A Building Block for Best Effort Communications

Raimo Kantola

Raimo.Kantola@aalto.fi

Aalto University/Comnet

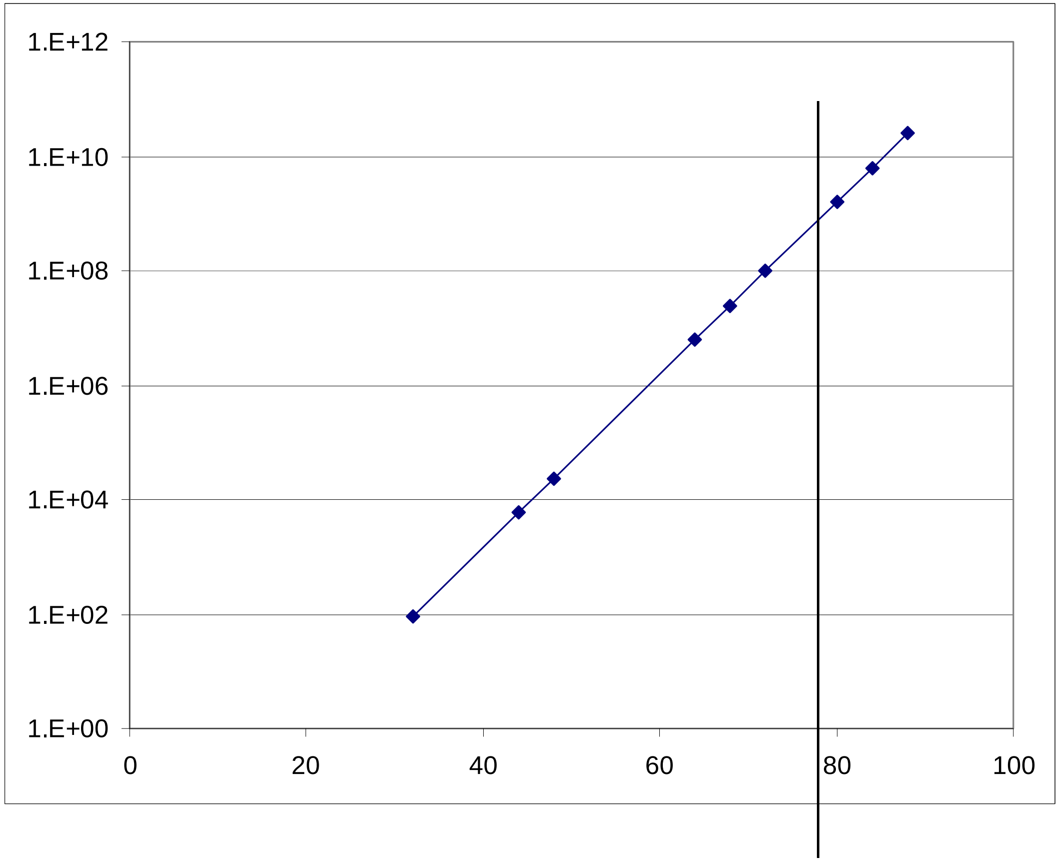
What kind of Communication IDs

- Globally unique deterministic IDs
  - high OPEX in case of a new type of ID
  - Could reuse mobile operator managed IDs
- Random ID is managed by the home network
- Temporary ID is managed by the visited network
  - Differ in how mobility is managed
  - national and cross border roaming breaks a session?

About IDs in the Internet

- Market is in the hands of Application developers and providers. Internet of Things area has its own proposed ID candidates.
- A generic communications ID is missing. IP address is used as the substitute
- Requirements
  - Ids are needed to *enhance trust* between communication partners, thus must be fairly stable, changes being governed by well understood rules potentially having legal enforcement (missing e.g. for IP address changes)
  - It must be possible to change an electronic communications ID when under attack or when there is evidence that the ID has been hijacked, hopefully with no impact on routing tables and similar distributed network entities
  - Needed level of legal enforcement/trust on an ID depends on the application (and the use case of an application)
  - Certificates are ”assured” IDs, assurance coming from an CA. CAs form a chain. Root of the chain is trusted by default (root certificate)

How many bits are needed for random communication identities used in CES nodes?

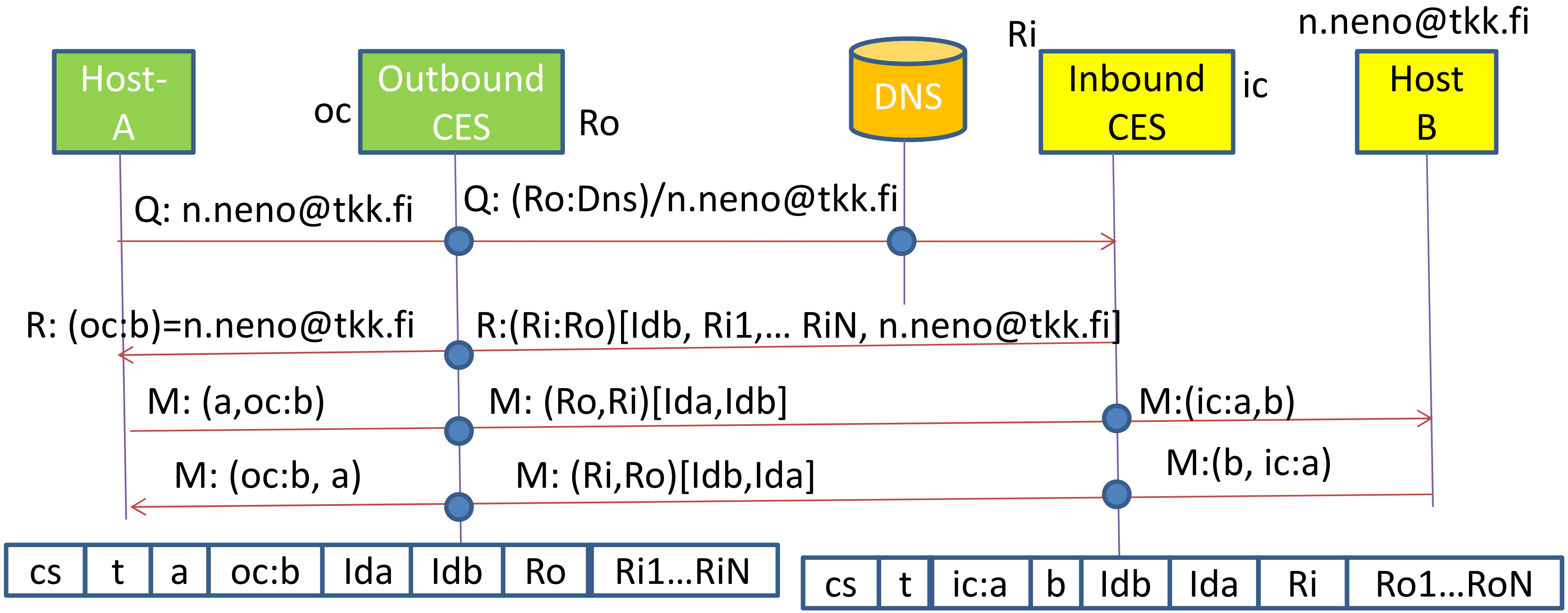


This is based on the birtday paradox. We assume that the probability of a clash of identities is < 1 in a million when all IDs are compared one by one. If ID dependent filtering akin to address dependent filtering in NATs is used, a pair of IDs is compared to another pair of IDs. This gives an additional safety margin.

Signaling Cases

	Legacy receiver	Receiver behind CES
Sender Behind CES (new Edge)	CES acts as NAT	**Customer Edge Traversal Protocol** used To tunnel packets Thru the core
Legacy IP sender	Traditional Internet	Inbound CES acts as ALG/Private Realm Gateway or server side NAT

Message Flow



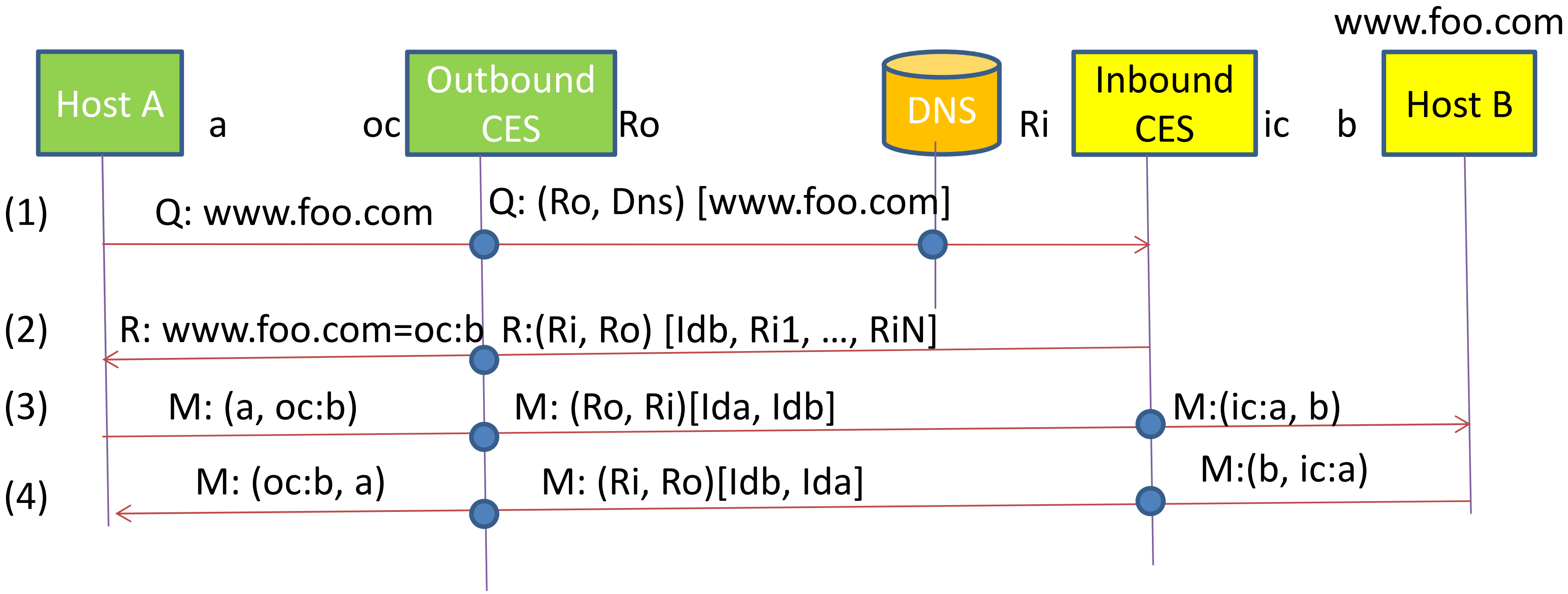
cs	t	a	oc:b	Ida	Idb	Ro	Ri1…RiN

cs	t	ic:a	b	Idb	Ida	Ri	Ro1…RoN

a – IP address of host a
oc – address pool of outbound CES
oc:b – IP address representing host b to host a
Ro (Ro1….RoN) – Routing locators of outbound CES
Ri (Ri1 …RiN) – Routing locators of inbound CES
cs – connection state, t - timeout

Ida – ID of host a
Idb – ID of host b
ic – address pool of inbound CES
ic:a – IP address representing
host a to host b

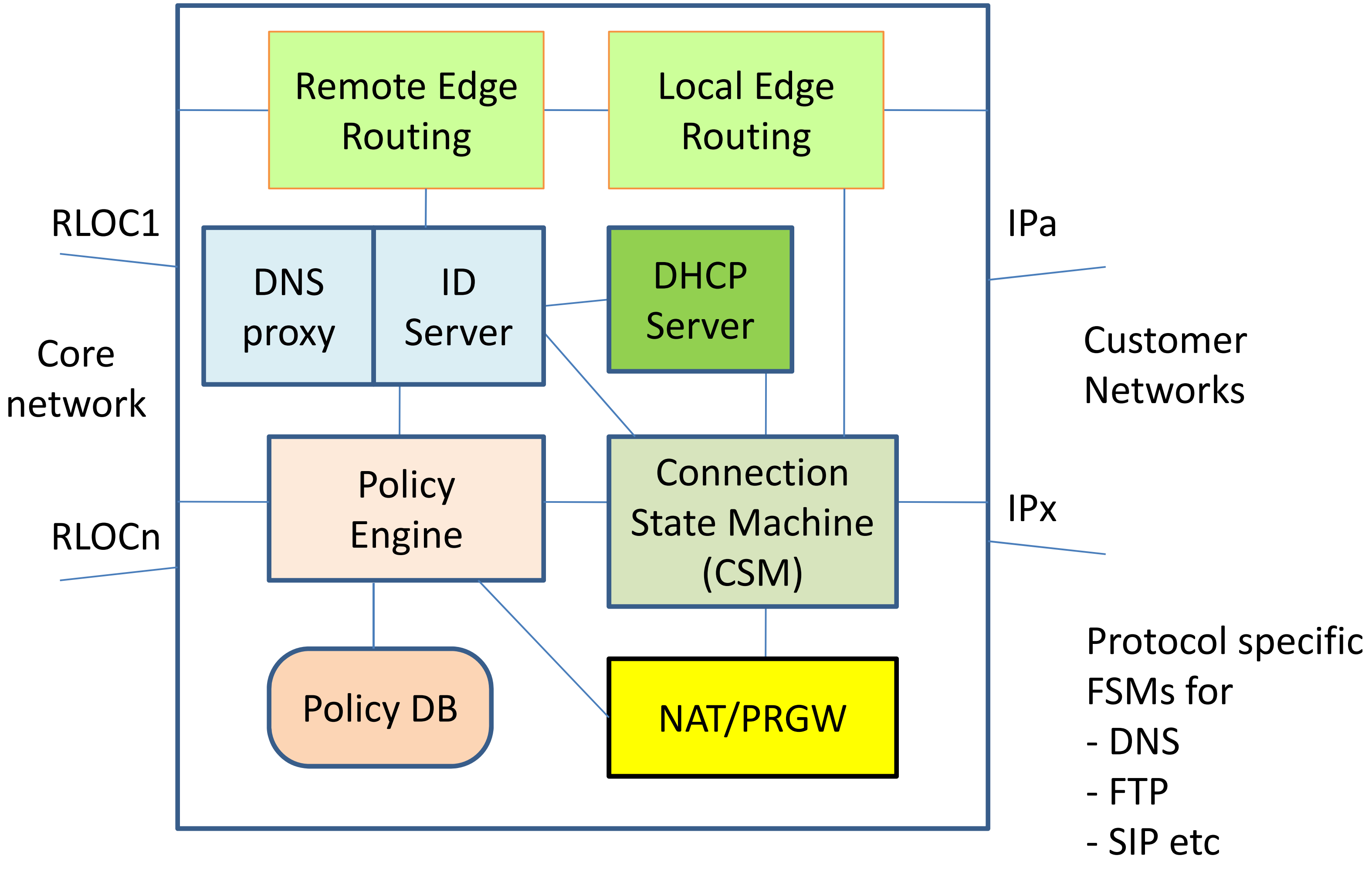
Message Flow



a – IP address of host a
b – IP address of host b
oc – address pool of outbound CES
oc:b – IP address representing host b to host a
Ro (Ro1….RoN) – Routing locators of outbound CES
Ri (Ri1 …RiN) – Routing locators of inbound CES

Ida – ID of host a
Idb – ID of host b
ic – address pool of inbound CES
ic:a – IP address representing host a to host b

Model of CES connected to IPv4 core



Protocol specific FSMs for
- DNS
- FTP
- SIP etc

PRGW = Private Realm Gateway

Policy Engine (or Trust Function)

- Uses Policy Rules in Policy DB to filter all traffic like a Collaborative stateful Firewall
- Upon new inbound flow, may admit and let CSM create state. May use e.g. criteria:
  - Not too many exising open tcp connections from the same source RLOC
  - Not too many established flows from this source RLOC
  - Execute return routability check on forwarding layer
  - Execute return routability check on naming (and forwarding) layer
  - Admit if return routability check is passed
  - Require assured ID of the initiator of communication
  - Etc. See e.g. CES paper in AINA Workshop
- For a flow returns: admit/deny/admit and log/deny and log/Require ID –type/Require TLVs from Peer CES(optionally offer own TLVs)/

Level of trustworthiness

- Return routability check in plain text can tackle RLOC spoofing by a host
  - Alternatives to it are Ingress Filtering and RPF checking
- If a domain’s routing can be hacked, RLOCs can not be trusted even with return routability check: PKI can be applied among CES devices and RLOCs in the return routability check can be signed by the sender’s secret key.

Connection State Machine

- For new inbound flow admitted by Policy Engine, creates mapping state
  - (remote RLOC, local RLOC, targetID → targetAddress; sourceID → LocalSourceAddress, timeout)
- For new outbound DNS query, creates state:
  - (local RLOC, sourceAddress → sourceID; DNS)
- Upon DNS response, gleans targetID and modifies state
  - (remote RLOC(s), local RLOC, sourceAddress → sourceID; targetID →localTargetAddress, timeout)
  - Informs Remote Edge Routing about new RLOCs
  - If target is non-CES, hands over to NAT
- Communicates with
  - local DHCP server to allocate addresses
  - Local edge routing to mirror CSM state (remoteID/localAddress mappings) to standby CES
  - Remote Edge Routing to modify remote active RLOC on the fly for failover
- Management interface: accepts CES switchover commands, modifies state → flow migrates to standby CES
- May supervise current flow or current remote RLOC, detect resending, move to remote edge failover automatically
- Inbound packet: decapsulates the packet, modifies the packet using state, re-calculates checksums
- Outbound packet: encapsulates the packet, modifies packet using state, recalculates checksums as needed

DHCP server

- Dynamically allocates IP addresses to local hosts
  - Informs ID server about the new host and its IP address
- Dynamically allocates *proxy IP addresses* to remote hosts that communicate with local hosts

ID server

- Is a leaf node of the DNS hierarchy
- For new IP addresses of local hosts, creates
  - ID → IP address mapping
- For names (whatever they are in the domain) creates Ids, possibly making use of parameters and operator services
- Upon ID protocol query (several protocols can be supported)
  - Returns an ID for a name, e.g www@foo.comnet.tkk.fi or SIP:raimo.kantola@comnet.tkk.fi using e.g. NAPTR format
  - May cache the query for trust processing
- Upon management command can phase off existing Ids and adopt new Ids for new flows.
  - This is a possible response to an attack
  - Existing flows keep on using the old Ids.
- Self-configures itself to DNS of the ISP providing the DNS service
  - Stores NS and A-records with idprotocol and rlocs.
- For roaming support may communicate with HSS using Diameter

DNS Proxy

- DNS proxy is configured as default DNS for all hosts served by CES
- Translates an address query for a name like www.foo.com to a NAPTR query
- Gleans the response for IDs, RLOCs, stores them into connection state and requests DHCP server to allocate a temporary proxy IP address that will represent the destination locally, returns that address as reponse to the Address query to the local host
- More complex scenarios and alternatives discussed elsewhere

Local Edge Routing

- Takes care of local multi-homing
  - If local network is IP routed, CES is the default gateway, selection of default gateway takes place as before
- May listen to interior routing protocol to learn about other CES nodes in the local domain
- May establish a session for monitoring the state of the other CES
  - If CSM mirrors its state: remoteID/remoteIP mapping to standby CES, next monitoring event is postponed upon each mirroring event

Remote Edge Routing

- Takes care of remote edge multi-homing
- May use e.g. LISP or CETP to query remote RLOC states at regular intervals
  - Learns about new remote RLOCs from CSM
  - Makes the remote RLOC state available to CSM for hot failover
  - May inform remote CES (its remote Edge Routing) about local RLOC states

NAT/Private Realm GW

- NAT = Network Address Translator;
- When the Outbound CES notices upon DNS query that the target is not behind a CES but rather a legacy IP destination, CSM hands over to NAT, from now on outbound CES acts as a NAT
- When Inbound CES sees a DNS query from a legacy IP source a (non-NAPTR query), ID Server maps the name in the Query to (RLOCi →, dest-IP=b, dest-Port), where RLOCi is the next available IP address in the sNAT circular pool and sets a short timeout (e.g. 2s)
  - RLOCi is locked for new flows during the timeout
  - Alternatively, CES could use SRV record to link the DNS query and the start of the application flow. Unfortunately very few applications support SRV –records
  - For many protocols this is not enough (e.g. http), instead a protocol proxy must be employed
- When Trust Function sees a non-encapsulated packet to (RLOCi,dest-Port) it removes the timeout making RLOCi available for new flows and directs the packets to the sNAT function
  - sNAT will create connection state: **cs, rloc, a, b, source-Port, dest-Port, t**
  - For a dest-port, sNAT may launch a protocol proxy (e.g. http and https)
  - Upon response from the target host, sNAT uses local addresses (a,b) and port numbers to map a to rloc

Unfortunately, very few applications support SRV –records, e.g. most browsers do not support them, although some SIP clients do support them!!!

Analysis of Server side NAT/PRGW

- Properties of RLOC pool for legacy interworking
  - Same RLOC is used by simultaneous connections with many hosts behind CES, packets can traverse CES using source address and source port and dest port
  - Makes it possible to allocate the same dest-Port simultaneously to many targets communicating with many sources
  - This may be desirable if the targets are e.g. mobiles (i.e. low flow density but numerous targets under one CES)
- Scalability
  - Limiting factor is how many new connections can be established per timeout (about 1…3 s). Number of simultaneous connections per RLOC is not limited.
  - Solution can serve servers with low arrival rates such as servers on Mobiles and IoT objects.
- For example a browser proactively opens many flows to a server to download objects on the page, as new flows are opened without DNS query and there may be a NAT on the client side, these new flows are difficult to identify on IP layer. Therefore, host-header in http is used instead to map the flow to a destination address. The result is an http-proxy in iCES.

Limitations of sNAT

- Forces CES to create state upon DNS query → makes the CES vulnerable to attacks
  - DDOS attack using spoofed IP source addresses → the attacker must know the names of the targets
  - Hijacking attack: for a short time RLOCi is open to a new flow from any IP address/port to the local target IP and dest-port
  - Attack scenarios and counter measures are for FS
- Multi-homing: sNAT does not support outbound nor inbound CES multihoming
  - Same as NAT in this respect
  - Multi-homing requires collaboration of the local and remote edge nodes, since the other end is legacy, there is not much we can do…

Protocol dependent state machines

- CSM MUST be able to spot protocols such as FTP, SIP
- These protocols use IP addresses as Identifiers
- CES has to modify IP addresses so that ”a” is replaced by ”ic-a” and ”b” is replaced by ”oc-b” or vice-versa.

CES Product Use Cases

- CES for fixed BB
  - Hierarchical: partly implemented in xDSL modem, partly in the fixed access network gateway
  - The Access Network CES may have several IP addresses (IPa….IPx) at the customer network side
  - If the Access network CES has many RLOCs, multi-interface access to the Internet can be supported
- CES for mobile BB
  - CES hosts trust services for mobiles
  - Resides in the Mobile ”Core” network (P-GW) or as an access gateway to the Internet after P-GW.
  - Similar Address allocation as the previous case
- CES for hosting trust services for corporate networks
  - Speeds up CES adoption
  - Probably MUST have many IP addresses at the customer side and MAY have many RLOCs
- A Corporate network CES
  - Large corporations only, because CES must have an RLOC and ISPs may want to adopt a conservative RLOC allocation policy

CES to ISP network connection

- A clear demarcation of responsibilities and a clean addressing priciple: ”none of the trust domains release their addresses to other domains” is achieved if
  - CES are connected to Provider Edge (PE) nodes
  - RLOCs are owned by PEs
  - CES sees link local addresses instead of RLOCs
  - PE has NAT-like connection state and maps between link local addresses and RLOCs
- Connection state in PE may be a scalability challenge!

State of prototyping

- In CETP definition we are on the 2nd round prototype.
- Core implementation of CETP with tunneling etc works
- Legacy interworking with several important protocols (http, ssh etc) works as well
  - We have a list of protocols for which ALGs are being developed

CES has protocol specific ALGs – how can we make the solution generic?

- Most new applications run over http. For such Apps, it suffices for CES to be fully compatible with http.
- By policy management
  - Apps developer can publish: application + policy + ALGs
  - Policy management infrastructure of Mobile Operators will take care of executing the policies and making ALGs available as needed while subscribers are roaming.