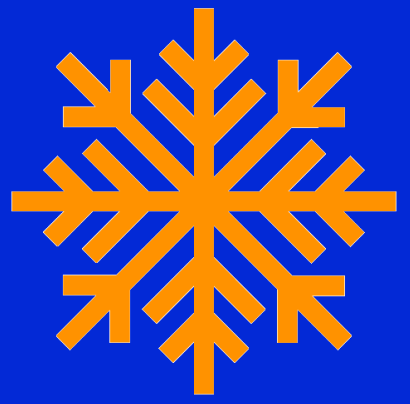
XML for Voice Services

Eric William Burger
IEEE Communications Society
IEEE Computer Society



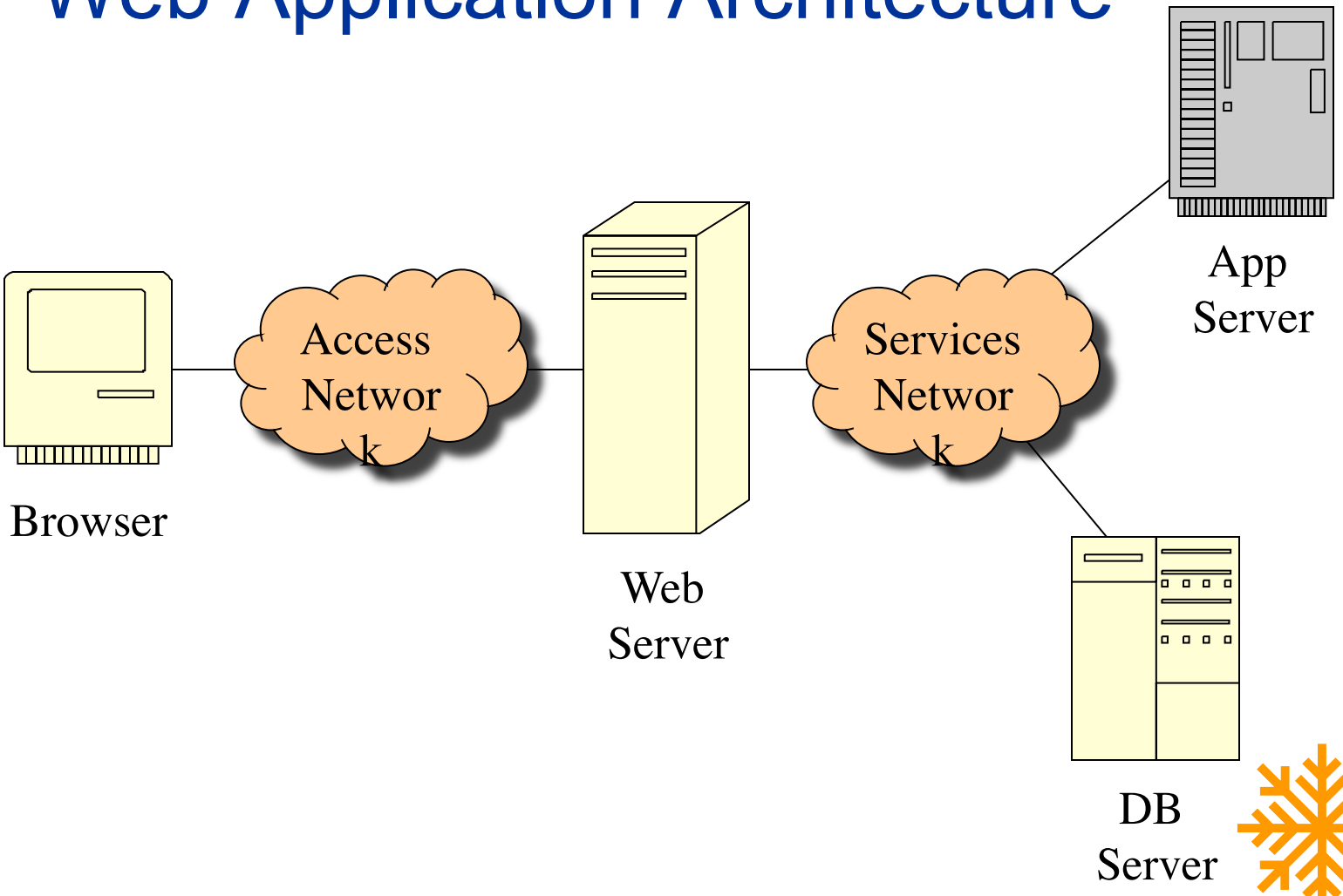
Disclaimer

- **This presentation will not disclose any private deliberation of the W3C Voice Browser Work Group.**
- **See http://www.w3c.org/Voice for publicly available information.**

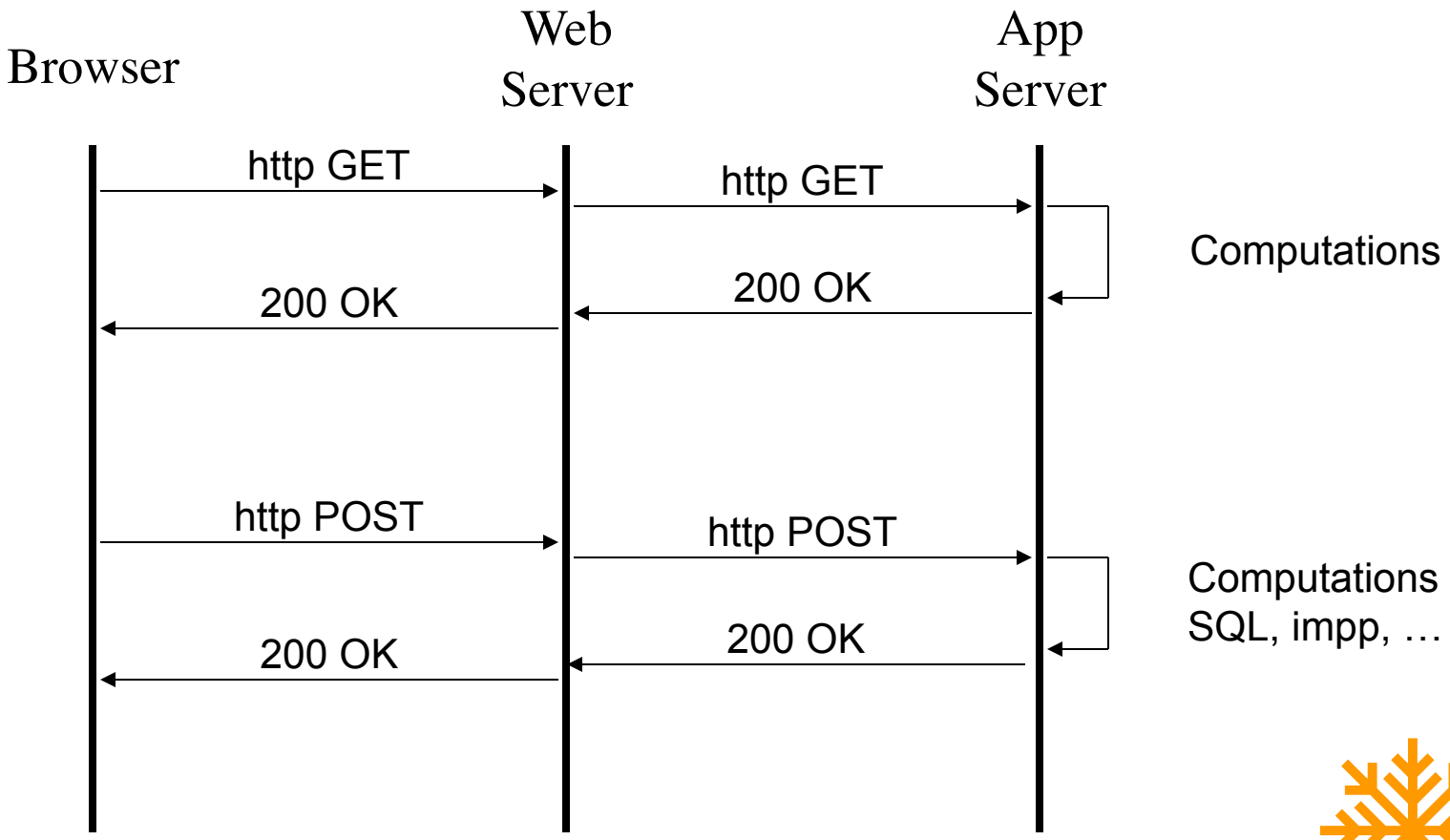
Roadmap

- Web Application Server Architecture
- Rendering Device Effects
- What is VoiceXML?
- Some Applications
- Architectures for Implementation
- Summary

Web Application Architecture



http Message Flow



SNOWSHORE NETWORKS

Web Application Architecture: Key Points

- **Everything Looks Like a Web Page**
  - Retrieved Through http
  - Rendered With HTML
- **Services Behind Application Server**
  - Hidden From Browser
  - Arbitrary Complexity
- **User at Browser Initiates Session and Clicks-Through to Different Web Pages**

Web Application Architecture: Division of Labor

- **Browser**
  - Rendering
  - 2D User Interface
- **Web Server**
  - Serve Static Pages (File Serving)
  - Connection to Application Server
- **Application Server**
  - Computation
  - Data Connectivity
  - Real-Time Interfaces

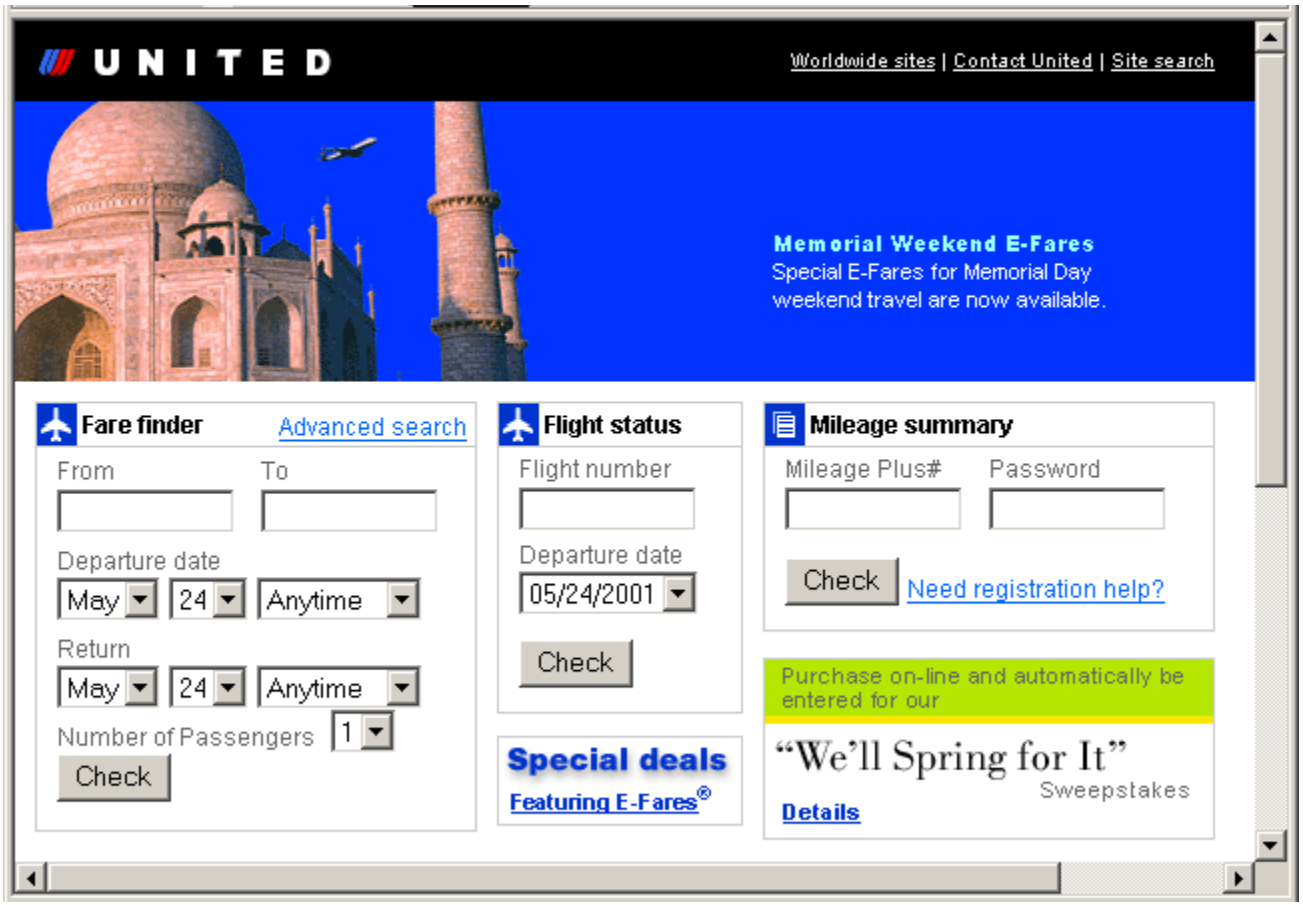
Roadmap

- **Web Application Server Architecture**
- ***Rendering Device Effects***
- **What is VoiceXML?**
- **Some Applications**
- **Architectures for Implementation**
- **Summary**

Multiple Device Types

- **PC Browser**
  - Large Bit-Mapped Display
  - Many Colors
- **Palm Device**
  - Small Bit-Mapped Display
  - Greyscale or 8-bit Color
- **Mobile Device**
  - Small, 4x40 Character Display
  - Monochrome Characters

PC Device



SNOWSHORE NETWORKS

WAP Device

Rendering Languages

- **HTML for Large 2D**
- **WML for Small 2D**
- **Optimized for Particular Display Type**
  - Generic Controls; Open Forms, vs.
  - Buttons, Scrolling, Alpha vs. Numeric Input
- **Still 2D Form**
  - See Whole Form at Once
  - WML Just “Smaller” Version

What About Voice?

- **Voice Fundamentally Different Than 2D**
  - 2D is Parallel: You See All Options At Once
  - Voice is Serial: You Hear Prompt Linearly
  - Temporal vs. Spatial Forms
- **Input Fundamentally Different**
  - HTML: Many options, scrollable
  - WML: Few options, hierarchical menus
  - Voice: “What Would You Like”, flat menus

Voice Response Types

(Jim Ferrans)

- **Basic Interactive Voice Response (IVR)**
  - Computer: “For stock quotes, press 1. For trading, press 2. …”
  - Human: (presses DTMF “1”)
- **Basic Speech IVR**
  - C: “Say the stock name for a price quote.”
  - H: “Sonus Networks”

SNOWSHORE NETWORKS

Voice Response Types

- **Advanced Speech IVR**
  - C: “Stock Services, how may I help you?”
  - H: “Uh, what’s Sonus trading at?”
- **“Near-Natural Language” IVR**
  - C: “How may I help you?”
  - H: “Um, yeah, I’d like to get the current price of Sonus”
  - C: “Sonus is up two at one hundred sixty eight and a half.”
  - H: “OK. I want to buy one hundred shares at market price.”
  - C: “…”

The Application

- **No Matter What the Rendering Device**
- **Application is the Same**
- **Airline Booking Example**
  - Very Different User Interfaces
  - Identical Application
- **Stock Quotes: Again, Same Application**
- **Reuse Servlets**
  - Application Needs Departure and Destination
  - Application Doesn't Care How to Get Input

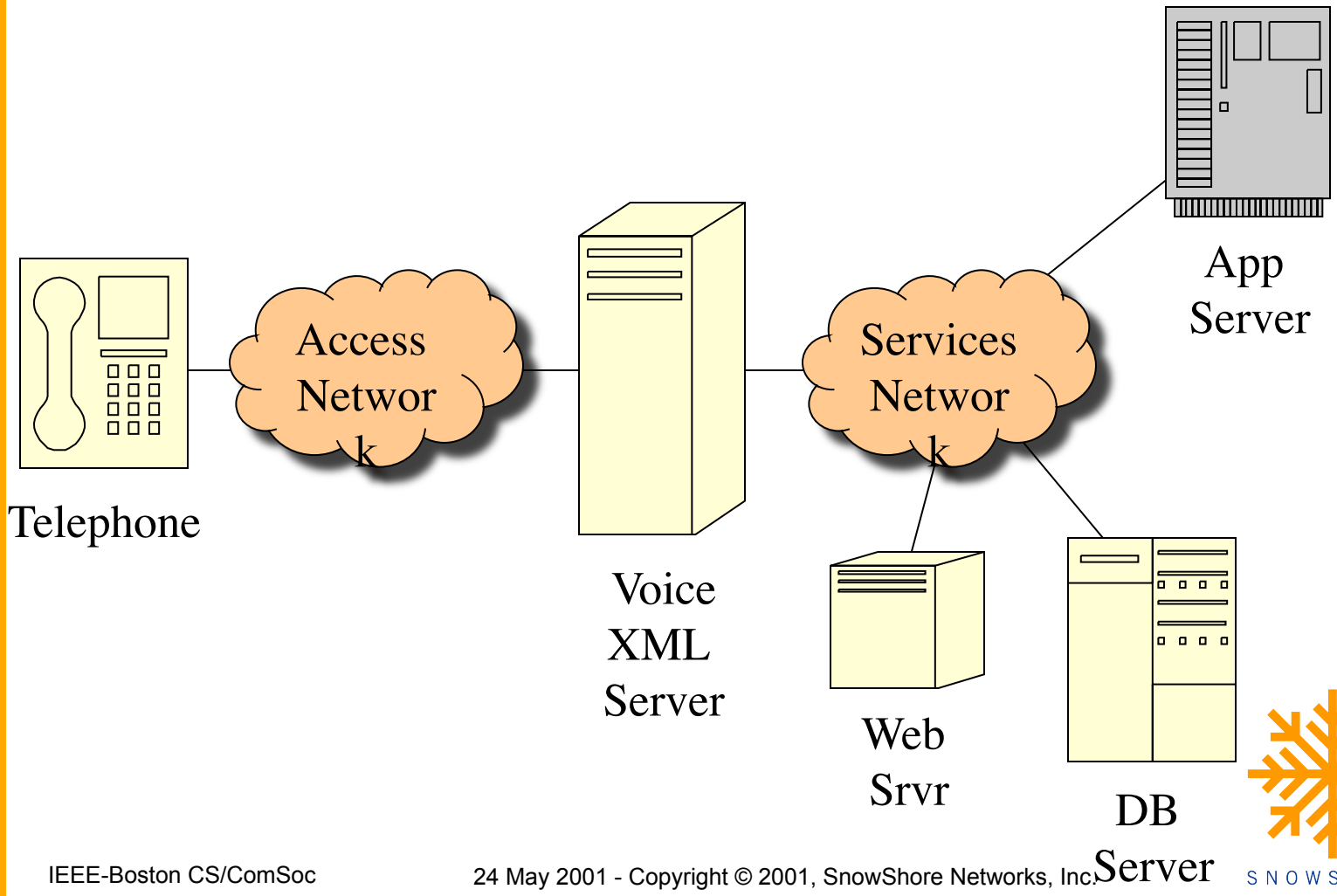
Roadmap

- **Web Application Server Architecture**
- **Rendering Device Effects**
- ***What is VoiceXML?***
- **Some Applications**
- **Architectures for Implementation**
- **Summary**

VoiceXML

- **VoiceXML Provides Specification for Audio User Interaction**
  - Input: Speech, DTMF
  - Output: Audio Files, TTS
- **Interaction Control**
  - Menus
  - Forms
  - Error Handling
- **Applications Think In Terms of**
  - Get Information
  - Convey Information

VoiceXML Application Architecture



Basic VoiceXML

```
<?xml version="1.0"?>
<vxml version="1.0">
   <form>
      <block>
         <prompt>
            <emp>Hello</emp>, World!
         </prompt>
      </block>
   </form>
</vxml>
```

Menus

```
<?xml version="1. 0"?>
<vxml version="1. 0">
   <menu>
        <prompt>Would you like <enumerate/></prompt>
        <choice next="http://…coffee.vxml">coffee</choice>
        <choice next="http://…tea.vxml">tea</choice>
        <choice next="http://…milk.vxml">milk</choice>
        <choice next="http://…nothing.vxml">nothing</choice>

        <nomatch>I didn't understand what you said.</nomatch>
        <noinput>You must say something.</noinput>
   </menu>
</vxml>
```

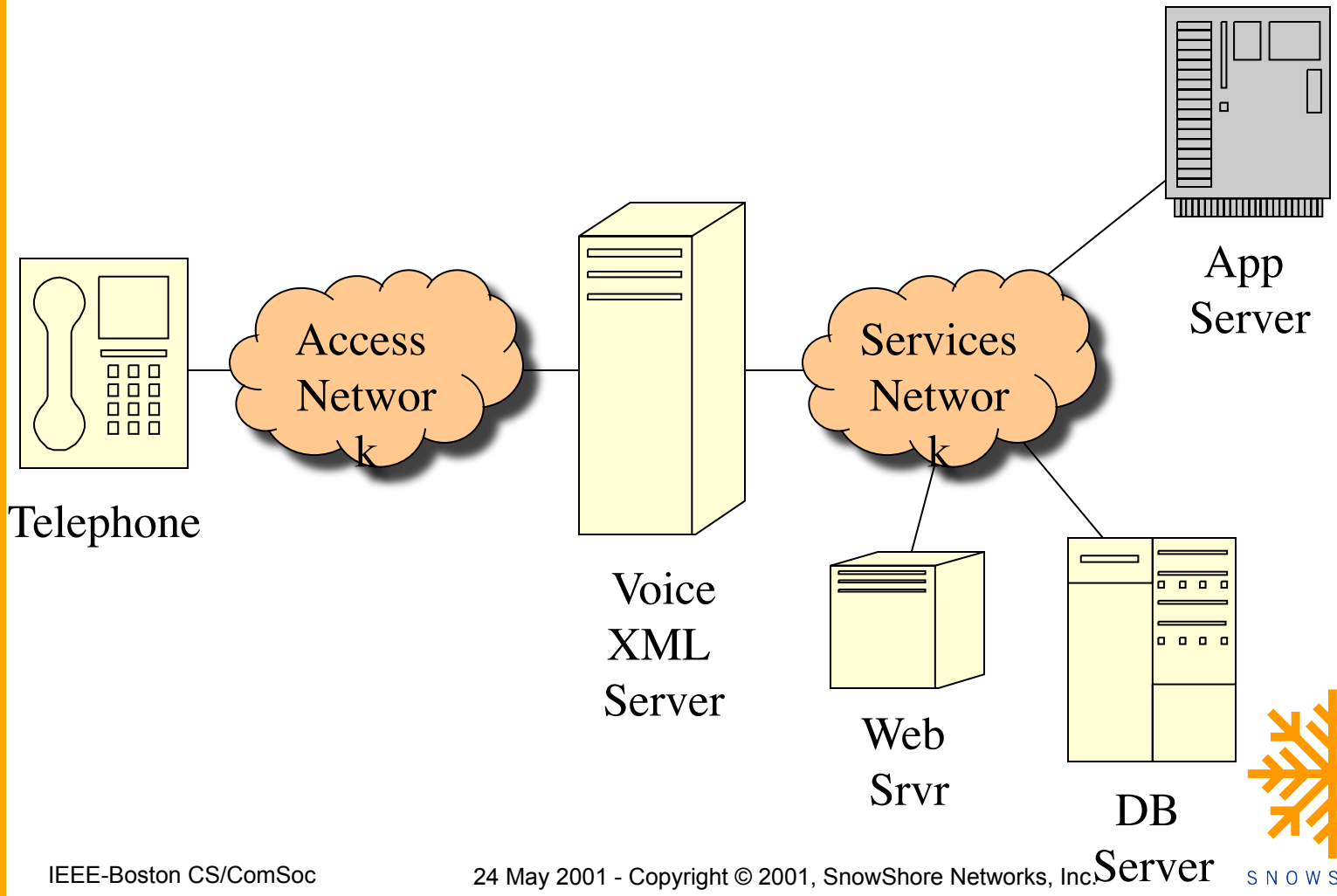
Roadmap

- **Web Application Server Architecture**
- **Rendering Device Effects**
- **What is VoiceXML?**
- ***Some Applications***
- **Architectures for Implementation**
- **Summary**

Dial-Out Conferencing

- **Presence Application**
- **Classic “When Everyone’s There, Call Them Into the Conference”**
- **Voice Dialog for Acceptance**
- **Voice Dialog for Control**

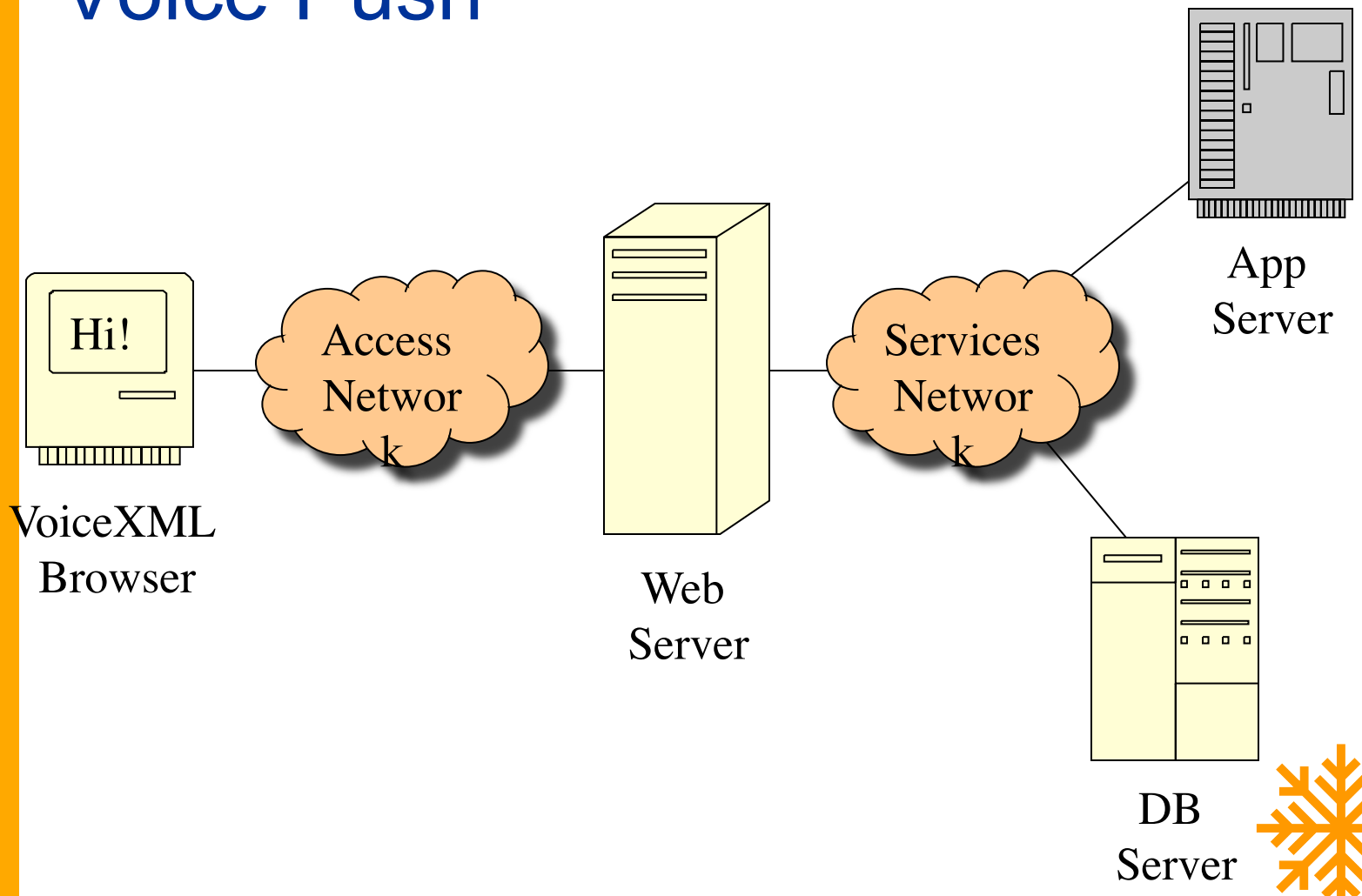
VoiceXML Application Architecture: Dial-Out Conferencing



Voice Form Push to End Device

- **Instant Message Application / Accessibility Application**
- **Render Voice at Terminal**
  - Not Just Text
  - Can Ask Questions: VoiceXML is a Form Language
- **Voice Synthesized, Recorded, or Streamed (VoIP)**

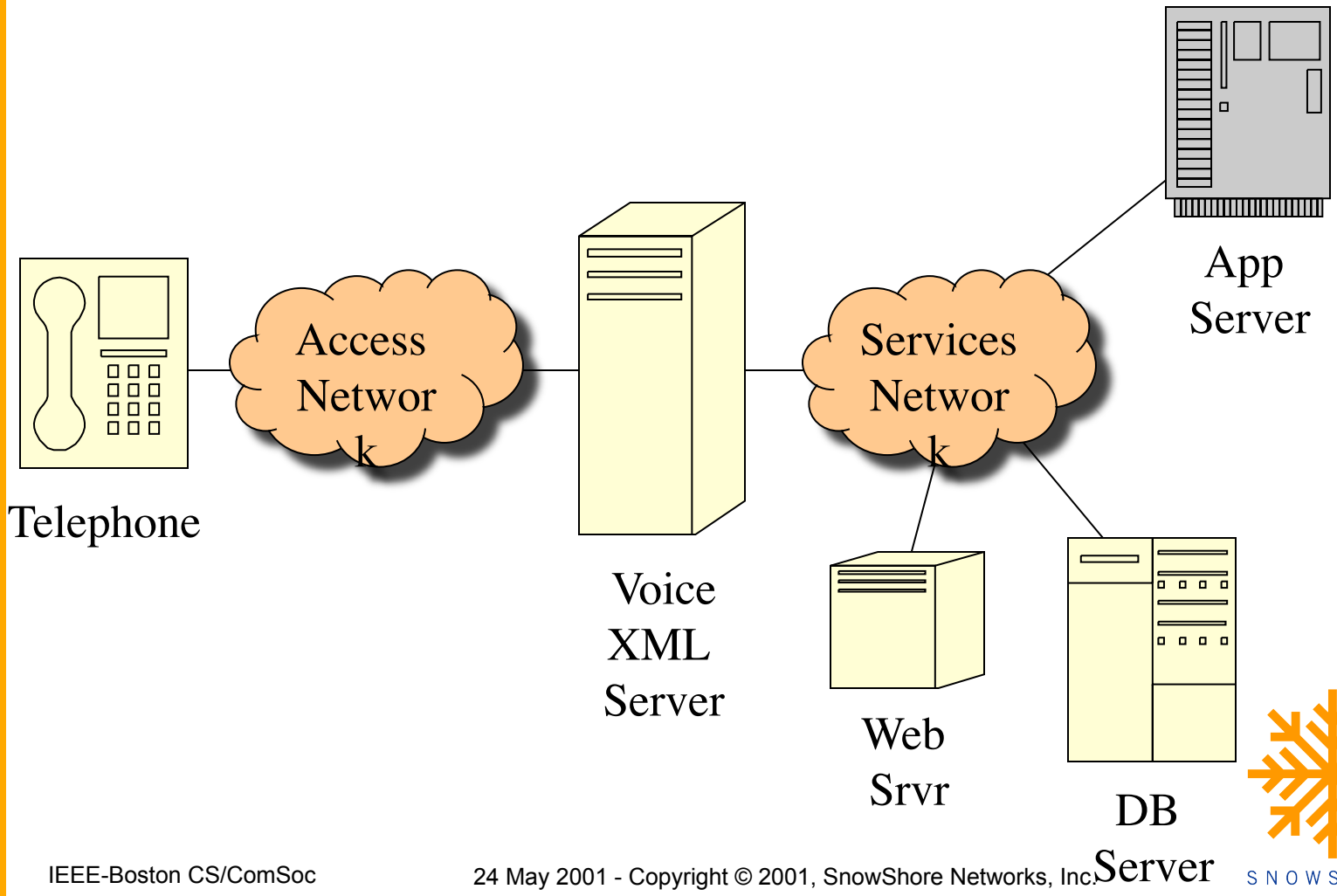
VoiceXML Application Architecture: Voice Push



Personal Assistant

- **Presence Application**
- **“Your boss is now available”**
- **“Your broker wants to speak with you”**

VoiceXML Application Architecture: Personal Assistant



Differences Between Voice and Visual Web

Visual Web

- Browser Is Edge Device
- Originally for Static Pages
- Browser Pulls Remote Pages
- Browser Can Render in Frames

Voice Web

- Telephone is Edge Device
- VoiceXML “Browser” is Intermediate Device
- Intermediate Device Pulls Pages
- Application Always Involved
- Can Pull Remote Pages or Transfer to Remote VoiceXML Browser

SNOWSHORE NETWORKS

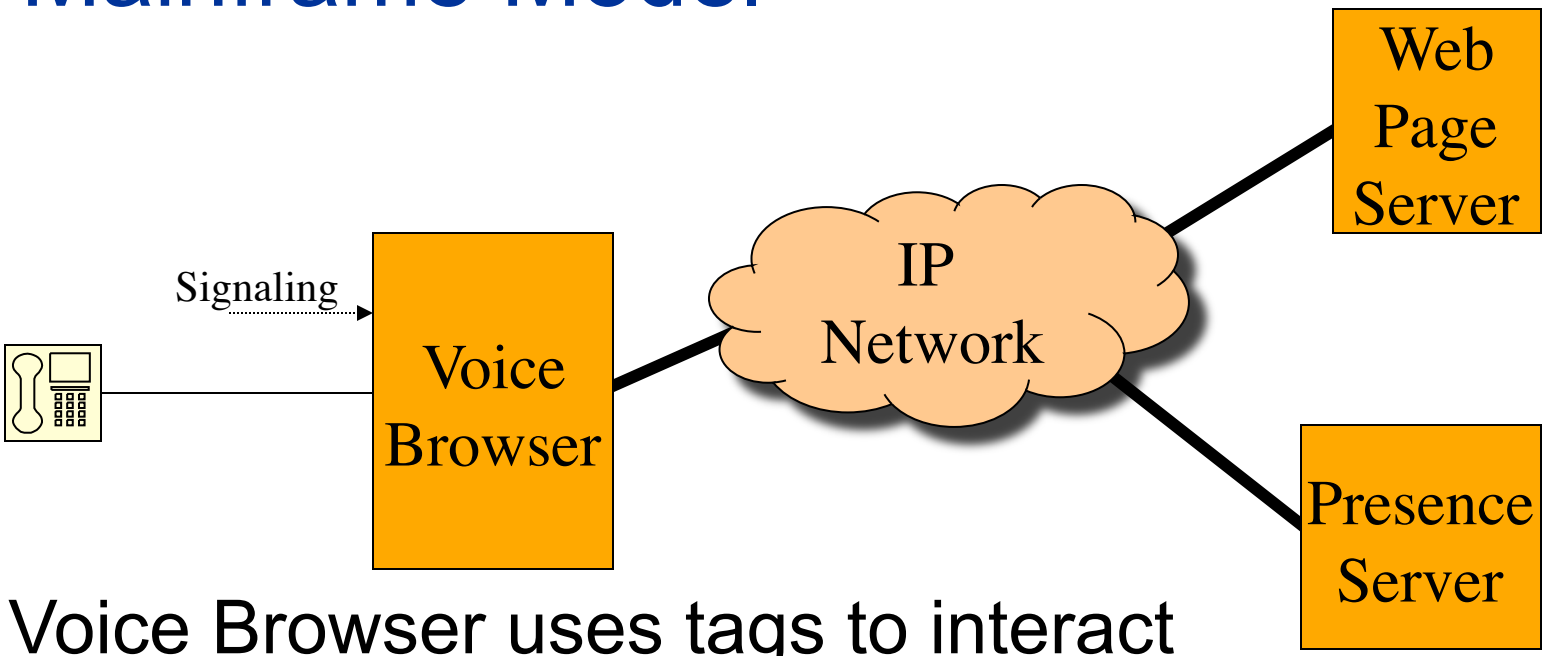
Roadmap

- **Web Application Server Architecture**
- **Rendering Device Effects**
- **What is VoiceXML?**
- **Some Applications**
- ***Architectures for Implementation***
- **Summary**

Voice UI Architecture

- **Mainframe Model**
- **Internet / WWW Model**

Mainframe Model



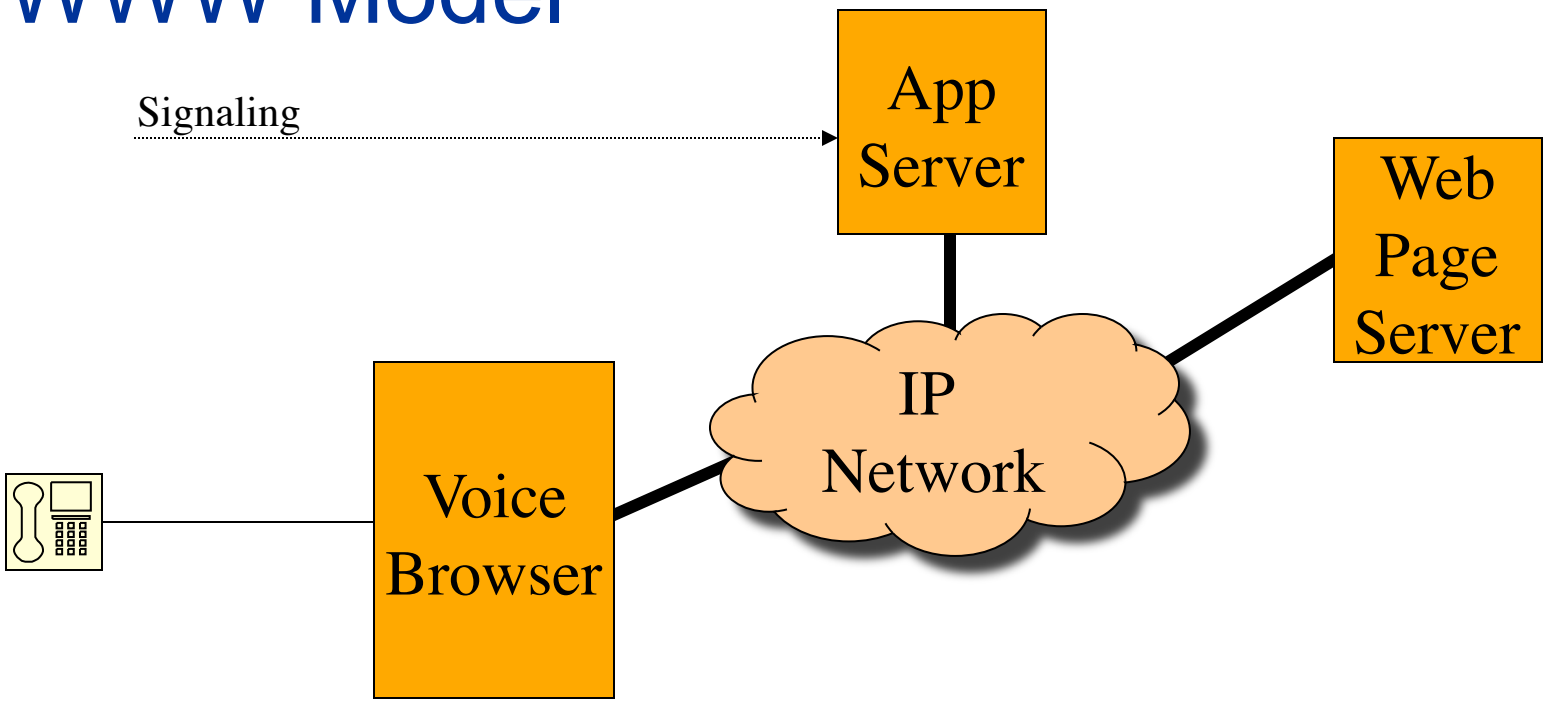
Voice Browser uses tags to interact directly with Presence Server.

```
<register event="present" src-uri="impp://foo…">
  <param user="Party B">
  …
</register>
```

Mainframe Model

- **All Interaction Through VoiceXML**
- **Pro: Linear Programming Model**
- **But**
  - New Capabilities Mean New Tags
  - New Tags Mean Language Extension
  - Language Extension Means Complexity
- ***Application Logic Intertwined With User Interface***

WWW Model



Application Server, Not Voice Browser, Interacts With Access Network.

WWW Model

- **New Features Don't Require New Tags**
- **Divides User Interface from Application**
  - 2D Web (HTML)
  - Limited 2D Web (WML)
  - Audio Web (VoiceXML)

Programming The WWW Model

- **Con: Distributed Programming Model**
  - Does Not Look as Easy as Linear Programming Model
- **But**
  - Distributed, WWW Model Is Well Understood
  - Lots of Tools Available, Especially for Application Server
  - Much More Powerful than Mainframe Model
  - Keeps VoiceXML Simple and Easy to Learn
- ***Application Logic Unchanged from Visual Application***

Summary

- Briefly Described Web Application Server Architecture
- Briefly Described VoiceXML
- Discussed Impact of Rendering Device on Application
- Showed How We Can Use Same Application with Different User Interfaces
- Discussed 3 “Out There” Applications
- Discussed Keeping VoiceXML a Display Language

The URL

- **http://briefcase.yahoo.com/awe3bz**