HA NOI UNIVERSITY OF SCIENCE AND TECHNOLOGY
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

DISTRIBUTED SYSTEMS

HA NOI UNIVERSITY OF SCIENCE AND TECHNOLOGY
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

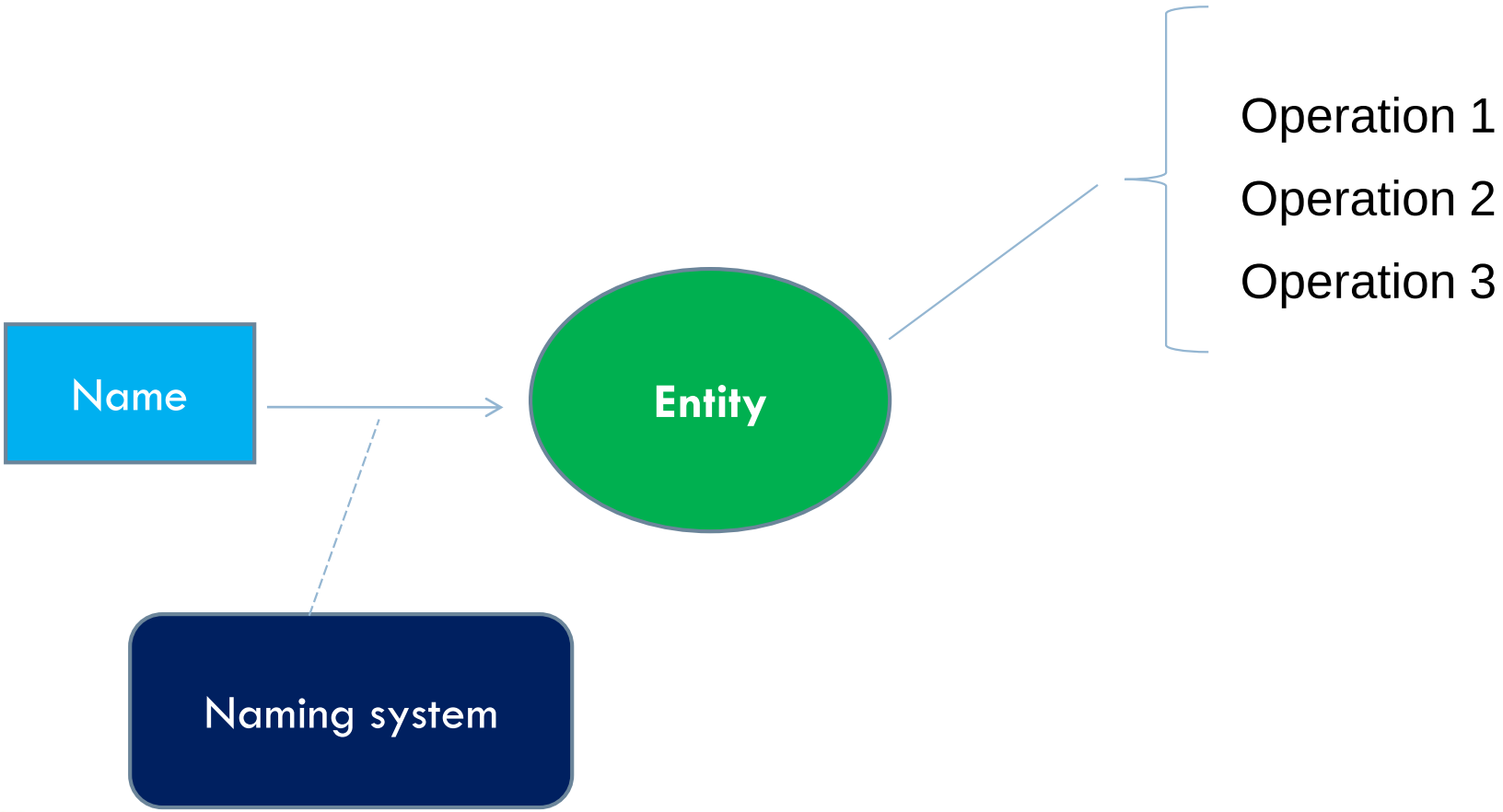
Chapter 3: Naming in Distributed Systems

Outline

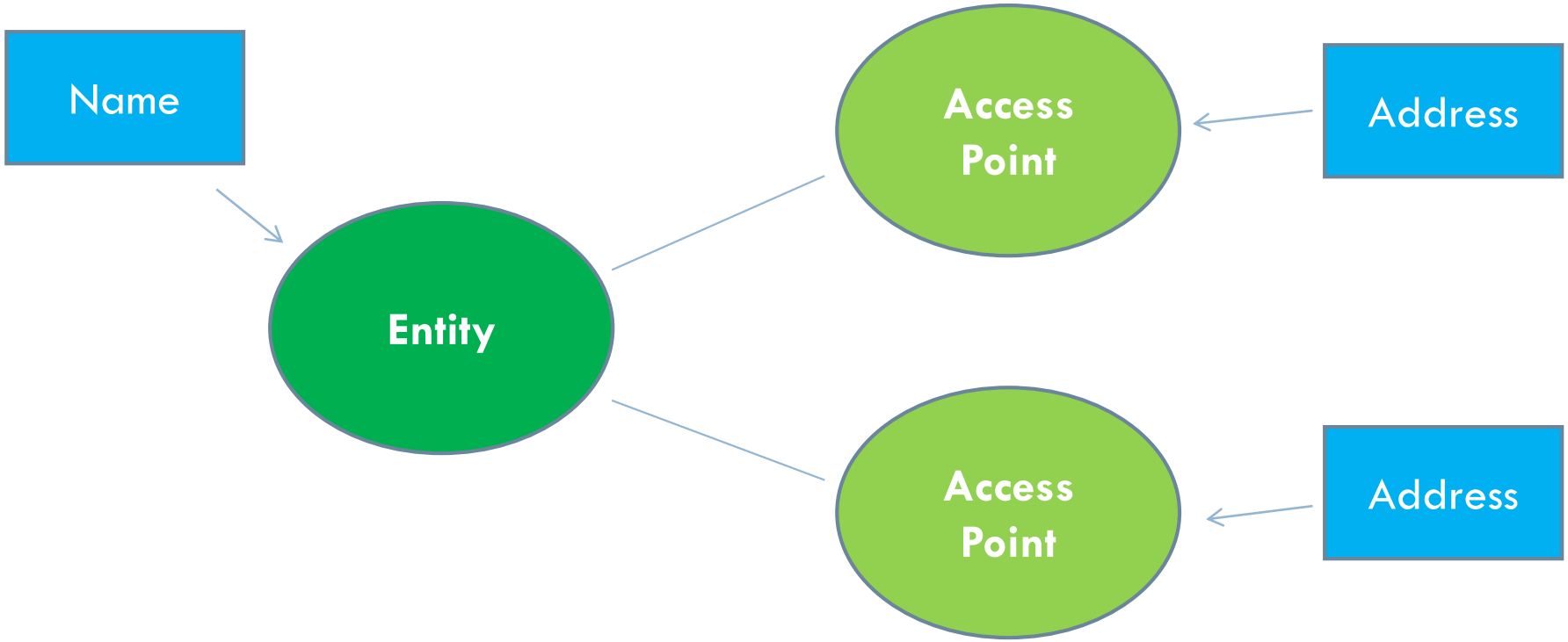
1. **Names. Identifiers and Address**
2. **Flat Naming**
3. **Structured Naming**

1. Names. Identifiers and Address

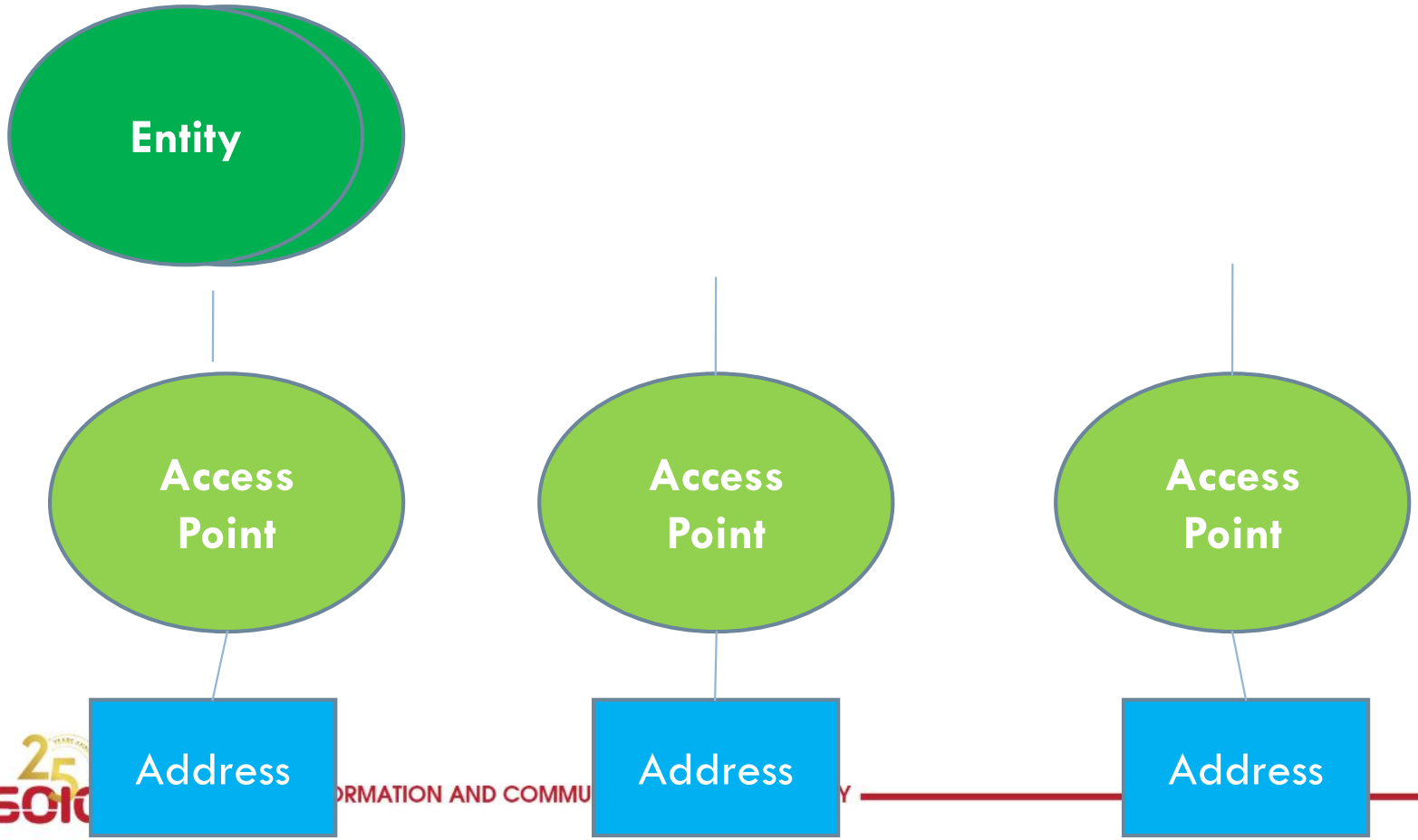
Entity & Name


Operation 1
Operation 2
Operation 3
Name
Entity
Naming system

Entity, A.P



Location independent



Identifier

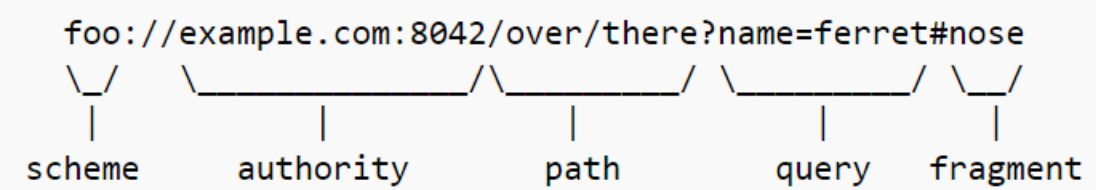
- 3 properties:
  1. An identifier refers to at most one entity.
  2. Each entity is referred to by at most one identifier.
  3. An identifier always refers to the same entity (it is never reused)
- Problems: The exhaustion of Identifier
- Solutions:
  - Extending the namespace
  - Re-assign identifier to new entities

Resolving names and identifiers to addresses

- Centralized approach
  - Name-to-address binding
  - Problem: not appropriate to large network
- Distributed naming systems

URI, URL và URN

- **URI:**
  - a string of characters used to identify a resource.
  - interact with representations of the resource over a network
  - URL and URN
  - It comprises 5 parts: scheme, authority, path, query and fragment



- **URN:**
  - ISBN 0486275574 (run:isbn:0-486-27557-4)
- **URL:**
  - file:///home/username/RomeoAndJuliet.pdf

2. Flat naming

2.1. Definition

- Identifiers are simply random bit strings (unstructured)
- It does not contain any information of location
- Goal: how flat names can be resolved
  1. Simple solutions
  2. Home-based Approaches
  3. Distributed Hash Tables
  4. Hierachical Approaches

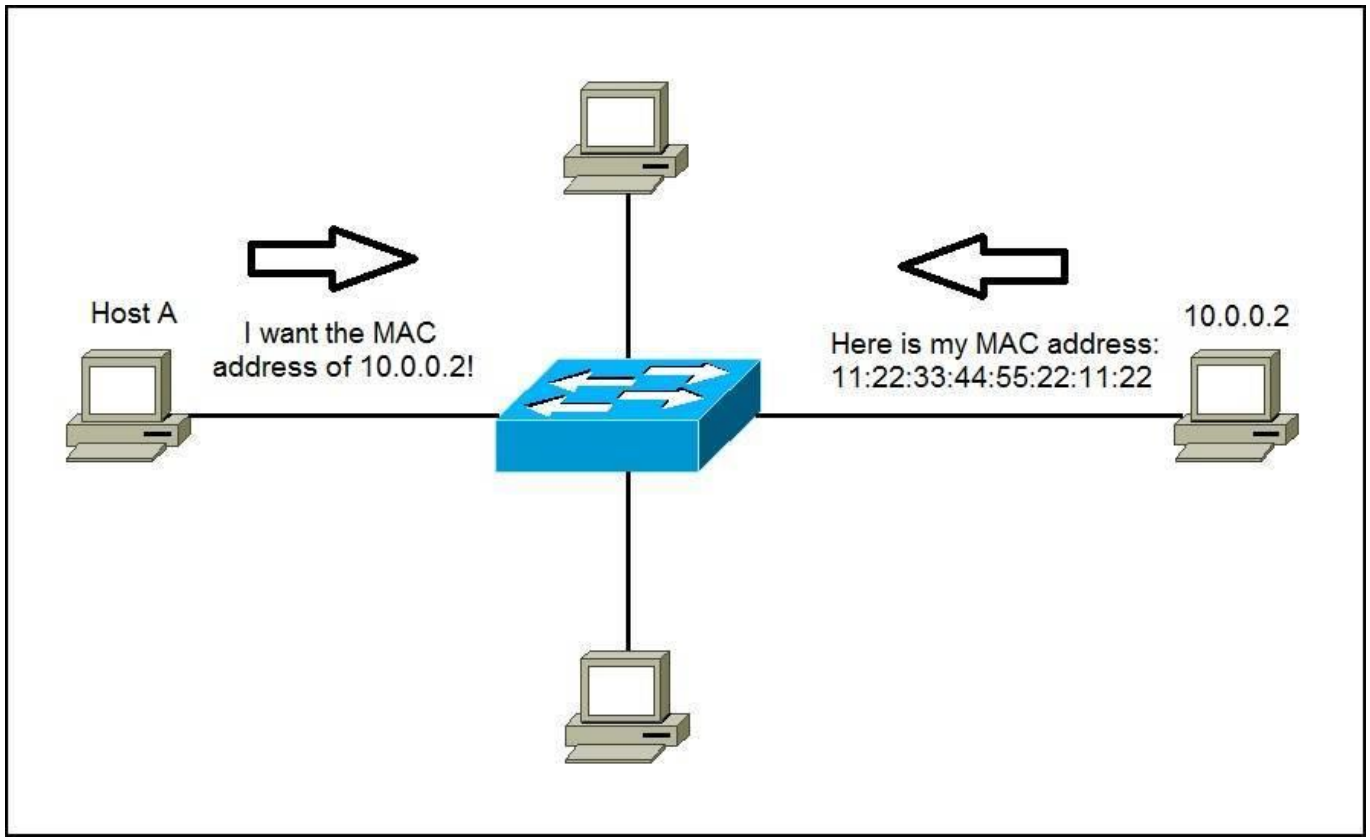
2.2. Simple Solutions

- **2.2.1. Broadcasting and Multicasting**
- **2.2.2. Forwarding pointers**

2.2.1. Broadcasting and Multicasting

- **Condition: System supports broadcasting facilities:**
  - A message containing the identifier of the entity is broadcasted to all other machines.
  - Each machine is requested to check whether it has that entity.
  - Only the machines that can offer an access point for the entity send a reply message containing the address of that access point.

Example: ARP


Host A
I want the MAC address of 10.0.0.2!
Here is my MAC address: 11:22:33:44:55:22:11:22
10.0.0.2

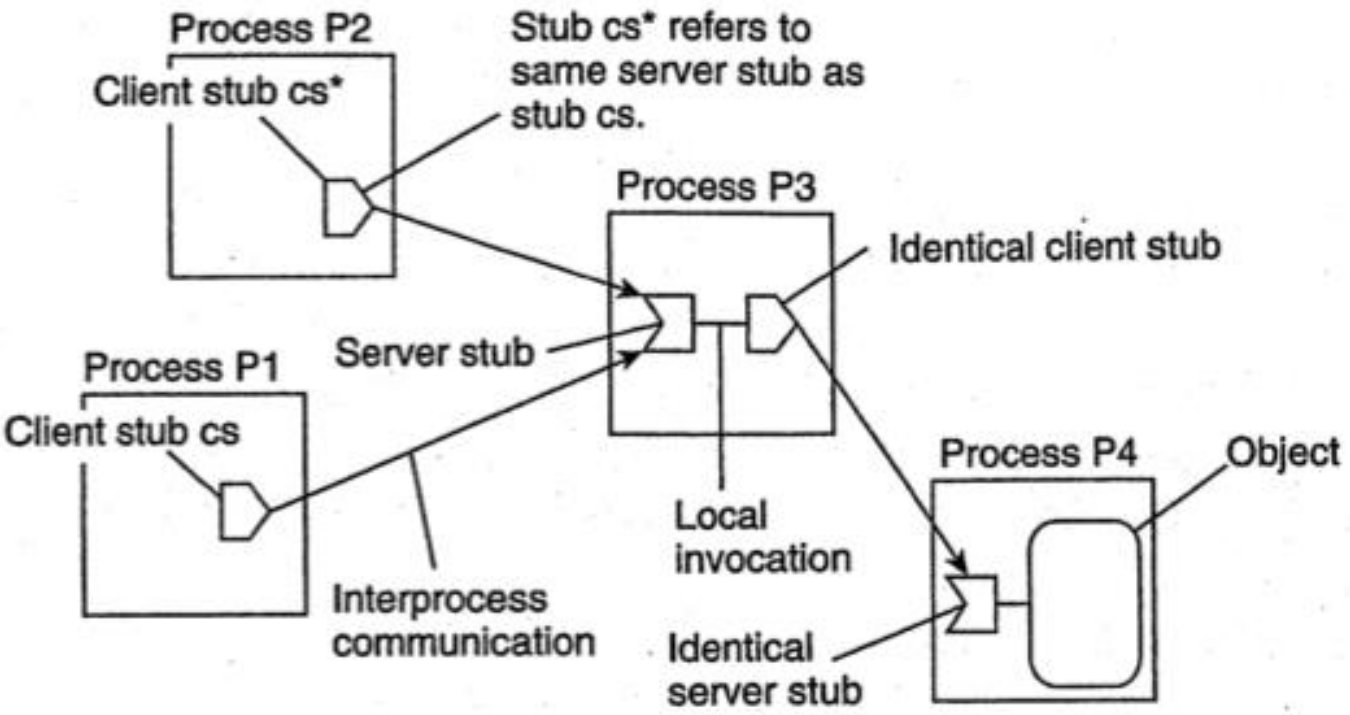
2.2.1. Broadcasting and Multicasting

- **Scalability problem:**
  - Wast network bandwidth by request messages
  - Too many hosts may be interrupted by requests they cannot answer.

2.2.2. Forwarding Pointer

- **When an entity moves from A to B, it leaves behind in A a reference to its new location at B.**

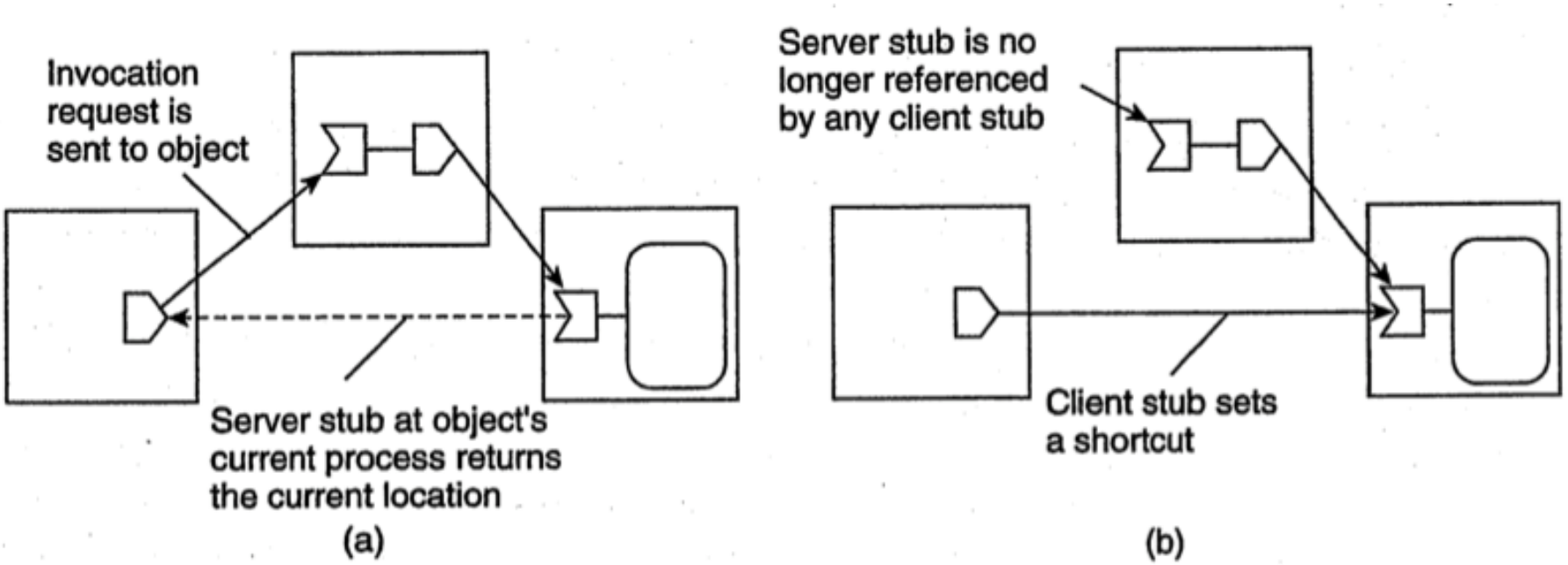
Forwarding Pointer mechanism



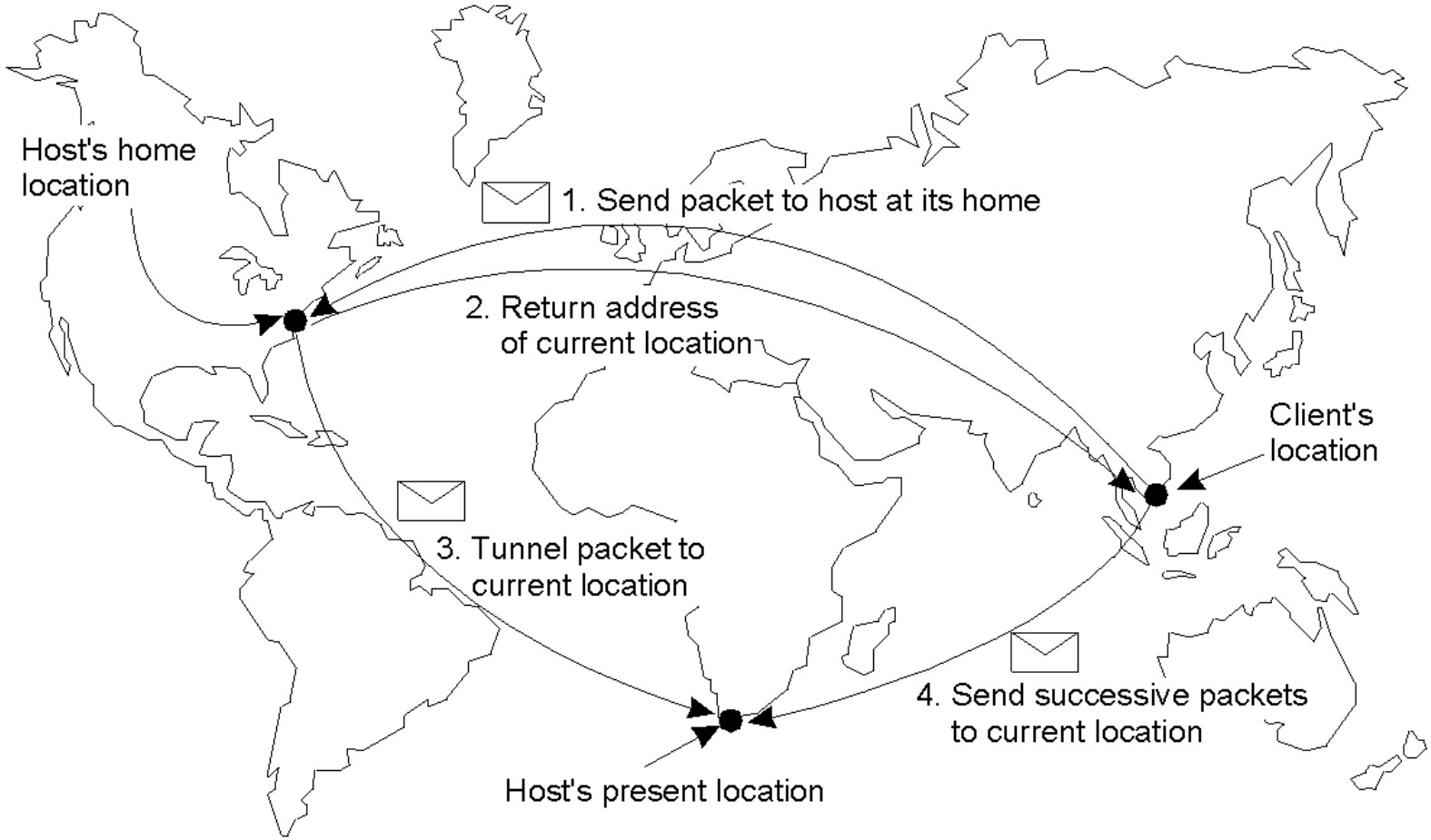
Forwarding Pointer mechanism

- **Advantage:**
  - Simplicity: By using a traditional naming service, a client can look up the current address by following the chain of forwarding pointers.
- **Drawbacks**
  - A chain of FP can become so long → locating that entity is expensive.
  - All intermediate nodes have to maintain their part of the chain.
  - Broken links → cannot reach the entity

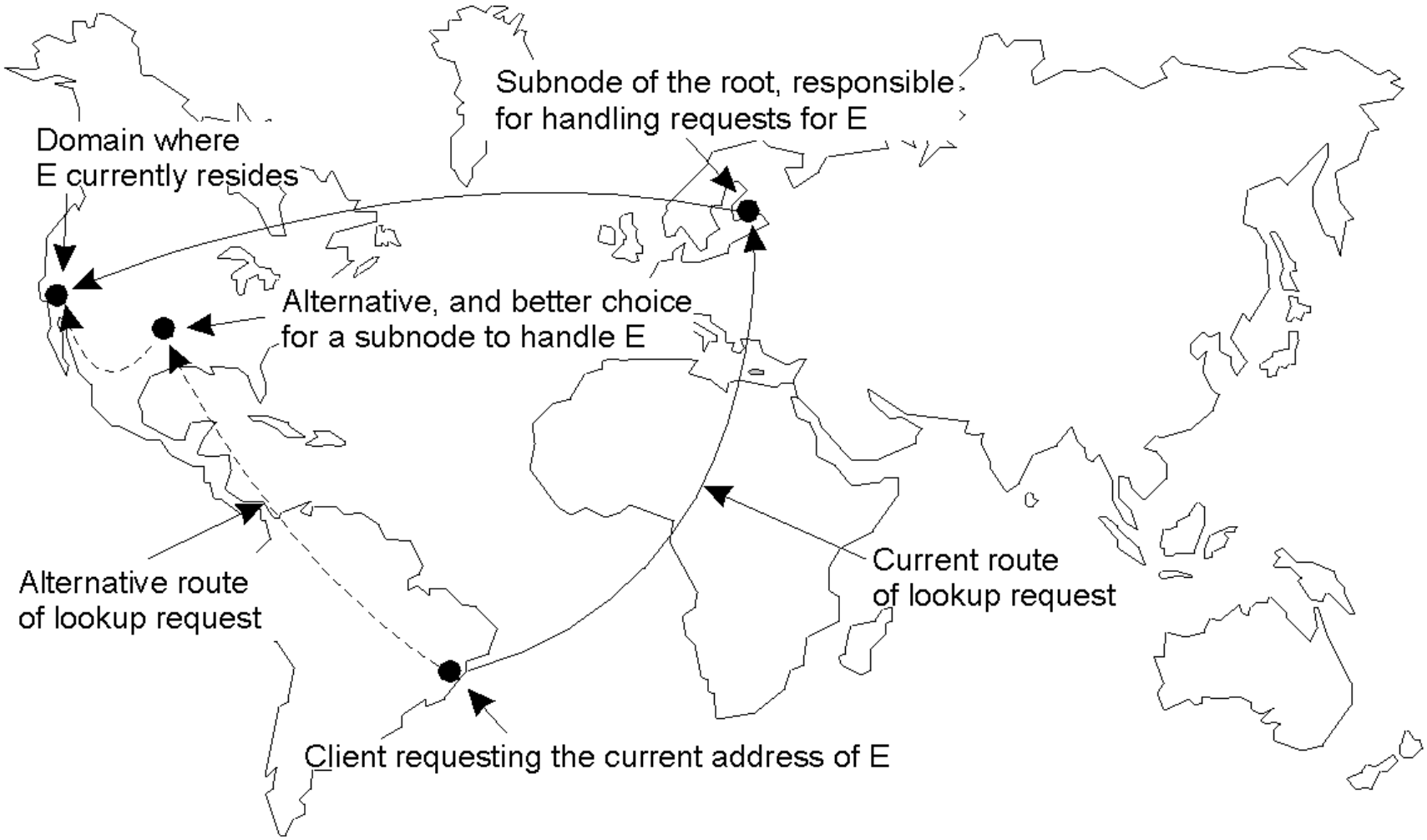
Solution: Redirecting a FP



2.3. Home-based Approaches



Solution for stable home problem



2.4. Distributed Hash Tables

- Chord system
- Create the ring with prev(n) and succ(n)
- Use finger table to determine the succ(k) of key k
- FTp is the finger table of node p:

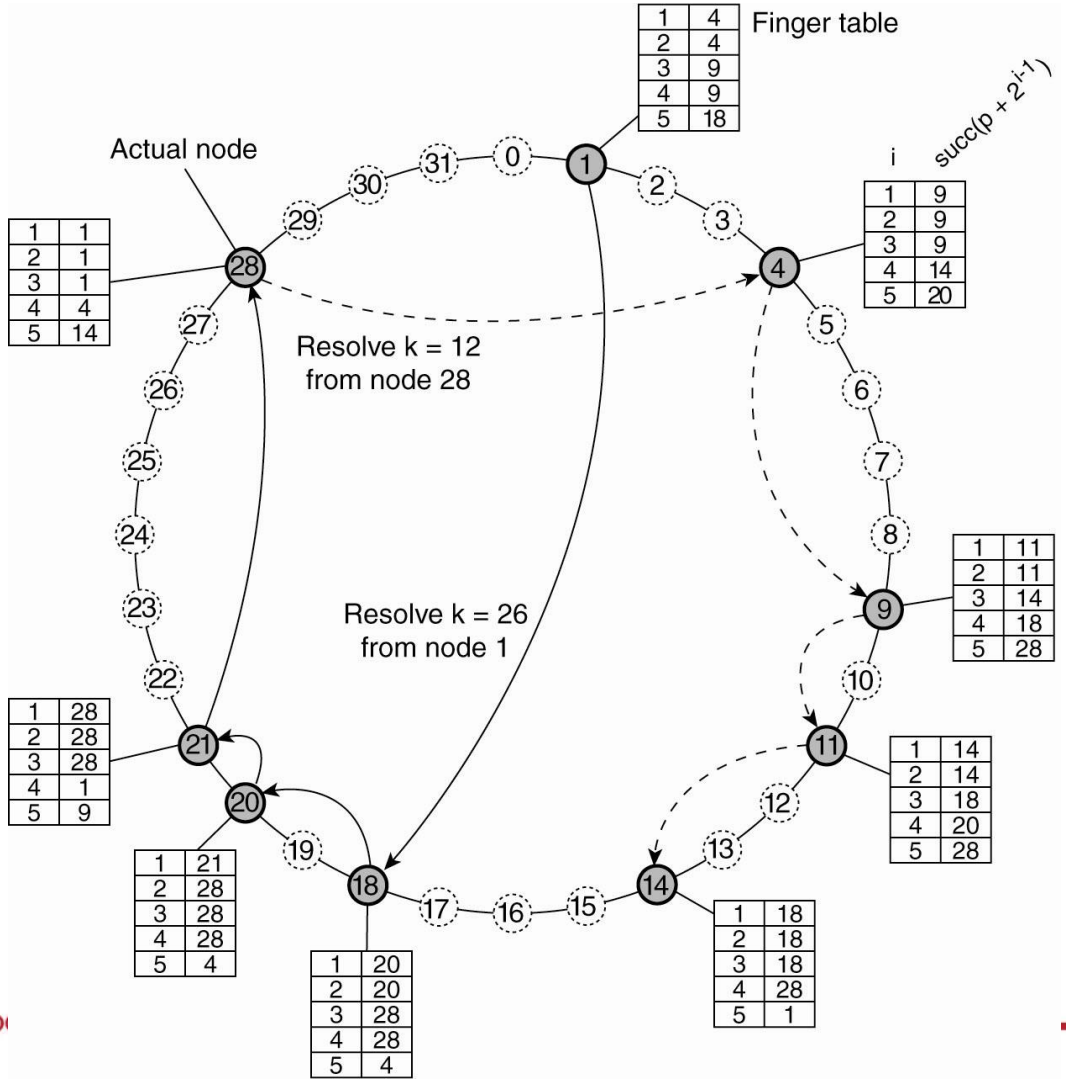
$$FT_p[i] = succ(p+2^{i-1})$$

- To look up a key *k*, node *p* will then immediately forward the request to node *q*:

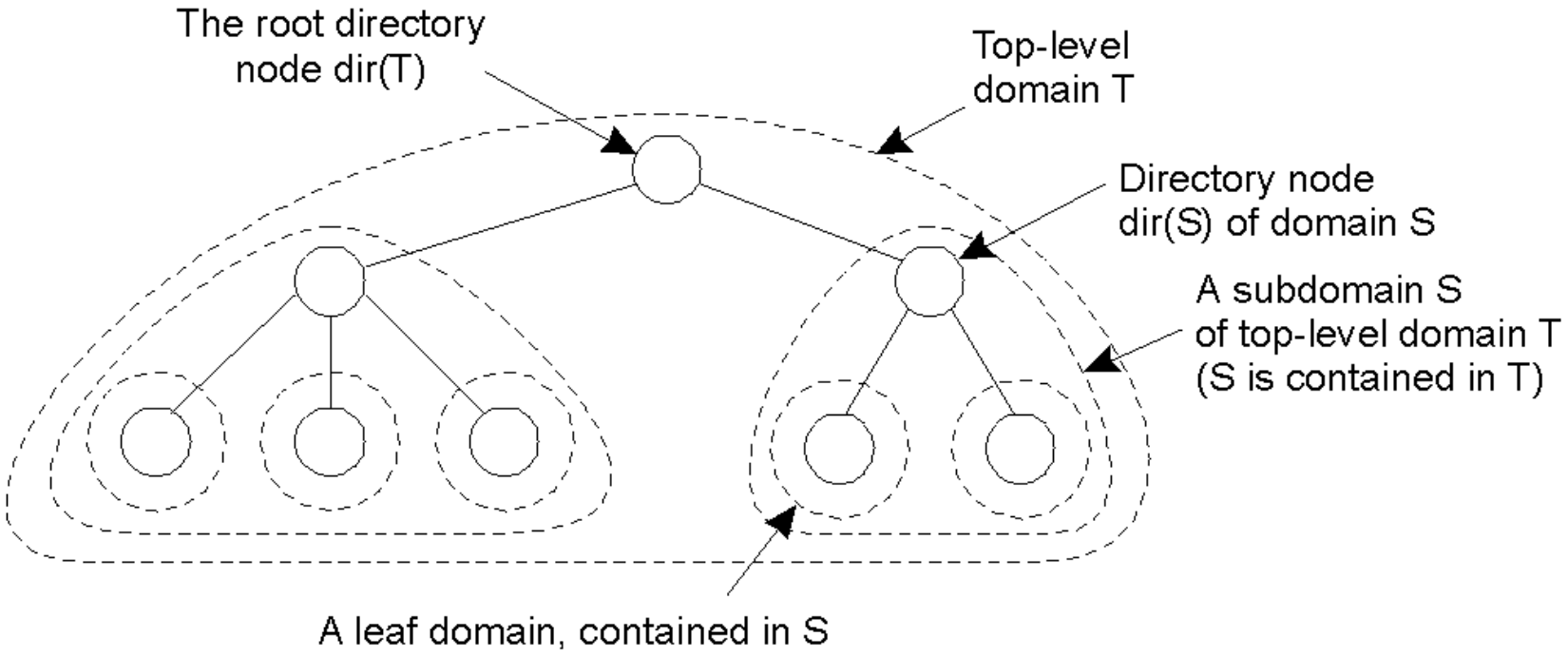
$$q = FT_p[j] \le k < FT_p[j+1]$$

- Update the finger tables after inserting a new node

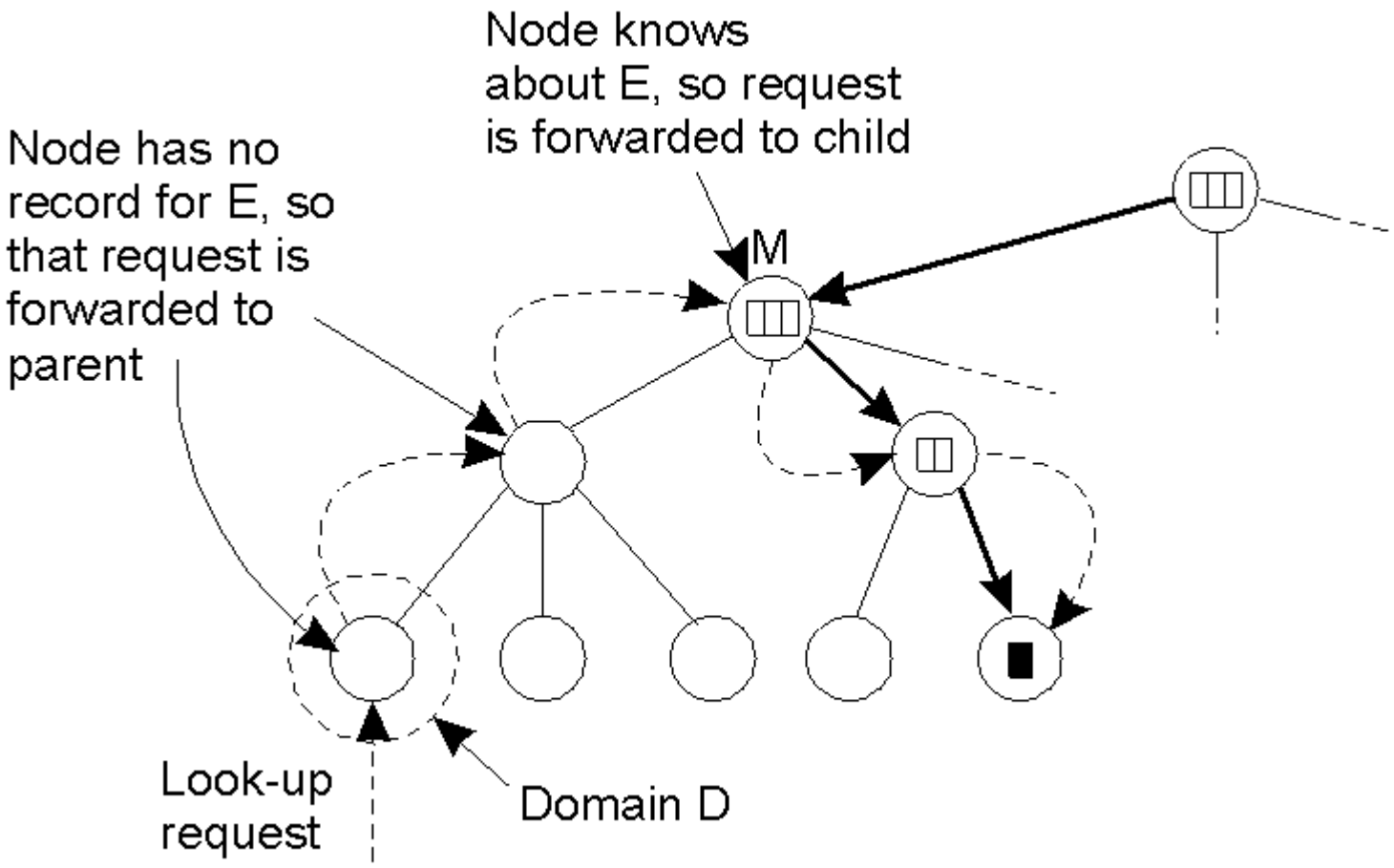
Chord system with finger tables



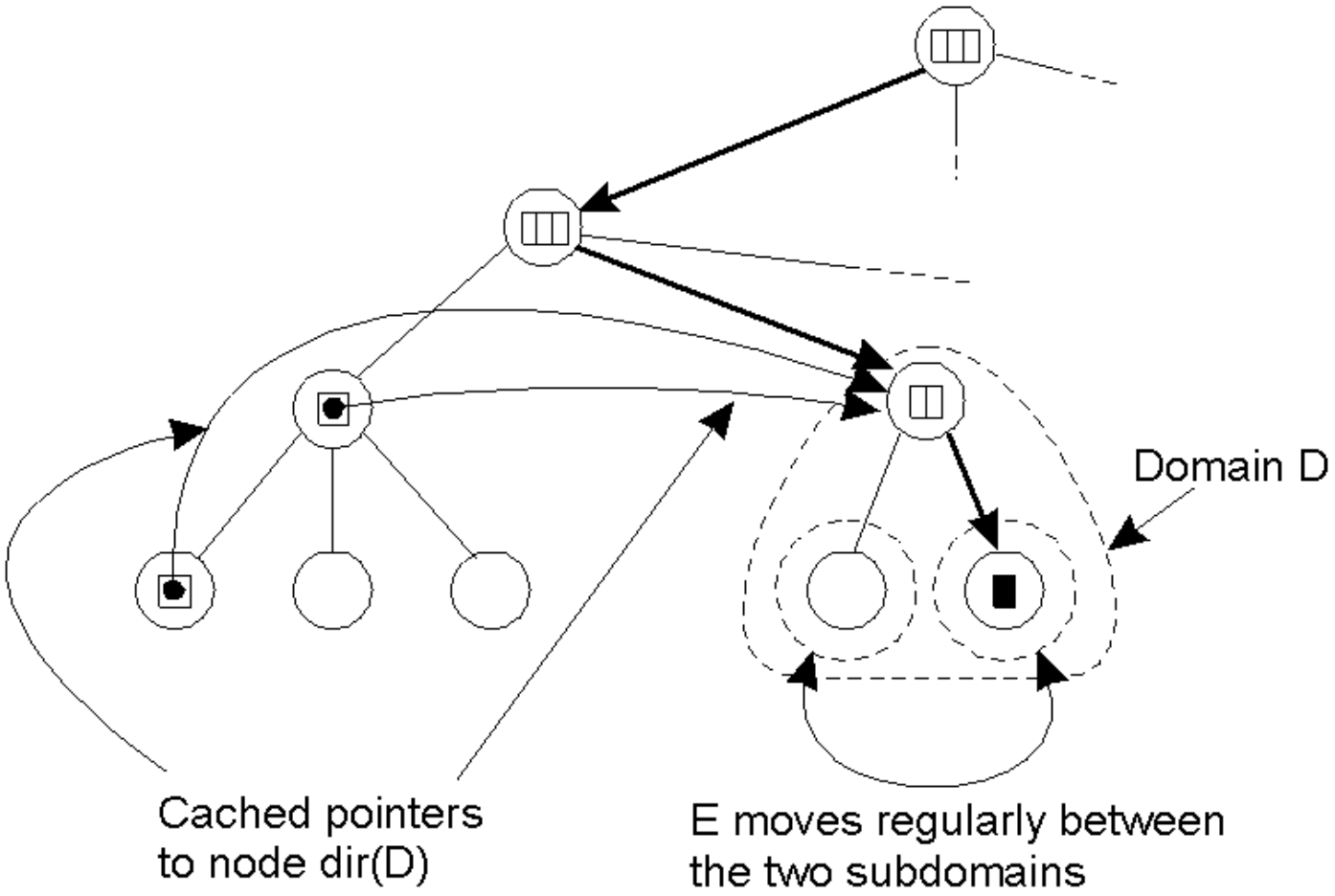
2.5. Hierarchical Approaches



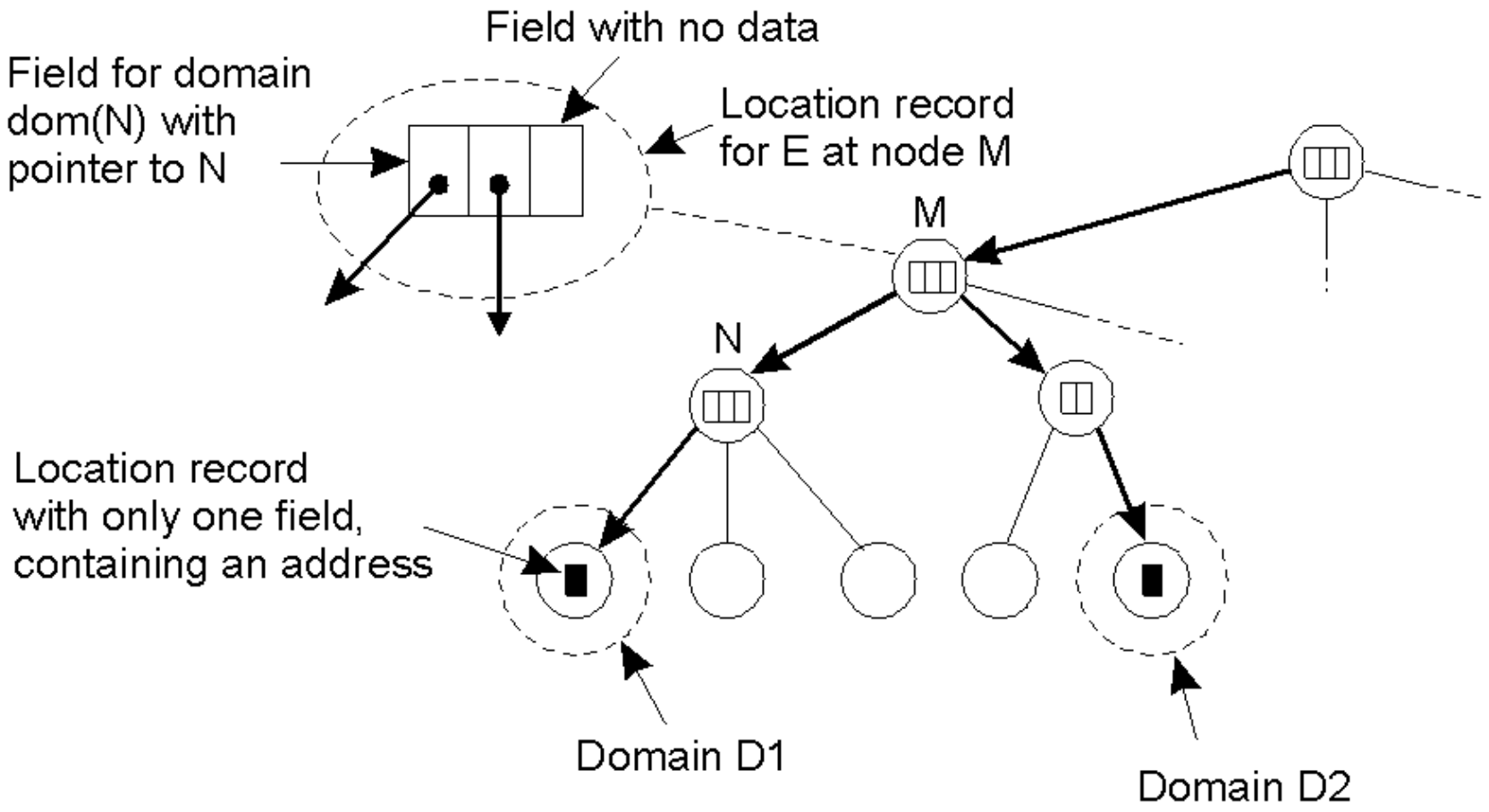
Looking-up


Node knows
about E, so request
is forwarded to child
Node has no
record for E, so
that request is
forwarded to
parent
M
Look-up
request
Domain D

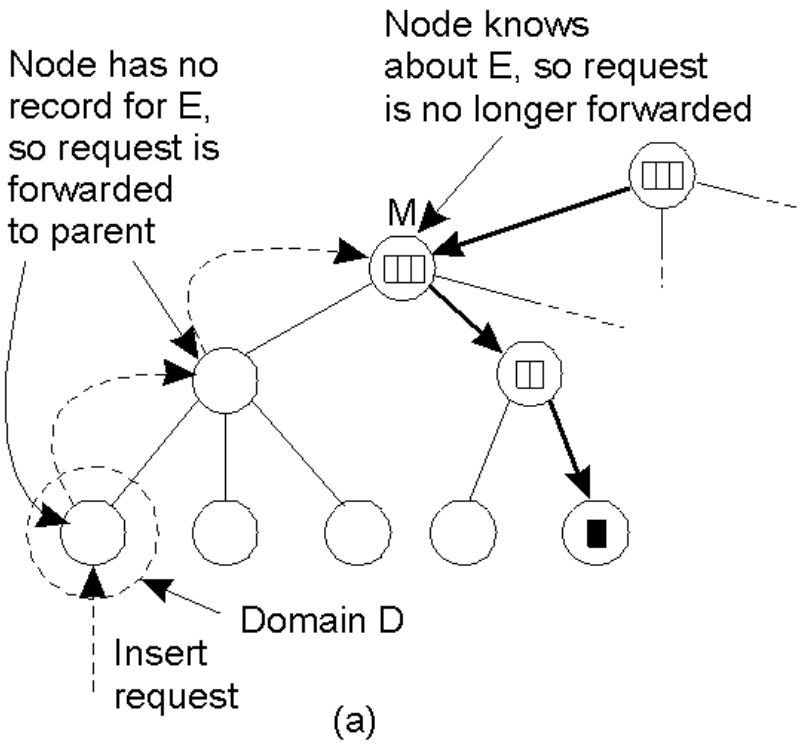
Caching

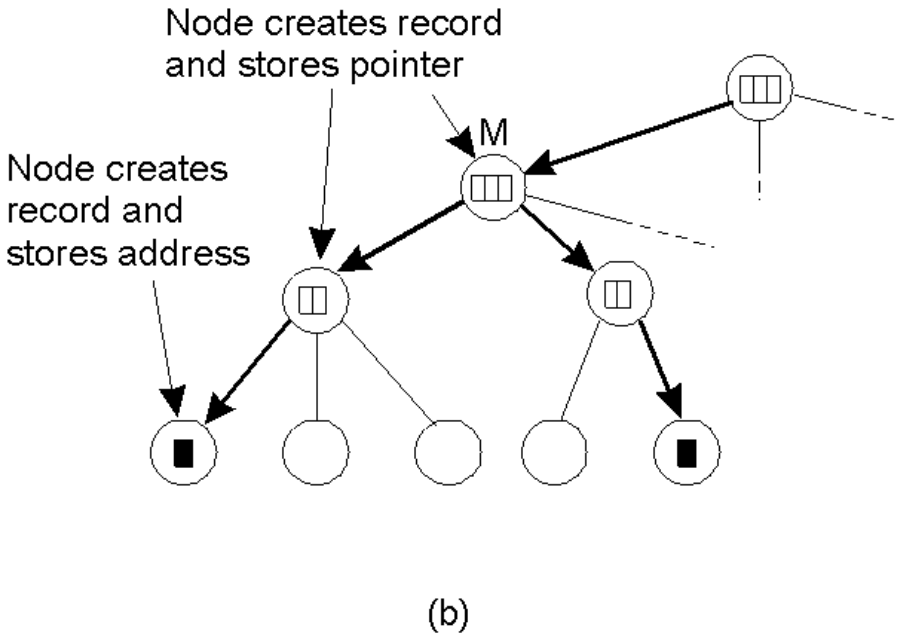

Domain D
Cached pointers
to node dir(D)
E moves regularly between
the two subdomains

An entity having two addresses in different leaf domains



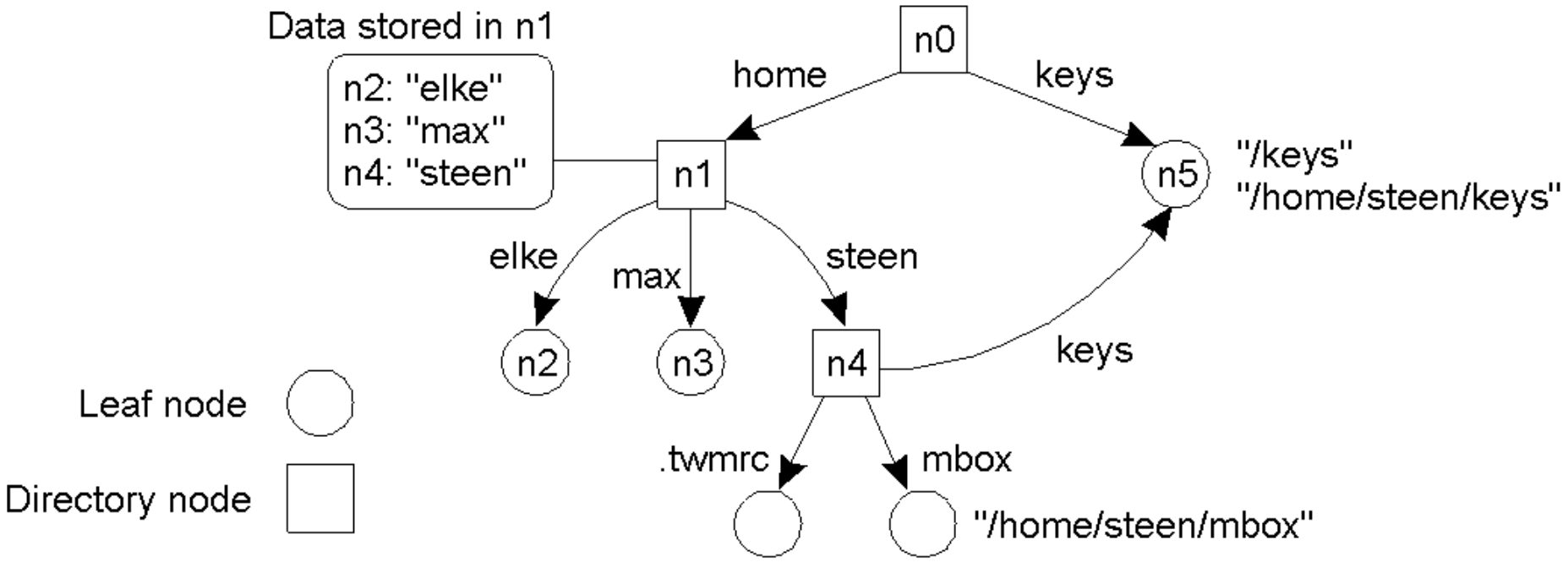
Updating


Node has no record for E, so request is forwarded to parent
Node knows about E, so request is no longer forwarded
M
Domain D
Insert request
(a)


Node creates record and stores pointer
M
Node creates record and stores address
(b)

3. Structured Naming

Structured Name Space



A general naming graph

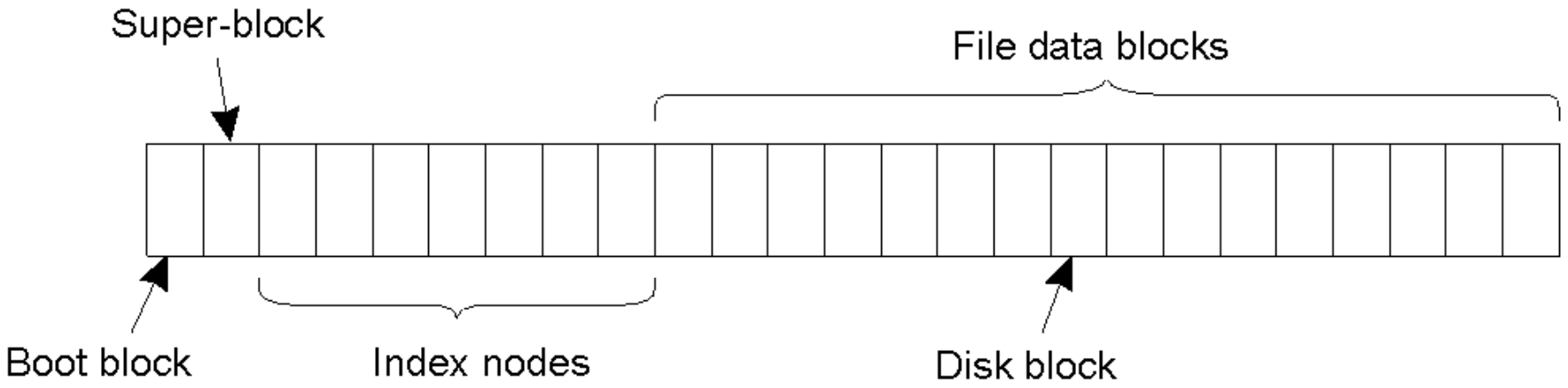
Name Spaces

- **Leaf node:**
  - No outgoing edge
  - Store information of its address
- **Directory node:**
  - Outgoing edge
  - Store a table with info (edge label, node identifier)
- **Path name: N: <label1, label2, label3, label4, ...>**
- **Absolute path name/Relative path name**

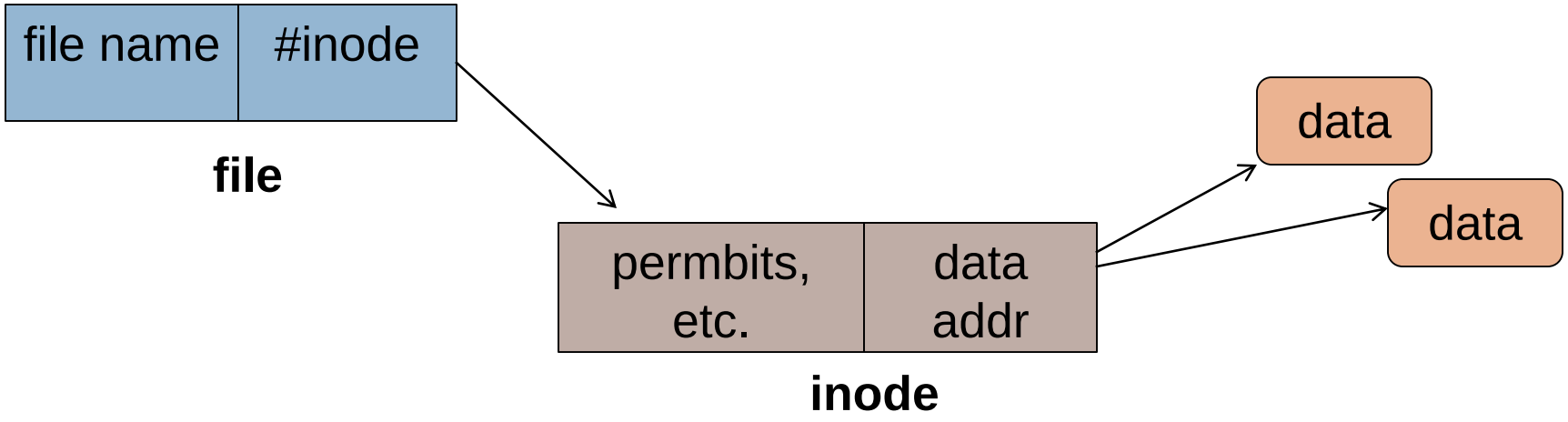
Name resolution

- Consider a path name: N:<label1, label2, ..., labeln>
- Start at node N of the naming graph, where the name label1 is looked up in the directory table, and which returns the identifier of the node to which label1 refers.
- Continue at the identified node by looking up the name label2
- So on ...
- Relatively with the UNIX file system

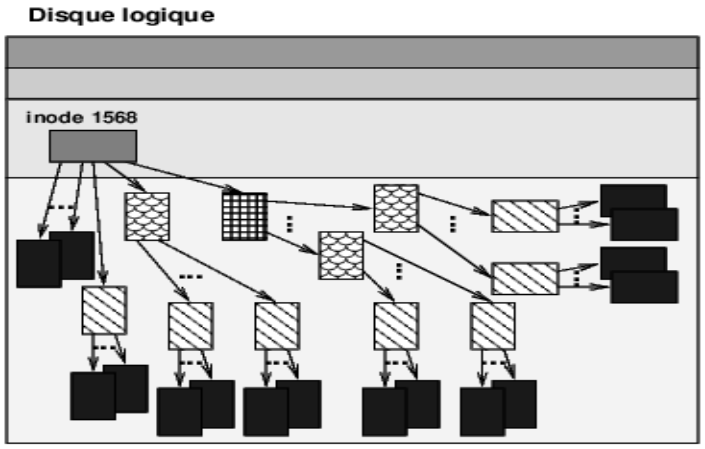
General organization of the UNIX file system

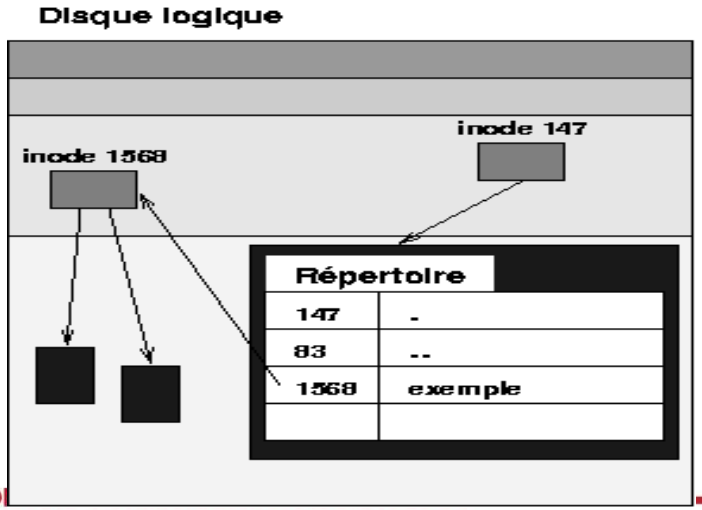


File system in UNIX

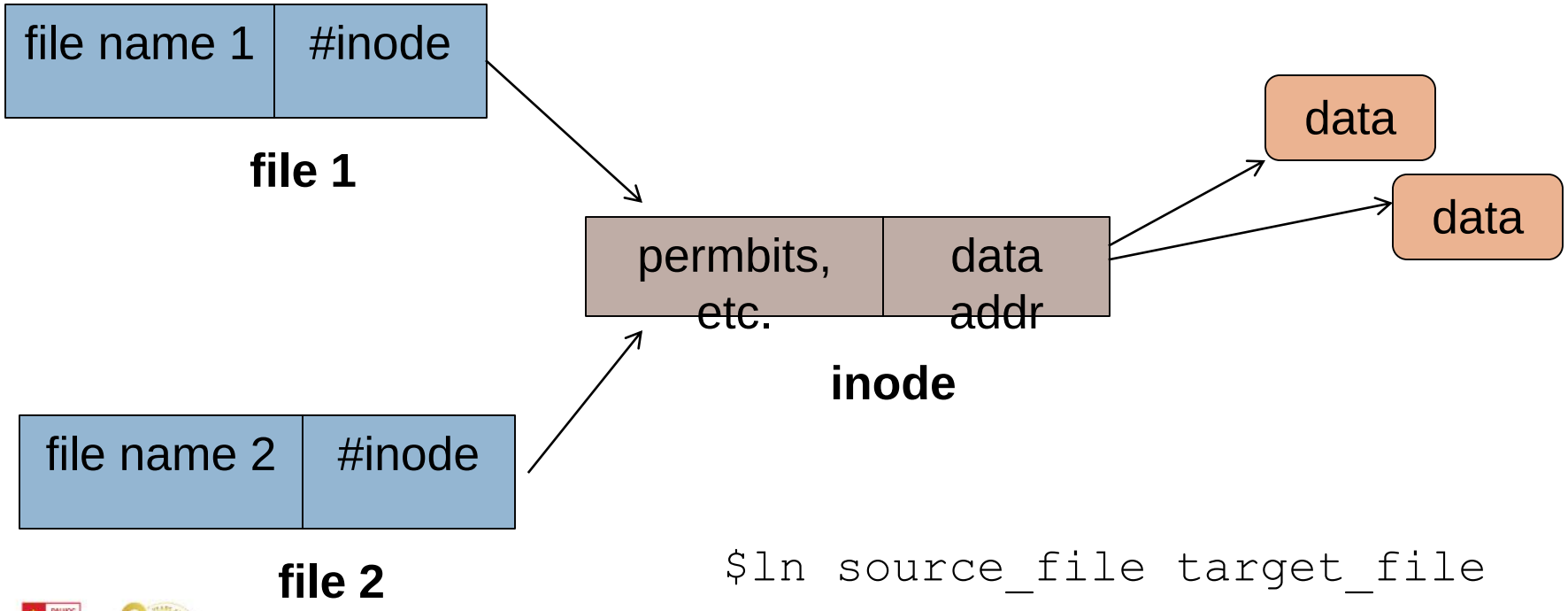


Directory node (folder)



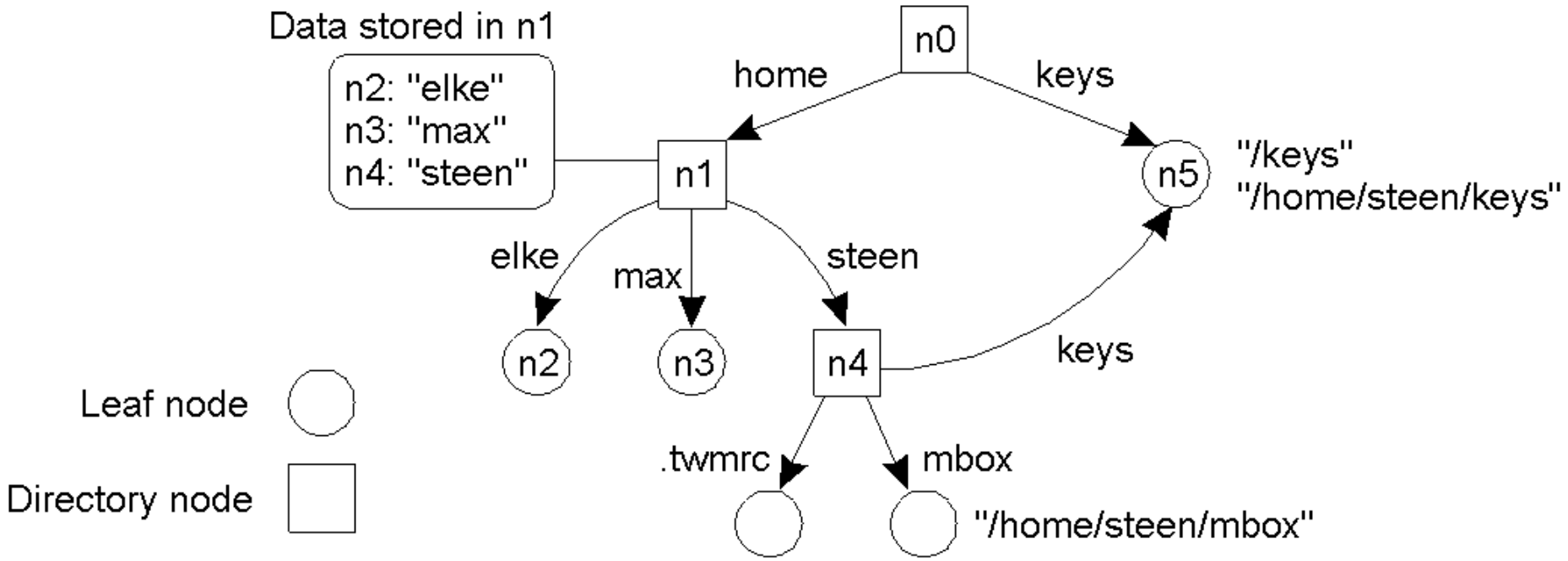


Hard link

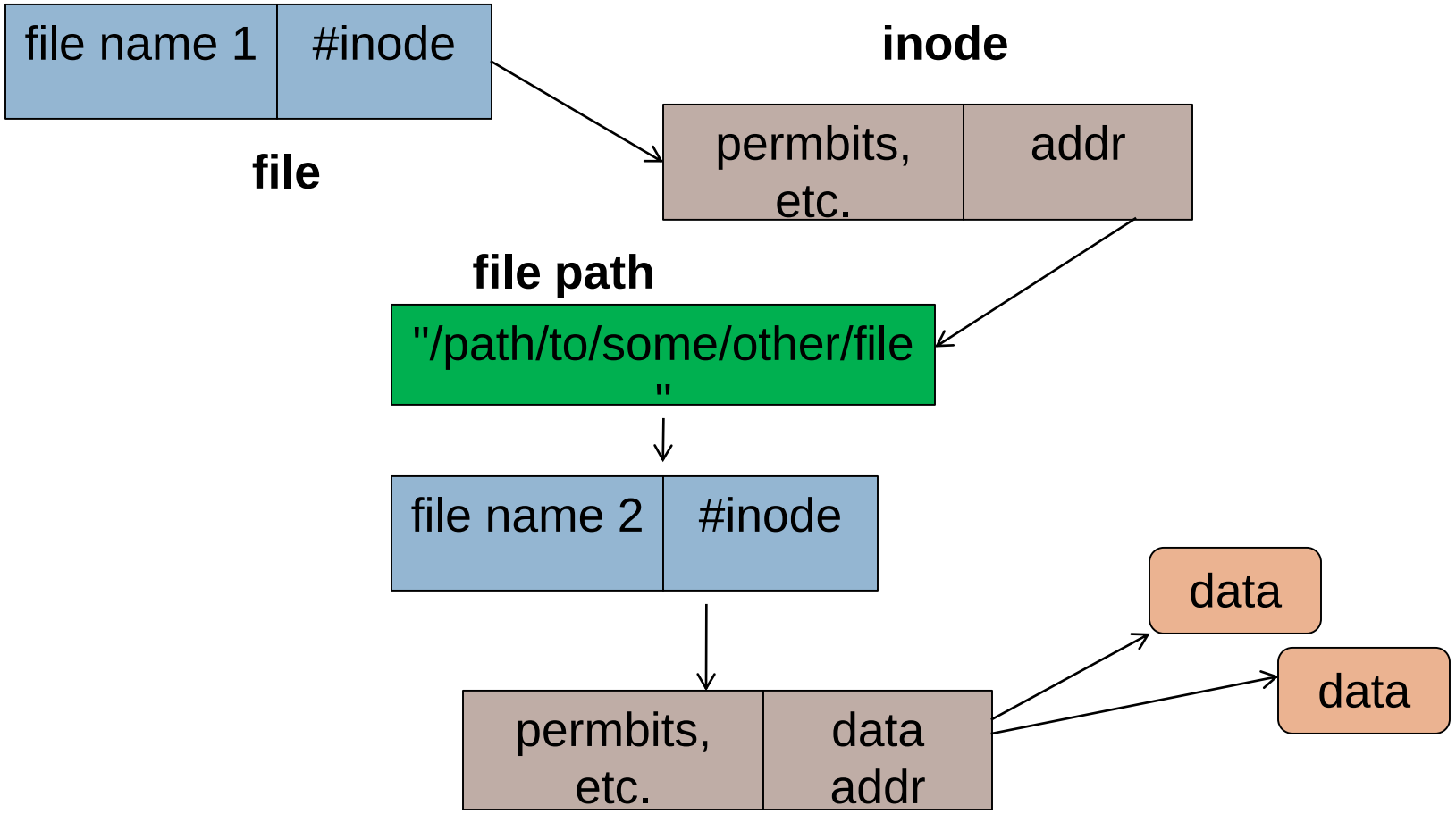



```
$ln source_file target_file
```

Hard link (cont.)

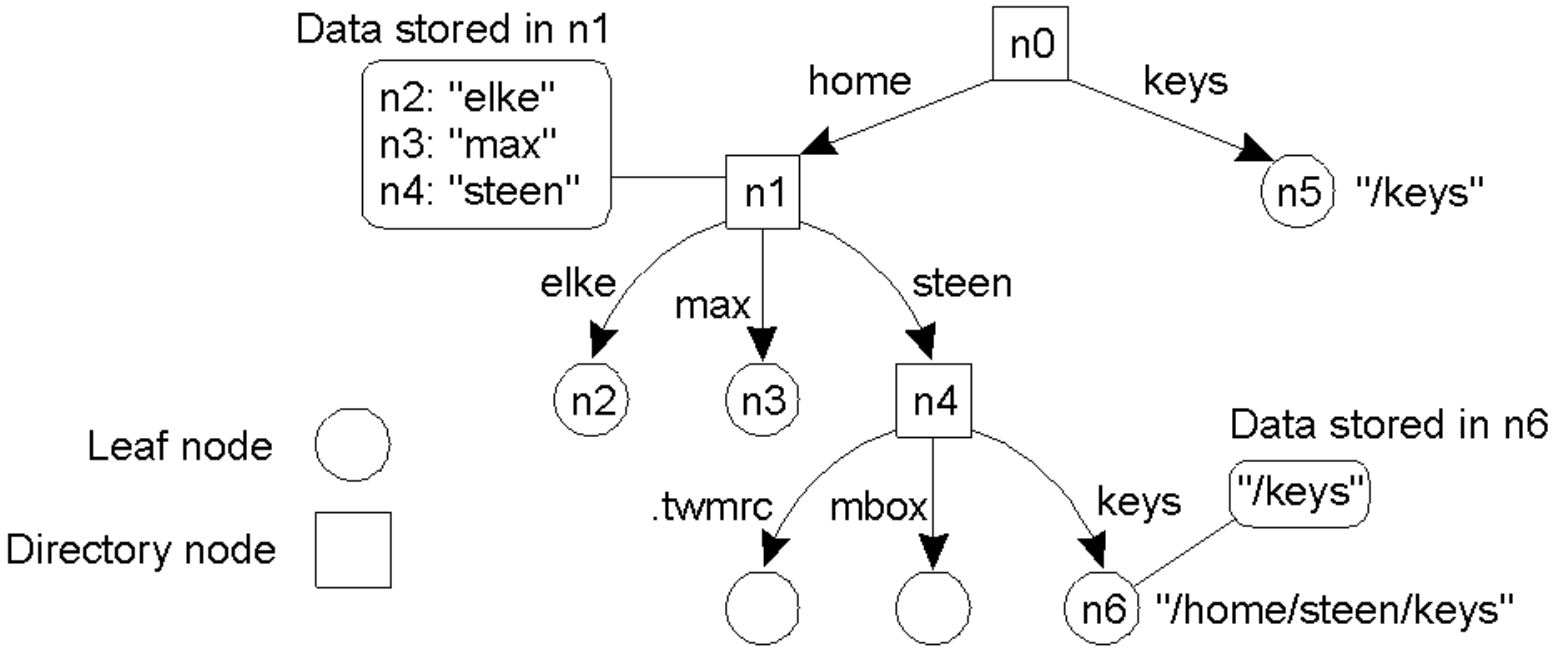


Soft link

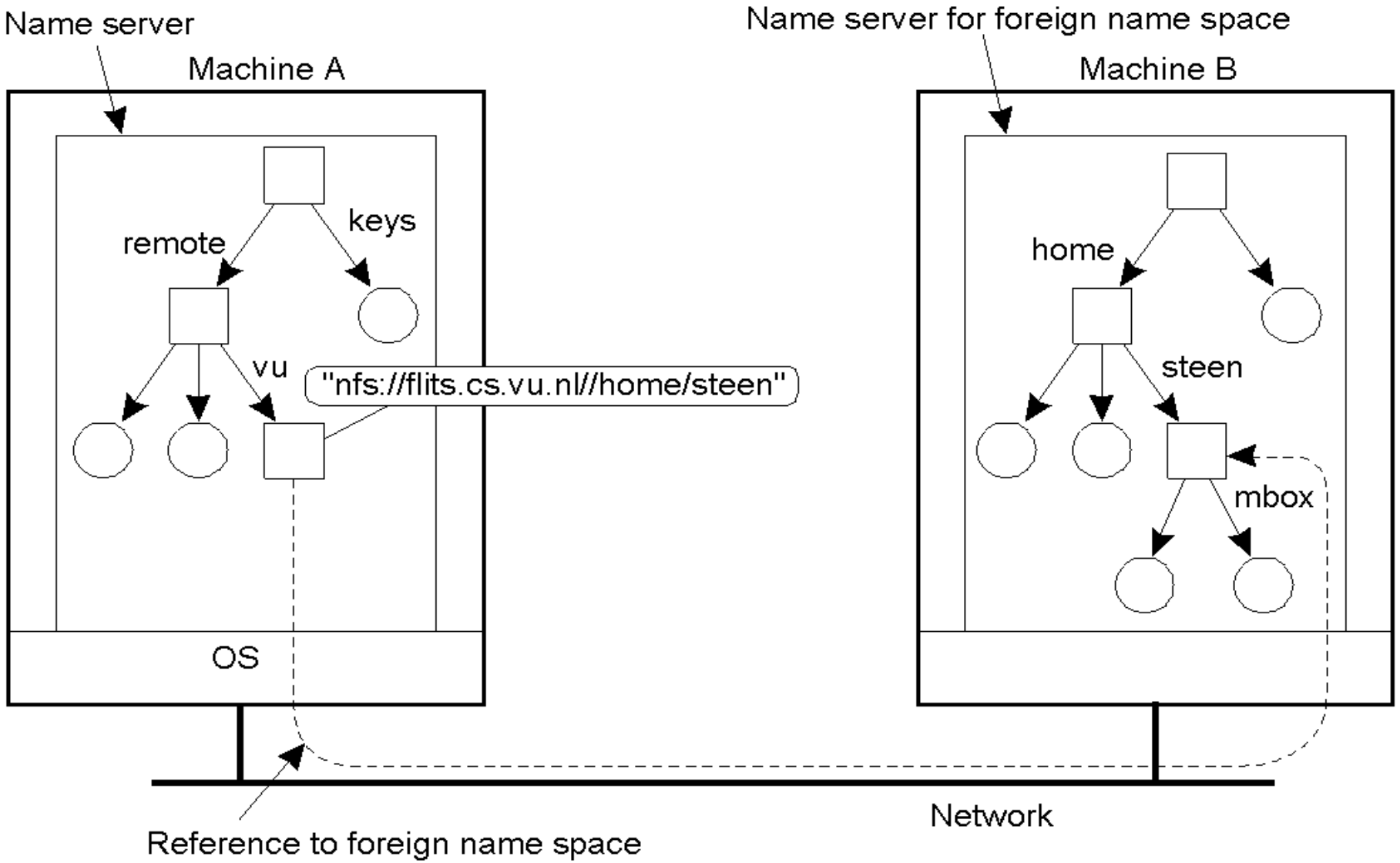



```
$ln -s source_file target_file
```

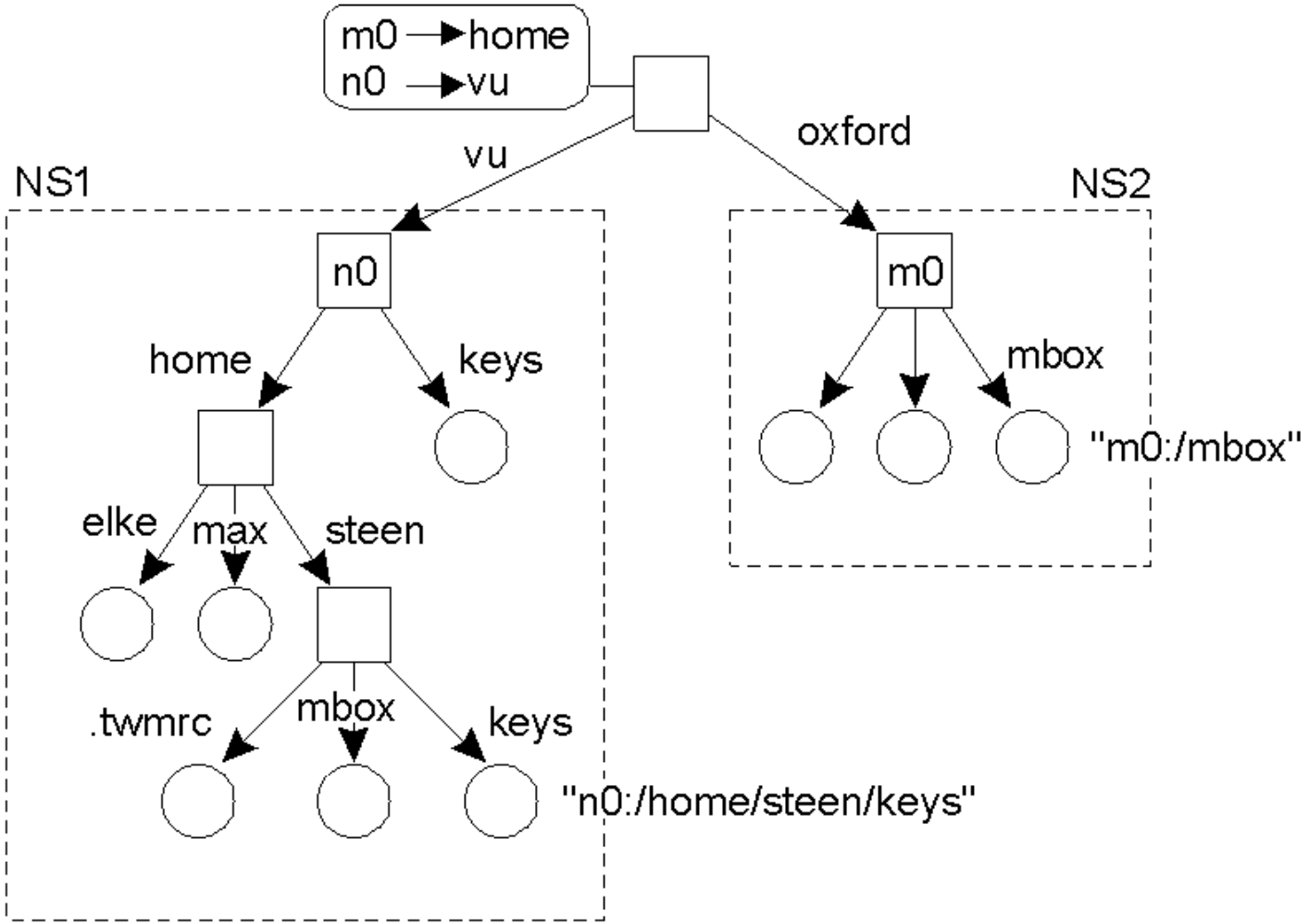
Soft link


Data stored in n1
n2: "elke"
n3: "max"
n4: "steen"
n0
home
keys
n1
n5
"/keys"
elke
max
steen
n2
n3
n4
Data stored in n6
"/keys"
.twmrc
mbox
keys
n6
"/home/steen/keys"
Leaf node
Directory node

Mounting


Name server
Machine A
Name server for foreign name space
Machine B
remote
keys
vu
"nfs://flits.cs.vu.nl//home/steen"
home
steen
mbox
OS
Network
Reference to foreign name space

Merging


m0 → home
n0 → vu
vu
oxford
NS1
NS2
n0
m0
home
keys
mbox
"m0:/mbox"
elke
max
steen
.twmrc
mbox
keys
"n0:/home/steen/keys"

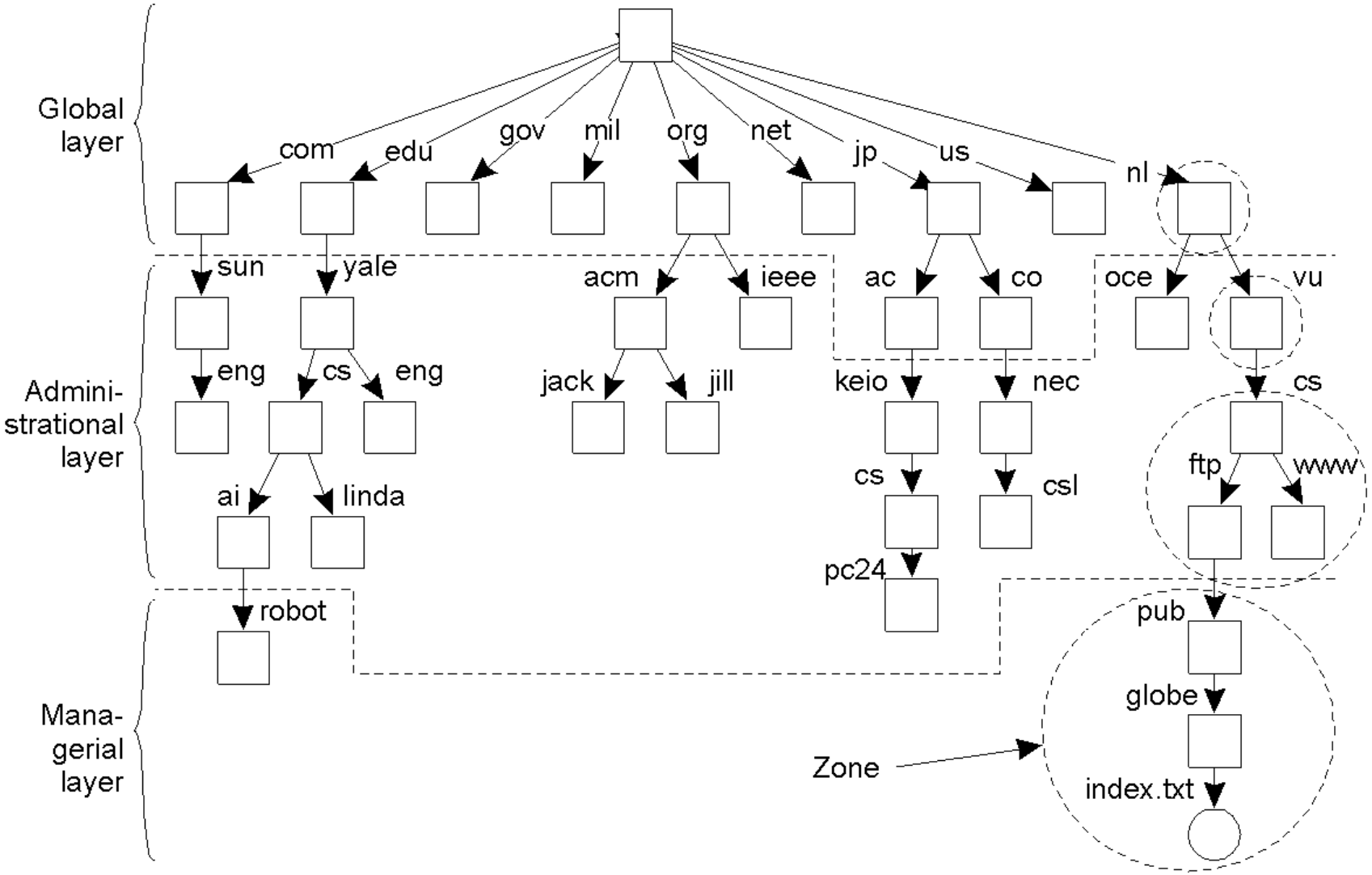
Naming service

- Functions:
  - Adding names
  - Removing names
  - Looking up names
- Naming service is implemented by name servers
- In large-scale distributed systems (many entities, large geographical area) → distribute the implementation of a name space over multiple name servers

Hierarchical organization

- **Global layer**
  - root node + directory nodes logically close to the root (children)
  - Stability (rarely changed)
  - represent organization, or group of organization
- **Administrational layer**
  - represent groups of entities that belong to the same organization
- **Managerial layer**
  - consist of nodes that may change regularly

DNS name space



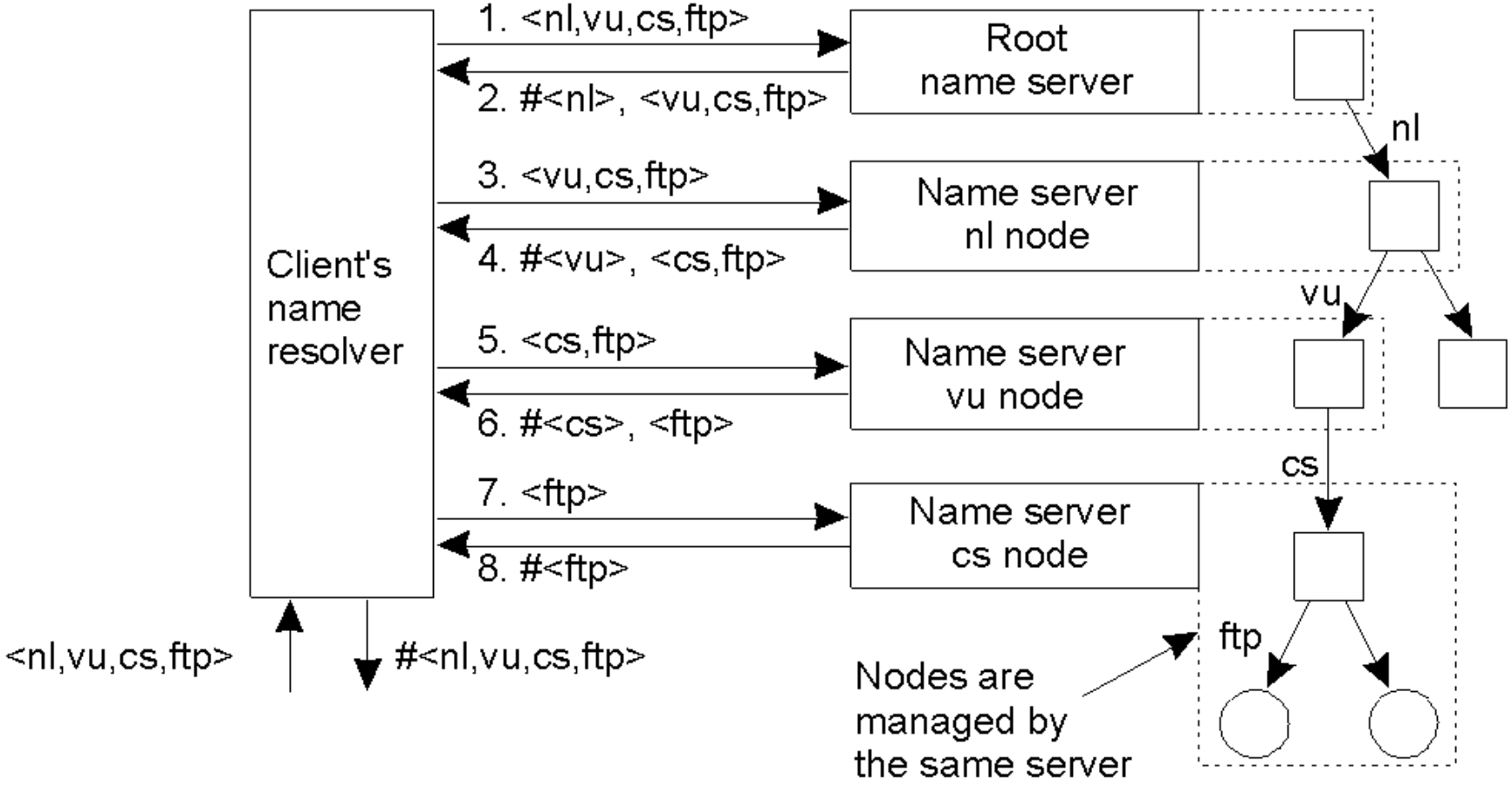
Comparison of three layers

Item	Global	Administrational	Managerial
Geographical scale of network	Worldwide	Organization	Department
Total number of nodes	Few	Many	Vast numbers
Responsiveness to lookups	Seconds	Milliseconds	Immediate
Update propagation	Lazy	Immediate	Immediate
Number of replicas	Many	None or few	None
Is client-side caching applied?	Yes	Yes	Sometimes

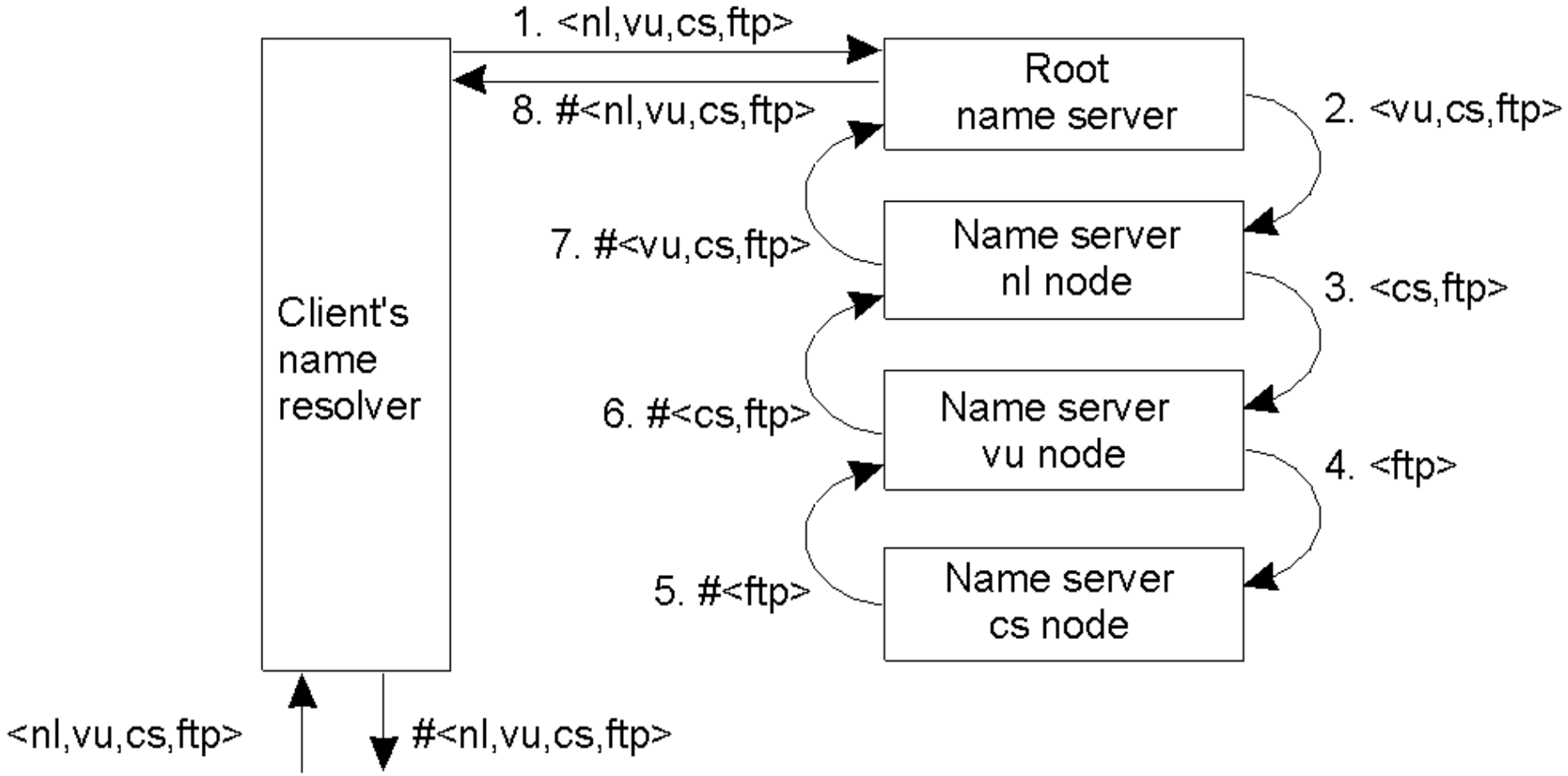
Implementation of Name Resolution

- **Depend on the distribution of a name space across multiple name servers**
- **Each client has a name resolver**
- **2 ways of implementation of name resolution:**
  - Iterative name resolution
  - Recursive name resolution

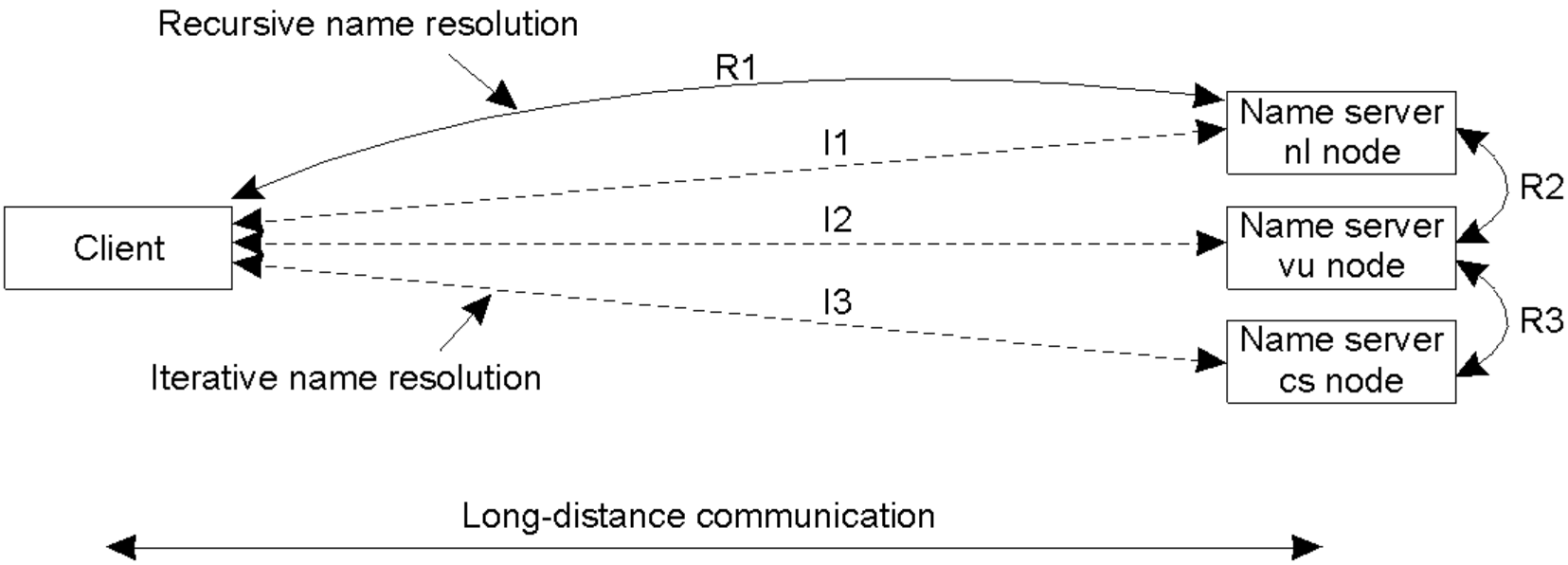
Iterative name resolution



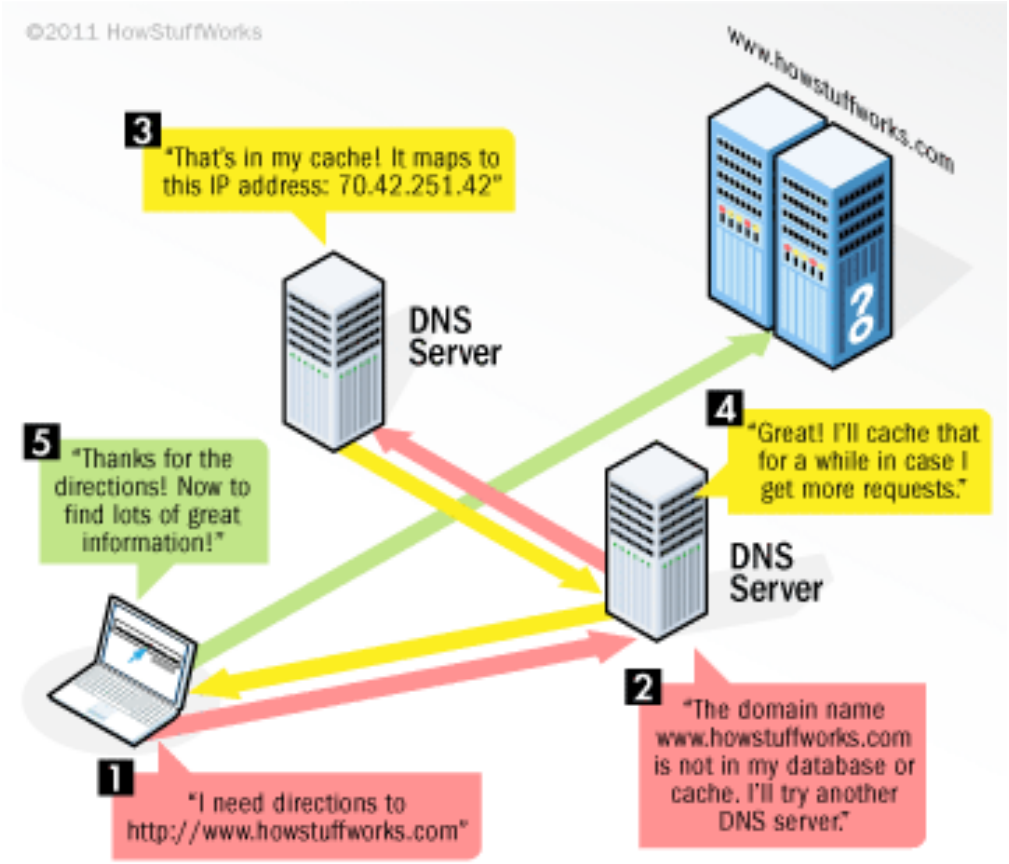
Recursive name resolution



Recursive vs. iterative name resolution



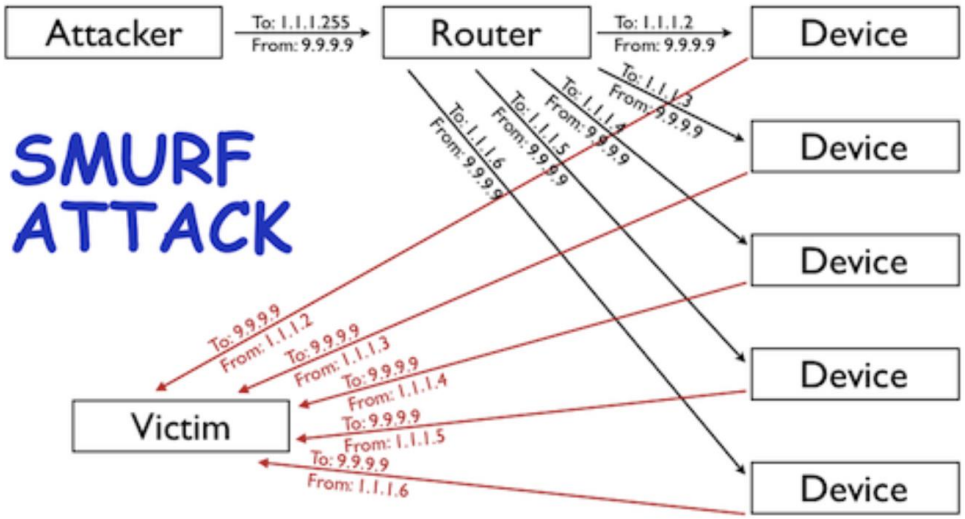
Example: DNS



[source: howstuffworks.com]

Security risk of recursive name resolution

Amplification attack



Problem: ICMP or UDP has no authentication mechanism

DNS Amplification attack


```
dig ANY isc.org @x.x.x.x
```

64 bytes query

```
; <<>> DiG 9.7.3 <<>> ANY isc.org @x.x.x.x
;; global options: +cmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 5147
;; flags: qr rd ra; QUERY: 1, ANSWER: 27, AUTHORITY: 4, ADDITIONAL: 5

;; QUESTION SECTION:
;isc.org.                       IN      ANY

;; ANSWER SECTION:
isc.org.                4084    IN      SOA     ns-int.isc.org. hostmaster.isc.org. 2012102700 7200 3600 24796800 3600
isc.org.                4084    IN      A       149.20.64.42
isc.org.                4084    IN      MX      10 mx.paol.isc.org.
isc.org.                4084    IN      MX      10 mx.amsl.isc.org.isc.org.             4084    IN      TXT     "v=spf1 a mx ip4:204.152.184.0/21 ip4:149.20.0.0/16 ip6:2001:04F8::0/32 ip6:2001:500:60::65/128 ~all"
isc.org.                4084    IN      TXT     "$Id: isc.org,v 1.1724 2012-10-23 00:36:09 bind Exp $"
isc.org.                4084    IN      AAAA    2001:4f8:0:2::d
isc.org.                4084    IN      NAPTR   20 0 "S" "SIP+D2U" "" _sip._udp.isc.org.
isc.org.                484     IN      NSEC    _kerberos.isc.org. A NS SOA MX TXT AAAA NAPTR RRSIG NSEC DNSKEY SPF
isc.org.                4084    IN      DNSKEY  256 3 5 BQEAAAAB2F1v2HWzCCE9vNsKfk0K8vd4EBwizNT9KO6WYXj0oxEL4eOJ aXbax/BzPFx+3qO8B8pu8E/JjkWH0oaYz4guUyTVmT5Eelg44Vb1kssy q8W27oQ+9qNiP8Jv6zdOj0uCB/N0fxfVL337lxbednFqoECfSFDZa6Hw jU1qzveSsW0=
isc.org.                4084    IN      DNSKEY  257 3 5 BEAAAAOhHQDBrhQbtphgq2wQUpEQ5t4DtUHxoMVPu2hWLDMvoOMRXjGr hhCeFvAZih7yJHf8ZGfW6hd38hXG/xylYCO6KrpbdojwxBYMXLA5/kA+ u50WIL8ZR1R6KTbsYVMf/Qx5RiNbPC1w+vT+U8eXEJmO2OjIS1ULgqy3 47cBBlzMnnz/4LJpA0da9CbKj3A254T515sNIMcwsB8/2+2E63/zZrQz BkjOBrN/9Bexjpiks3jRhZatEaXn3dTy47R09Uix5WcJt+xzqZ7+ysyL KOoedS39Z7Dmsn2eA0FKtQpwA6LXeG2w+jxmw3oA8lVUgEf/rzeC/hB yBNsO70aEFTd
isc.org.                4084    IN      SPF     "v=spf1 a mx ip4:204.152.184.0/21 ip4:149.20.0.0/16 ip6:2001:04F8::0/32 ip6:2001:500:60::65/128 ~all"
isc.org.                484     IN      RRSIG   NS 5 2 7200 20121125230752 20121026230752 4442 isc.org. oFeNy69Pn+/JnnltGPUZQnYzo1YGglMhS/SZKnlgyMbz+tT2r/2v+X1j AkUl9GRW9JAZU+x0oEj5oNAkRiQqK+D6DC+PGdM2/JHa0X41lnMIE2NX UHDAKMmbqk529fUy3MvA/ZwR9FXurcfYQ5fnpEEaawNS0bKxomw48dcp Aco=
isc.org.                484     IN      RRSIG   SOA 5 2 7200 20121125230752 20121026230752 4442 isc.org. S+DLHzE/8WQbnSl70geMYoKvGlIuKARVlxmssce+MX6DO/JlxdK9xGac XCuAhRpTMKElKq2dIhKp8vnS2e+JTZLrGl4q/bnrrmhQ9eBS7lFmrQ6s OcKEEyuijumOPlKCCN9QX7ds4siiTIrEOGhCaamEgRJqVxqCsgldBUrR hKk=
isc.org.                484     IN      RRSIG   MX 5 2 7200 20121125230752 20121026230752 4442 isc.org. VFqFWRPyulIT8VsldXKMpMRJTYpdggoGgOjKJzKJs/6ZrxmbJtmAxgEu /rkwD6Q9JwsUCepNC74EYxzXFvDaNnKp/Qdmt2139h/xoZsw0JVA4Z+b zNQ3kNiDjdV6zl6ELtCVDqj3SiWDZhYB/CR9pNnolFAF2joIjYSwiwbS Lcw=
isc.org.                484     IN      RRSIG   TXT 5 2 7200 20121125230752 20121026230752 4442 isc.org. Ojj8YCZf3jYL9eO8w4Tl9HjWKP3CKXQRFed8s9xeh5TR3KI3tQTKsSeI JRQaCXkADiRwHt0j7VaJ3xUHa5LCkzetcVgJNPmhovVa1w87Hz4DU6q9 k9bbshvbYtxOF8xny/FCiR5c6NVeLmvvu4xeOqSwIpoo2zvIEfFP9deR UhA=
isc.org.                484     IN      RRSIG   AAAA 5 2 7200 20121125230752 20121026230752 4442 isc.org. hutAcroONBMvKU/m+21F8sgIYyIVWORTp/utIn8KsFlWOwwM2QMGa5C9 /rH/ZQBQqN46ZMmiEm4LxH6mtaKxMsBGZwgzUEdfsvVtr+fS5NUoAlrF wg92eBbInNdCvTOif8m1Sldx5/hSqKn8EAscKfg5BMQp5YDFsllsTauA 8Y4=
isc.org.                484     IN      RRSIG   NAPTR 5 2 7200 20121125230752 20121026230752 4442 isc.org. ZDl4qEHRTjVXn5uJUn6XR9Lvt5Pa7YTEW94hNAn9Lm3TlnkglIAeZiOU 3woQlpg+esCQepKCiBlplPLcag3LHlQl9OdACrHGUzxM+rnHY50Rr/H4 XQTqDWHBF2Cs0CvfqRxlvAl5AY6P2kb/iUQ6hV8GoO0PvmMEkJOnxPPw 5i4=
isc.org.                484     IN      RRSIG   NSEC 5 2 3600 20121125230752 20121026230752 4442 isc.org. rYlhqZAtyM045vv3bMY0wqJhxHJQofkXLeRLk20LaUlmVTyu7uair7jb MwDVCVhxF7gfRdgu8x7LP8vJKUl6sn73lY80CnGwszXBp6tVpgw6oOcr Pi0rsnzC6lIarXLwNBFmLZqZAza6SSirrOPCbnmK6PLQCdmaVAPrVJQs FHY=
isc.org.                484     IN      RRSIG   DNSKEY 5 2 7200 20121125230126 20121026230126 4442 isc.org. i0S2MFqvHB3wOhv2IPozE/IQABM/eDDCV2D7dJ3AuOwilA3sbYQ29XUd BK82+mxxsET2U6hv64crpbGTNJP3OsMxNOAFA0QYphoMnt0jg3OYg+AC L2j92kx8ZdEhxKiE6pm+cFVBHLLLmXGKLDaVnffLv1GQIl5YrIyy4jiw hOA=
isc.org.                484     IN      RRSIG   DNSKEY 5 2 7200 20121125230126 20121026230126 12892 isc.org. jlkgWw+wFFw01E2z2kXq+biTGlrrnGlXoPl7pIOToZHElgpy7F6kEgyj fN6e2C+qvXxOAABQ+qr76o+P+ZUHrLUEI0ewtC3v4HziME10Z2/NE0MH qAEdmEemezKn9OlEAOC7gZ4nU5psmuYlqxcCkUDbW0qhLd+u/8+d6L1S nlrD/vEi4RlSL12bD5VBtakczOz+2BEQLveUt/UusSlqhYcFjdCYbHqF JGQziTJv9ssbEDHT7COc05gG+AlAv5tNN5ag7QHWa0VE+Ux0nH7JUy0N ch1kVecPbXJVHRF97CEH5wCDEgcFKAyyhaXXh02fqBGfON8R5mIcgO/F DRdXjA==
isc.org.                484     IN      RRSIG   SPF 5 2 7200 20121125230752 20121026230752 4442 isc.org. IB/bo9HPjr6aZqPRkzf9bXyK8TpBFj3HNQloqhrguMSBfcMfmJqHxKyD ZoLKZkQk9kPeztau6hj2YnyBoTd0zIVJ5fVSqJPuNqxwm2h9HMsl40r3 9HmbnkO7Fe+Lu5AD0s6+E9qayi3wOOwunBgUkkFsC8BjiiGrRKcY8GhC kak=
isc.org.                484     IN      RRSIG   A 5 2 7200 20121125230752 20121026230752 4442 isc.org. ViS+qq95DibkkZ5kbL8vCBpRUqI2/M9UwthPVCXl8ciglLftiMC9WUzq Ul3FBbri5CKD/YNXqyvjxyvmZfkQLDUmffjDB+ZGqBxSpG8jlfDwK6nl hWbKf7QSe4LuJZyEgXFEkP16CmVyZCTITUh2TNDmRgsoxrvrOqOePWhp 8+E=
isc.org.                4084    IN      NS      ns.isc.afilias-nst.info.
isc.org.                4084    IN      NS      ams.sns-pb.isc.org.
isc.org.                4084    IN      NS      ord.sns-pb.isc.org.
isc.org.                4084    IN      NS      sfba.sns-pb.isc.org.

;; AUTHORITY SECTION:
isc.org.                4084    IN      NS      ns.isc.afilias-nst.info.
isc.org.                4084    IN      NS      ams.sns-pb.isc.org.
isc.org.                4084    IN      NS      ord.sns-pb.isc.org.
isc.org.                4084    IN      NS      sfba.sns-pb.isc.org.

;; ADDITIONAL SECTION:
mx.amsl.isc.org.        484     IN      A       199.6.1.65
mx.amsl.isc.org.        484     IN      AAAA    2001:500:60::65
mx.paol.isc.org.        484     IN      A       149.20.64.53
mx.paol.isc.org.        484     IN      AAAA    2001:4f8:0:2::2b
_sip._udp.isc.org.      4084    IN      SRV     0 1 5060 asterisk.isc.org.

;; Query time: 176 msec
;;SERVER: x.x.x.x#53(x.x.x.x)
;; WHEN: Tue Oct 30 01:14:32 2012
;; MSG SIZE  rcvd: 3223
```

3,223 byte response

DNS Terminology, Components, and Concepts

Top-Level Domain

Hosts

SubDomain

Fully Qualified Domain Name (FQDN)

Name Server

Zone File

Records

Record types

Start of Authority (SOA)

```
domain.com.   IN SOA ns1.domain.com. admin.domain.com. (
                                              12083    ; serial number
                                              3h       ; refresh interval
                                              30m      ; retry interval
                                              3w       ; expiry period
                                              1h       ; negative TTL
)
```

A and AAAA Records

```
host      IN      A       IPv4_address
host      IN      AAAA    IPv6_address
```

CNAME records

```
server1     IN  A       111.111.111.111
www         IN  CNAME   server1
```

Record types

MX records

```
        IN  MX  10  mail1.domain.com.
        IN  MX  50  mail2.domain.com.
mail1   IN  A       111.111.111.111
mail2   IN  A       222.222.222.222
```

NS records

```
        IN  NS      ns1.domain.com.
        IN  NS      ns2.domain.com.
ns1     IN  A       111.222.111.111
ns2     IN  A       123.211.111.233
```