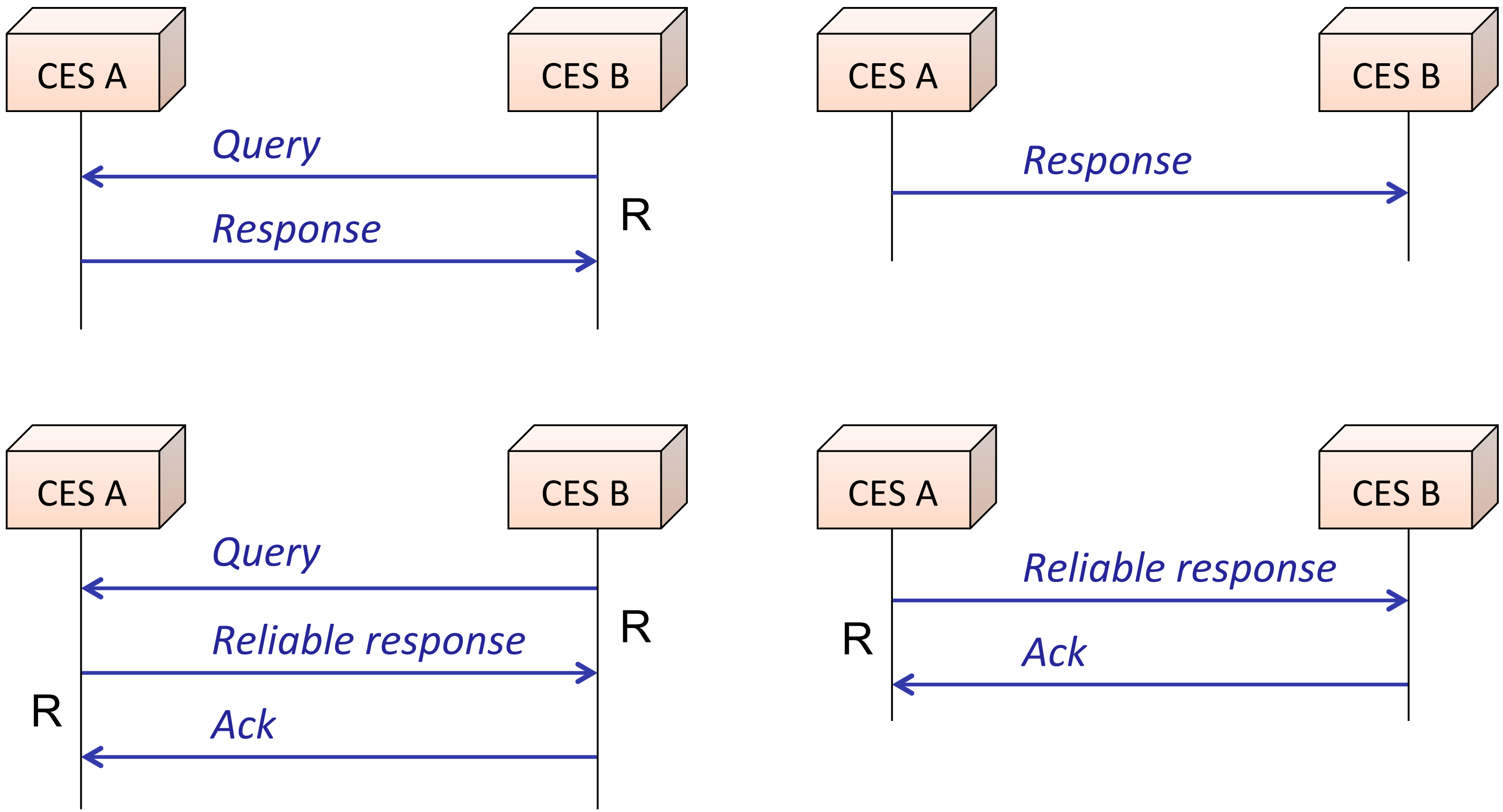
TLV structure

I,D	Res.	TLV Code	Q,R	Length	Value

type

- Message types (Q,R-bits)
  - QR=00 – Query ”Q” (also conveys own value)
  - QR=01 – Response ”R”
  - QR=10 – Acknowledged response ”RR”
  - QR=11 – Acknowledgment ”A”
- Backward compatibility (I,D-bits)
  - Ignore only this or all TLVs, send reply or be silent
- Shorter format available for payload TLVs

Queries and responses



R = Retransmission state

Control TLV overview

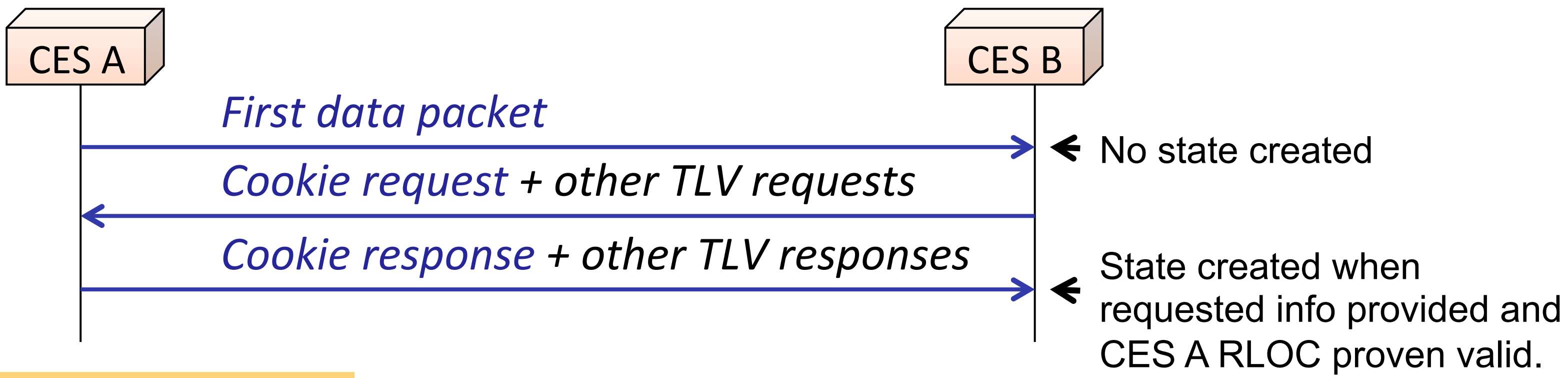
- Reachability TLVs
  - Coveys a list of RLOCs
- Timeout of customer edge state
  - For syncronizing the timers used for removing inactive connection state
- Cookie
  - For postponing creating connection state at inbound edge
- New ID type request
  - For requesting the peer to use a different type of ID
- Address of Certification Authority (such as HSS)
  - Gives a HSS/CA address with which the inbound CES can check the validity of the ID
- FQDN
  - Conveys the FQDN associated with a user (e.g. for reverse DNS queries)
- Header signature
  - Signs the message (over header and all TLVs) to prevent modification
- Unexpected message report
  - Prevents reflector attacks
- Backoff Codes
  - Conveys the reason why the connection was not accepted

Reachability TLVs

- Conveys a list of RLOCs
  - Multiple RLOCs for robustness and multihoming
  - Outbound edge obtains RLOCs of inbound edge from DNS, but signaling needed for RLOCs of outbound edge
- TLVs
  - IPv4 reachability info
  - IPv6 reachability info
  - Ethernet reachability info
- For each group of RLOCs
  - Order and preference
  - List of addresses

Postponing state creation

- The Cookie TLV is used to postpone state creation
  - State information is stored in cookie instead of as connection state
  - Similar to SCTP
- Cookie TLV sent by inbound edge when a new connection is received. The same cookie must be returned by outbound edge.
- Cookie is signed so that it cannot be modified
- Cookie algorithm decided by inbound edge

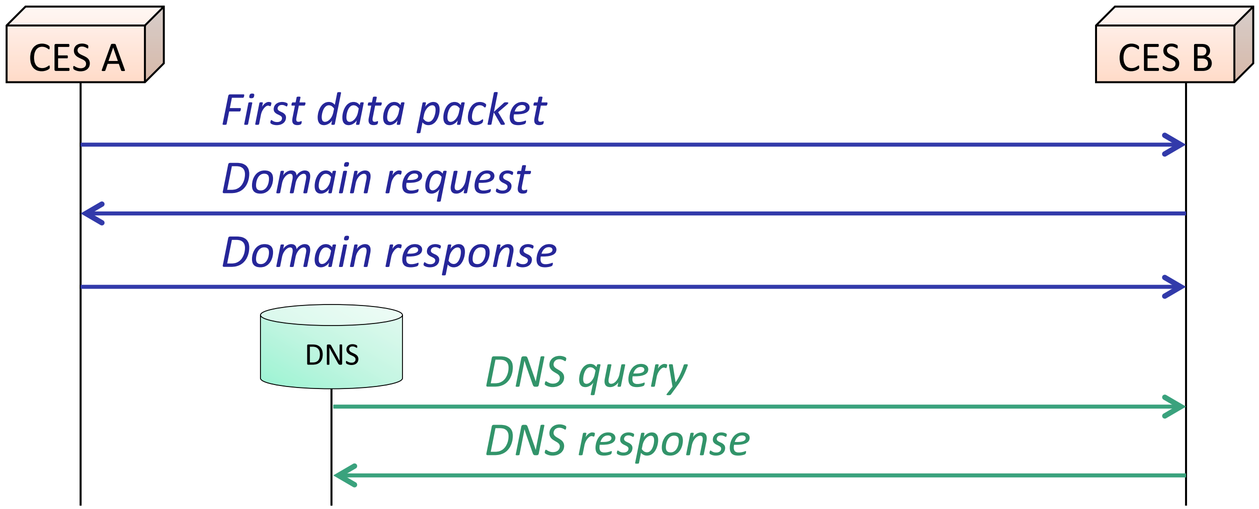


Avoiding source RLOC spoofing

- Spoofed source addresses used in several types of attacks
- Can be done on two levels
  - Forwarding level: if a message sent to the RLOC is returned, then the RLOC must be valid
  - Naming level: check that the RLOC for the given FQDN in DNS (which is trusted) is the same as the used RLOC
- Forwarding level: Reverse routability check with Cookie TLV
  - Checks that the outbound edge's RLOC is correct
  - Connection state created only when the cookie is returned

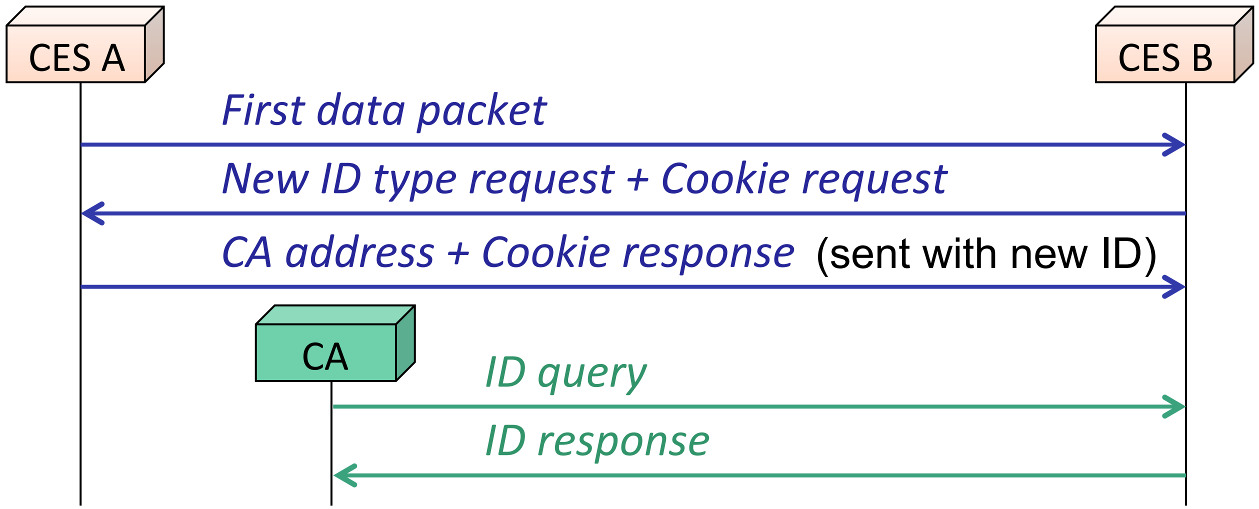
Naming level RLOC check

- Naming level check with Domain TLV
  - Inbound edge requests Domain TLV
  - With the received FQDN, the inbound edge can query DNS and check that the outbound edge's FQDN is one of the ones received from DNS

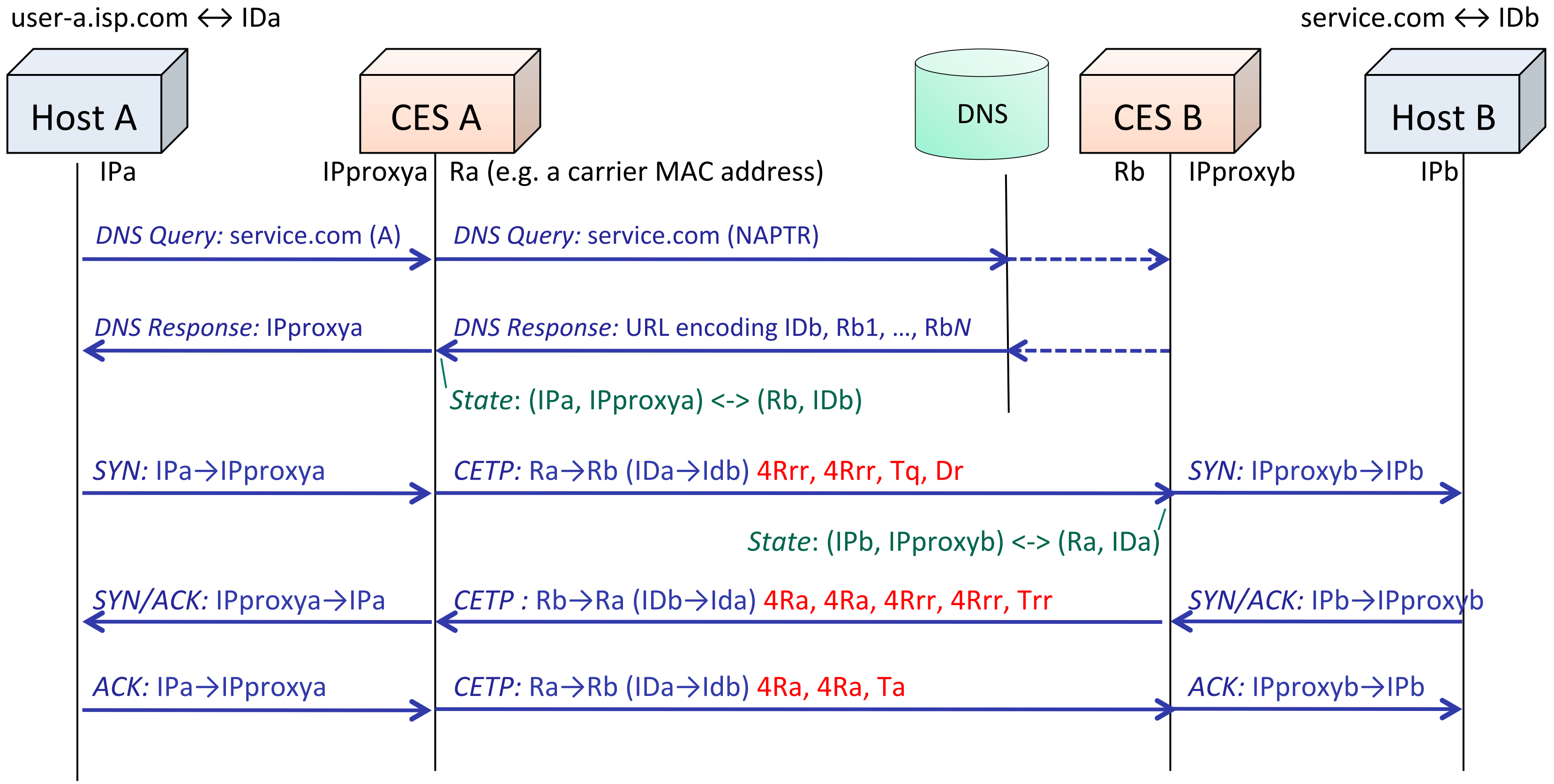


Requesting a given type of ID

- Inbound edge may require a given type of ID
- The CA address TLV returned by outbound edge allows the inbound edge to check the validity of the ID
- Cookie TLV is used to tie together sessions with old and new IDs

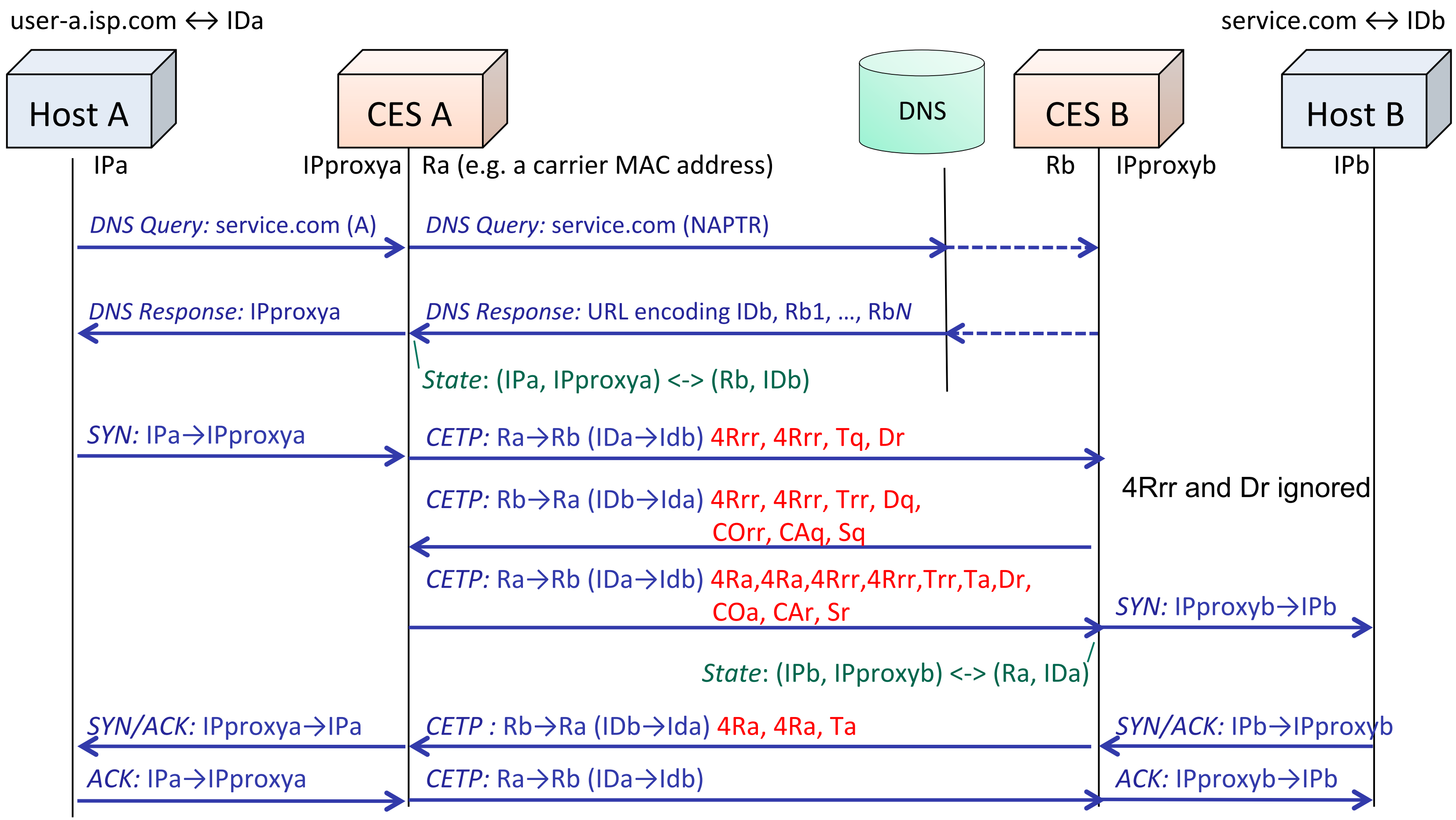


Signaling example (lax policy)



4Rrr=RLOC Reliable reply, 4Ra=RLOC Ack, Tq=Timeout query, Trr=Timeout reliable reply, Ta=Timeout ack, Dr=Domain reply

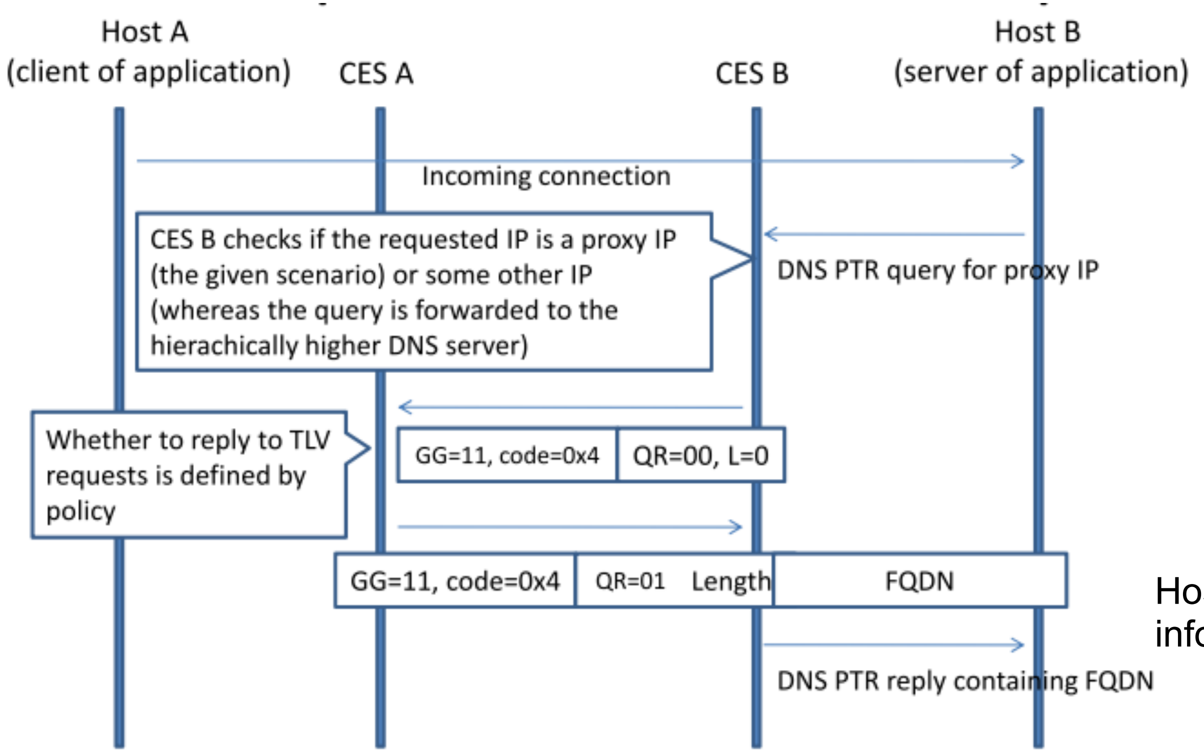
Signaling example (strict policy)



4Rrr=RLOC Reliable reply, 4Ra=RLOC Ack, Tq=Timeout query, Trr=Timeout reliable reply, Ta=Timeout ack, Dr=Domain reply, COrr=Cookie reliable reply, COa=Cookie Ack, CAq=CA Address query, CAr=CA Address reply, Sq=Signature query, Sr=Signature reply

Inbound edge can request FQDN on demand

- If not required by policy, but needed by application



Host B gets reliable information

Policies

- The policy determines what is required before accepting a connection
  - Reverse routability check
  - Domain name checking
  - Certificates
  - Given type of ID
- The policy also determines what information is provided to the peer
- Firewall based on ID instead of just IP
- User define policies and firewall rules on a high level which are translated into a low-level policy

Low-level policy definition

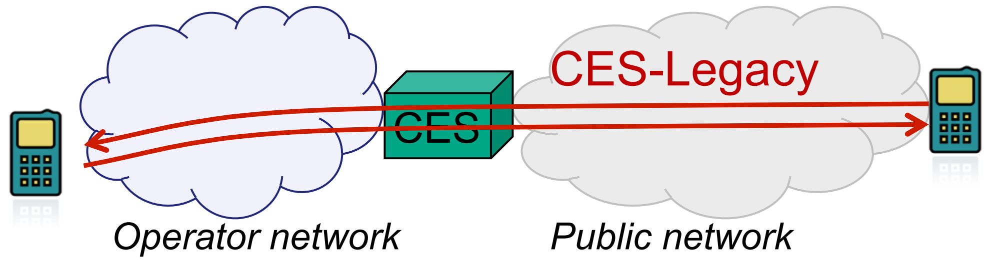
- Required TLVs in the role of iCES
- Required TLVs in the role of oCES
- Offered TLVs in the role of oCES
- Offered TLVs in the role of iCES
- Reliability policy for TLVs
  - Replies or reliable replies
- ID type policy
  - May depend on application, etc

Tutorial on Customer Edge Switching

Private Realm Gateway (PRGW)

For MEVICO internal use

Interworking Between CES Enabled and Legacy Networks



- For communication with networks without CES ("Legacy networks")
  – Deployment of the CES concept one network at a time
- To enable outbound and inbound connectivity and some of the security features

Outbound and inbound connectivity to legacy networks

- Outbound connectivity
  - Like a normal NAT
  - Choices: give real IP address of destination to private host *or* show a proxy address instead
- Inbound connectivity
  - Goal: no NAT traversal mechanisms!
  - Some protocols, such as HTTP, are feasible with reverse proxy
    - Each HTTP request contains the target domain
    - Could also be applied to SIP
  - For all other protocols, CES integrates a Private Realm Gateway (PRGW)

Inbound connectivity with Circular Pool of Public Addresses

- New concept developed within Mevico
- Useful also without the full CES
- Uses a pool of public addresses for inbound connections
- Matches a DNS query with the data traffic
- One public address reserved for each connection being setup (DNS query to first packet)
- An unlimited number of simultaneous established connections can share the same public address
  - Several sources, several destinations, all can use same port

Pairing FQDN with traffic

- The FQDN is in the DNS query
  - Reserve a public address and send it in DNS reply
  - At the time of DNS query, the sender's address is unknown and connection state cannot be created
  - PRGW stores the FQDN and the allocated address in Waiting State
- When first packet is received, the connection state is created
  - FQDN obtained from Waiting State identified by the public address
- Waiting State times out if no traffic within a timeout

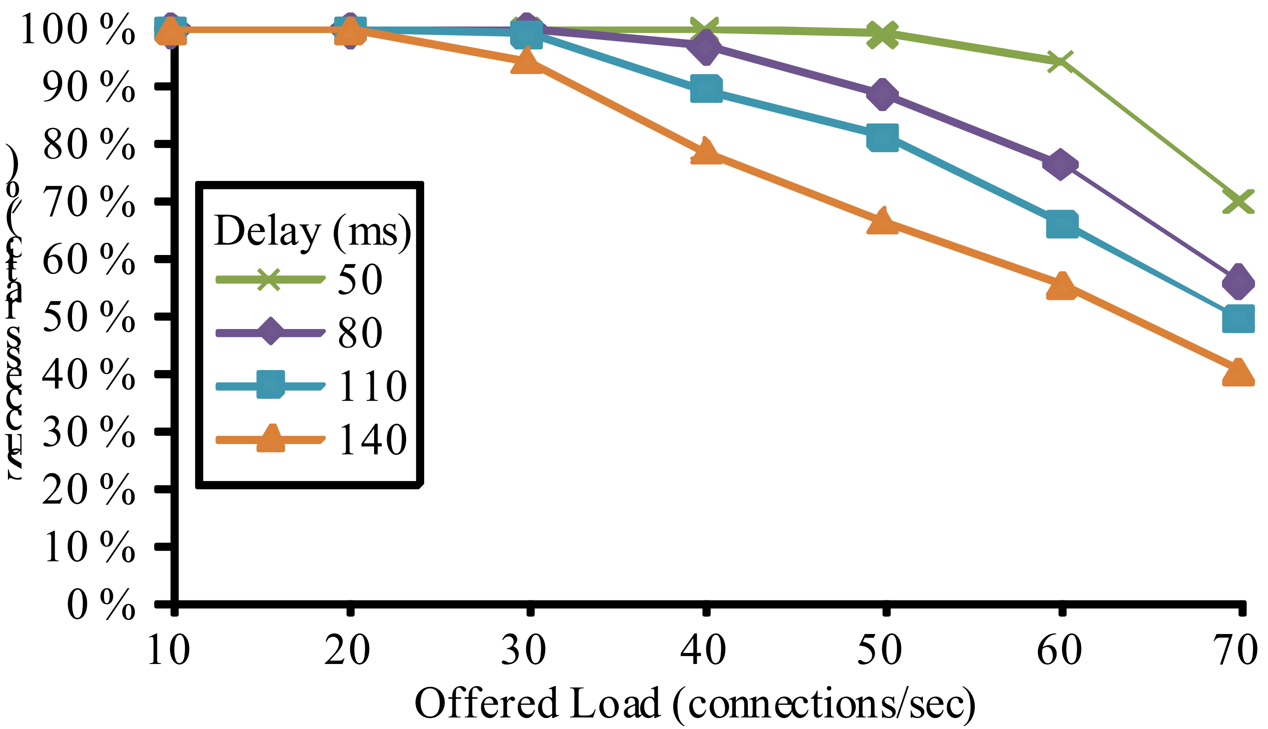
Scalability of Circular Pool

- Practically unlimited number of connections per address once the connection is established
- Each public address can only be used for one connection establishment at a time (0..2 sec)
- Time of connection establishment depends on round-trip delay
- Max capacity (connections/s) = pool size / delay

→ Pool size is determined by the rate of arriving connections and round-trip delay

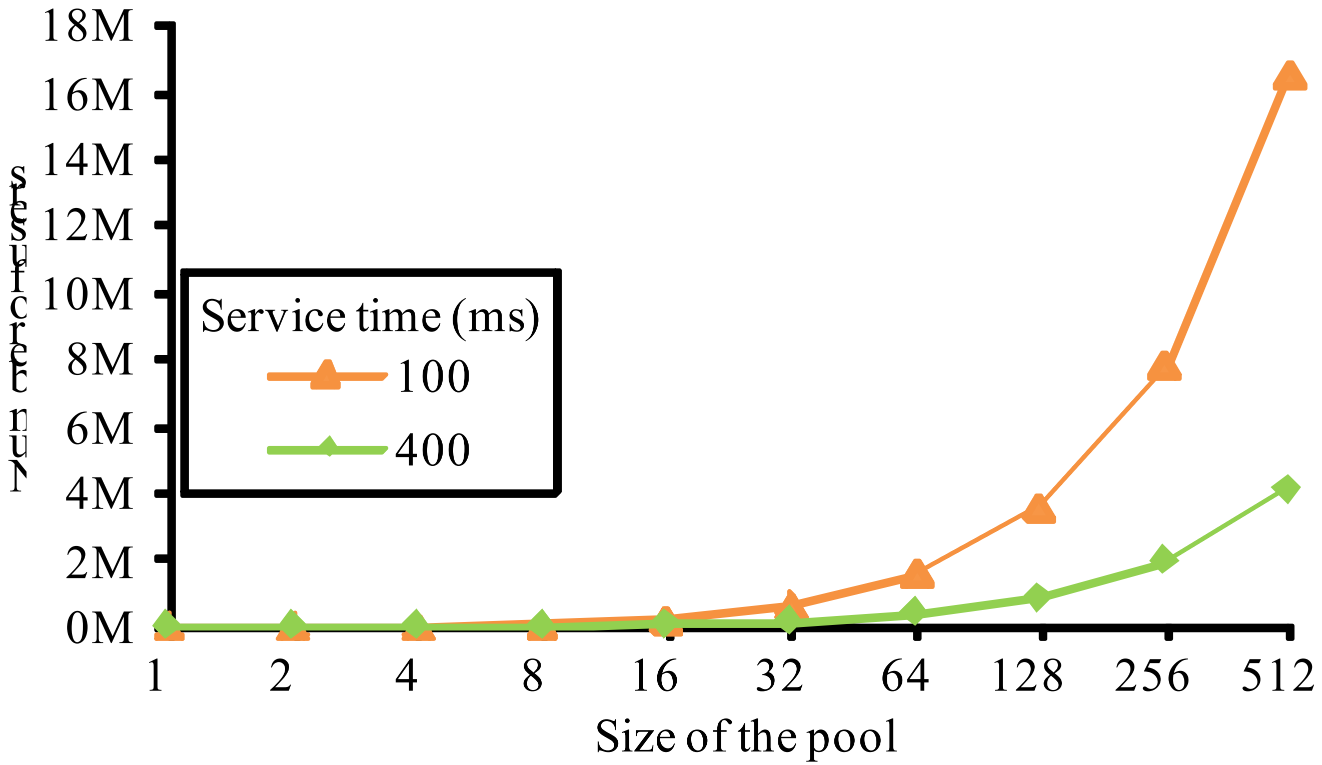
- Distribution also affects performance

Circular address pool performance



Success ratio for a pool size of 5 addresses

Scalability analysis of Circular Pool



Erlang-B with blocking rate 0.1% of offered connections

Tutorial on Customer Edge Switching

Customer Edge Switching in EPC

For MEVICO internal use

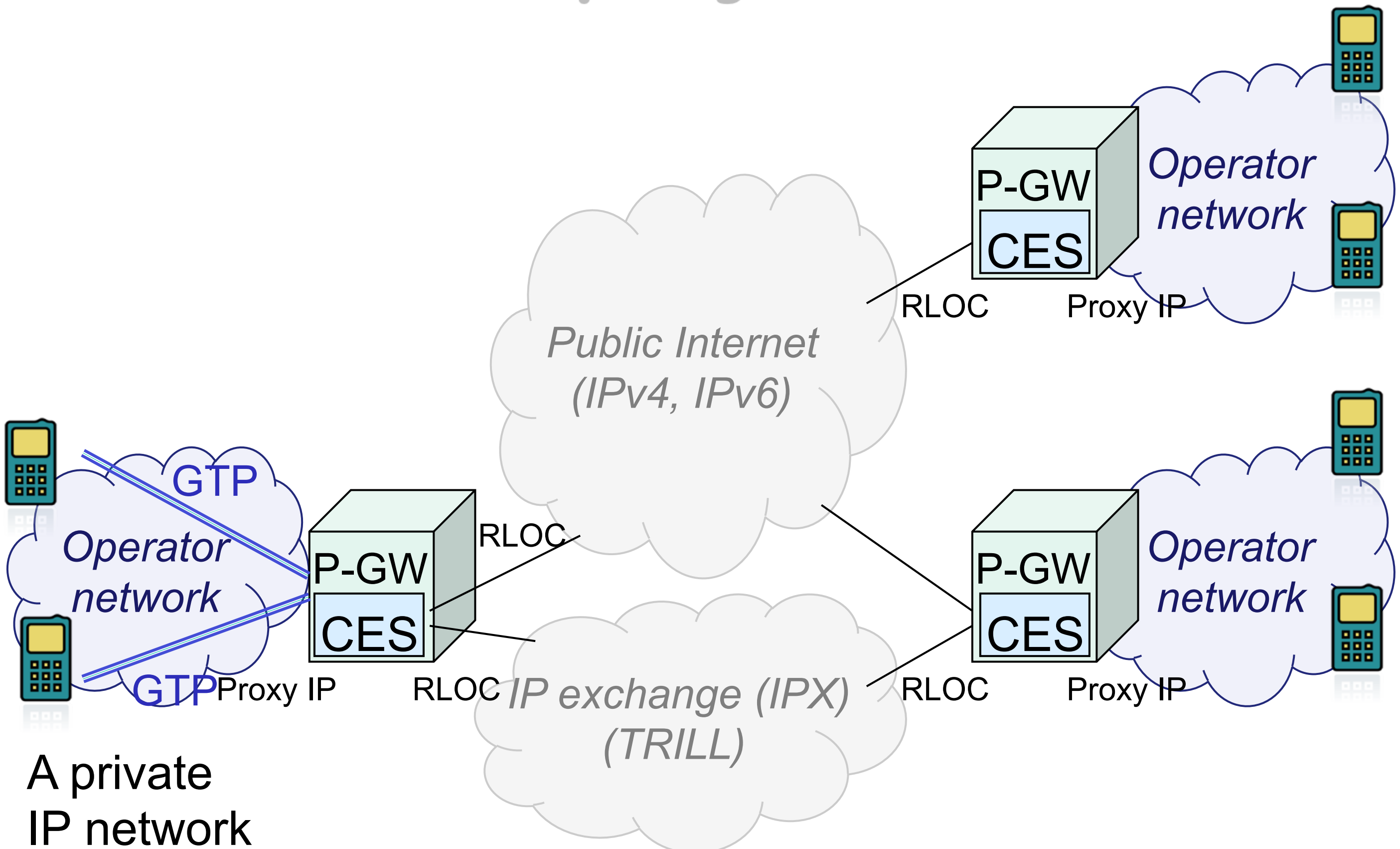
CES represents both user and operator

- CES was designed as a user network device
- In mobile network, CES is maintained by operator
- Operator benefits from security improvements, scalability improvements and reduction of unwanted traffic
- User benefits from inbound connectivity without NAT traversal and possibility to define reliable policies

Policies

- User policies
  - Configured e.g. through web interface
  - Examples
    - Allow traffic only from certain IDs
    - Calls and messages only from identified sources (avoid SPIM)
    - Only user's own devices can access content on home network
- Operator policies
  - For example, prevent spoofed addresses

Possible deployment scenario

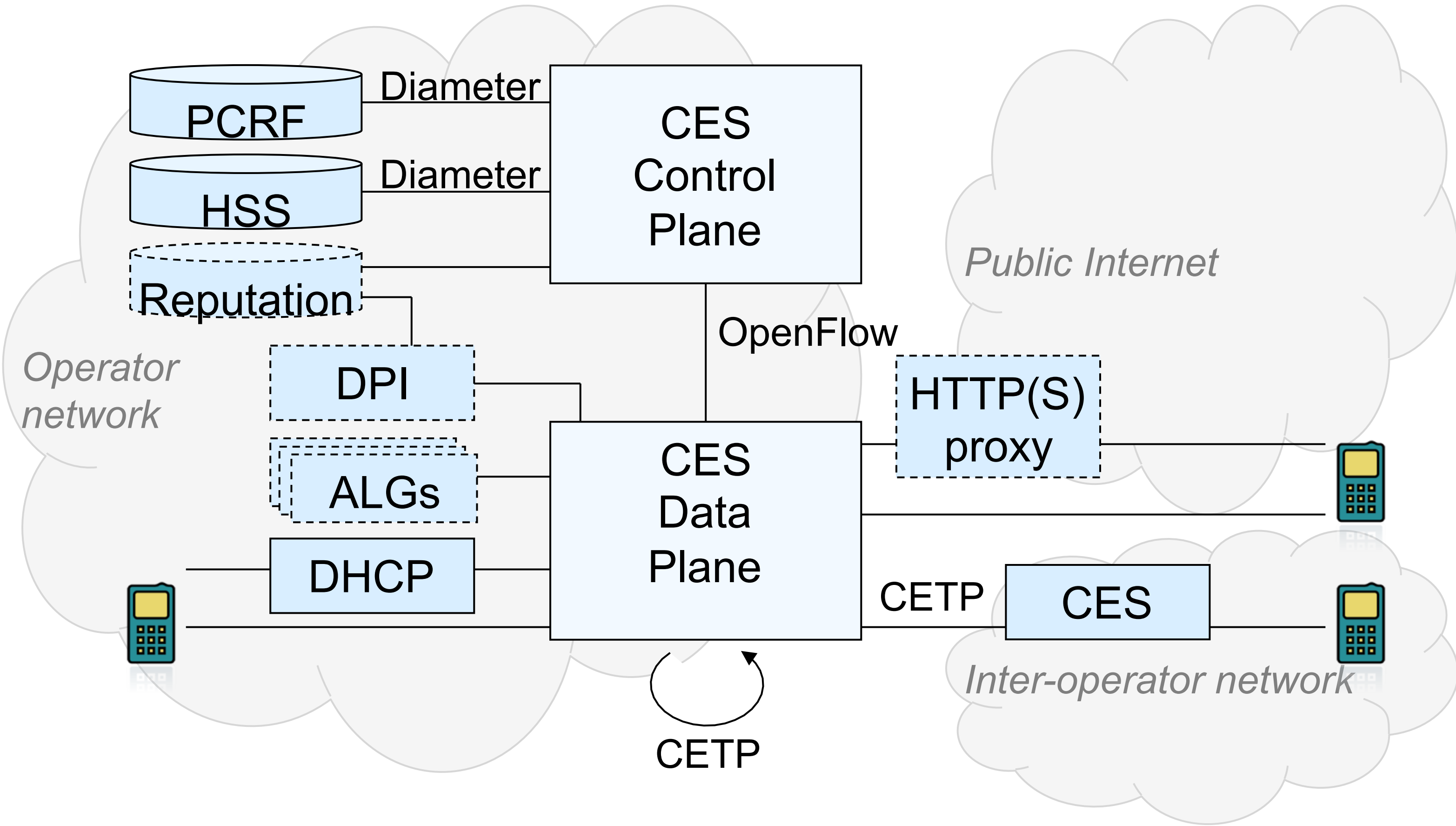


A private IP network per UE

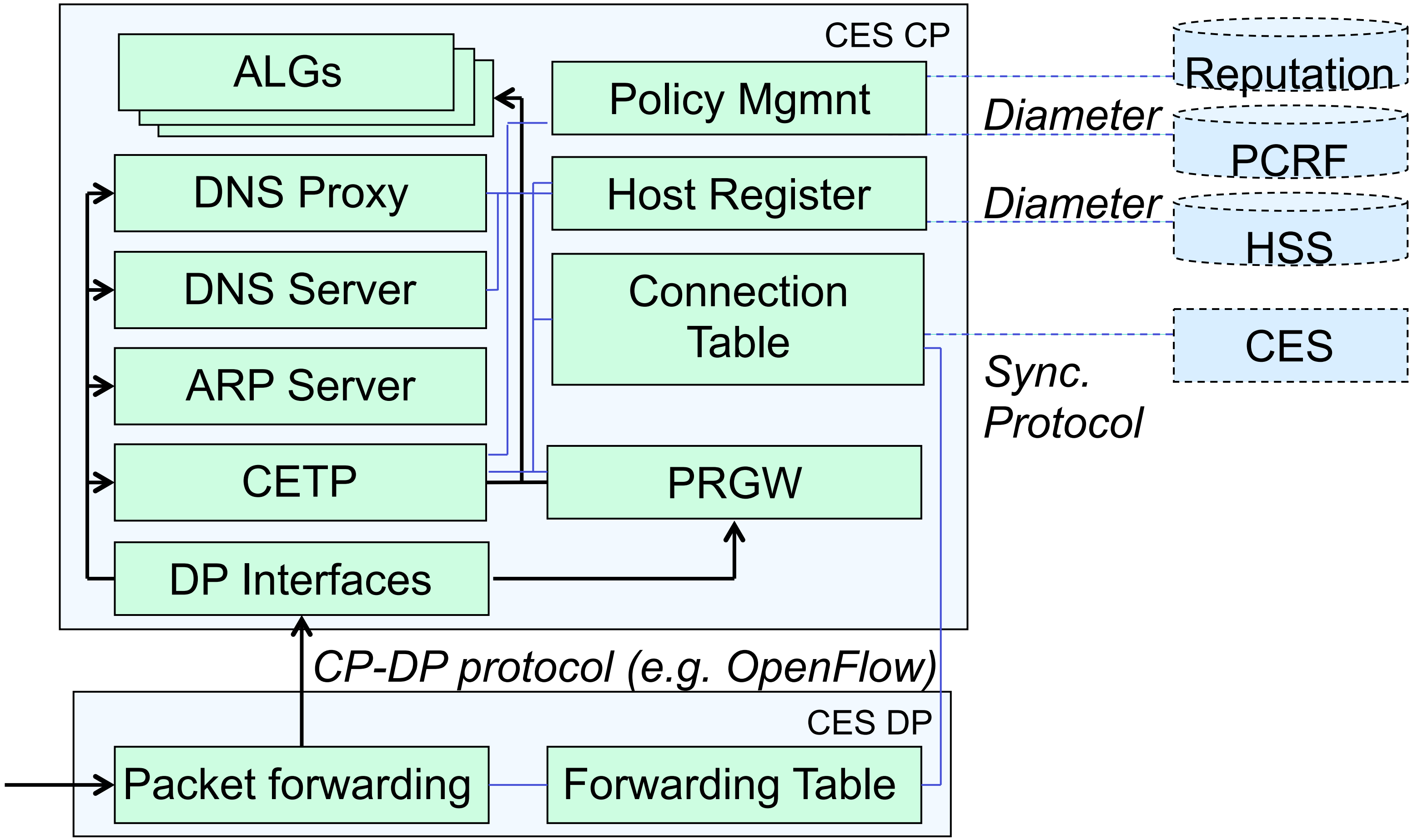
Control/data plane separation

- CES separates between control and data plane
- The DNS queries and first data packet go to the control plane, which creates state in data plane
- Data plane terminates GTP and CETP tunnels

Connectivity



Internal logical structure

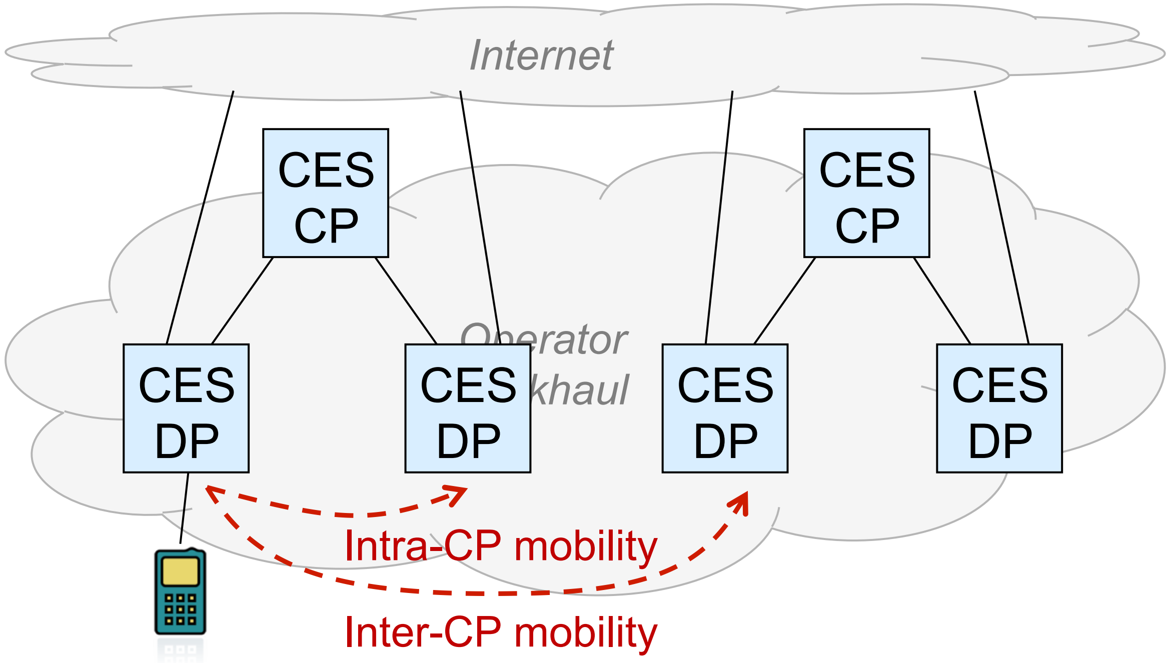


Deployment

- No changes to
  - UE
  - Applications
  - Transport between UE and GW
  - IP connectivity from GW
- Changes
  - Integrate CES in GW
  - Diameter interface CES to HSS and PCRF
  - HSS and PCRF updates
  - CES acts as DNS proxy
  - PRGW requires CES to be a DNS leaf

Local breakout and mobility

- CES-CES connections can be transferred
- CES-Legacy connection bound to a public IP address
  - Tunneling between data planes
- Connection state transfer between control planes



Integration with DPI and reputation systems

- Reputation calculated based on DPI input
- Sharing in reputation systems
- Reputation level affects policy, e.g. a more strict policy may be selected for sources with bad reputation or unidentified sources

Tutorial on Customer Edge Switching

Protocol Compatibility and Application Layer Gateways

For MEVICO internal use

Requirements for protocols

- A protocol must address the endpoint using a FQDN.
  - The protocol cannot transport IP addresses between domains Same as for NATs
- An application must perform a DNS lookup on the FQDN before sending traffic to a given destination. Specific to CES

Possible problems

- Application starts communication by sending traffic to an IP address directly without performing a DNS query first.
  - In practice, this scenario is rather uncommon, since users mostly use domain names to specify destinations.
- Application stores the IP addresses of IP addresses between sessions and reuse these in later sessions.
  - Connection state timed out
- Application sends its IP address to a peer device and expect the peer to send traffic to this address.
  - Typical in applications where the control connection is separated from the data connection, e.g. in FTP and SIP.
  - This problem is common both to ordinary NATs and CES.

Tested protocols

- Most common client-server protocols
  - Server behind CES
  - HTTP, HTTPS, FTP, SSH, ICMP
- Messaging, voice/video calls and file transfer directly between users
  - Inbound connectivity to users
  - Typically separate connections for media
  - SIP, XMPP, IRC, MSNP, Skype, Oscar, YMSG
  - Additionally web interfaces to these

Results

- Possible outcomes:
  1. The protocol works without problems
  2. The protocol works because of NAT traversal, which could be replaced by an ALG
  3. The protocol does not work but the problems can be solved using an ALG
  4. The protocol does not work and an ALG cannot be implemented (e.g. because of encryption)

Results

Protocol	Scenario	Operation	Result	Reason
HTTP	CES-Legacy	Page retrieval	Success	Optimization with proxy
	CES-CES	Page retrieval	Success	
HTTPS	CES-Legacy	Page retrieval	Success	Optimization with proxy
	CES-CES	Page retrieval	Success	
SSH	CES-Legacy	Interactive	Success	
	CES-CES	Interactive	Success	
ICMP	CES-Legacy	Ping	Success	
	CES-CES	Ping	Success	
SIP	CES-Legacy	Calls	ALG required	Private IP used
	CES-CES	Calls	ALG required	
FTP	CES-Legacy	File transfer	ALG required	Private IP used
	CES-CES	File transfer	ALG required	Private IP used
IRC	CES-CES	Messaging	Success	
		File transfer	ALG required	Private IP used
MSN	CES-Legacy	Messaging	Success	
		File transfer	Success	
Skype	CES-Legacy	Messaging	Success	
		Calls	Success	
XMPP	CES-Legacy	Messaging	Application dependent	Private IP used
		File transfer	Application dependent	Private IP used
	CES-CES	Messaging	ALG required	Private IP used
		File transfer	ALG required	Private IP used
Oscar (AIM)	CES-Legacy	Messaging	Success	
	CES-Legacy	File transfer	Application dependent	Private IP used
Oscar (ICQ)	CES-Legacy	Messaging	Success	
	CES-Legacy	File transfer	Application dependent	Private IP used
	CES-Legacy	Calls	Application dependent	Private IP used
YMSG	CES-Legacy	Messaging	Success	
	CES-Legacy	File transfer	Application dependent	Private IP used
	CES-Legacy	Calls	Application dependent	Private IP used

Application Layer Gateways

- Protocols that are not natively working with CES are handled by an Application Layer Gateway (ALG)
  - modifies protocol messages on the application layer.
- Adapting CES to application (ALGs) vs. adapting application to CES ("NAT traversal")
  - NAT traversal mechanisms have lots of drawbacks!
- Successfully implemented ALGs for FTP and SIP
- Guidelines for other ALGs

SIP ALG

- SIP transports IP addresses in the SIP header and in SDP, mappings needed for signaling and media flow
- The ALG adapts the IP addresses and the ports to achieve connectivity
- No global IP address → FQDN is a better alternative
  - FQDNs are allowed by SDP [RFC 4566] but usually applications use IP addresses
- Using FQDN is more straight forward approach than IP
  - No need to store temporary information
  - Algorithms/code easier to understand

SIP ALG

- CES-Legacy scenario
  - Adapts the scope of the IP addresses and port numbers conveyed in the SIP messages
  - Create mappings dynamically for the media and media control connections.
- In CES-CES scenario
  - In this case, the IP addresses are replaced by FQDNs of the hosts
  - End hosts issue new DNS queries for media addresses that allocate state in CES
  - No need to create additional mappings or modify the port numbers

Translation in CES-Local scenario:

$\blacksquare Source:@Destination:@[Media]:$
$[\blacksquare IPs:Ps@IPd:Pd@IPm:Pm\ ]\rightarrow$
$[\blacksquare\blacksquare FQDNs:Ps@FQDNd:Pd@FQDNm:Pm\ \ ]$

Outbound translation in CES-CES scenario:

$\blacksquare Source:@Destination:@[Media]:$
$[\blacksquare IPs:Ps@IPd:Pd@IPm:Pm\ ]\rightarrow$
$[\blacksquare\blacksquare FQDNs:Ps@FQDNd:Pd@FQDNm:Pm\ \ ]$

Inbound translation in CES-CES scenario:

$\blacksquare Source:@Destination:@[Media]:$
$[\blacksquare FQDNs:Ps@FQDNd:Pd@FQDNm:Pm\ ]\rightarrow$
$[\blacksquare\blacksquare FQDNs:Ps@FQDNd:Pd@FQDNm:Pm\ \ ]$

Outbound translation in CES-Legacy scenario:

$\blacksquare Source:@Destination:@[Media]:$
$[\blacksquare IPs:Ps@IPd:Pd@IPm:Pm\ ]\rightarrow$

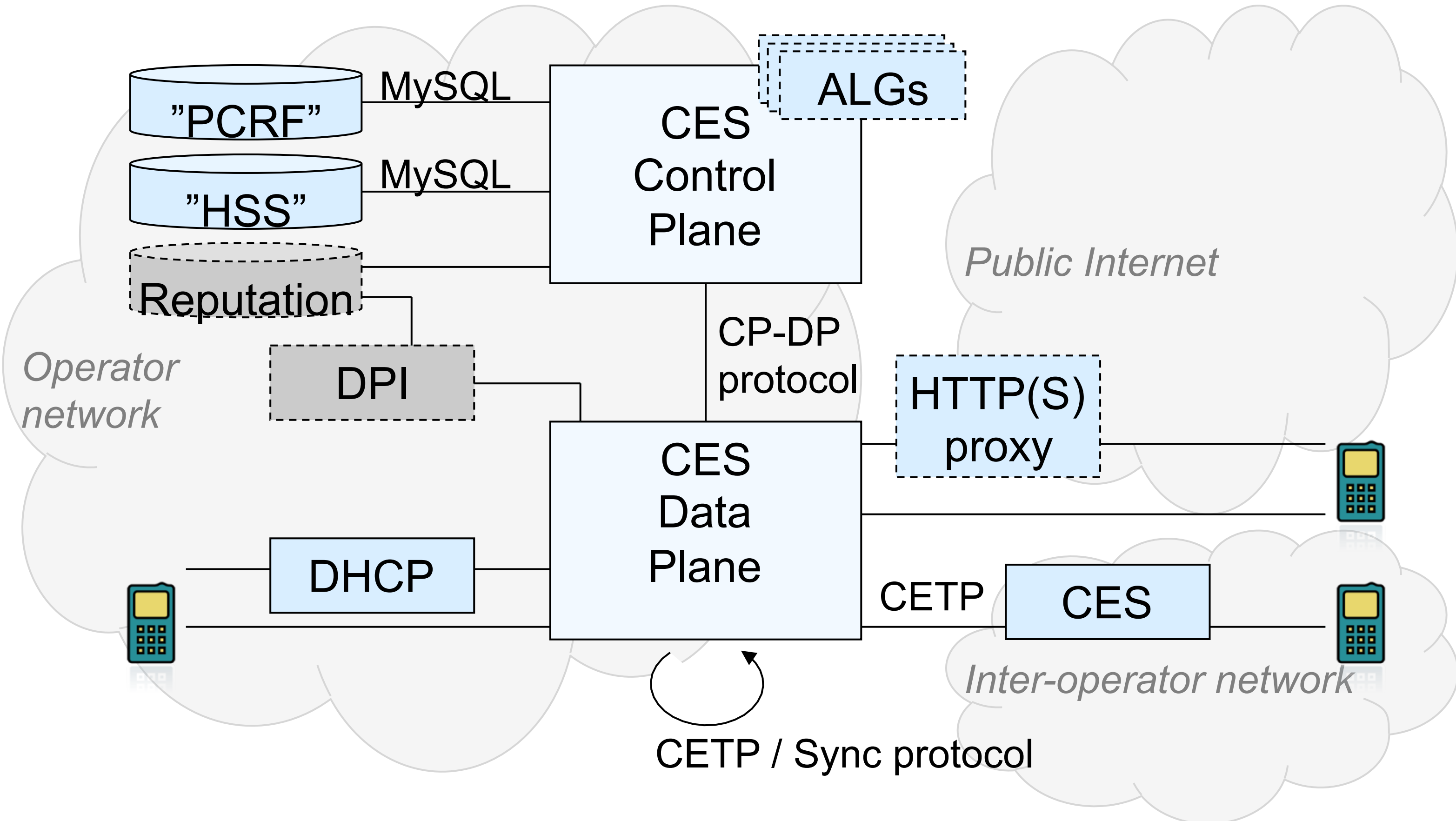
Tutorial on Customer Edge Switching

Prototype Implementation

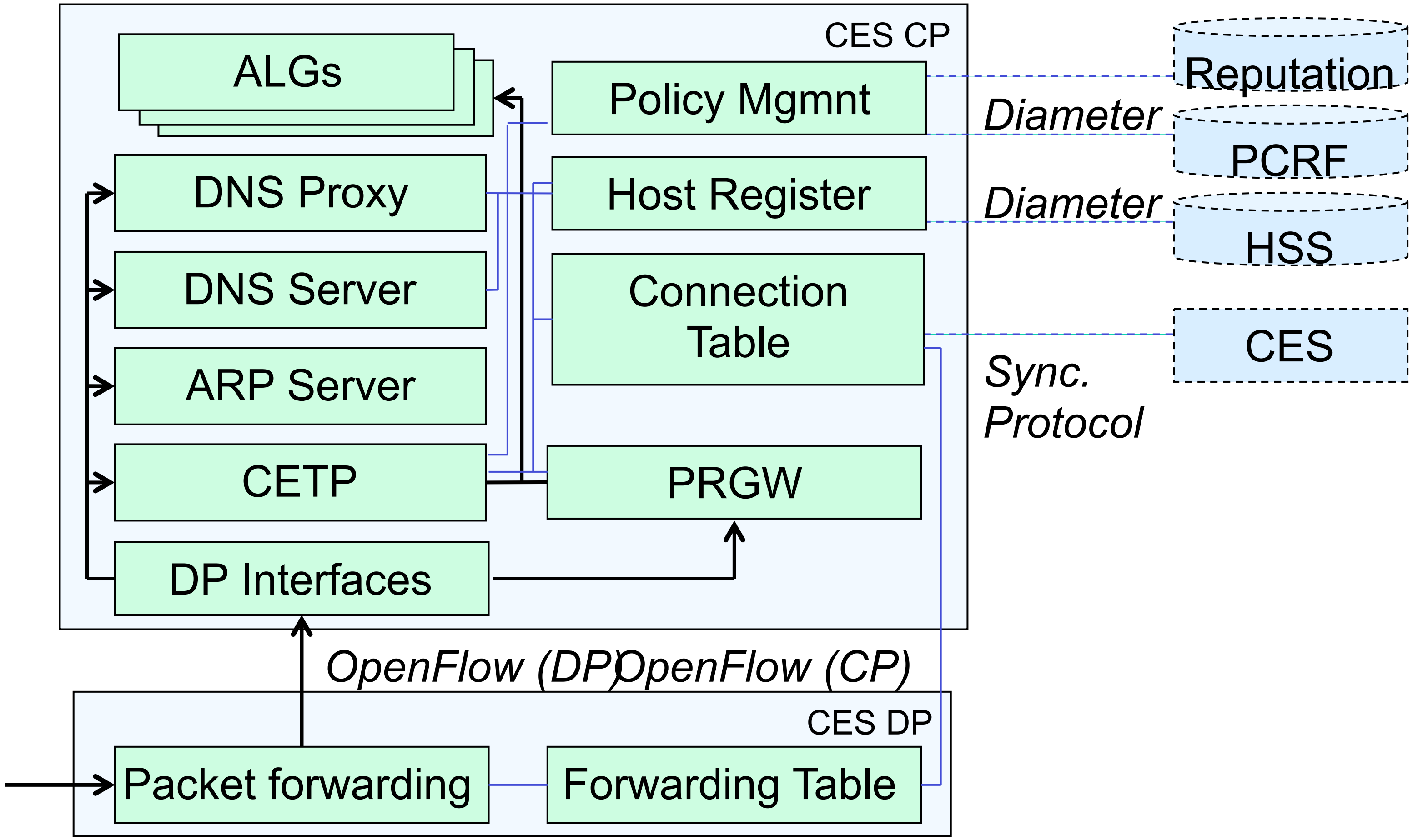
Prototype implementation

- Control plane
  - Implemented in Python
  - Policy management, DNS, connection table, host register, ALGs, state management
  - Interface to HSS and PCRF (currently modeled with mySQL)
- Data plane
  - Python Data plane for quick prototyping
  - C Data plane for performance (limited functionality)
  - Packet capture with Libpcap (C) and Scapy (Python)
- Proprietary protocol between planes
  - Replaced by OpenFlow/ForCES in future
- Network emulation with Netem

Prototype connectivity



Prototype structure



Tutorial on Customer Edge Switching

Conclusions

Conclusions

- CES enhances firewalls/NATs with the possibility to accept incoming connections in a controlled way
  - Allows end-to-end traffic, not only client-server
- Best security obtained when both endpoints are behind CES devices
  - The CEPT protocol allows security related signaling
- Some of the features can be provided even when only one endpoint is behind a CES

Publications

Presentations on CES and CETP

- http://www.re2ee.org/

Papers

- Jesús Llorente Santos, Raimo A. Kantola, Nicklas Beijar and Petri Leppäaho. Implementing NAT Traversal with Private Realm Gateway. Submitted to IEEE International Conference on Communications (ICC), 9-13 Jun 2013.
- Petri Leppäaho, Nicklas Beijar, Raimo Kantola, Jesús Llorente Santos. Traversal of the Customer Edge with NAT-Unfriendly Protocols. Submitted to IEEE International Conference on Communications (ICC), 9-13 Jun 2013.

Theses

- Petri Leppäaho, Design of Application Layer Gateways for Collaborative Firewalls, May 2012.
- Jesús Llorente Santos, Private Realm Gateway, November 2012.

QUESTIONS?