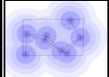
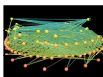



CS525
Advanced Topics in Distributed Systems
Spring 2008

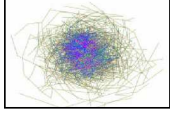
Indranil Gupta (Indy)
 Lecture 1
 January 15, 2008

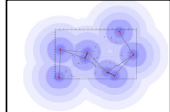
What is a Distributed System? (examples)



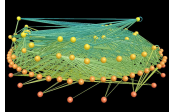
The Internet



Gnutella peer to peer system



A Sensor Network



Food Web of
Little Rock Lake, WI

Can you name some examples of Operating Systems?

Can you name some examples of Operating Systems?

...

Linux WinXP Unix FreeBSD Mac
2K Aegis Scout Hydra Mach SPIN
OS/2 Express Flux Hope Spring
AntaresOS EOS LOS SQOS LittleOS TINOS
PalmOS WinCE TinyOS

...

What is an Operating System?

What is an Operating System?

- User interface to hardware (device driver)
- Provides abstractions (processes, file system)
- Resource manager (scheduler)
- Means of communication (networking)
- ...

FOLDOC definition

- The low-level software which handles the interface to peripheral hardware, schedules tasks, allocates storage, and presents a default interface to the user when no application program is running.
- The OS may be split into a kernel which is always present and various system programs which use facilities provided by the kernel to perform higher-level house-keeping tasks, often acting as servers in a client-server relationship.
- Some would include a graphical user interface and window system as part of the OS, others would not. The operating system loader, BIOS, or other firmware required at boot time or when installing the operating system would generally not be considered part of the operating system, though this distinction is unclear in the case of a roamable operating system such as RISC OS.
- The facilities an operating system provides and its general design philosophy exert an extremely strong influence on programming style and on the technical cultures that grow up around the machines on which it runs.

Can you name some examples of Distributed Systems?

Can you name some examples of Distributed Systems?

- Client-server (e.g., NFS)
- The Internet
- The Web
- An ad-hoc network
- A sensor network
- DNS
- Kazaa (peer to peer overlays)

What is a Distributed System?

FOLDOC definition

A collection of (probably heterogeneous) automata whose distribution is transparent to the user so that the system appears as one local machine. This is in contrast to a network, where the user is aware that there are several machines, and their location, storage replication, load balancing and functionality is not transparent. Distributed systems usually use some kind of client-server organization.

Textbook definitions

- A distributed system is a collection of independent computers that appear to the users of the system as a single computer
[Andrew Tanenbaum]
- A distributed system is several computers doing something together. Thus, a distributed system has three primary characteristics: multiple computers, interconnections, and shared state
[Michael Schroeder]

Unsatisfactory

- Why are these definitions short?
- Why do these definitions look inadequate to us?
- Because we are interested in the insides of a distributed system
 - algorithmics
 - design and implementation
 - maintenance
 - study

I shall not today attempt further to define the kinds of material I understand to be embraced within that shorthand description; and perhaps I could never succeed in intelligibly doing so. But I know it when I see it, and the motion picture involved in this case is not that.

[Potter Stewart, Associate Justice, US Supreme Court (talking about his interpretation of a technical term laid down in the law, case Jacobellis versus Ohio 1964)]

A working definition for us

*A distributed system is a collection of entities, each of which is **autonomous**, **programmable**, **asynchronous** and **failure-prone**, and which communicate through an **unreliable** communication medium.*

- Our interest in distributed systems involves
 - algorithmics, design and implementation, maintenance, study
- Entity=a process on a device (PC, PDA, mote)
- Communication Medium=Wired or wireless network

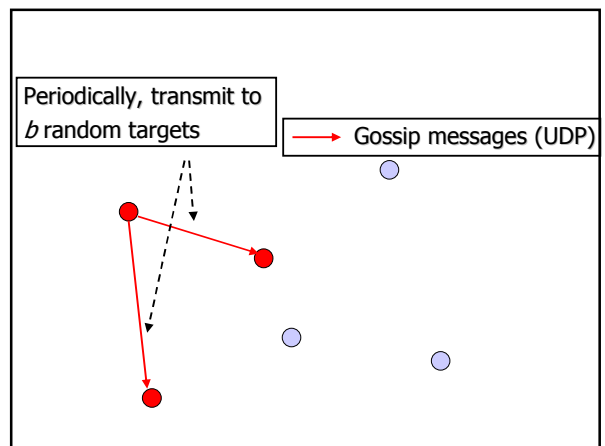
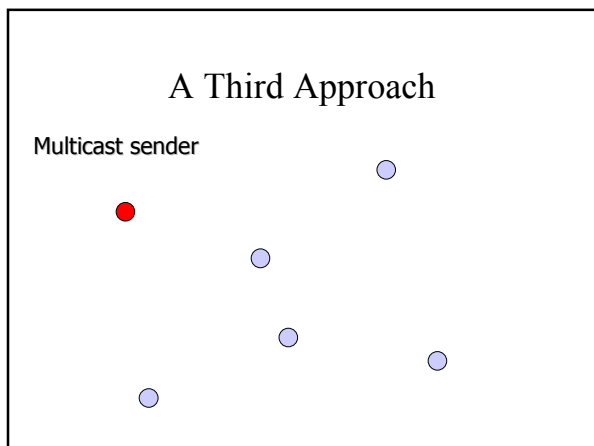
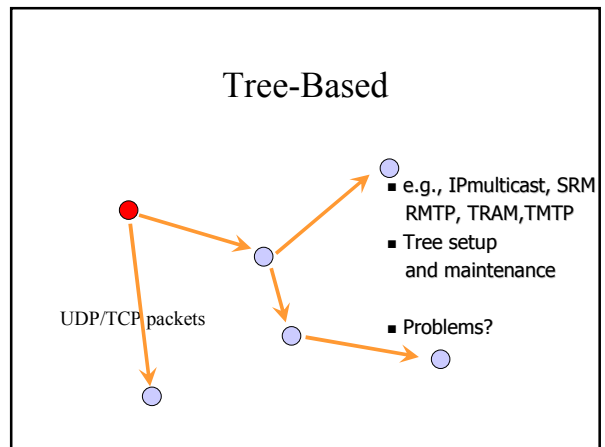
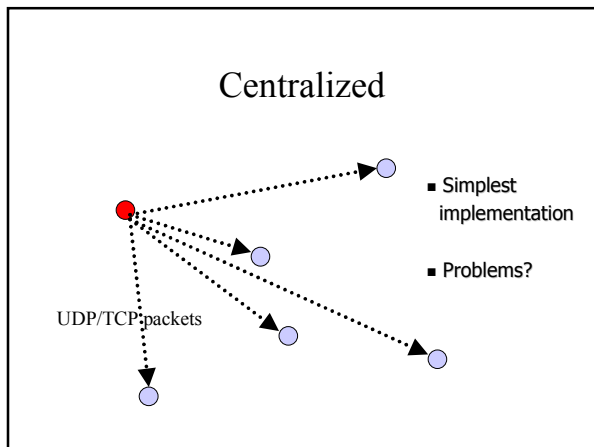
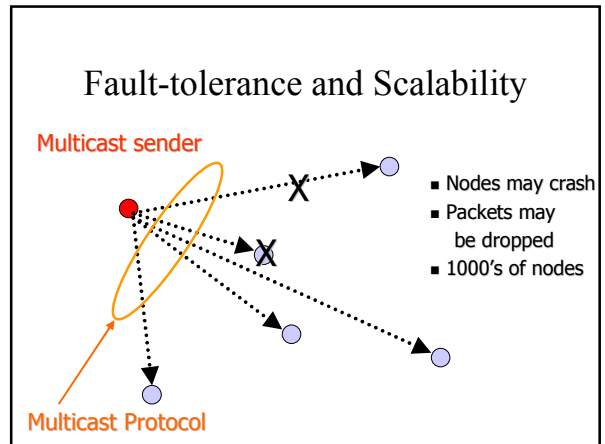
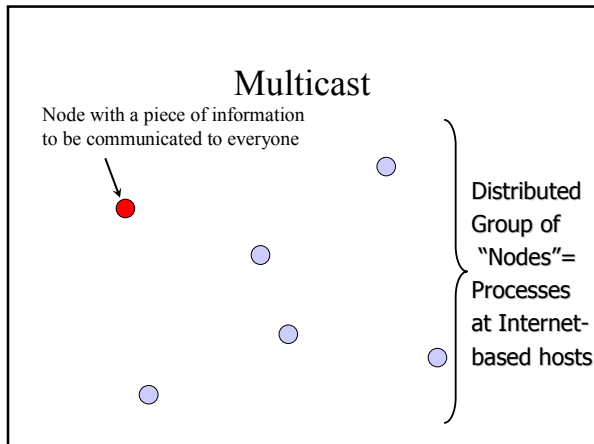
A range of interesting problems for Distributed System designers

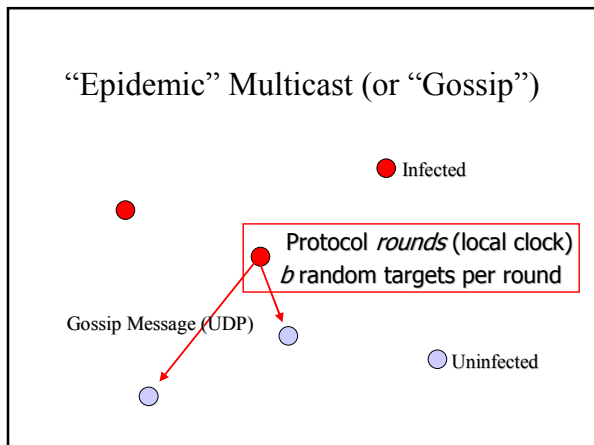
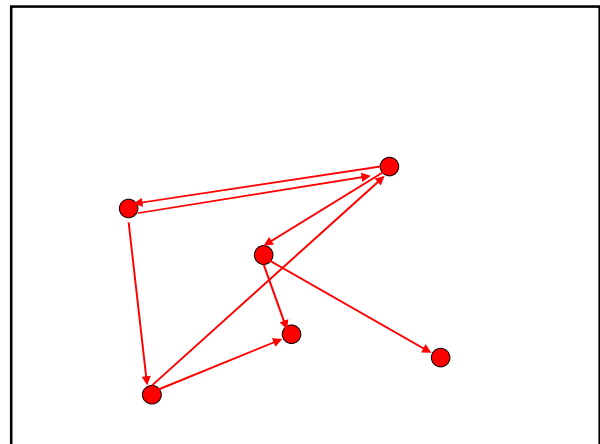
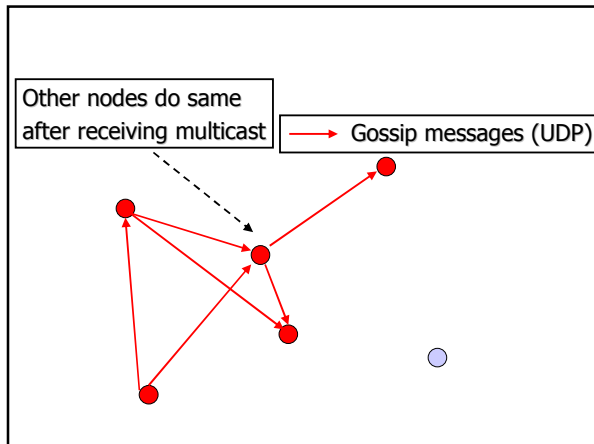
-
- Routing [IP,BGP]
- Multicast [IP multicast, SRM, RMTP]
- Post and retrieve [Usenet]
- Search [Kazaa, Google]
- Storage [Databases]
- Coordination [SETI@Home]
-
-

A range of challenges

-
- Failures
- Asynchrony
- Scalability
- Security
-

Multicast





Properties

Claim that this simple protocol

- Is lightweight in large groups
- Spreads a multicast quickly
- Is highly fault-tolerant

Analysis

From old mathematical branch of *Epidemiology* [Bailey 75]

- Population of $(n+1)$ individuals mixing homogeneously
- Contact rate between any individual pair is β
- At any time, each individual is either uninfected (numbering x) or infected (numbering y)
- Then, $x_0 = n, y_0 = 1$
and at all times $x + y = n + 1$
- Infected–uninfected contact turns latter infected

Analysis (contd.)

- Continuous time process
- Then

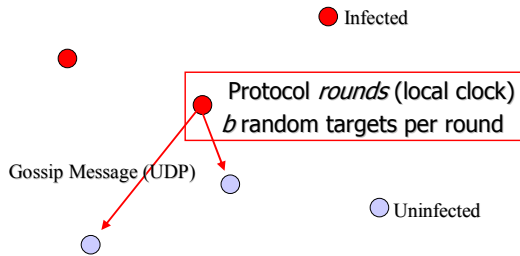
$$\frac{dx}{dt} = -\beta xy \quad (\text{why?})$$

with solution

$$x = \frac{n(n+1)}{n + e^{\beta(n+1)t}}, y = \frac{(n+1)}{1 + ne^{-\beta(n+1)t}}$$

(correct? can you derive it?)

Epidemic Multicast



Epidemic Multicast Analysis

$$\beta = \frac{b}{n} \quad (\text{why?})$$

Substituting, at time $t = c \log(n)$, num. infected is

$$y \approx (n+1) - \frac{1}{n^{cb-2}}$$

(correct? can you derive it?)

Analysis (contd.)

- Set c, b to be small numbers independent of n
- Within $c \log(n)$ rounds, [**low latency**]
 - all but $\frac{1}{n^{cb-2}}$ of nodes receive the multicast [**reliability**]
- each node has transmitted no more than $cb \log(n)$ gossip messages [**lightweight**]

Fault-tolerance

- Packet loss
 - 50% packet loss: analyze with b replaced with $b/2$
 - To achieve same reliability as 0% packet loss, takes twice as many rounds
- Node failure
 - 50% of nodes fail: analyze with n replaced with $n/2$ and b replaced with $b/2$
 - Same as above

Fault-tolerance

- With failures, is it possible that the epidemic might die out quickly?
- Possible, but improbable:
 - Once a few nodes are infected, with high probability, the epidemic will not die out
 - So the analysis we saw in the previous slides is actually behavior *with high probability* [Galey and Dani 98]
- Think: why do rumors spread so fast? why do infectious diseases cascade quickly into epidemics? why does a worm like Blaster spread rapidly?

So,...

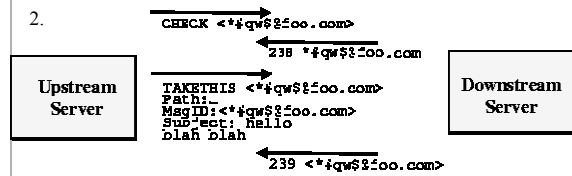
- Is this all theory and a bunch of equations?
- Or are there implementations yet?

Some implementations

- Clearinghouse project: email and database transactions [PODC '87]
- refDBMS system [Usenix '94]
- Bimodal Multicast [ACM TOCS '99]
- Ad-hoc networks [Li Li et al, Infocom '02]
- **Usenet NNTP (Network News Transport Protocol) ! ['79]**

NNTP Inter-server Protocol

1. Each client uploads and downloads news posts from a news server

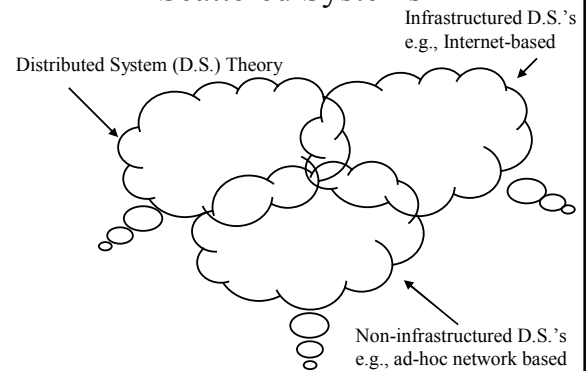


Server retains news posts for a while,
transmits them lazily, deletes them after a while

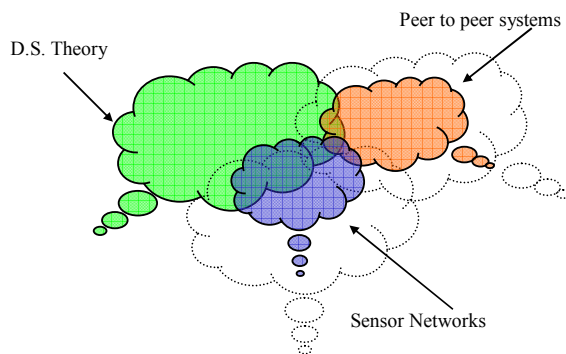
We'll cover some of these other
implementations during the course

- But let's dwell on the big picture of the course

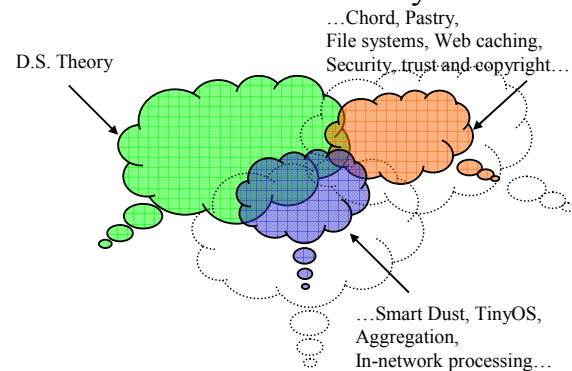
Scattered Systems



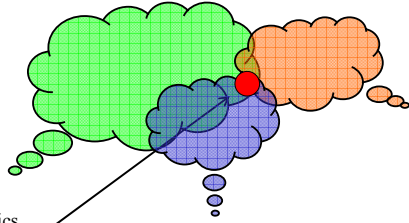
CS 525 and Scattered Systems



CS 525 and Scattered Systems



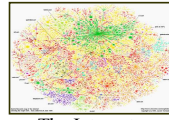
Interesting: Area Overlaps



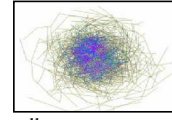
Epidemics
NNTP
Gossip-based ad-hoc routing

Interesting: Area Overlaps

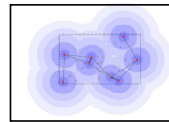
Do projects and write papers in these overlap areas!



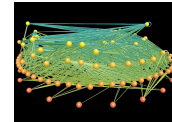
The Internet



Gnutella peer to peer system



A Sensor Network



Food Web of
Little Rock Lake, WI

Let's Look at the Course Information Sheet...

- Papers
 - Presentations
 - Reviews
- Project
 - Conference-quality paper
 - Novel idea solving useful problem, backed up with good evaluation
- No exams
- Class Participation a must (and fun!)
- My office hours: right after lecture/class (3112 SC)

Things for you to do today

- Look at the course website
- Follow “Schedule / Papers and Presentations link” and read instructions
 - Need to sign up for a presentation slot by Jan 31
- Take a look at conference papers arising out of previous versions of this course (CS598IG/CS525)
 - Fall 03: 9/12 project papers in conferences
 - Fall 04, Spring 06, Spring 07: Many under review in conferences and workshops, similar success rates expected

Next Lecture

- Internals of the Gnutella peer to peer system
 - Read paper handed out to you (no reviews required)

Epidemic Multicast Analysis

$$\beta = \frac{b}{n} \quad (\text{why?})$$

Substituting, at time $t=c \log(n)$

$$\begin{aligned} y &= \frac{n+1}{1 + ne^{-\frac{b}{n}(n+1)c \log(n)}} \approx \frac{n+1}{1 + \frac{1}{n^{cb-1}}} \\ &\approx (n+1) \left(1 - \frac{1}{n^{cb-1}}\right) \\ &\approx (n+1) - \frac{1}{n^{cb-2}} \end{aligned}$$