G-Lab

A Naming Scheme for Identifiers in HiiMap

SPONSORED BY THE
Federal Ministry of Education and Research

Outline

- Locator/Identifier-Split
- HiiMap Architecture
- Naming Identifiers
  - Requirements
  - Naming schemes
- Search Mechanism

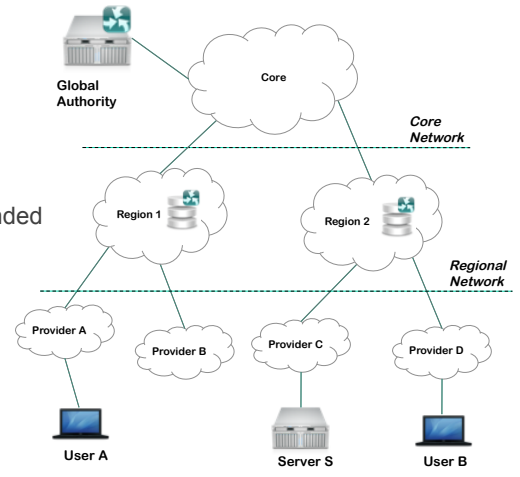
Motivation

Why split IP into Locator and Identifier?

- Addresses for completely different purposes:
  - Identifier (ID) – *Who* am I?
  - Locator – *Where* am I?
- Supports mobility inherently
- Helps to reduce the size of BGP routing tables
  - Readdressing is possible
  - Multihoming and relocation do not increase routing tables any more
- Suitable not only for host addressing, but also for content/persons
- However: Mapping System needed

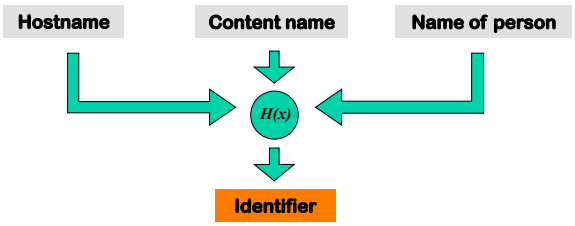
HiiMap Architecture

- Two-tier locator
  - Local Part (LTA)
  - Global Part (gUID)
- Mapping System divided into regions
- Unique Identifiers with prepended region prefix (RP)
- Global Authority (GA) for PKI infrastructure
- Mapping System realized with DHTs
  - Highly scalable
  - Easily extendible



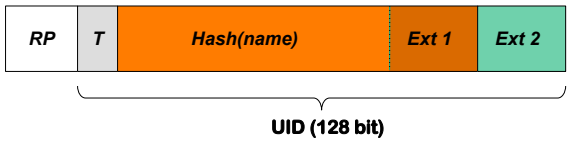
How to name Identifiers?

- Requirements for IDs:
  - User friendly (no one remembers fixed-length bit strings)
  - Suitable for DHT storage
  - Easy generation out of plaintext name
  - Usable for hosts, content and persons
  - Storable in one database
  - Easy registration process
  - No additional Domain Name System, only a Mapping System
- Possible Solution
  - Apply Hash function $H(x)$ to a plain text name



How to name Identifiers?

- General scheme for an Identifier



  - Type T (4 bit):
    - Denotes the type of the Identifier
    - Extendible depending on length of T
  - Hash (92 bit):
    - Main part of Identifier
    - Hash function from a plain text string
  - Two extension fields (each 16 bit):
    - Separate information and function depending on Identifier-Type

Identifier for Hosts

- Type is set to host
- Usage of human readable hostname as input string for hash function
- Ext 1 is merged with hashed hostname: 108 bit addressable hosts
- Ext 2: Service Identifier
  - Denotes a specific service running on that machine
  - Not stored in mapping system
  - Set to zero when queryin the mapping system
  - Only used during communication

Identifier for Content

- Content and information is in the focus of the future Internet
- Content can be of any kind
  - Webpage
  - Media information (Audio/Video/Picture)
  - News

- Usage of a meaningful name for hashing
- Ext 1: used for subinformation that directly belongs to main content
  - Songs in an album
  - Articles in an newspaper
- Ext 2 is used for versioning information
  - Value zero returns current version
  - Different versions can be stored ond different locations

Identifier for Persons

- Web 2.0, VoIP, social networks: person is in the focus of interest

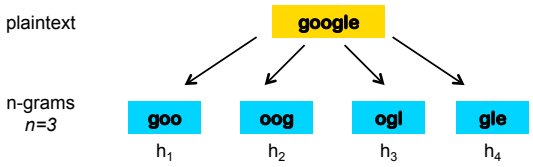
- Persons full name (first name + last name) used for hashing
- Ext 1: random value
  - Different people can have the same name
  - Mapping entry provides additional information to distinguish person
- Ext 2: Communication channel
  - Mailbox, VoIP, Social network...
  - Different locators for different channels possible
  - Value 0 returns locator of current machine (if desired and updated)

Search Strategy

- Problem: Correct Identifier is only found if **exactly** known!
  - Spelling mistakes can occur
  - Sometimes exact spelling not known, only phonetics
- Today: Google can find information even if spelling is incorrect
  - „Did you mean...?“
  - Works with any kind of information
- Solution: Use an n-gram based search mechanism
  - Suitable for DHTs
  - Build n-grams out of plaintext string as indexing scheme
  - Store n-grams in mapping system
  - Use n-grams only if UID did not match
  - N-gram results sorted according their occurence

Search Strategy

- Build n-gram database
  - Generate *i* n-grams h_i out of plaintext string



  - Store tuples <*Hash*(h_i); *plaintext*> in mapping system
  - No n-gram update necessary when object locator(s) change
- Search with n-grams
  - If initial ID-search did not match, build n-grams of plaintext string
  - Seach for n-gram tuples in mapping system
  - Sort results according occurence and relevance

Conclusion

- Generalized Identifier scheme
  - For hosts
  - For content and information objects
  - For persons
  - Easily extendible
- Mapping system can return mapping entries suitable to identifier-type
  - Direct locators for hosts and machines
  - Information models for Content and informations
  - Person-specific data for humans
- Search mechanism capable of correcting spelling mistakes