

What's in a name?

Noah Mendelsohn

Tufts University

Email: noah@cs.tufts.edu

Web: <http://www.cs.tufts.edu/~noah>

Goals

- **Understand why names are fundamental to the design of almost every computer system**
- **Notice and learn from examples of names in everything from CPU hardware to Internet-scale distributed systems**
- **Explore choices and tradeoffs relating to naming – how they impact what systems can do**
- **Understand the unique importance of URIs as names for resources on the Web**

There are only two hard things in
Computer Science: cache
invalidation and naming things.

-- Phil Karlton

Why is this so important?

Why naming systems are important

- **What you can name and how you name it tends to be fundamental to what a system can do**
- **If you have a uniform naming systems, and uniform operations – then you can operate on things you haven't seen before**

To use the phone system we need phone numbers and to know how to dial them

Phone numbers

Telephone numbers

1-781-555-1234

There's more going on here than you think

For the details, see: http://en.wikipedia.org/wiki/North_American_Numbering_Plan

Telephone numbers

Technically, the country code is part of the number..the 10 digit forms we use in the USA are actually short forms.

1-781-555-1234

There's more going on here than you think

For the details, see: http://en.wikipedia.org/wiki/North_American_Numbering_Plan

Telephone numbers

...but within the US, the '1' is often used as a prefix to indicate that a 10 digit number w/area-code as opposed to 7 digit local number is to follow.

1-781-555-1234

There's more going on here than you think

For the details, see: http://en.wikipedia.org/wiki/North_American_Numbering_Plan

Telephone numbers

No good...
First digit of area code can't be 1 -
causes confusion with 1 as prefix

1-181-555-1234

There's more going on here than you think

Telephone numbers

No good...
Either of these can be "1", but
not both. 511 is reserved for
traffic information (who
knew?) but avoiding 911 is
more obvious.

781-511-1234

There's more going on here than you think

For the details, see: http://en.wikipedia.org/wiki/North_American_Numbering_Plan

Telephone numbers

...and finally, though this attempts to be a fictitious number, it's technically not.

1-781-555-1234

*"Telephone numbers with the prefix **555** are widely used for fictitious telephone numbers in North American television shows, films, video games, and other media.... In fact, only 555-0100 through 555-0199 are now specifically reserved for fictional use "*

Historical tidbit: one of the first known uses of a 555 phone number was in the 1962 movie "Panic in Year Zero".

For the details, see: http://en.wikipedia.org/wiki/555_%28telephone_number%29

A bit about mobile phones

- **Q. What's the most important thing about mobile phone numbers?**
- **A. They're the same as landline numbers!**

Think a minute about how important this is!

Because mobile & landline numbers are the same

- **Mobile phones can call all the landlines that were installed years before mobile phones were conceived**
- **Because the space is not partitioned, you can reassign a landline number to a mobile phone, etc.**

These aren't details...these are fundamental to why mobile phones have been successful!

... and your point is?

- **The structure of names is important**
- **We need to consider:**
 - What can be named ← *this determines what the system can access!*
 - Who can generate new names
 - Whether we can tell anything about what is named by looking at the name
- **Some names are *locators***
 - Locator names *help you find* things (e.g. the area code tells you which phone switch to contact)
 - Another name for locator is *address*
- **The structure of names often reflects the structure and capabilities of the system itself**

Are we having fun yet?

*Why even go through this
in a course about computer systems?*

This History and Impact of the Phone Numbering System

Why consider phone numbers?

- **Names are crucial to the behavior and success of systems**
- **The phone system is very much like the Web:**
 - Scale (worldwide)
 - Time horizon (centuries)
 - Need to evolve to unanticipated uses (Phones: fax, mobile, modems; Web: TBD)
 - Need for federated administration by separate organizations and countries

Some things to note:

- **The nature of phone numbers determines who can be called, and how easily**
- **Lots of information is embedded in the phone number**
 - Country where phone is hosted
 - Well known numbers have special use (911)
- **Assignment is delegated to countries, etc**
- **The structure of a phone number helps with call routing**
- **The phone system has evolved in ways unimagined 100 years ago, and phone numbers mostly work**

We see the phone network as a "space" of phone numbers

Interesting Reference

CRAFT, E. B. , MOREHOUSE, L. F. and CHARLESWORTH, H. P., *Machine Switching Telephone System for Large Metropolitan Areas*, Bell System Technical Journal, 1923

This is the technical paper that describes the introduction of automated dialing into the US telephone system. There is lots of information from nearly 100 years ago on phone numbers and their use. It includes a 19th century quote from Alexander Graham Bell:

"It is conceivable that cables of telephone wires could be laid underground, or suspended overhead, communicating by branch wires with private dwellings, country houses, shops, manufactories, etc., etc.—uniting them through the main cable with a central office where the wires could be connected as desired, establishing direct communication between any two places in the city. Such a plan as this, though impracticable at the present moment, will, I firmly believe, be the outcome of the introduction of the telephone to the public. Not only so, but I believe in the future wires will unite the head offices in different cities, and a man in one part of the country may communicate by word of mouth with another in a distant part.

"Believing, as I do, that such a scheme will be the ultimate result of the telephone to the public, I will impress upon you all the advisability of keeping this end in view, that all present arrangements of the telephone may be eventually realized in this grand system."

So, naming can play a huge
role in determining how a
system works and what it can
do...

...let's consider some examples
from the computing world

C++ Class and member Names

```
Class MyClass {  
    Someclass* myFunction (int someArg);  
    int myMember;  
}
```

- Arbitrarily long...characters and digits (no digits first)
- No keywords like *int*
- User convention: classes Capitalized...members not
- Compilers can punt on noticing differences at the end of very long names!
- No characters like hyphen (-) in names – any idea why?

C++ Pointers

```
int *myPointer;
```

- **Mostly opaque to us as programmers...just bits**
- **Resetting the pointer (myPointer = NULL) does not free the resource**
 - ...but in Java it does!
- **On most machines, pointers really are just numbers (opaque identifiers), but...**
- **...there have been machines in which pointers had internal structure or content restrictions...usually these caused trouble**
 - CS 40: Universal Machine pointers are in two parts: segment ID and word offset within segment
 - Example: IBM 360 pointers were 32 bits long, but only the low order 24 bits were used for addressing. Memory was expensive, so programmers would stick flags or other data in the high order byte. When memory got cheap and machines got bigger, 31 bit addressing was introduced -- there was lots of trouble keeping those 24 bit programs working

Unix file names

`/u/noah/comp117/samples/test.html`

- **Any character except / in path segment**
- **Hierarchical names support hierarchical navigation**
- **Removing the name deletes the file**
 - `rm /u/noah/comp150/samples/test.html` *removes the name and deletes the file*
- **Actually, it's more complicated...files can have multiple names**
 - The file itself only disappears when the last of its names is removed:
 - * In `/u/noah/comp150/samples/test.html /u/noah/comp150/samples/name2.html`
 - * `rm /u/noah/comp150/samples/test.html` ← file still exists with name
 - * `rm /u/noah/comp150/samples/name2.html` ← file deleted
 - Unless we use symbolic links (`ln -s`) which *don't* protect the file...symbolic links can result in dangling references

Naming in Internet-scale Systems

Do we need really big names for really big systems?

- **You'd think that for a system as big as the Web, which can link "everything", we'd need really big names**
- **Actually, that's not true:**
 - A 34 byte binary number can encode values up to 10^{80}
 - That's enough to provide a unique id for every atom in the universe!*
 - If we were willing to use unstructured binary URIs, 20 bytes would probably be plenty
- **There are other reasons for long names**
 - Text names are convenient...not as compact as binary
 - Extra name structure used to encode: hierarch (/), metadata, etc.
 - We need ways of letting lots of organizations allocate names – that tends to be easier if we pre-allocate lots of unused names for each such organization
 - Etc.

*Source: <http://www.wolframalpha.com/input/?i=number+of+atoms+in+the+universe>

Summary

Key messages

- **There are an amazing set of subtle tradeoffs in designing names for your systems**
- **To a surprising degree, the choices you make determine:**
 - The structure of your system
 - What information it can integrate
 - What your system can do
 - How it can evolve
 - How robust and secure it is
- **Computer scientists and engineers have years of experience trying to get this right**
 - There are tons of great examples to study
 - We all still struggle to make good choices
- **Naming with URIs is fundamental to what the Web is – if you have a URI for something you can link it, and you can interact with it “on the Web”**