Infososphere:
Smart Delivery of Fresh Information

Calton Pu
Professor and John P. Imlay, Jr. Chair in Software
Georgia Institute of Technology

with
(OGI) Walpole (GT) Liu, Schwan, Abowd

© 2000 Calton Pu
DARPA/ITO Expeditions

- MIT’s Oxygen project (Sci. Am. 99/08)
- UC Berkeley’s Endeavour (Wired 00/01)
- CMU’s Aura project
- Washington’s Portolano project (with Xerox PARC)
- Georgia Tech & OGI’s *Infosphere* project
Moore’s Law

- Gordon Moore, computer pioneer
  - CPU speed doubles every 18 months
  - Memory density doubles every 18 months
  - Disk storage density doubles every 12 months

- Computers almost free
Network Is The Computer

- Next generation Internet
  - OC12 (622 Mb/s) connections

- Wireless networks
  - Megabit/second wireless connections

- Computer interconnects
  - Gigabit Ethernet, affordable clusters

- *Networks almost free*
Network Everywhere

- High speed backbone wide area networks
  - Next generation Internet around the world
- Metropolitan and local area networks
  - Wired example: Portland, Oregon
  - Wireless: cellular networks, satellites
- Home and personal networks
  - Bus-based smart homes, Bluetooth
Computers Everywhere

◆ Traditional computers
  • Mainframes, desktops, notebooks, etc

◆ Embedded systems
  • Cars, PDAs, cell phones, smart appliances
  • Micro-electronic mechanical (MEM) systems

◆ Smart environments
  • Sensors, automated controls, monitoring
Ubiquitous Computing

- Plenty of computers
  - Are Everywhere
  - Know Everything
  - Almost Free

- Scarce resource is human
  - User attention span
  - Programmer time
Information Growth

Storage capacity growth
- Disk capacity sold per year: Exabytes

Information content growth
- All of human published information: Exabytes
- Computer-generated information: web robots
- Sensor-generated information
Infosphere Focus Area

- Too many heterogeneous sources
- Too much data
  - Internet data
  - Digital Earth (100TB/day)
  - Smart Dust sensors (thousands p/km2)
Traditional Computer Science

- Computational models
  - Theory: finite state machines, Turing machines
  - Programming languages: programs
  - Operating systems: processes
  - Computer architecture: instruction sets
- Distributed computations scale poorly
  - Example: agreement protocols
Information Flow Computing

- Internet applications are primarily *information flow applications* (DL, EC)
- Networking success
  - Massive information delivery, QoS
- Historical examples (centralized systems)
  - Dataflow machines in computer architecture
  - Dataflow diagrams in software engineering
  - Unix pipes
Comp. Models & Info Flow

- Focus on algorithms
  - Distributed programs
  - Global algorithms
  - Scalability problems

- Info flow “on the side”
  - Implicit or separate description of I/O

- Localized algorithms
- Global info flow
  - Flow composition
  - Composable properties

- Focus on info flow
  - Explicit description of syntax, semantics, and QoS properties
Infosphere Meta-Approach

◆ Focus on distributed information flow
  - In contrast to computation-centric computing
  - Infopipe as the central abstraction
  - Hypothesis: composable, predictable, scalable distributed software systems

◆ Missing link: the systems software
  - OS kernel, middleware, data management
Problem: too many sources, too much information

Infosphere

Clean, Reliable, Timely Information, Anywhere

Personalized Filtering & Info. Delivery

Resource Adaptation

Property Mgmt

specialization

Continuous Queues

Information Quality

Microfeedback

Infopipes

Internet: Information Jungle

Digital Earth

Sensors
Infopipes: Backbone of Infosphere

Research Challenges:

- Property preserving composition of Infopipes
- Timely delivery of high quality fresh information
Infopipe Abstraction

- Several reasonable definitions
- Component Infopipes
  - Ends: Typespec, property specifications
  - Middle: processing, buffering, active
- Composition of Infopipes
  - End-to-end property preservation
  - Multiplex ends and middles
Impact on OS Research

- Computation-centric
  - Process abstraction
  - Inter-process communications
  - Synchronization
  - Memory and I/O are “on the side”

- Information-driven
  - Infopipe abstraction
  - Infopipe connection and composition
  - Flow control
  - I/Os are natural flows
Impact on DB Research

- Traditional DB
  - Focus: data reservoir
  - Closed world
  - Homogeneity and slow evolution
  - Static control (DBA)
  - Data warehouse

- Infosphere/Infopipe
  - Focus: data flow
  - Link to real world
  - Heterogeneity and rapid changes
  - Dynamic adaptation
  - Fresh information

Person. Filtering
Preserv. Info Quality
Fresh Information Applications

◆ Near Term
  - Georgia Tech Aware Home
  - I/O-Intensive embedded systems

◆ Long term
  - Electronic commerce
  - Personal guidance
  - Environmental observation and forecasting
Georgia Tech Aware Home

♦ GRA, other funds
♦ Many sensors
  • 20 video cameras
  • microphones, vibes, …
♦ Big backend
  • 2 Gigabit connections
  • 128-CPU cluster
♦ Real-time sensor information laboratory
Aware Home Projects

◆ Ubiquitous sensing
  - Recognize people based on their footsteps
  - Multiple camera/multiple person tracking
  - Sensor fault-tolerant scene modeling

◆ Potential applications
  - Long term elderly care
  - Emergency rescue missions
Classic Embedded Systems

- Embedded = Closed
  - Small, independent, self-contained
  - Custom hardware and custom software
  - Small footprint, predictable performance, reliability

- Closed = Limited Evolution
  - Limited hardware and software life expectancy
  - Each generation is a new product
I/O-Intensive Embedded Systems

◆ Computers and communications
  • Sensor-actuators, PDAs, remote control
  • Network is the embedded system

◆ Network embedded software requirements
  • Traditional: small footprint, reliability, predictable performance
  • New: composability, adaptability, extensibility, end-to-end performance
3-Tier Client/Server Today
N-Tier Systems Tomorrow

Infotaps & Fat Clients

Variety of Servers

Sensors

Many sources

Database Server

Email: Calton.Pu@cc.gatech.edu
E-Commerce Applications

- N-Tier electronic commerce systems
  - Predictable end-to-end latency
- Logistics, real-time decision support
  - Predictable, reliable, real-time information flow
  - Recovery from faults, accidents
  - Adaptation to environmental changes
Personal Guidance Today

- Tele Aid (Mercedes) and OnStar (Cadillac)
  - Cell phone link
  - GPS navigation
  - Human operator
- Palm VII
  - Radio link (cities)
  - Simple web browser
Personal Guidance Tomorrow

- Traffic-aware road navigation
  - Car receives news, suggests alternative routes

- Cell phones with GPS, WAP
  - Heidelberg tour guide, Digital Kyoto
Environ. Observation & Forecast

- Columbia River
  - 2nd in No. America
- Observe (nowcast)
  - 12 data stations in Columbia River Mouth
- Calculate (forecast)
  - Observation-driven model of unpredictable environments
Tomorrow’s Precise Forecasts

◆ Many sensors everywhere
  ● Video cameras, MEMs, satellites, ...

◆ Weather-adaptive applications
  ● Safe river and shallow water navigation
  ● Disaster prevention in precision farming
  ● Airline crew and airplane optimization

◆ Monitoring of environment
  ● Amazon illegal timber traffic detection
Infosphere: Current State

- **Infopipe basic research**
  - Infopipe concepts and specification (ISL)
  - Infopipe “stub generators” (ISL compiler)

- **Infopipe technology development**
  - Personalized filtering (Continual Queries)
  - Middleware Infopipes (Event Channels)
  - Kernel QoS Support (Quasar/Microfeedback)
Infopipe Concepts

Serial
(1 source, 1 sink)

Multiplex
(n sinks)

Demultiplex
(n sources)

Buffer
(storage)

Filter
(transformation)
Infopipe Specification

- Syntax of info flow
  - Java class, C record
- Semantics of info flow
  - Currently XML (placeholder)
- QoS requirements of info flow
  - Performance (bandwidth, latency, jitter)
  - Security (level of encryption)
Infopipe “Stub Generators”

- Translate the Infopipe specification into executable code and OS run-time support

![Diagram showing the process from XML parser to XML generator](image)

(1) XML parser
(2) DOM tree
(3) User function
(4) DOM tree
(5) XML generator

Input-end

Output-end

Calton.Pu@cc.gatech.edu
InfoFilters

♦ Personalized filtering
  • Interesting, important, urgent (IIU)

♦ Continual Queries
  • Monitoring of IIU updates on the Internet
  • Event-based filtering of new information
  • Pro-active delivery of IIU information
Recent InfoFilter Results

- **WebCQ**: web update monitoring
  - Built on OpenCQ
  - Currently a service
- **XWRAP Elite**: Wrapper generator
  - Semi-automated generation of wrapper code

http://www.cc.gatech.edu/~lingliu

[Diagram with arrows and boxes for WebCQ, HTML, Wrapper, and Infopipe]
Middleware Infopipes

- Event channels as InfoEvents
  - Publish/subscribe
  - Push/pull
- Quality of Service
  - Performance, security, availability, ...
  - Freshness, timeliness
Recent InfoEvent Results

◆ **DataExchange** software release
  - BPIO: Low-overhead data interchange format
  - ECho: Event Channel code generator
  - JECho: Java Event Channels
  - Heterogeneous distributed platforms

◆ **Event channels** as InfoEvents
  - QoS being added: real-time, security, etc...
Real-Rate Infopipes

◆ Support applications w/ real-world events
  ● Sensor-actuator control and sensor information
  ● Routers and active network nodes
  ● Multimedia (bandwidth, latency, jitter)

◆ Rate-matching and quality degradation
  ● Real-rate OS kernel research
  ● Microfeedback-based adaptive mechanisms
  ● Utility-based QoS degradation policies
Recent Real-Rate OS Results

- Video Camera
- Wireless Internet Connection
- Pentium Motherboard, 233MHz, 128MB RAM, no disk.
- 28V battery pack, power supply, serial-port driven servo controller, steering servos and speed control.
Current Collaborations

- Georgia Tech and OGI
- Specialization of systems software
  - France: INRIA, IRISA, Univ. Bordeaux
  - Japan: Univ. Tokyo, Univ. Tsukuba
- Info flow software, Internet data mgmt
  - Germany: GMD/IPSII
  - Japan: Sony Corp.
Infosphere Summary

- The ubiquitous computing vision
  - Many computers everywhere, out of the way
  - Too many sources, too much data

- Systems software missing link
  - Information flow perspective
  - Composing Infopipes w/ predictable properties
  - Smart delivery of fresh information
Fresh Information On the World