the leading edge of information technology
INTERNATIONAL FEDERATION FOR INFORMATION PROCESSING

Working Group 10.4
DEPENDABLE COMPUTING
AND FAULT TOLERANCE
http://www.dependability.org/wg10.4

48TH MEETING — HAKONE, JAPAN

JULY 1 – 5, 2005

View of Mount Fuji while Rowing on Lake Ashi

Toulouse, December 2005

Credits: Cover photographs by Jay Lala (front) and Johan Karlsson (back)

the leading edge of information technology

INTERNATIONAL FEDERATION FOR INFORMATION PROCESSING

Working Group 10.4
DEPENDABLE COMPUTING
AND FAULT TOLERANCE
http://www.dependability.org/wg10.4

Chairperson:

Jean Arlat

LAAS-CNRS
7, Avenue du Colonel Roche
31077 Toulouse Cedex 4
France

Phone: +33 5 61 33 62 33
Fax: +33 5 61 33 64 11
EMail: jean.arlat@laas.fr

Vice Chairpersons:

Takashi Nanya

RCAST
University of Tokyo
4-6-1, Komaba, Meguro-ku
Tokyo 153-8904
Japan

Phone: +81 3 5452 5160
Fax: +81 3 5452 5161
EMail: nanya@hal.rcast.u-tokyo.ac.jp

William H. Sanders

CRHC - Coordinated Science Lab.
University of Illinois at Urbana-Champaign
1308 West Main Street
Urbana, IL 61801
USA

Phone: +1 217 333 0345
Fax: +1 217 244 3359
EMail: whs@crhc.uiuc.edu

Organizers:

48th Meeting of IFIP WG 10.4

Hakone, Japan

July 1-5, 2005

Yoshihiro Tohma
4-21-15-1404 Shimomaruko
Ohta-Ku, Tokyo 146-0092 Japan
Tel : +81 50 7518 2566
Fax. +81 50 7518 2566
Email: tohma@cocoa.plala.org.jp

Takashi Nanya

W. Kent Fuchs
College of Engineering
Cornell University
242 Carpenter Hall
ITHACA, NY 14853-2201 USA
Tel: +1 607 255 9679
Fax: +1 607 255 9606
Email: kent-fuchs@cornell.edu

CONTENTS

Program of the Meeting

48th Meeting, Hakone, Japan, July 1-5, 2005

Meeting Host: Takashi Nanya, University of Tokyo, Japan

Workshop on Grid Computing and Dependability *Saturday, July 2*

Coordinator: **Yoshihiro Tohma**, Tokyo Denki University, Japan

Yoshihiro Tohma,
Introduction to the Workshop

Session 1 – Evolution of Grid Computing and its Dependability — Moderator: Hirokazu Ihara, Hiro Systems Lab, Japan

Matti A. Hiltunen, AT&T Labs - Research, Florham Park, NJ, USA
Dependability Issues in the Emerging Web Services-Based Grid Computing Standards

T. Basil Smith, IBM Research, Hawthorne, NY, USA
Grid on Future Blade Data Center Infrastructure

Luca Simoncini, University of Pisa, Italy
Grid Computing Evolution and Challenges for Resilience, Performance and Scalability

Session 2 – Practice and Experiments — Moderator: Jaynarayan H. Lala, Raytheon Company, Arlington, VA, USA

Takanori Seki, IBM, Tokyo, Japan
Customer Interest, Expectation, and Requirement for Grid in Dependability Context

Nobutoshi Sagawa, Hitachi System Development Labs, Kawasaki, Japan
Japanese Business Grid Project – Objectives & Key Technical Issues

Session 3 – Fault Tolerance in Grid Computing — Moderator: Richard D. Schlichting, AT&T Labs - Research, Florham Park, NJ, USA

Xavier Défago, Japan Advanced Institute of Science and Technology, Ishikawa, Japan
Revisiting Failure Detection for Grid Systems

Franck Cappello, LRI – University of Paris-Sud, Orsay, France
Fault Tolerance in Grid and Grid 5000

Session 4 – Security in Grid Computing — Moderator: Paulo J. E. Veríssimo, University of Lisbon, Portugal

Jon Kim, Pohang University of Science & Technology, Korea
Security Issues in Grid: Authentication and Authorization

Ravishankar K. Iyer, University of Illinois at Urbana-Champaign, USA
Dependability and Security Issues in Measurement-Based Design of Grid

Carl Landwehr, NSF, Arlington, VA, USA
Secure Grid Computing: An Empirical View

Session 5 – Synthesis and Wrap Up — Moderator: Yoshihiro Tohma

Hirokazu Ihara
Summary of Session 1

Jaynarayan H. Lala
Summary of Session 2

Richard D. Schlichting
Summary of Session 3

Paulo J. E. Veríssimo
Summary of Session 4

Yoshihiro Tohma
Wrap Up

Workshop on Nomadic Computing and Dependability — *Monday, July 4*

Coordinator: **W. Kent Fuchs,** Cornell University, Ithaca, NY, USA

Session 1 – Nomadic Devices and Dependability — Moderator: Yoshiaki Koga, Acad. & Educ. Foundation for NDA, Yokohama, Japan

W. Kent Fuchs
1) Workshop Introduction and Overview of Issues — 2) Assignments

Marc-Olivier Killijian, LAAS-CNRS, Toulouse, France
Cooperative Backup for Nomadic Devices

Session 2 – Challenges in Mobile Distributed Systems — Moderator: Karama Kanoun, LAAS-CNRS, Toulouse, France

Yoshiaki Kakuda, Hiroshima City University, Hiroshima, Japan
Autonomous Clustering and Hierarchical Routing for Mobile Ad Hoc Networks

Farnam Jahanian, Arbor Networks and University of Michigan, Ann Arbor, Michigan, USA
The Crumbling Perimeter: Mobile Networking and Internal Security Issues

Christof Fetzer, Dresden University of Technology, Germany
Timed Asynchronous Models for Mobile Systems

Session 3 – Mobility and Ubiquitous Computing — Moderator: Henrique Madeira, University of Coimbra, Portugal

Emin Gün Sirer, Cornell University, Ithaca, NY, USA
A Comprehensive Localization Framework for Self-Organizing Nomadic Systems

Richard D. Schlichting, AT&T Research, Florham Park, NJ, USA
A Network Service Provider's View of Ubiquitous Computing

Session 4 – Synthesis and Wrap Up — Moderator: W. Kent Fuchs

Yoshiaki Koga
Summary of Session 1

Karama Kanoun
Summary of Session 2

Henrique Madeira
Summary of Session 3

Carl E. Landwehr, NSF, Arlington, VA, USA
Contribution to Assignment: Terminology

David Powell, LAAS-CNRS, Toulouse, France
Contribution to Assignment: Terminology

Paulo J. E. Veríssimo, University of Lisbon, Portugal
Contribution to Assignment: Find Your Way in the Jungle of Perviquitous Systems

W. Kent Fuchs
Wrap Up

IFIP WG 10.4 Business Meeting

Jean Arlat, LAAS-CNRS, Toulouse, France

Jean Arlat
Overall Presentation and News

Takashi Nanya
Report on IEEE/IFIP DSN-2005

Chandra M.R. Kintala, Stevens Institute of Technology, Hoboken, NJ, USA
Update on IEEE/IFIP DSN-2006 (see also www.dsn.org)

Tom Anderson, University of Newcastle, UK (presented by Jean Arlat)
Update on 52th IFIP WG 10.4 Meeting (Edinburgh, Scotland)

Richard D. Schlichting
Update on 49th IFIP WG 10.4 Meeting (Tucson, AZ, USA)

Jaynarayan H. Lala, Raytheon Company, Arlington, VA, USA
Update on 50th IFIP WG 10.4 Meeting (Annapolis, MD, USA)

João Gabriel Silva, University of Coimbra, Portugal
Update on EDCC-6 (Coimbra, Portugal)

Research Reports *Tuesday, July 5*

Session 1 – Moderator: William H. Sanders, UIUC, Urbana-Champaign, IL, USA

Elias P. Duarte Jr., Federal University of Parana, Curitiba, Brazil
Dependable TCP/IP Networking

Hiroshi Nakamura, University of Tokyo, Japan
Megascale Project: A Low-Power and Compact Cluster for High-Performance

Jie Xu, University of Leeds, UK
Provenance-Aware Fault Tolerance for Grid Computing

António Casimiro, University of Lisbon, Portugal
A New Programming Model for Dependable Adaptive Real-Time Applications

Paulo J. E. Veríssimo, University of Lisbon, Portugal
***FLP is Back!* Or A Forgotten Dimension of Time in Distributed Systems Problems**

Rainer Knauf, Technical University of Ilmenau, Germany
Human Expertise in Fault Detection and Adjustment: An Empirical Case Study

Session 2 – Moderator: Takashi Nanya

Nobuyasu Kanekawa, Hitachi Research Laboratory, Hitachi Ltd, Ibaraki, Japan
X-by-Wire Systems

Setsuo Tsuruta, Tokyo Denki University, Japan
Reflection Oriented Dependable Planning Concept (RDPC)
and its Application to the Learning in Education and in Intelligent Agent

Henrique Madeira, University of Coimbra, Portugal
Experimental Software Risk Assessment

Kevin Driscoll, Honeywell Laboratories, Minneapolis, MN, USA
Real Time Cryptography

Yuji Hirao, JR Railway Technical Research Institute, Tokyo, Japan
A Railway Maintenance Staff Protection System

Attendance List

Prof. ALVISI Lorenzo
Taylor Hall 2.124
University of Texas at Austin
AUSTIN, TX 78712-1014 USA
Tel. (+1) 512 471 9792
Fax. (+1) 512 232 7886
Email. lorenzo@cs.utexas.edu

Dr. ARLAT Jean
LAAS - CNRS
7, avenue du Colonel Roche
31077 TOULOUSE Cedex 4 FRANCE
Tel. (+33) 05 61 33 62 33
Fax. (+33) 05 61 33 64 11
Email. Jean.Arlat@laas.fr

Prof. BONDAVALLI Andrea
Dip. di Sistemi e Informatica
Universita di Firenze
Viale Morgani, 65
I - 50134 FIRENZE ITALY
Tel. (+39 0) 32 943 09838
Fax. (+39 0) 55 423 7436
Email. BONDAVALLI@UNIFI.IT

Dr.CAPPELO Franck
INRIA – LRI
Laboratoire de Recherche Informatique
Bâtiment 490 - Université Paris Sud
91405 ORSAY CEDEX FRANCE
Tel. (+33) 6 70 31 03 39
Fax. (+33) 1 69 15 65 86
Email. fci@iri.fr

Mr. CASIMIRO COSTA Antonio
Bloco C5, Piso 1
Dept. of Informatics
University of Lisboa
Campo Grande
1749-016 LISBOA PORTUGAL
Tel. (+351) 217 500 612
Fax. (+351) 217 500 533
Email. casim@di.fc.ul.pt

Prof. DEFAGO Xavier
Graduate School of information Science
JAIST
1-1 Asahidai, NOMI-GUN
ISHIKAWA, 923-1292 JAPAN
Tel. (+81) 761 51 1224
Fax. (+81) 761 51 1149
Email. defago@jaist.ac.jp

Mr. DRISCOLL Kevin R.
MN65-2200
Honeywell Laboratories
3660 Technology Drive
MINNEAPOLIS, MN 55418-1006 USA
Tel : (+1) 612 951 7263
Fax : (+1) 612 951 7438
Email : kevin.driscoll@honeywell.com

Prof. DUARTE JR. Elias P.
Dept. Informatics
UFPR
Caixa Postal 19018
81531-990 CURITIBA, PR BRAZIL
Tel. (+55) 41 3361 3656
Fax. (+55) 41 3361 3205
Email. elias@inf.ufpr.br

Dr. ELNOZAHY Mootaz
M/S 9460, System Software Dept.
IBM
11400 Burnet Rd.
AUSTIN, TX 78758 USA
Tel. (+1) 512 823 6738
Fax. (+1) 305 402 2432
Email. mootaz@us.ibm.com

Prof. FETZER Christof
Dept. of Computer Science
Technische Universität Dresden
Mommsenstr. 13
D - 01062 DRESDEN GERMANY
Tel. (+49) 351 463 39709
Fax. (+49) 351 463 39710
Email. cf2@inf.tu-dresden.de

Prof. FUCHS W. Kent
College of Engineering
Cornell University
242 Carpenter Hall
ITHACA, NY 14853-2201 USA
Tel. (+1) 607 255 9679
Fax. (+1) 607 255 9606
Email. Kent-Fuchs@cornell.edu

Dr. HEIMERDINGER Walter
MN 65-2200
Honeywell Laboratories
3660 Technology Drive
MINNEAPOLIS, MN 55418-1006 USA
Tel. (+1) 612 951 7333
Fax. (+1) 612 951 7438
Email. walt.heimerdinger@honeywell.com

Mr. HEINER Günter
Bartningallee 16
D - 10557 BERLIN GERMANY
Tel. (+49) 30 392 5720
Fax. (+49)
Email. guenter.heiner@ieee.org

Dr. HILTUNEN Matti
Room E211
AT&T Labs Research
180 Park Avenue
FLORHAM PARK, NJ 07932 USA
Tel. (+1) 973 360 5504
Fax. (+1) 973 360 8077
Email. hiltunen@research.att.com

Dr. HIRAO Yuji
Signalling & Telecom. Technology Division
Railway Technical Research Inst.
2-8-38 Hikari-Cho
Kokubunji-Shi
TOKYO, 185-8540 JAPAN
Tel. (+81) 42 573 7326
Fax. (+81) 42 573 7328
Email. hirao@rtri.or.jp

Prof. IHARA Hirokazu
School of Information Environment
Tokyo Denki University
1-15-11 Minaminaruse
MACHIDA, TOKYO, 194-0045 JAPAN
Tel. (+81) 42 723 4043
Fax. (+81) 42 723 4043
Email. ihara@coral.ocn.ne.jp

Prof. IYER Ravishankar K.
Center for Reliable and High Performance Computing
University of Illinois
1308 West Main Street
URBANA, IL 61801 USA
Tel. (+1) 217 333 2510
Fax. (+1) 217 244 1764
Email. iyer@crhc.uiuc.edu

Prof. JAHANIAN Farnam
Dept. of EECS
University of Michigan
1301 Beal Ave.
ANN ARBOR, MI 48109-2122 USA
Tel. (+1) 734 936 2974
Fax. (+1) 734 763 8094
Email. farnam@umich.edu

Prof. KAKUDA Yoshiaki
Faculty of Information Sciences
Hiroshima City University
3-4-1, Ozuka-Higashi
Asaminami-ku
HIROSHIMA, 731-3194 JAPAN
Tel. (+81) 82 830 1696
Fax. (+81) 82 830 1792
Email. kakuda@ce.hiroshima-cu.ac.jp

Dr. KANEKAWA Nobuyasu
MD : # 104
Hitachi Ltd.
2520 Takaba, Hitachi-naka-city
IBARAKI, 312-8503 JAPAN
Tel. (+81) 29 276 6856
Fax. (+81) 29 274 9811
Email. kanekawa@hrl.hitachi.co.jp

Dr. KANOUN Karama
LAAS - CNRS
7, avenue du Colonel Roche
31077 TOULOUSE Cedex 4 FRANCE
Tel. (+33) 05 61 33 62 35
Fax. (+33) 05 61 33 64 11
Email. karama.kanoun@laas.fr

Prof. KARLSSON Johan
Dept. of Computer Science & Engineering
Chalmers University of Technology
SE-412 96 GOTEBORG SWEDEN
Tel. (+46) 31 772 1670
Fax. (+46) 31 772 3663
Email. johan@ce.chalmers.se

Dr. KILLIJIAN Marc-Olivier
LAAS - CNRS
7, avenue du Colonel Roche
31077 TOULOUSE Cedex 4
FRANCE
Tel. (+33) 05 61 33 62 41
Fax. (+33) 05 61 33 64 11
Email. Marco.Killijian@laas.fr

Prof. KIM Jong
Dept. of CSE
Pohang Univ. Science & Technology
San-31, Hyoja-dong
790-784 POHANG KOREA (REP. OF)
Tel. (+82) 54 279 2257
Fax. (+82) 54 279 2299
Email. jkim@postech.ac.kr

Prof. KINTALA Chandra M.R.
Electrical & Computer Engineering
Stevens Institute of Technology
Castle Point on Hudson
HOBOKEN, NJ 07030 USA
Tel. (+1) 201 216 8057
Fax. (+1) 201 216 8246
Email. chandra@kintala.com

Dr. KNAUF Rainer
School of Comp. Science & Automation
Technical University of Ilmenau
Postfach 10 05 65
D - 98684 ILMENAU GERMANY
Tel. (+49) 3677 69 1445
Fax. (+49) 3677 69 1665
Email. rainer.knauf@tu-ilmenau.de

Prof. KOGA Yoshiaki
Academic & Education Foundation for NDA
1-1-8 Noukendai Kanazawaku
YOKOHAMA, 236-0057 JAPAN
Tel. (+81) 45 771 9311
Fax. (+81) 45 771 9323
Email. yoshi.koga@nifty.com

Dr. KONDO Masaaki
RCAST
University of Tokyo
4-6-1 Komaba, Meguro-ku
TOKYO, 153-8904 JAPAN
Tel. (+81) 3 5452 5167
Fax. (+81) 3 5452 5165
Email. kondo@hal.rcast.u-tokyo.ac.jp

Prof. KOOPMAN Philip
ECE Dept. - HH A-308
Carnegie Mellon University
PITTSBURGH, PA 15213 USA
Tel. (+1) 412 268 5225
Fax. (+1) 412 268 6353
Email. koopman@cmu.edu

Prof. KUO Sy-Yen
Dept. of Electrical Eng. BL - 522
National Taiwan University
106 TAIPEI TAIWAN
Tel. (+886) 2 3766 3577
Fax. (+886) 2 2368 9172
Email. sykuo@cc.ee.ntu.edu.tw

Dr. LALA Jaynarayan H.
Crystal Center 2 - Suite 1000
Raytheon Company
2461 South Clark St.
ARLINGTON, VA 22202 USA
Tel. (+1) 703 419 1401
Fax. (+1) 703 419 1310
Email. Jay_Lala@raytheon.com

Dr. LANDWEHR Carl E.
Program Director, Cyber Trust
CISE/CNS
National Science Foundation
4201 Wilson Boulevard
ARLINGTON, VA 22230 USA
Tel. (+1) 703 292 8950
Fax. (+1) 703 292 9059
Email. clandweh@nsf.gov

Dr. LAPRIE Jean-Claude
LAAS - CNRS
7, avenue du Colonel Roche
31077 TOULOUSE Cedex 4 FRANCE
Tel. (+33) 05 61 33 78 85
Fax. (+33) 05 61 33 64 11
Email. Jean-Claude.Laprie@laas.fr

Prof. MADEIRA Henrique
Dep. Eng. Informatica
Universidade de Coimbra
Pinhal de Marrocos
P - 3030-290 COIMBRA PORTUGAL
Tel. (+351) 2 39 790 003
Fax. (+351) 2 39 701 266
Email. henrique@dei.uc.pt

Prof. MIYAHO Noriharu
Dpt Information Environnment Engineering
Tokyo Denki University
2-1200 Muzai Gakuendai Inzai
CHIBA, 270-1382 JAPAN
Tel. (+81) 476 46 8632
Fax. (+81) 476 46 8632
Email. miyaho@sie.dendai.ac.jp

Prof. NAKAMURA Hiroshi
RCAST
University of Tokyo
4-6-1 Komaba, Meguro-ku
TOKYO, 153-8904 JAPAN
Tel. (+81) 3 5452 5162
Fax. (+81) 3 5452 5163
Email. nakamura@hal.rcast.u-tokyo.ac.jp

Prof. NANYA Takashi
RCAST
University of Tokyo
4-6-1 Komaba, Meguro-ku
TOKYO, 153-8904 JAPAN
Tel. (+81) 3 5452 5160
Fax. (+81) 3 5452 5161
Email. nanya@hal.rcast.u-tokyo.ac.jp

Dr. POWELL David
LAAS - CNRS
7, avenue du Colonel Roche
31077 TOULOUSE Cedex 4 FRANCE
Tel. (+33) 05 61 33 62 87
Fax. (+33) 05 61 33 64 11
Email. david.powell@laas.fr

Mr. SAGAWA Noburoshi
System Development Laboratory
Hitachi Ltd.
1099 Ohzenji Asao Kawasaki
KANAGAWA, 215 JAPAN
Tel. (+81) 44 959 0232
Fax. (+81) 44 959 0853
Email. sagawa@sdl.hitachi.co.jp

Prof. SANDERS William H.
Coordinated Science Lab.
University of Illinois
1308 West Main Street
URBANA, IL 61801 USA
Tel. (+1) 217 333 0345
Fax. (+1) 217 244 3359
Email. whs@uiuc.edu

Dr. SCHLICHTING Richard D.
Shannon Laboratory, E221
AT&T Labs Research
180 Park Avenue
FLORHAM PARK, NJ 07932 USA
Tel. (+1) 973 360 8234
Fax. (+1) 973 360 8077
Email. rick@research.att.com

Mr. SEKI Taka
IBM Japan, Ltd.
19-21 Nihonbashi Hakozaki-cho
103-8510, Chio-Ku – Tokyo JAPAN
Tel: (+81) 3 5644 2500
Fax. (+81) 3 3664 4999
Email: ts@jp.ibm.com

Prof. SILVA João Gabriel
Dep. Eng. Informatica
Universidade de Coimbra
Pinhal de Marrocos
P - 3030-290 COIMBRA PORTUGAL
Tel. (+351) 2 39 790 005
Fax. (+351) 2 39 701 266
Email. jgabriel@dei.uc.pt

Prof. SIMONCINI Luca
Dept. of Information Engineering
Dip. Ing. dell'Informazione
Universita di Pisa
Via Girolamo Caruso
I-56122 PISA ITALY
Tel. (+39 0) 50 315 2983
Fax. (+39 0) 50 970 456
Email. luca.simoncini@isti.cnr.it

Prof. SIRER Emin Gun
Computer Science Dpt
Cornell University
4119A Upson Hall
ITHACA, NY 14853 USA
Tel. (+1) 607 255 7673
Fax. (+1) 607 255 4428
Email. egs@cs.cornelle.edu

Dr. SMITH T. Basil
M.S. 4S-A26
IBM
19 Skyline Drive
HAWTHORNE, NY 10532 USA
Tel. (+1) 914 784 7018
Fax. (+1) 914 784 6201
Email. tbsmith@us.ibm.com

Prof. SURI Neeraj
Fachbereich Informatik/Dept. of Computer Science
Technische Universität Darmstadt
Hochschulst. 10
D-64289 DARMSTADT GERMANY
Tel. (+49) 6151 16 3513
Fax. (+49) 6151 16 4310
Email. suri@informatik.tu-darmstadt.de

Prof. TOHMA Yoshihiro
4-21-15-1404 Shimomaruko
Ohta-Ku, Tokyo 146-0092 JAPAN
Tel. (+81) 50 7518 2566
Fax. (+81) 50 7518 2566
Email. tohma@cocoa.plala.org.jp

Prof. TSURUTA Setsuo
School of Information Environment
Tokyo Denki University
2-1200 Muzai Gakuendai Inzai
CHIBA, 270-1382 JAPAN
Tel. (+81) 476 46 8491
Fax. (+81) 476 46 8449
Email. tsuruta@sie.dendai.ac.jp

Prof. VERISSIMO Paulo
Bloco C6.3.10
Dept. of Informatics
University of Lisboa
Campo Grande
1749-016 LISBOA PORTUGAL
Tel. (+351) 21 750 01 03
Fax. (+351) 21 750 00 84
Email. pjv@di.fc.ul.pt

Prof. XU Jie
School of Computing
University of Leeds
LEEDS, LS2 9JT UK
Tel. (+44) 113 343 5193
Fax. (+44) 113 343 5468
Email. jxu@comp.leeds.ac.uk

Prof. YOKOTA Haruo
Tokyo Institute of Technology
2-12-1, Ookayama, Meguro-ku
TOKYO, 152-8552 JAPAN
Tel. (+81) 3 5735 3505
Fax. (+81) 3 5734 3504
Email. yokota@cs.titech.ac.jp

Workshop 1

Grid Computing and Dependability

Coordinator

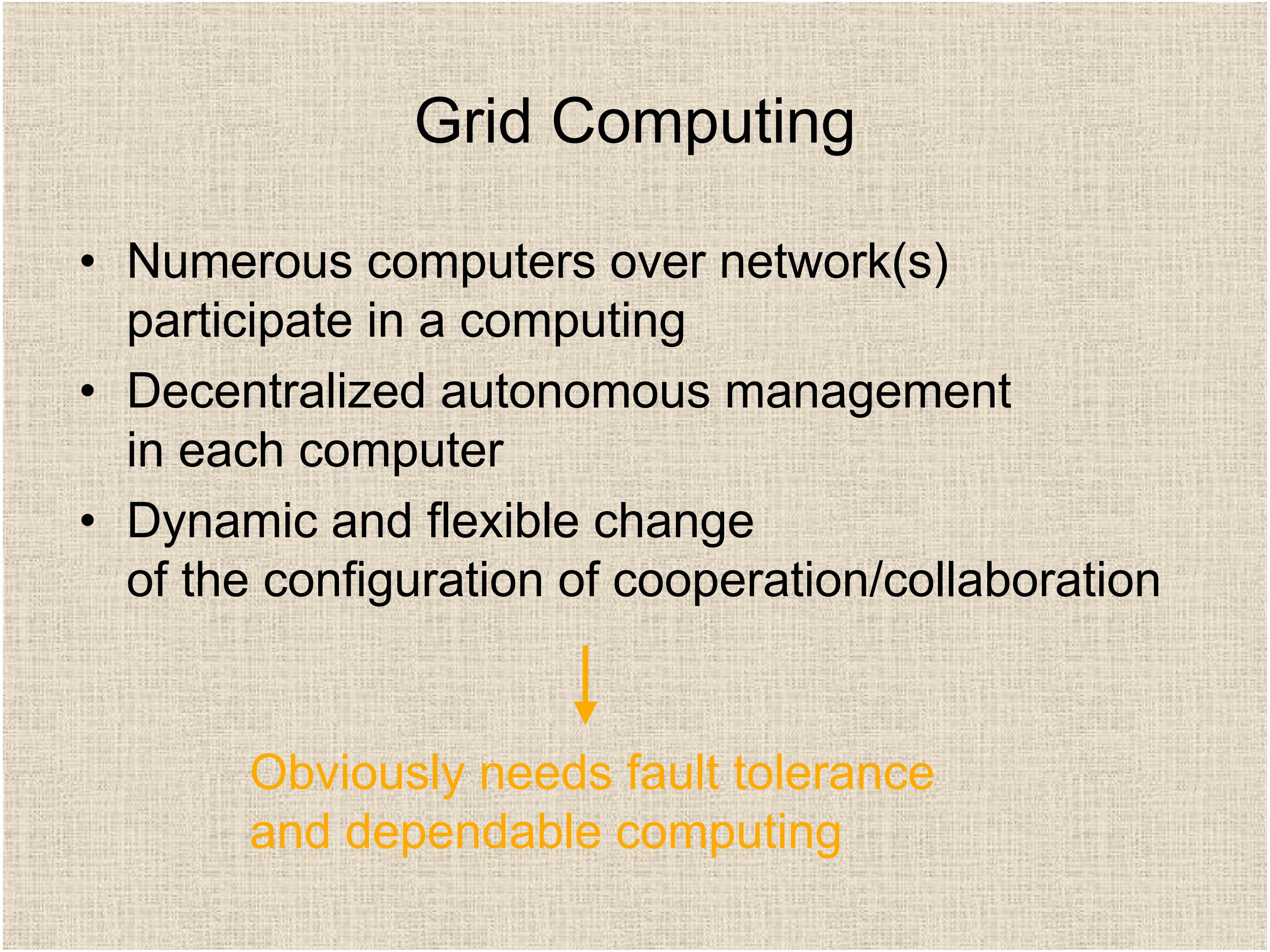
Yoshihiro Tohma, Tokyo Denki University, Japan

Challenges to Dependable Computing

- Level-up of requirements
  - even in traditional computing
- New components
  - in devices
  - in operational modes
- New environment
  - of exploding network(s)
  - of emerging computation paradigm(s)

↓

Grid Computing



Grid Computing

- Numerous computers over network(s) participate in a computing
- Decentralized autonomous management in each computer
- Dynamic and flexible change of the configuration of cooperation/collaboration

↓

Obviously needs fault tolerance
and dependable computing

Need of Interaction between Two Communities

- Dependable Computing People must know more about Grid.
- Grid Computing people must know more about Dependable Computing.
- The interaction of both communities is beneficial for the improvement of Dependability of Grid.

↓

Motivation of the Workshop

Program

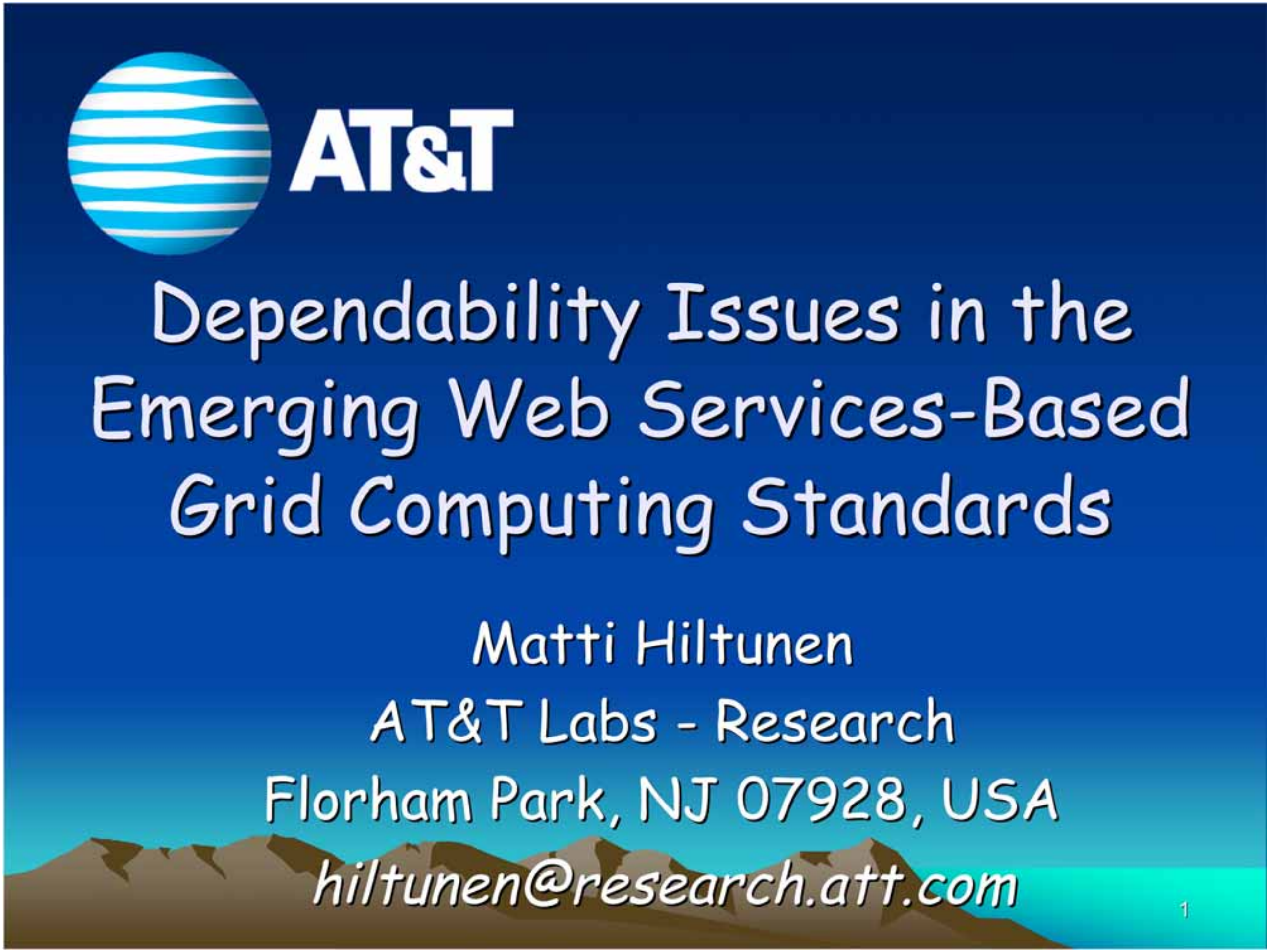
- 9 presentations (4 sessions + Synthesis & Wrap-Up)
- 40 min for each presentation + discussion
- Morning sessions relate more closely to Grid Computing itself
- Afternoon sessions relate more closely to Dependability itself

Session 1.1

Evolution of Grid Computing and its Dependability

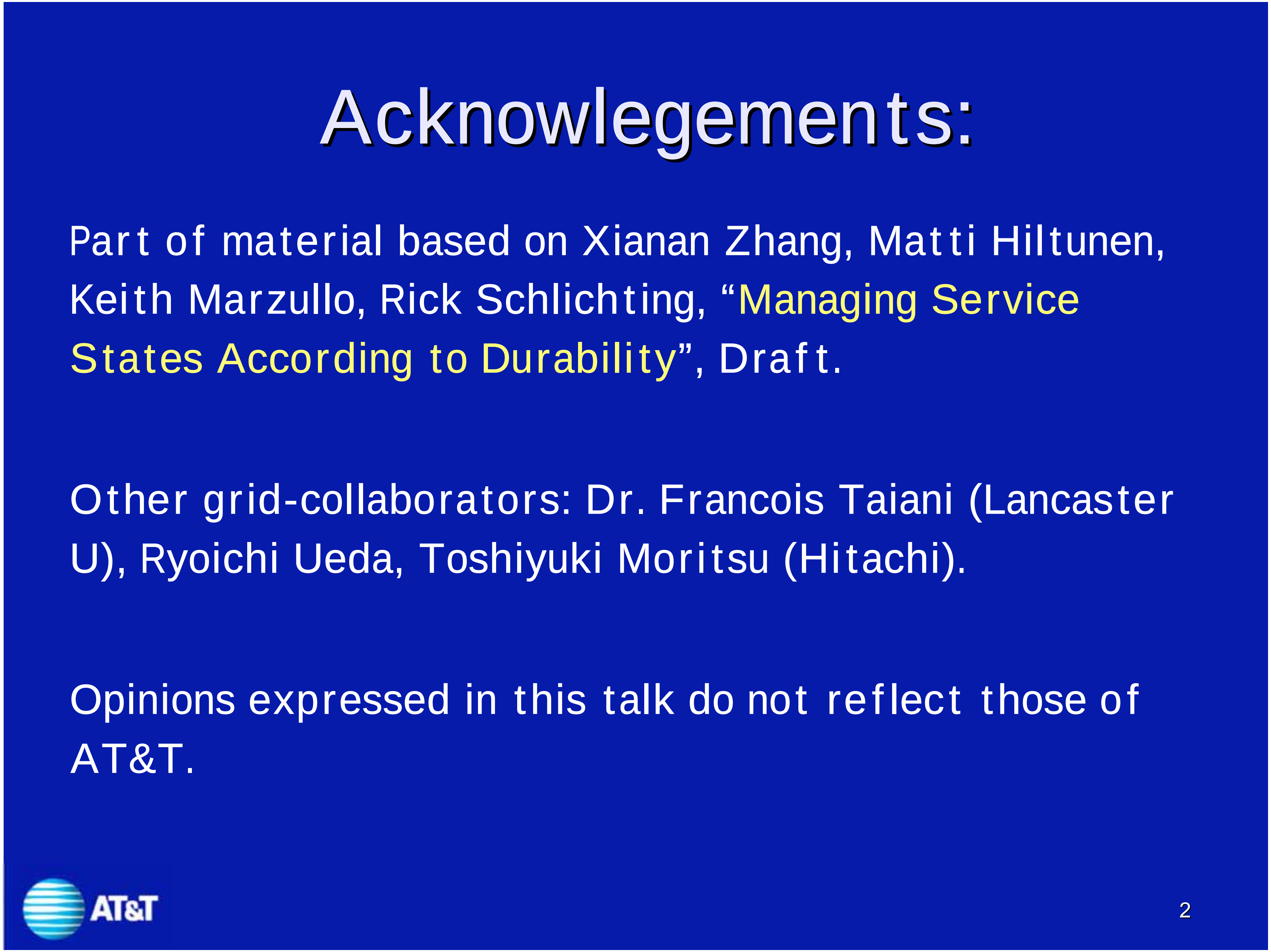
Moderator and Rapporteur

Hirokazu Ihara, Hiro Systems Lab, Japan



Dependability Issues in the Emerging Web Services-Based Grid Computing Standards

Matti Hiltunen
AT&T Labs - Research
Florham Park, NJ 07928, USA
hiltunen@research.att.com

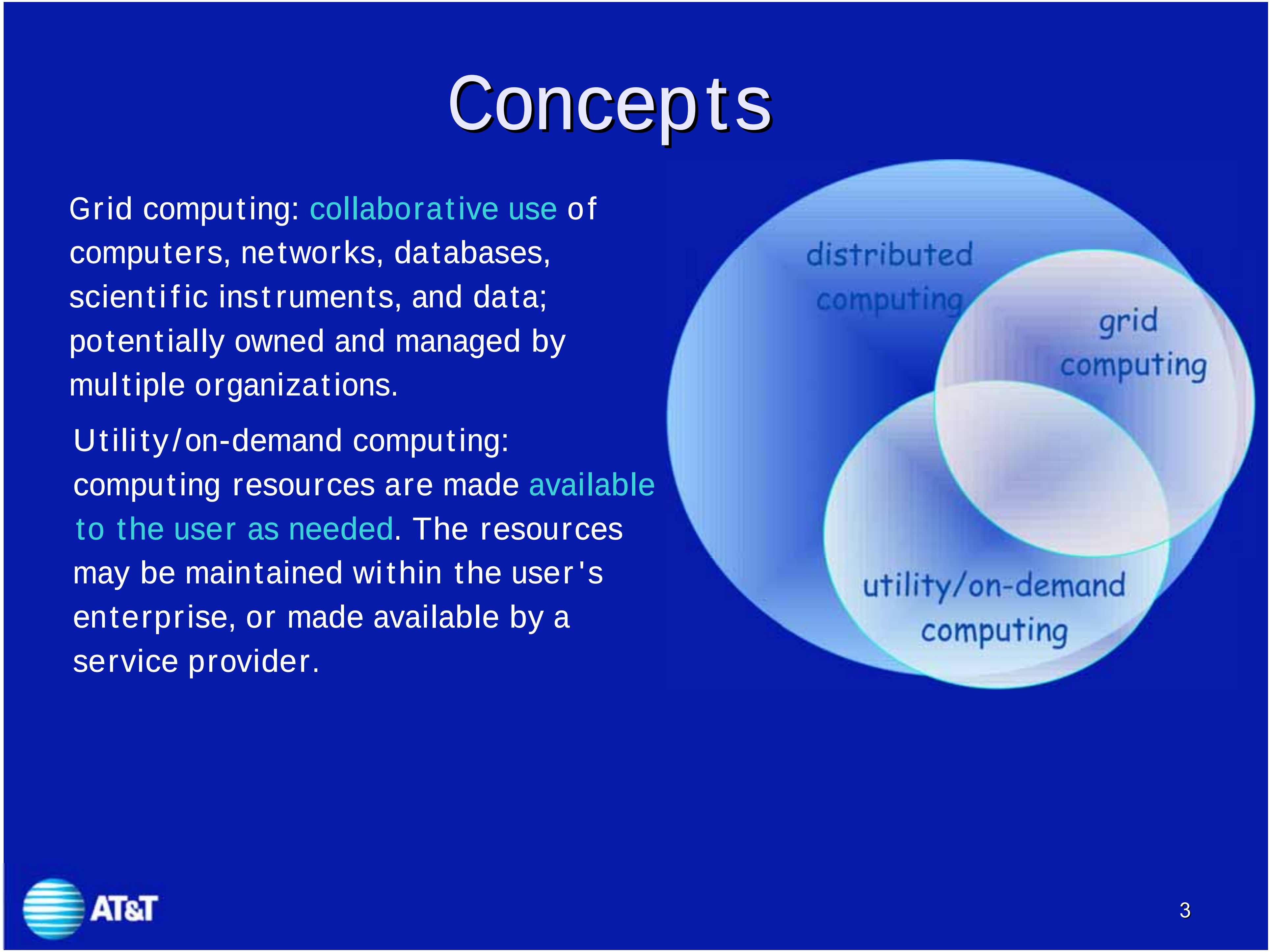


Acknowlegements:

Part of material based on Xianan Zhang, Matti Hiltunen, Keith Marzullo, Rick Schlichting, “Managing Service States According to Durability”, Draft.

Other grid-collaborators: Dr. Francois Taiani (Lancaster U), Ryoichi Ueda, Toshiyuki Moritsu (Hitachi).

Opinions expressed in this talk do not reflect those of AT&T.



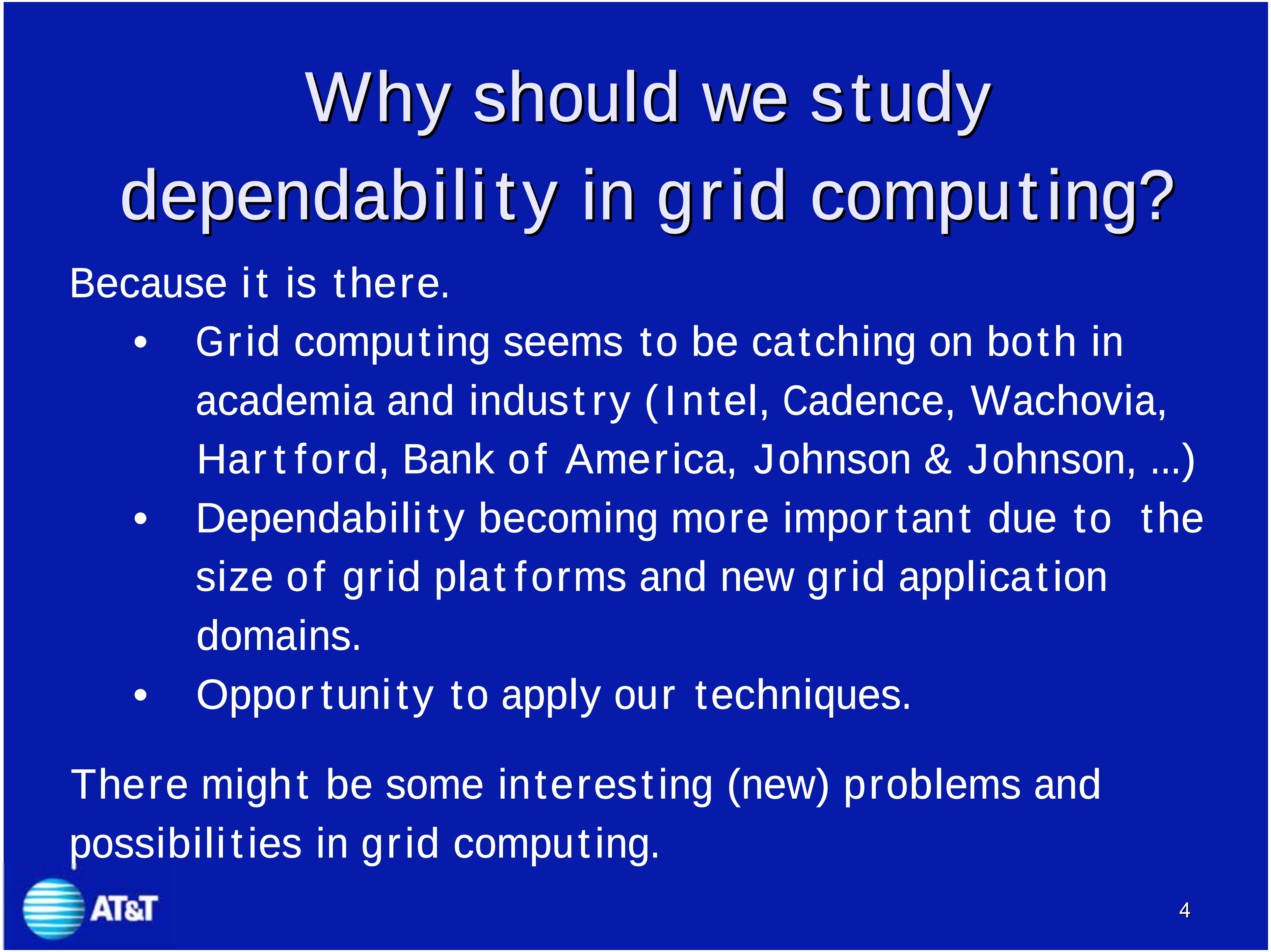
Concepts

Grid computing: collaborative use of computers, networks, databases, scientific instruments, and data; potentially owned and managed by multiple organizations.

Utility/on-demand computing: computing resources are made available to the user as needed. The resources may be maintained within the user's enterprise, or made available by a service provider.

AT&T

3

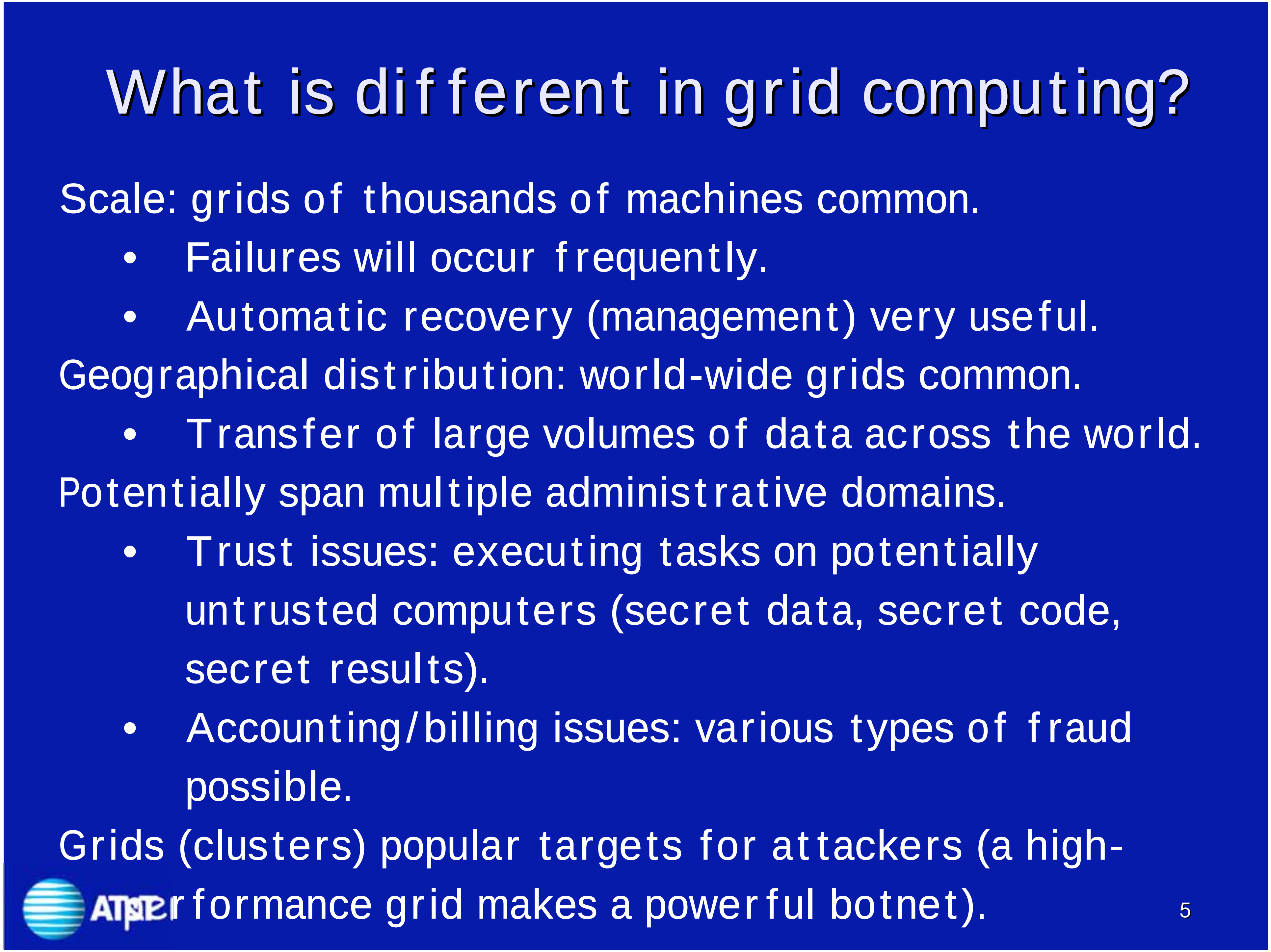


Why should we study dependability in grid computing?

Because it is there.

- Grid computing seems to be catching on both in academia and industry (Intel, Cadence, Wachovia, Hartford, Bank of America, Johnson & Johnson, ...)
- Dependability becoming more important due to the size of grid platforms and new grid application domains.
- Opportunity to apply our techniques.

There might be some interesting (new) problems and possibilities in grid computing.



What is different in grid computing?

Scale: grids of thousands of machines common.

- Failures will occur frequently.
- Automatic recovery (management) very useful.

Geographical distribution: world-wide grids common.

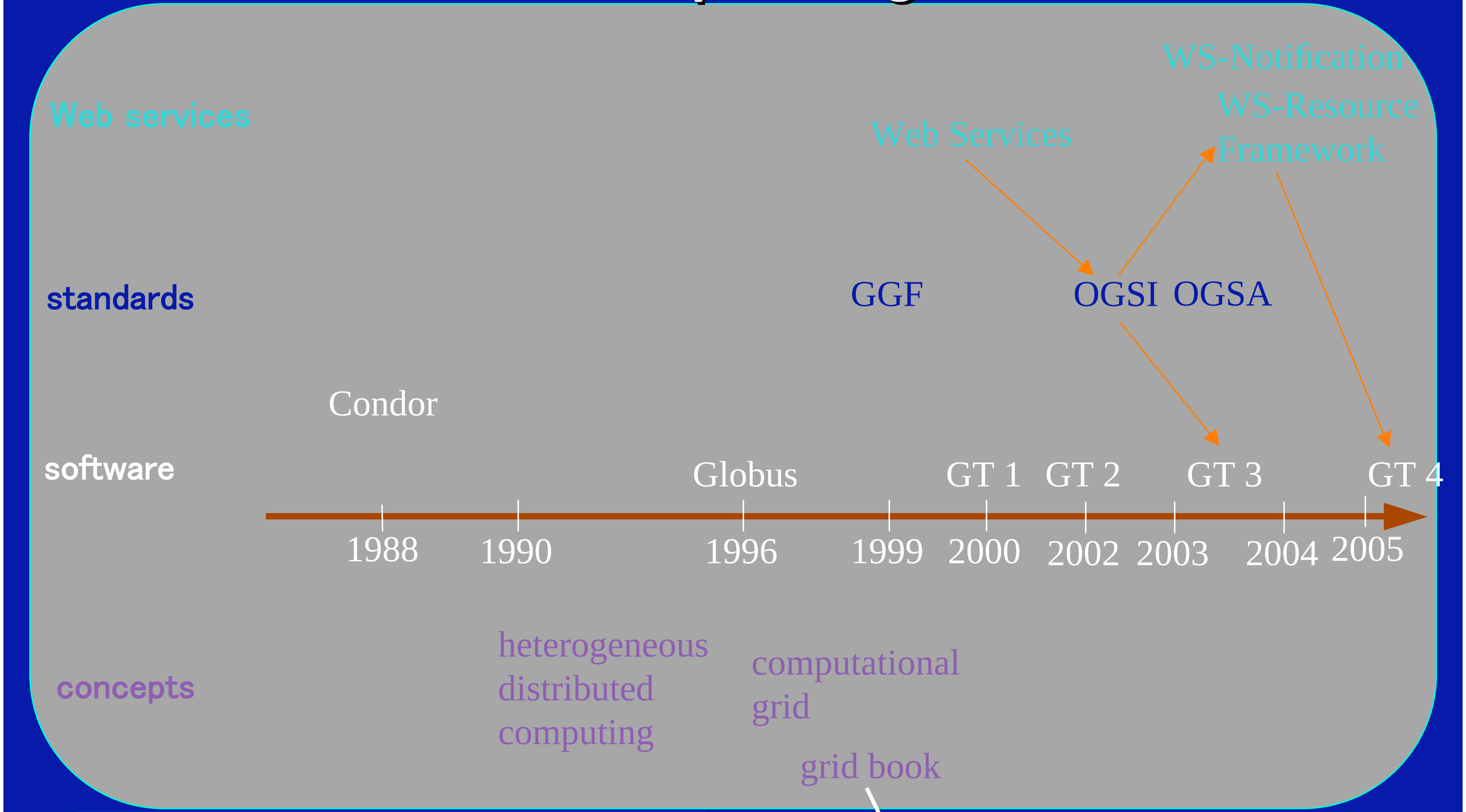
- Transfer of large volumes of data across the world.

Potentially span multiple administrative domains.

- Trust issues: executing tasks on potentially untrusted computers (secret data, secret code, secret results).
- Accounting/billing issues: various types of fraud possible.

Grids (clusters) popular targets for attackers (a high-performance grid makes a powerful botnet).

Grid computing timeline

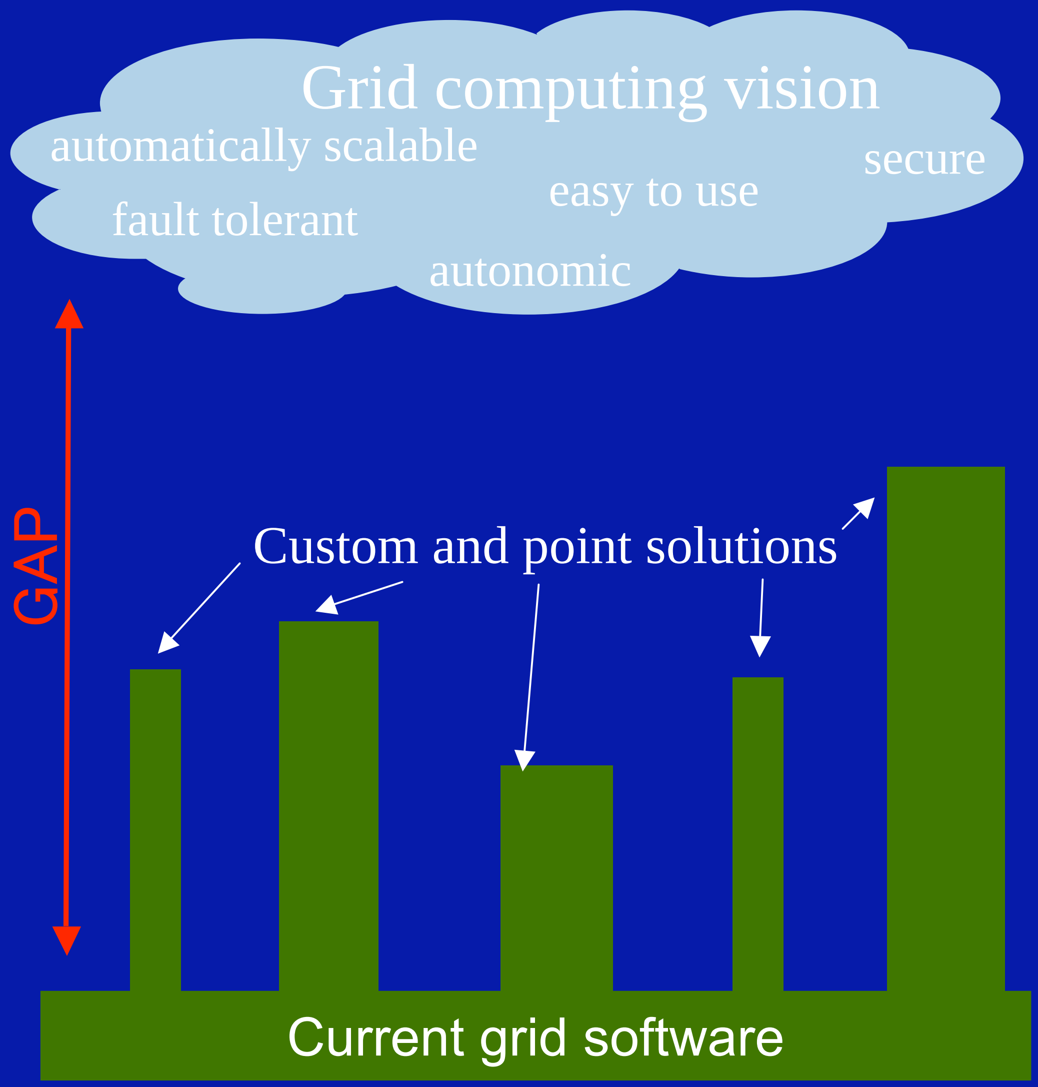


WS-Notification
WS-Resource Framework
Web services
Web Services
standards
GGF
OGSI
OGSA
Condor
software
Globus
GT 1
GT 2
GT 3
GT 4
1988
1990
1996
1999
2000
2002
2003
2004
2005
concepts
heterogeneous distributed computing
computational grid
grid book

GGF: Global Grid Forum
OGSI: Open Grid Services Infrastructure
OGSA: Open Grid Services Architecture

The Grid: Blueprint for a New Computing Infrastructure, Foster and Kesselman

Vision vs. current status



Grid computing vision

automatically scalable

fault tolerant

easy to use

secure

autonomic

GAP

Custom and point solutions

Current grid software

Why should we care about standards?

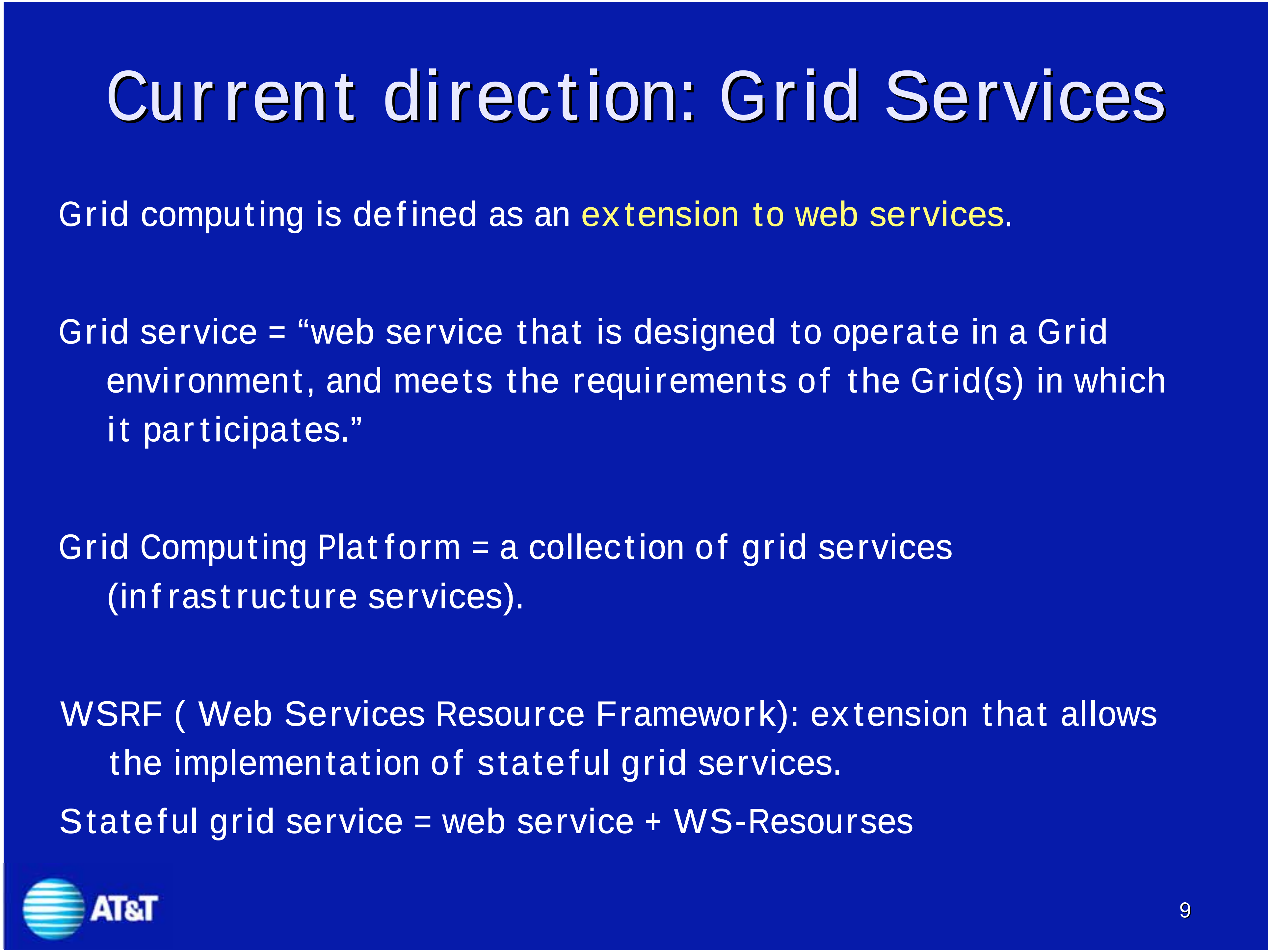
The concept of grid computing is not based on, or require, any standards.

However ..

- Interoperability requires standards (can your grid platform talk to mine).
- Commercial users of grid computing demand standard compliance to avoid locking in with one vendor.
- Basing your work on existing standards and existing implementations can speed up your work (do not need to implement everything from scratch – just the parts that you are interested in).
- Publishability (think transport protocols vs. TCP).

Opportunity for impact:

- The specifications at GGF are still in early stages – it is still possible (easy) to define or refine these specifications.
- It is possible to add your pieces into open source grid platforms such as Globus.



Current direction: Grid Services

Grid computing is defined as an extension to web services.

Grid service = “web service that is designed to operate in a Grid environment, and meets the requirements of the Grid(s) in which it participates.”

Grid Computing Platform = a collection of grid services (infrastructure services).

WSRF (Web Services Resource Framework): extension that allows the implementation of stateful grid services.

Stateful grid service = web service + WS-Resourses

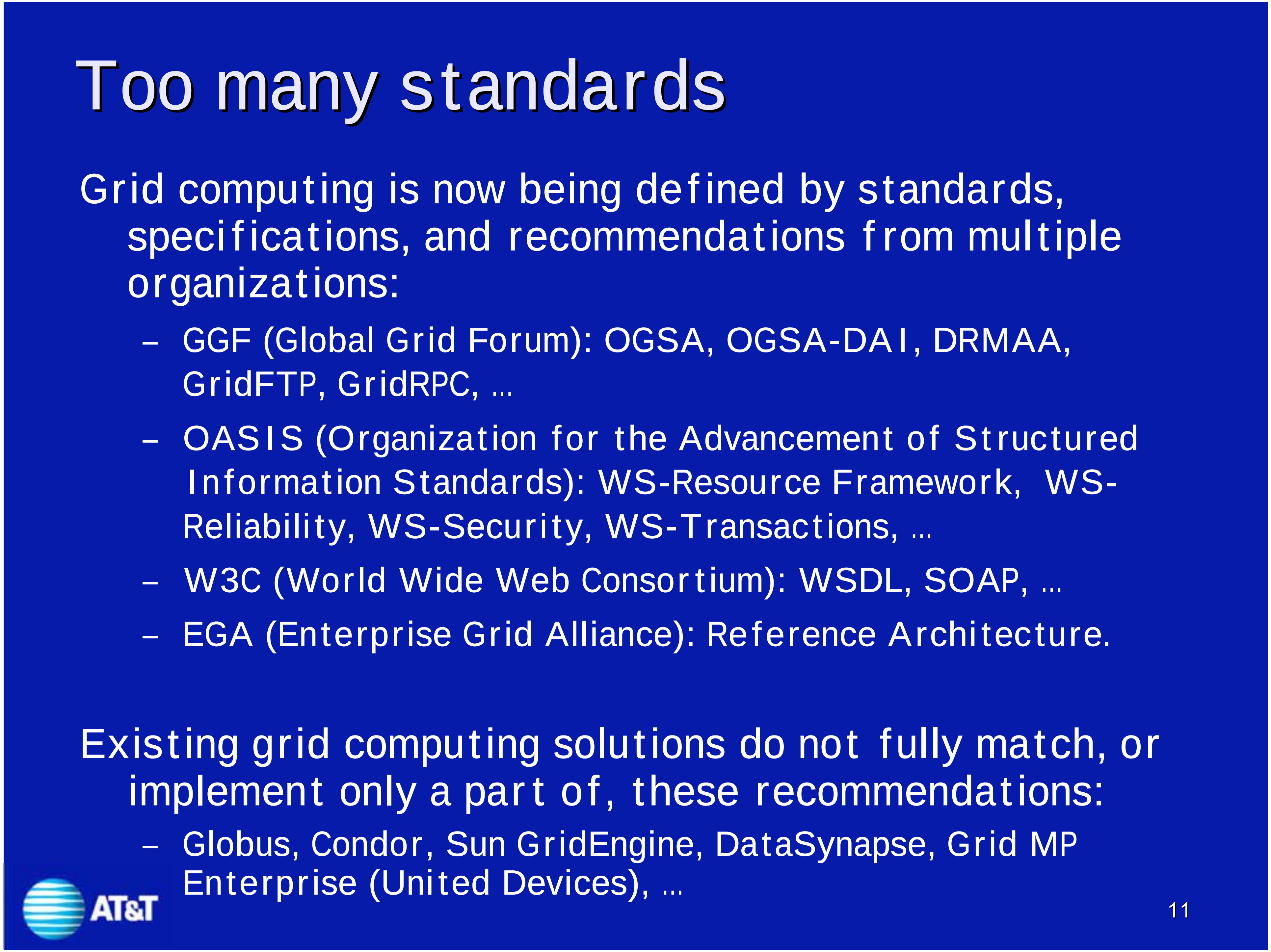
Is this a good idea?

Positives:

- Can leverage existing web service platforms and web service standards.
- Ride on the popularity of web services – easier acceptance.

Negatives:

- Large performance impact (response time from 100+ms to 10s of seconds for trivial grid services in Globus 3.9.4).
  - Note that web service protocols are only needed for interaction between different grid services (not between nodes in a grid application).
- Complexity of the resulting grid middleware (number of layers).
- WS specifications are still evolving and competing.



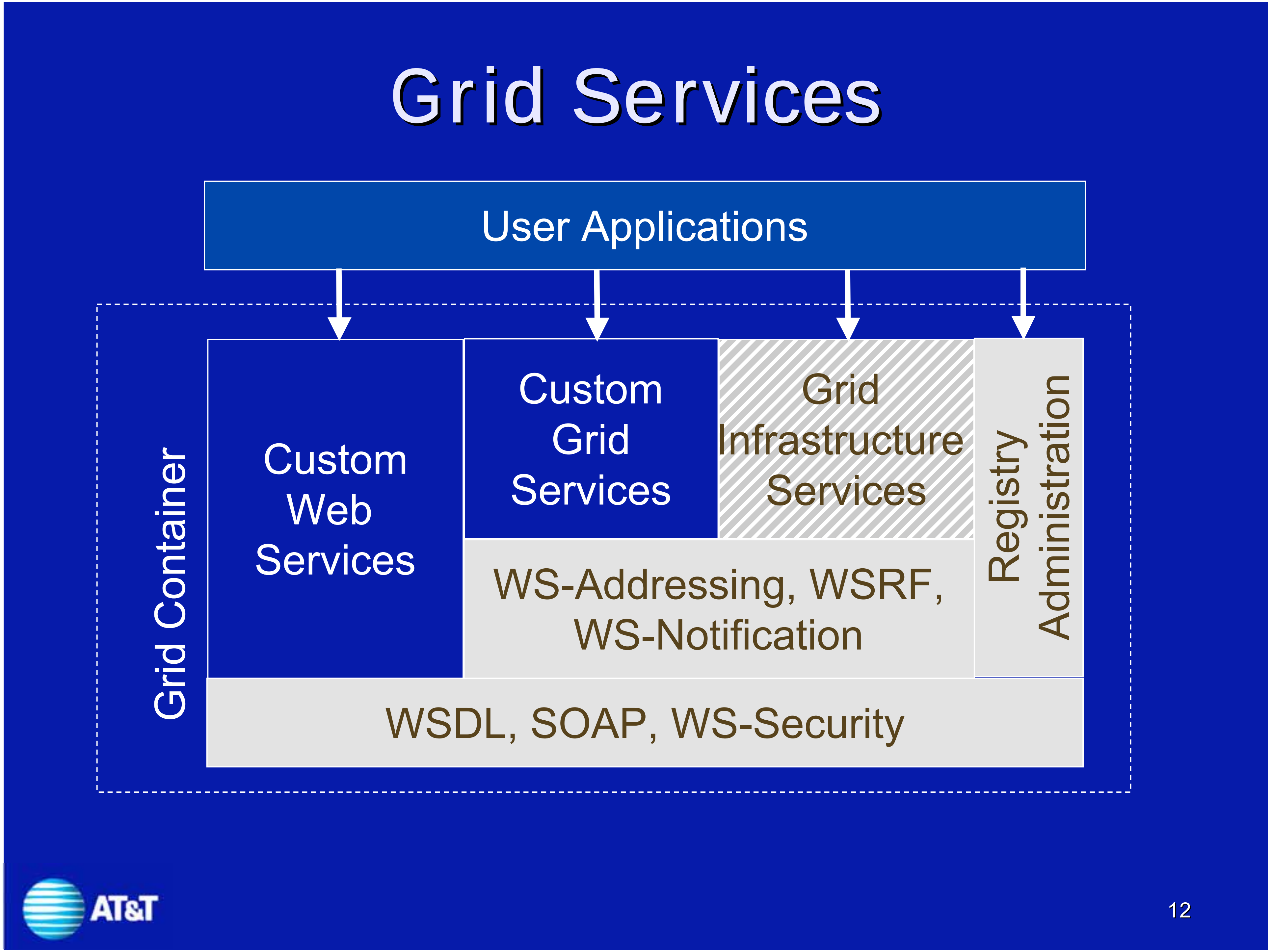
Too many standards

Grid computing is now being defined by standards, specifications, and recommendations from multiple organizations:

- GGF (Global Grid Forum): OGSA, OGSA-DAI, DRMAA, GridFTP, GridRPC, ...
- OASIS (Organization for the Advancement of Structured Information Standards): WS-Resource Framework, WS-Reliability, WS-Security, WS-Transactions, ...
- W3C (World Wide Web Consortium): WSDL, SOAP, ...
- EGA (Enterprise Grid Alliance): Reference Architecture.

Existing grid computing solutions do not fully match, or implement only a part of, these recommendations:

- Globus, Condor, Sun GridEngine, DataSynapse, Grid MP Enterprise (United Devices), ...



Grid Services

User Applications

Grid Container

Custom Web Services

Custom Grid Services

Grid Infrastructure Services

Registry Administration

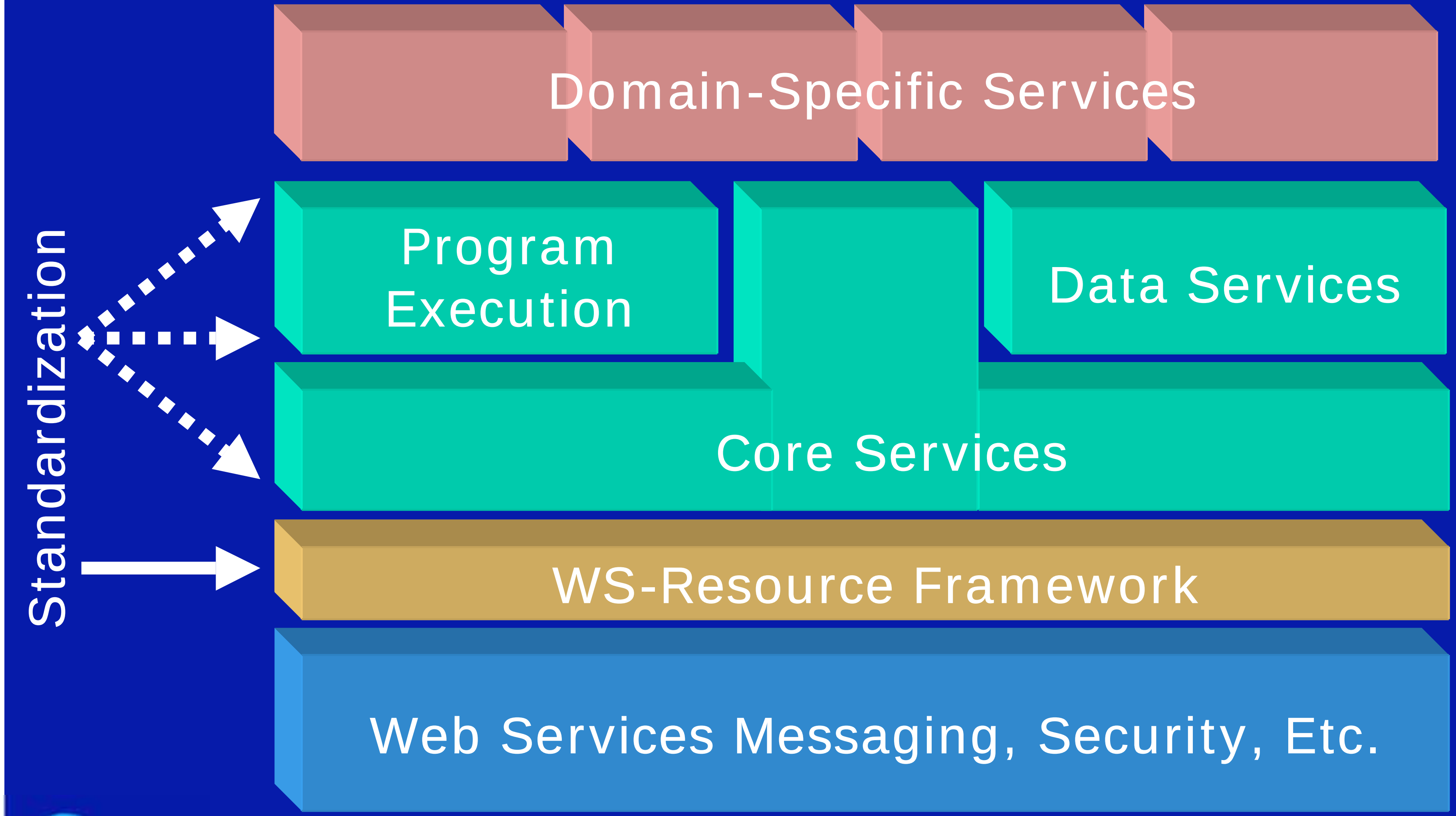
WS-Addressing, WSRF, WS-Notification

WSDL, SOAP, WS-Security

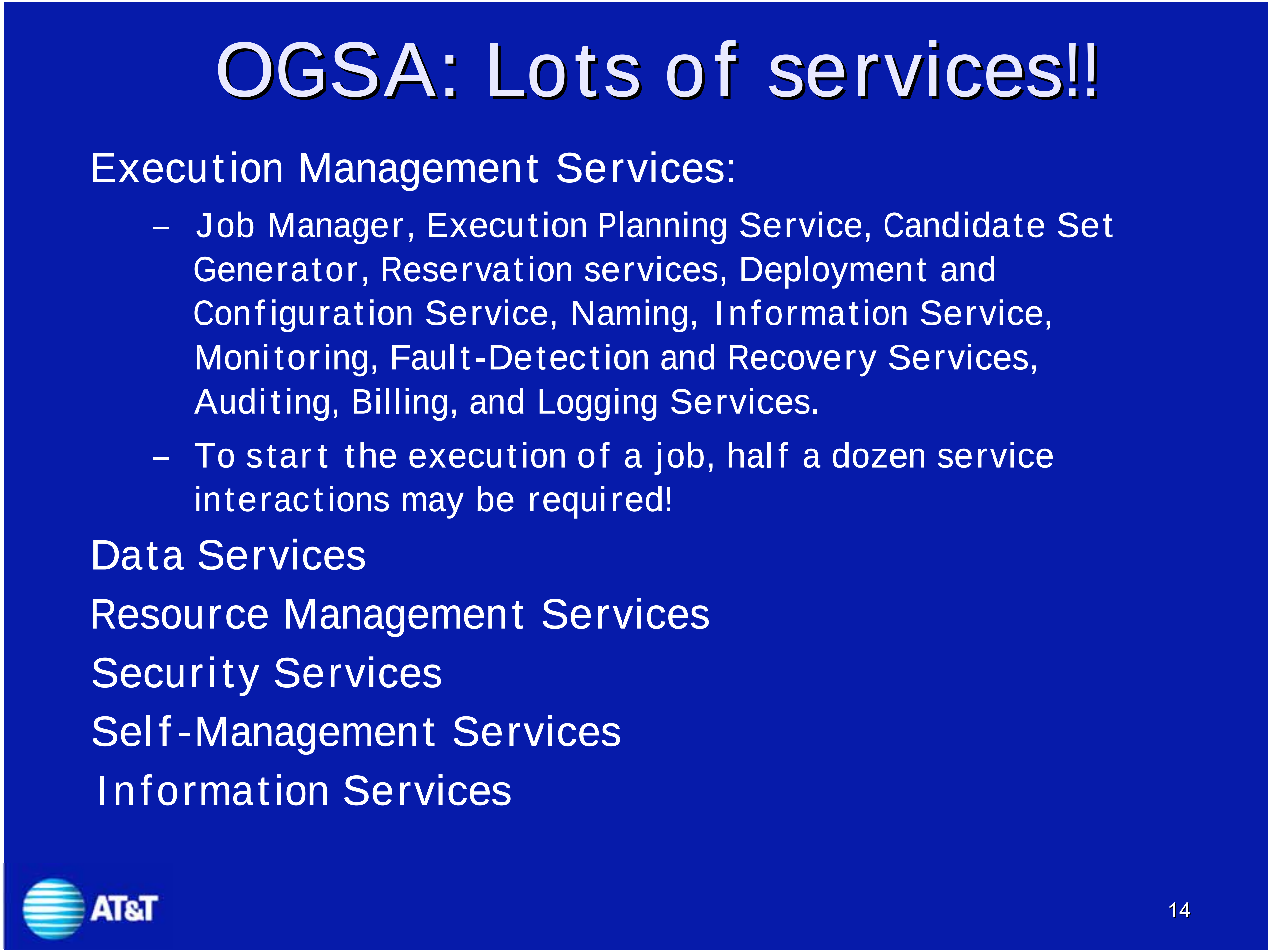
AT&T

12

Open Grid Services Architecture



AT&T



OGSA: Lots of services!!

Execution Management Services:

- Job Manager, Execution Planning Service, Candidate Set Generator, Reservation services, Deployment and Configuration Service, Naming, Information Service, Monitoring, Fault-Detection and Recovery Services, Auditing, Billing, and Logging Services.
- To start the execution of a job, half a dozen service interactions may be required!

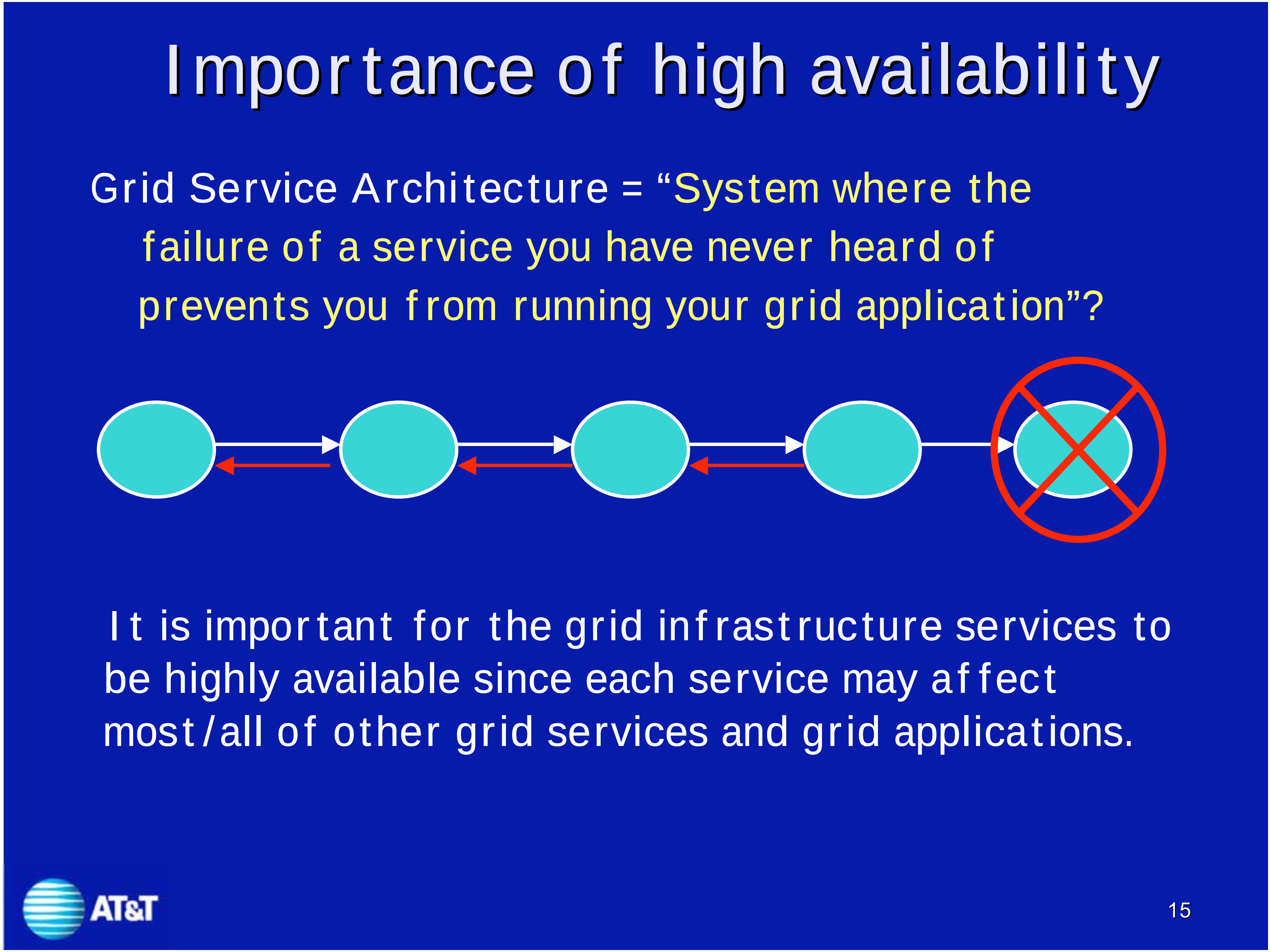
Data Services

Resource Management Services

Security Services

Self-Management Services

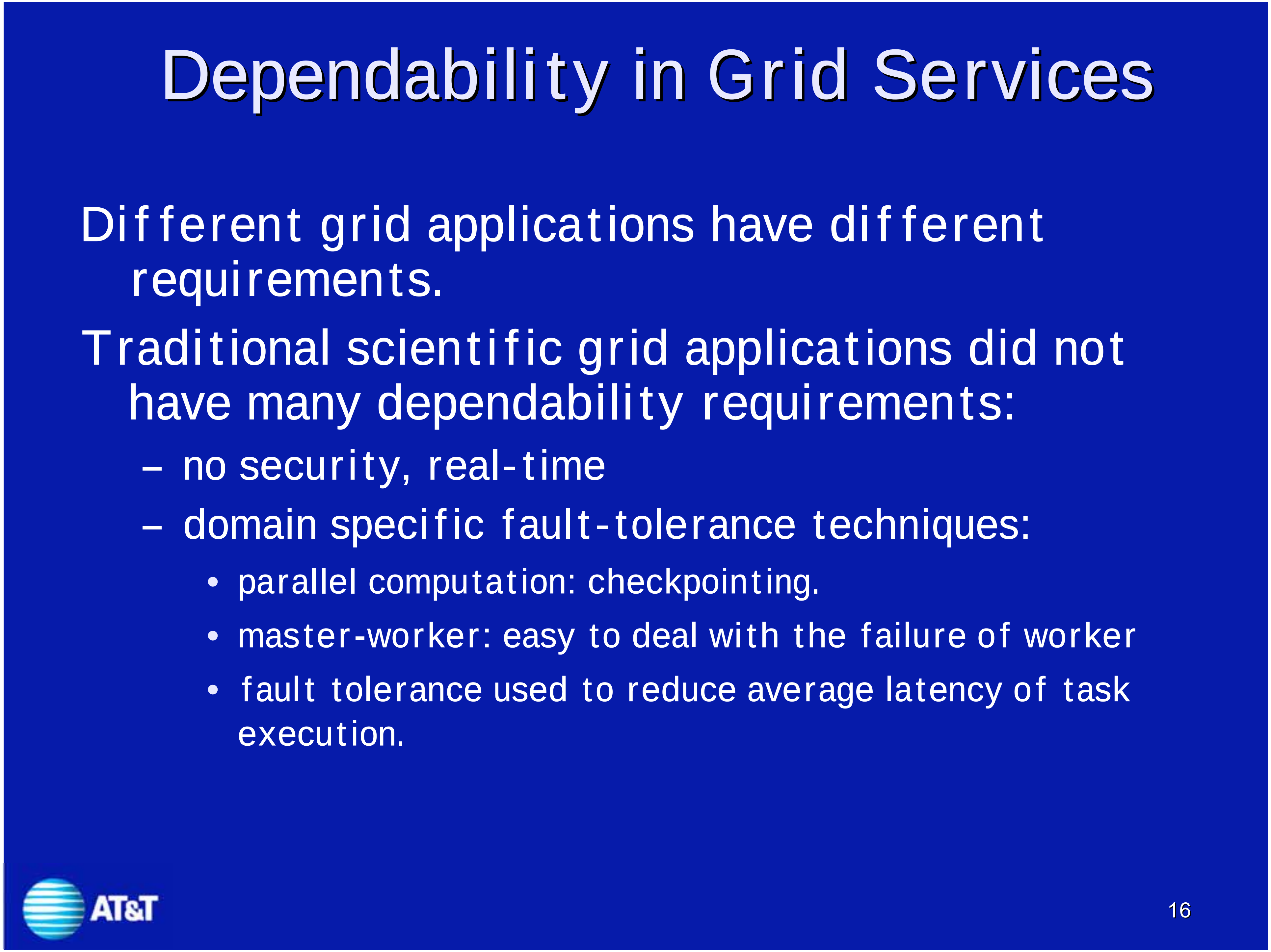
Information Services



Importance of high availability

Grid Service Architecture = “System where the failure of a service you have never heard of prevents you from running your grid application”?

It is important for the grid infrastructure services to be highly available since each service may affect most/all of other grid services and grid applications.

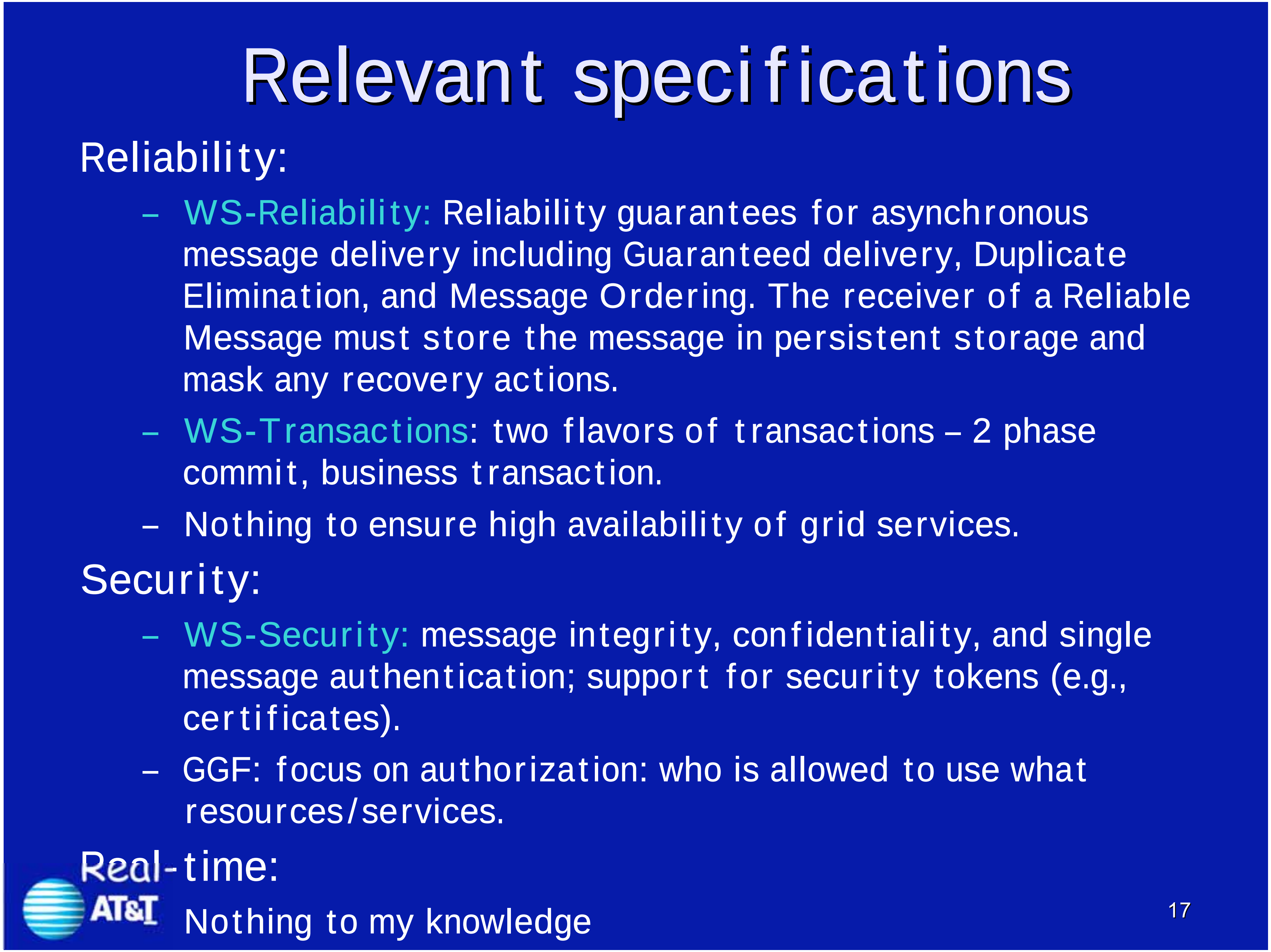


Dependability in Grid Services

Different grid applications have different requirements.

Traditional scientific grid applications did not have many dependability requirements:

- no security, real-time
- domain specific fault-tolerance techniques:
  - parallel computation: checkpointing.
  - master-worker: easy to deal with the failure of worker
  - fault tolerance used to reduce average latency of task execution.



Relevant specifications

Reliability:

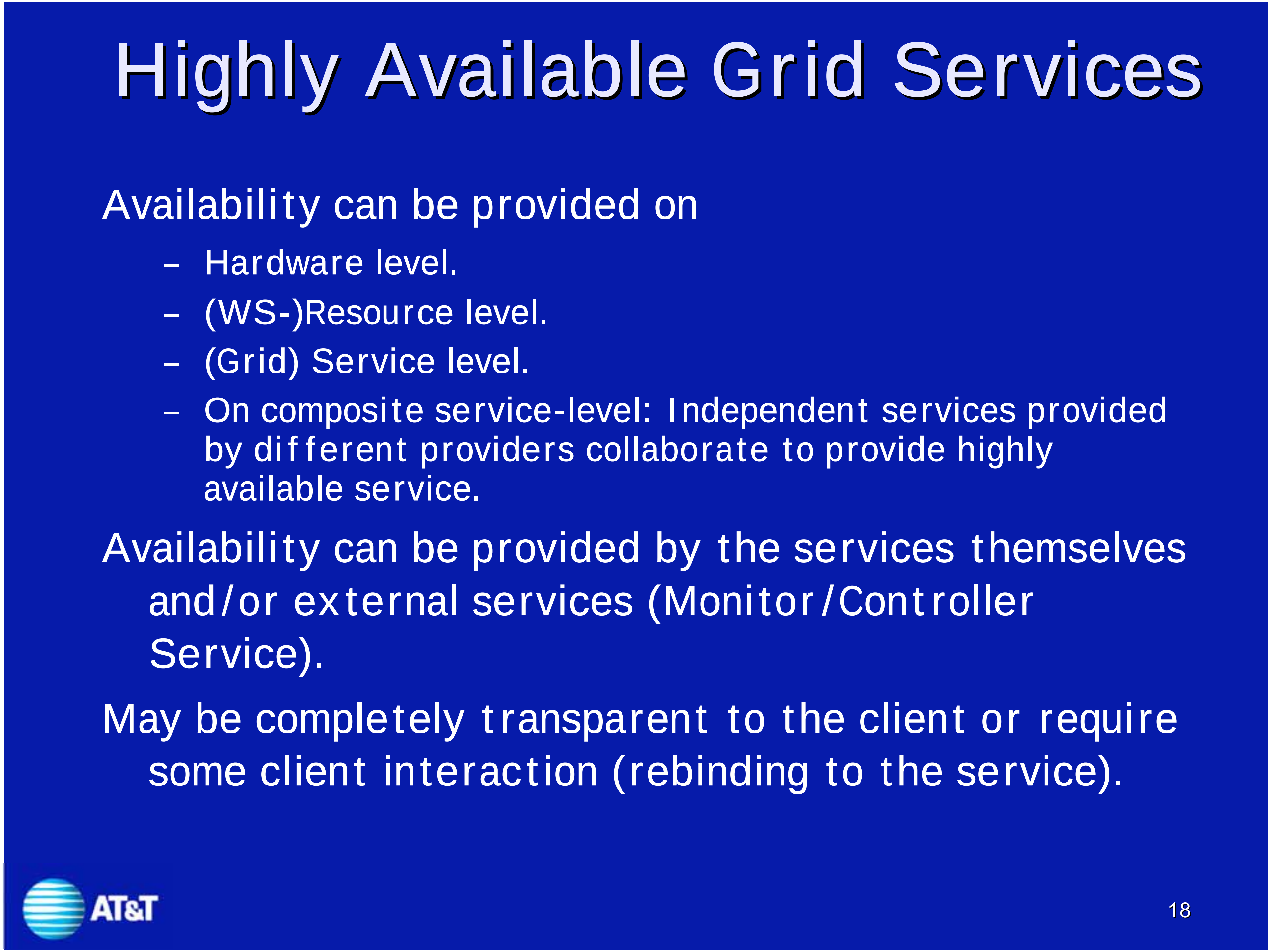
- WS-Reliability: Reliability guarantees for asynchronous message delivery including Guaranteed delivery, Duplicate Elimination, and Message Ordering. The receiver of a Reliable Message must store the message in persistent storage and mask any recovery actions.
- WS-Transactions: two flavors of transactions – 2 phase commit, business transaction.
- Nothing to ensure high availability of grid services.

Security:

- WS-Security: message integrity, confidentiality, and single message authentication; support for security tokens (e.g., certificates).
- GGF: focus on authorization: who is allowed to use what resources/services.

Real-time:

Nothing to my knowledge



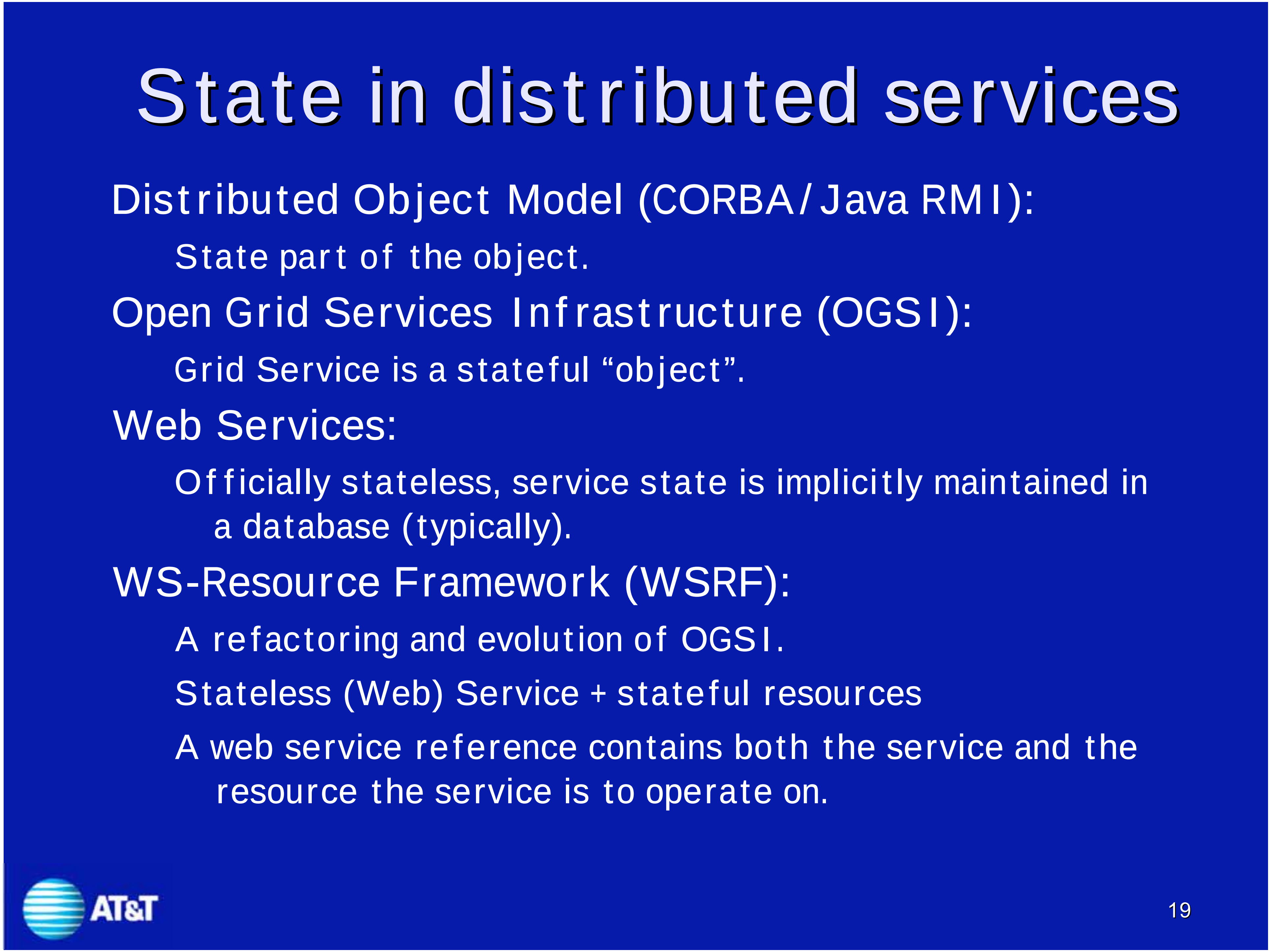
Highly Available Grid Services

Availability can be provided on

- Hardware level.
- (WS-)Resource level.
- (Grid) Service level.
- On composite service-level: Independent services provided by different providers collaborate to provide highly available service.

Availability can be provided by the services themselves and/or external services (Monitor/Controller Service).

May be completely transparent to the client or require some client interaction (rebinding to the service).



State in distributed services

Distributed Object Model (CORBA/Java RMI):

- State part of the object.

Open Grid Services Infrastructure (OGSI):

- Grid Service is a stateful “object”.

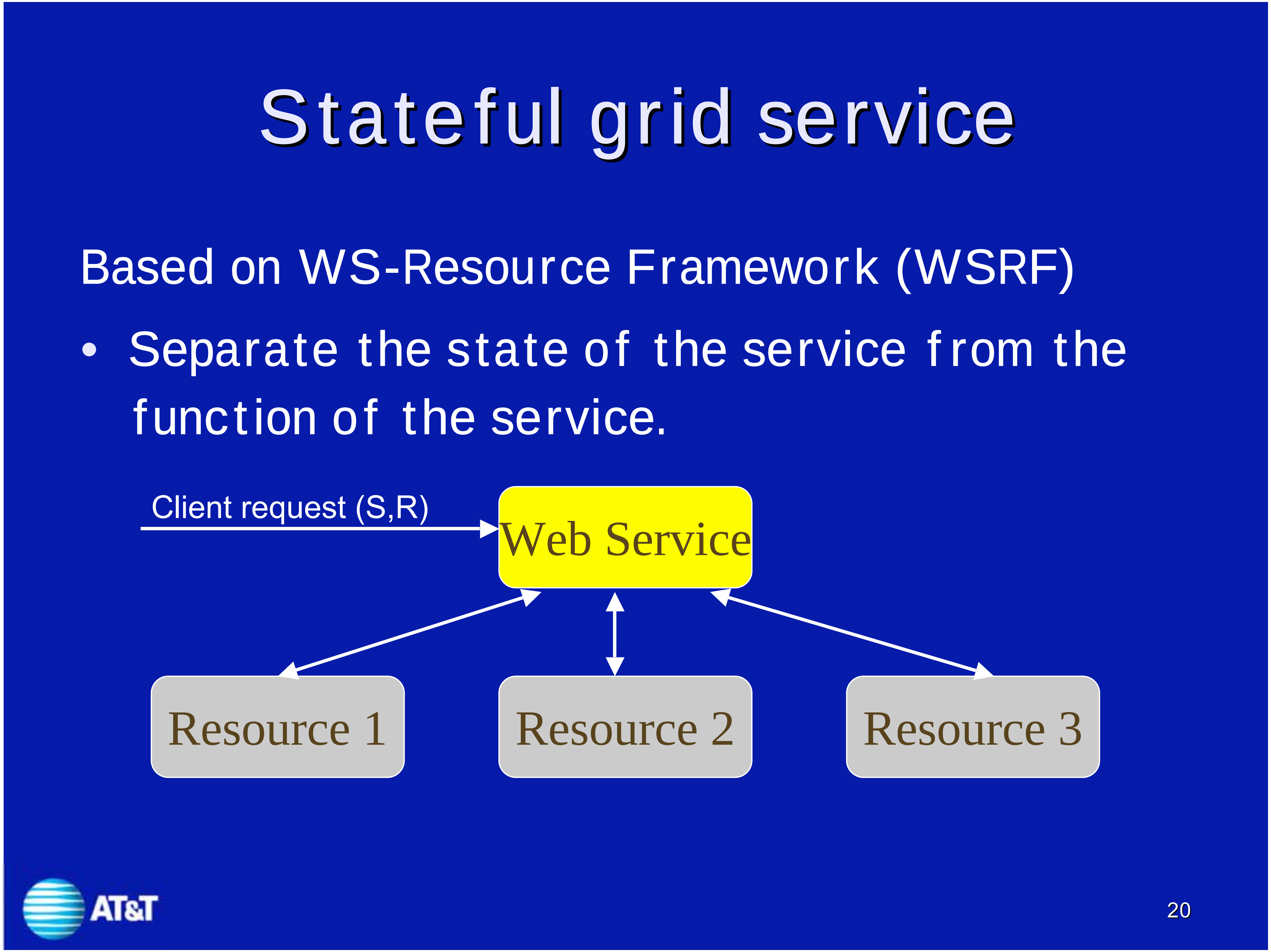
Web Services:

- Officially stateless, service state is implicitly maintained in a database (typically).

WS-Resource Framework (WSRF):

- A refactoring and evolution of OGSI.
- Stateless (Web) Service + stateful resources
- A web service reference contains both the service and the resource the service is to operate on.

AT&T

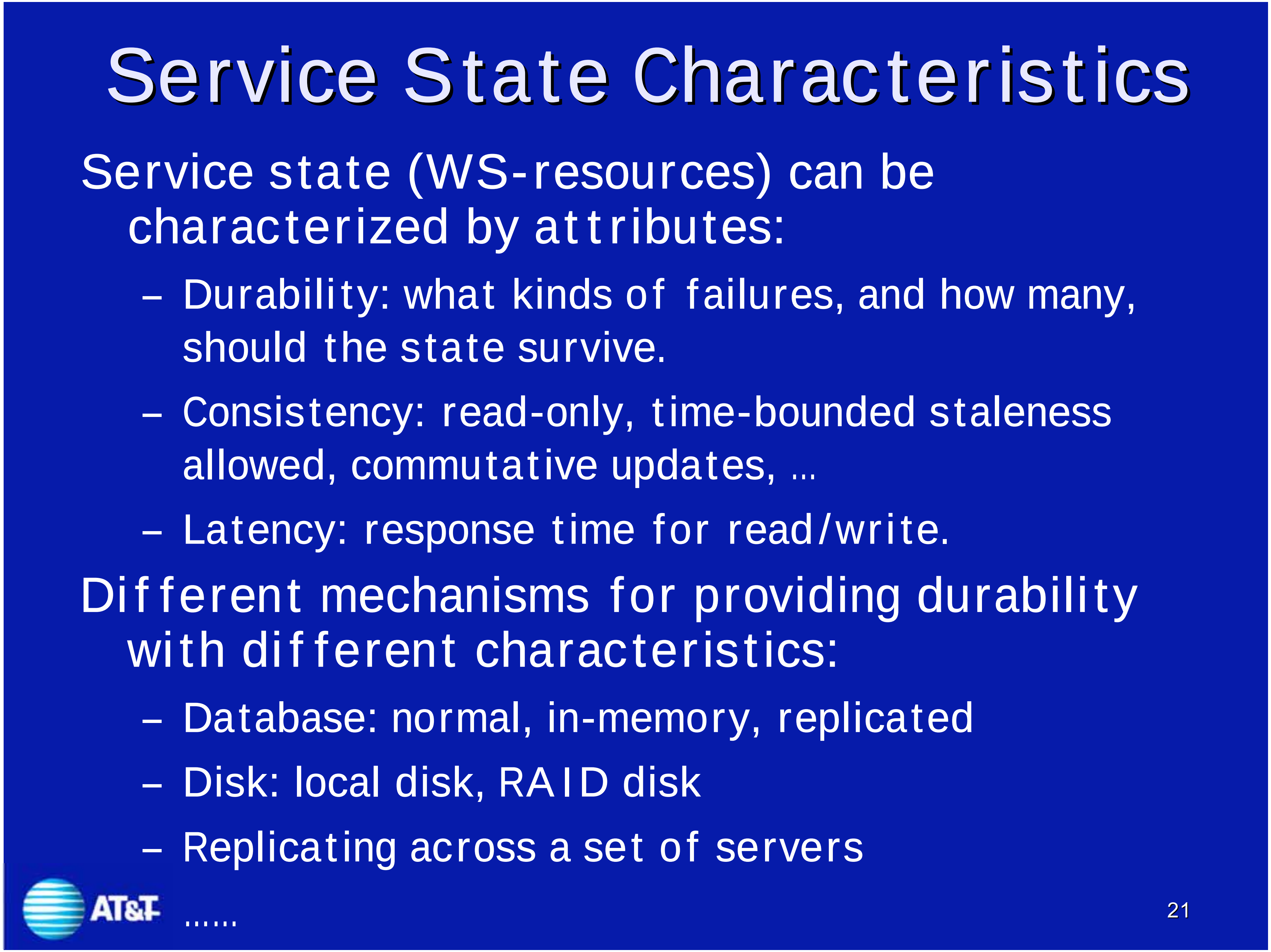


Stateful grid service

Based on WS-Resource Framework (WSRF)

- Separate the state of the service from the function of the service.

20



Service State Characteristics

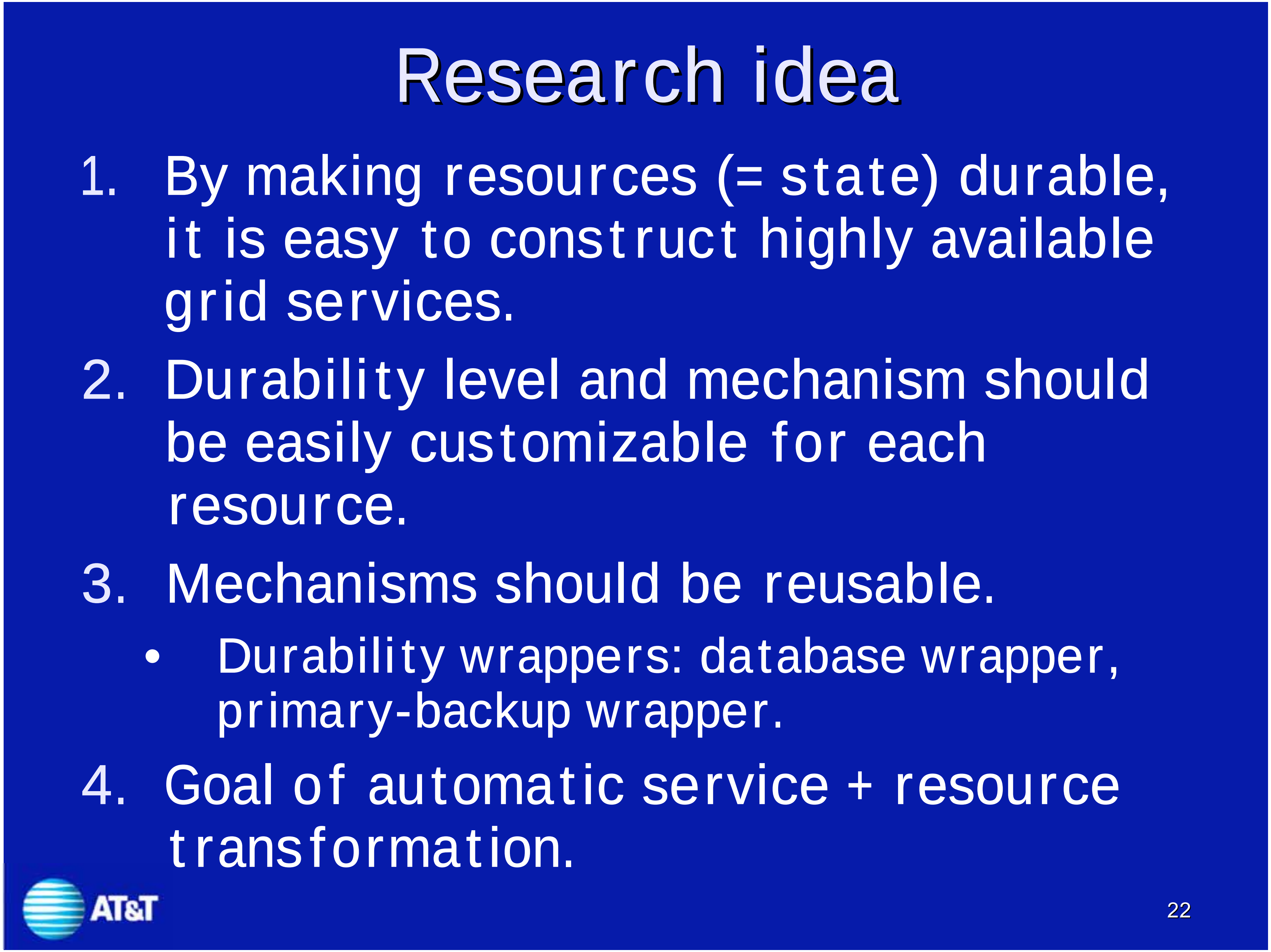
Service state (WS-resources) can be characterized by attributes:

- Durability: what kinds of failures, and how many, should the state survive.
- Consistency: read-only, time-bounded staleness allowed, commutative updates, …
- Latency: response time for read/write.

Different mechanisms for providing durability with different characteristics:

- Database: normal, in-memory, replicated
- Disk: local disk, RAID disk
- Replicating across a set of servers

……

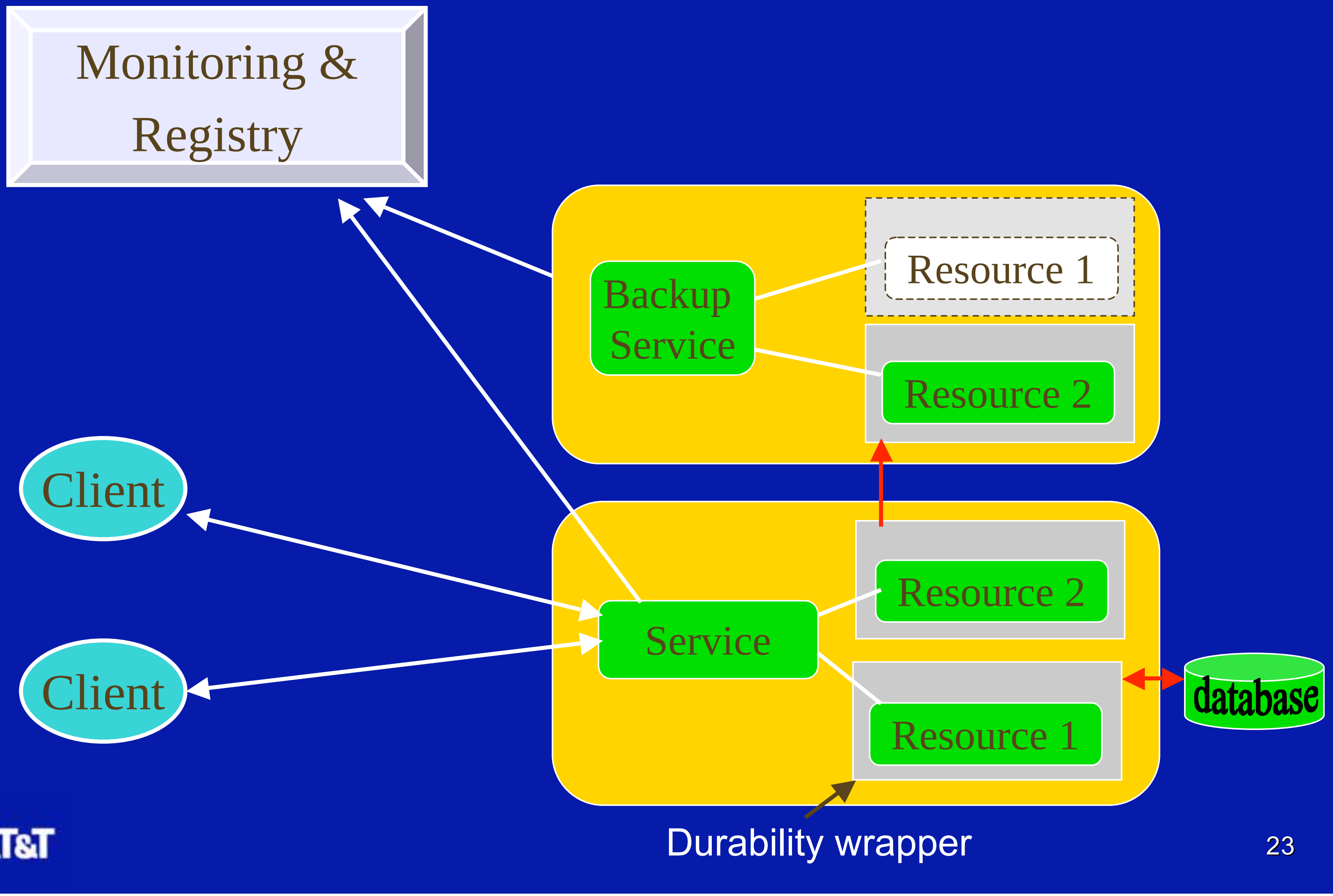


Research idea

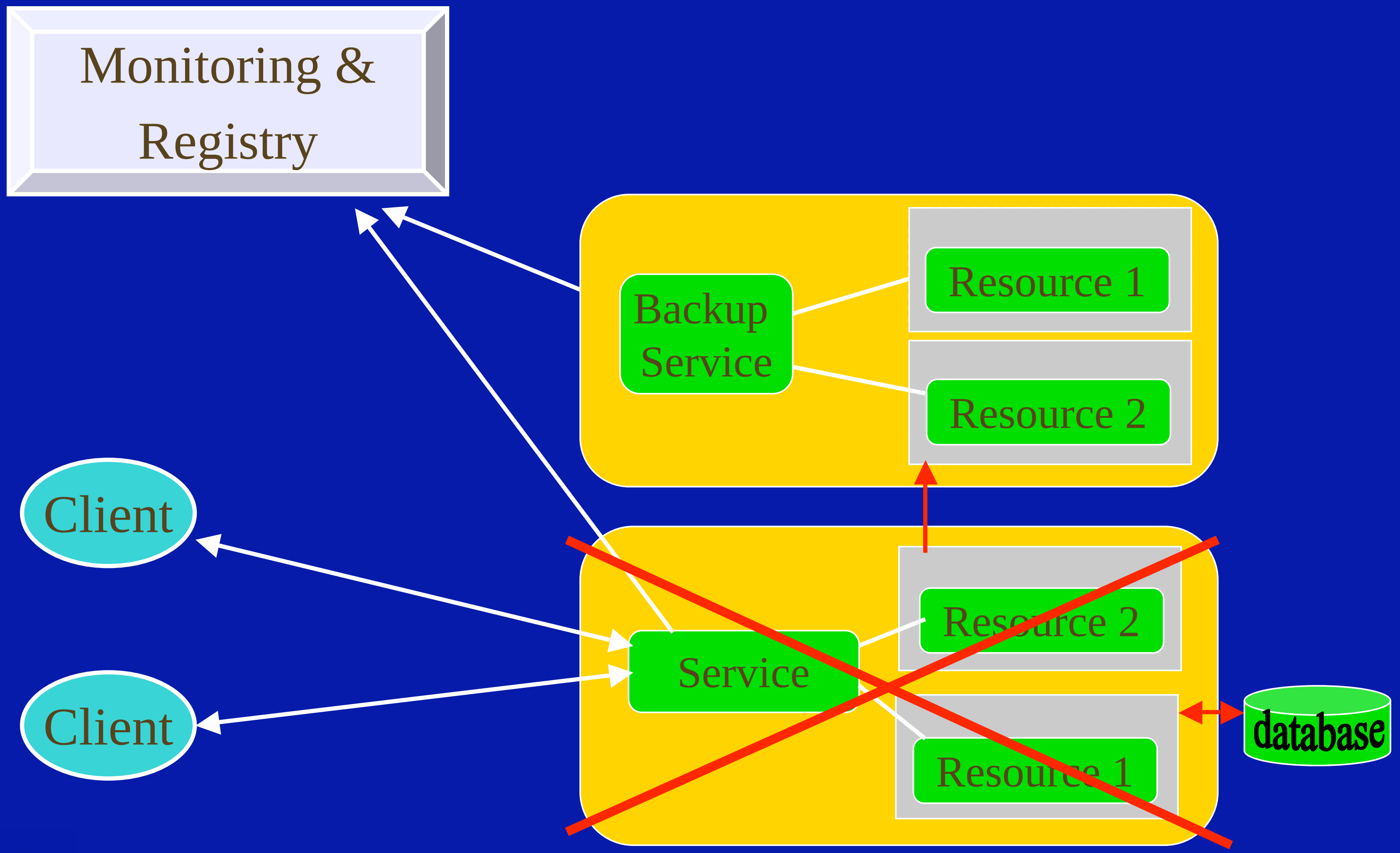
1. By making resources (= state) durable, it is easy to construct highly available grid services.
2. Durability level and mechanism should be easily customizable for each resource.
3. Mechanisms should be reusable.
   - Durability wrappers: database wrapper, primary-backup wrapper.
4. Goal of automatic service + resource transformation.

AT&T

Proposed Architecture



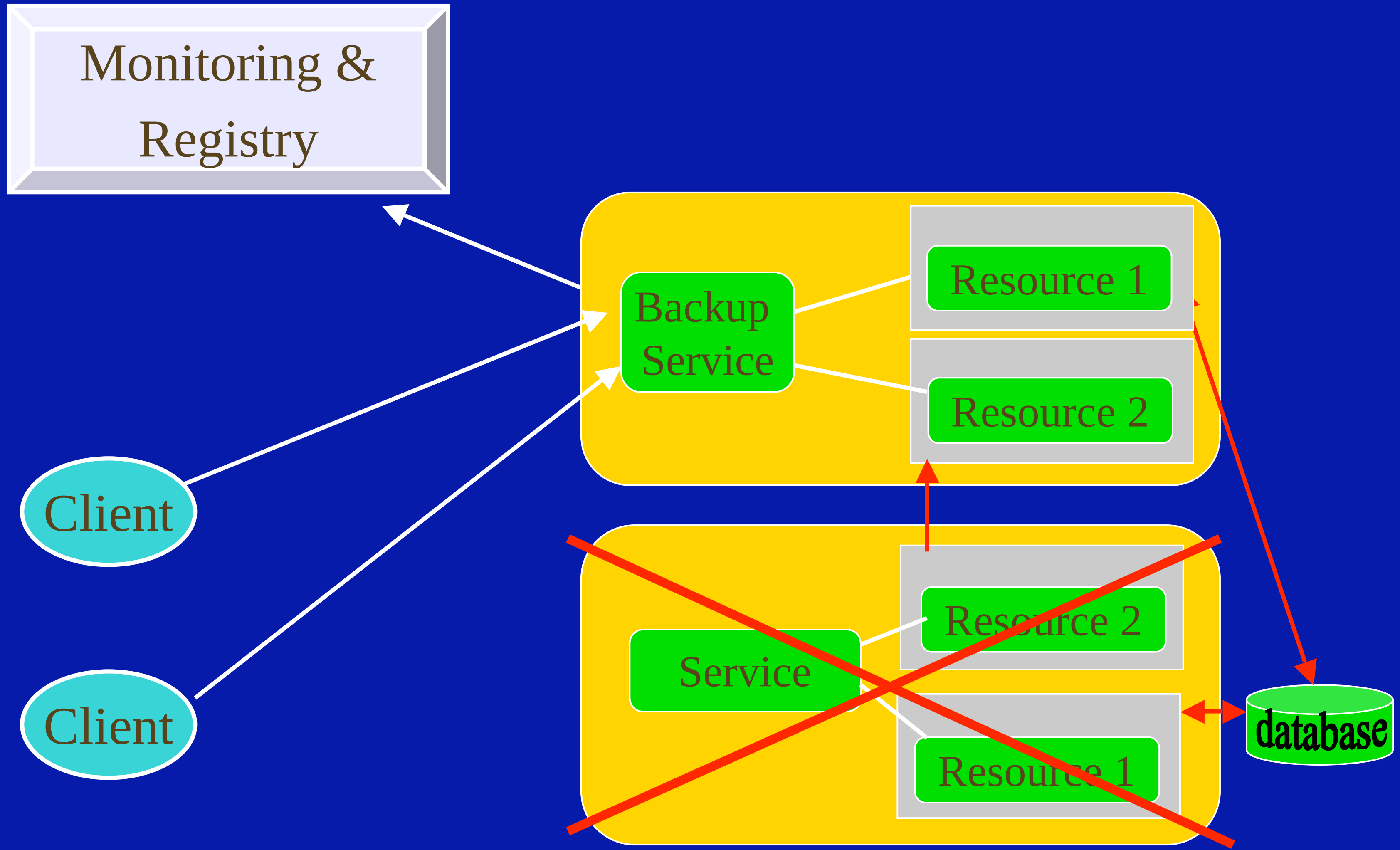
Failure

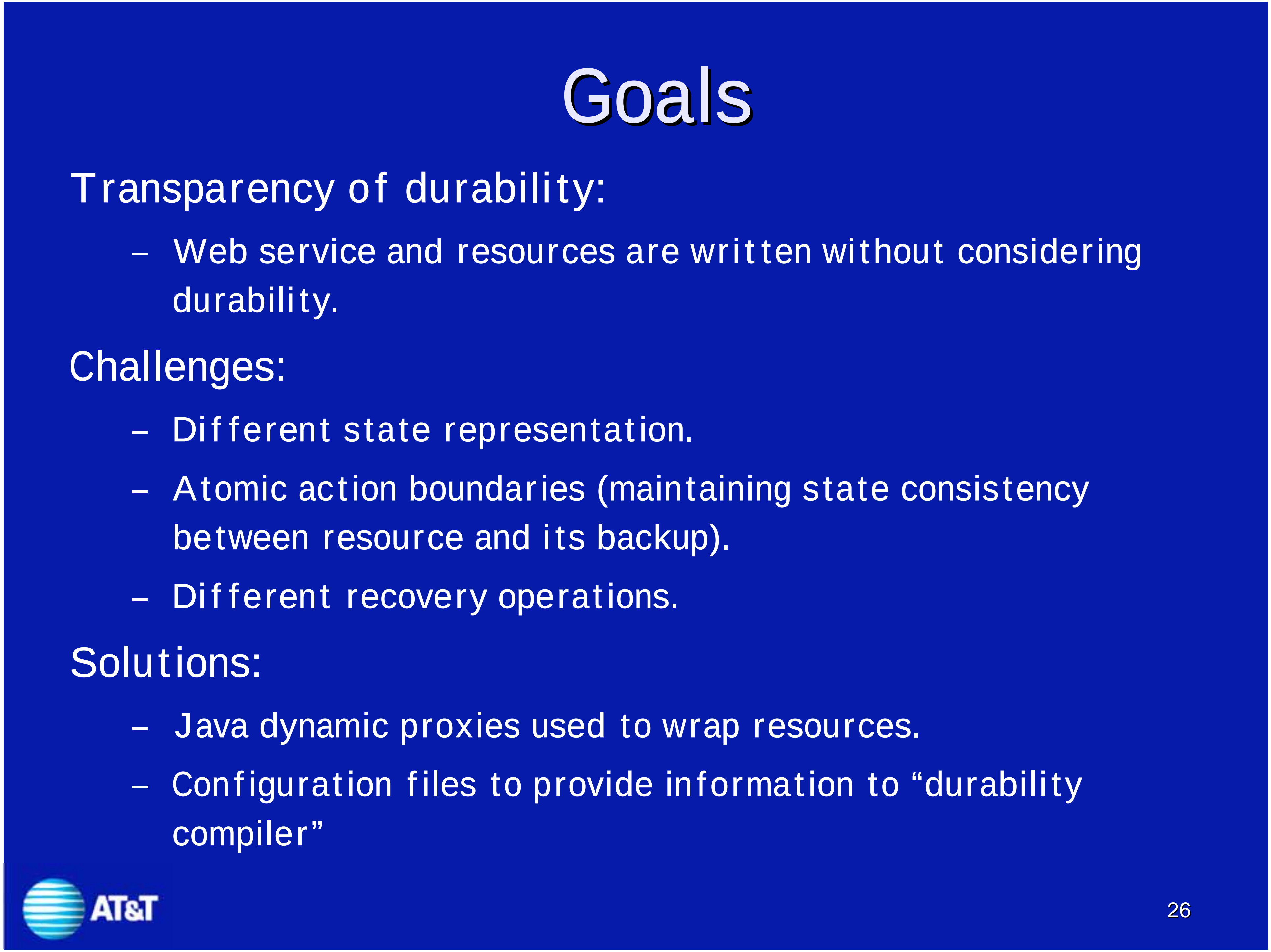


AT&T

24

Recovery


Monitoring & Registry
Backup Service
Resource 1
Resource 2
Client
Client
Service
Resource 2
Resource 1
database



Goals

Transparency of durability:

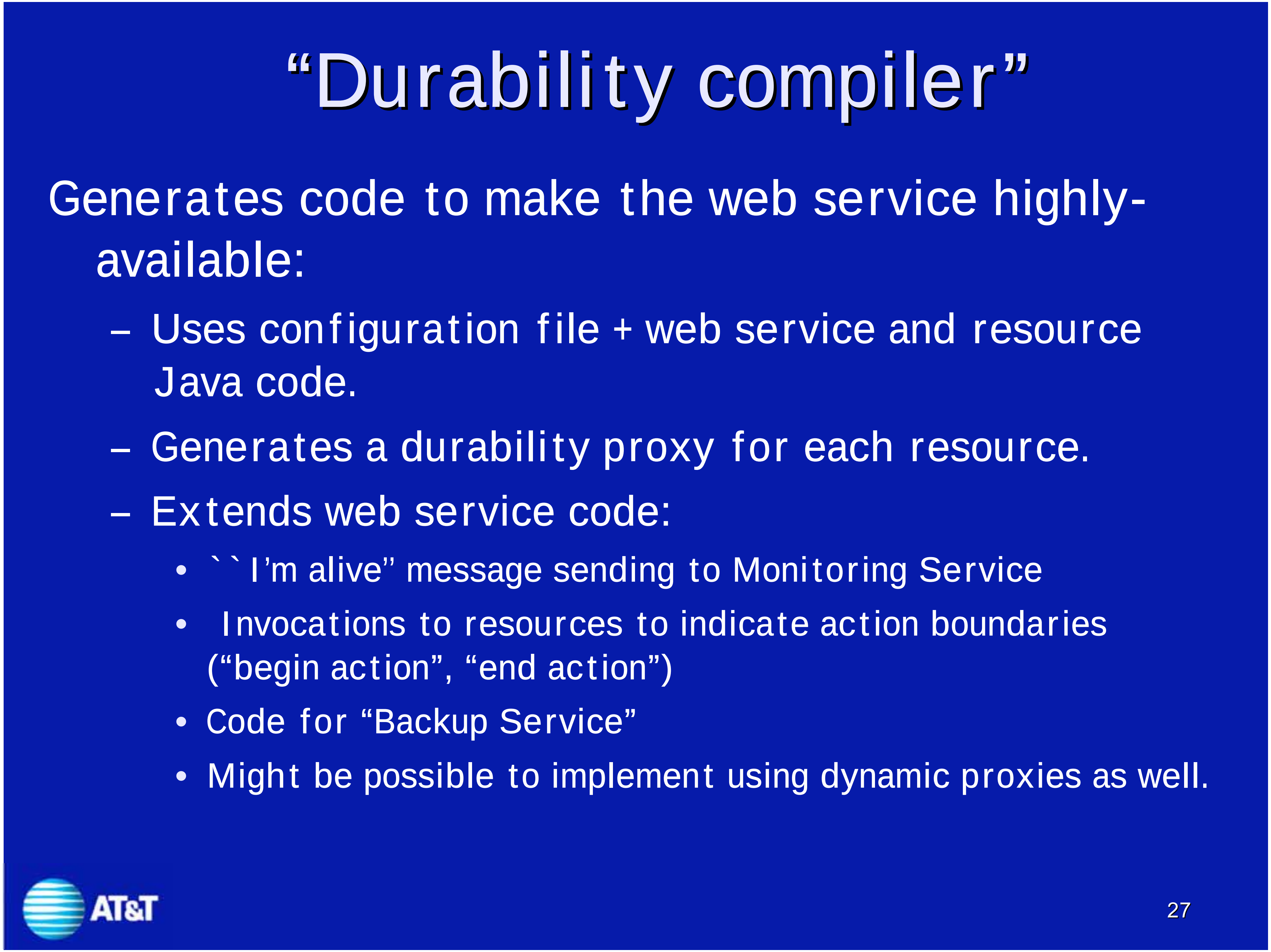
- Web service and resources are written without considering durability.

Challenges:

- Different state representation.
- Atomic action boundaries (maintaining state consistency between resource and its backup).
- Different recovery operations.

Solutions:

- Java dynamic proxies used to wrap resources.
- Configuration files to provide information to “durability compiler”



“Durability compiler”

Generates code to make the web service highly-available:

- Uses configuration file + web service and resource Java code.
- Generates a durability proxy for each resource.
- Extends web service code:
  - ``I’m alive’’ message sending to Monitoring Service
  - Invocations to resources to indicate action boundaries (“begin action”, “end action”)
  - Code for “Backup Service”
  - Might be possible to implement using dynamic proxies as well.

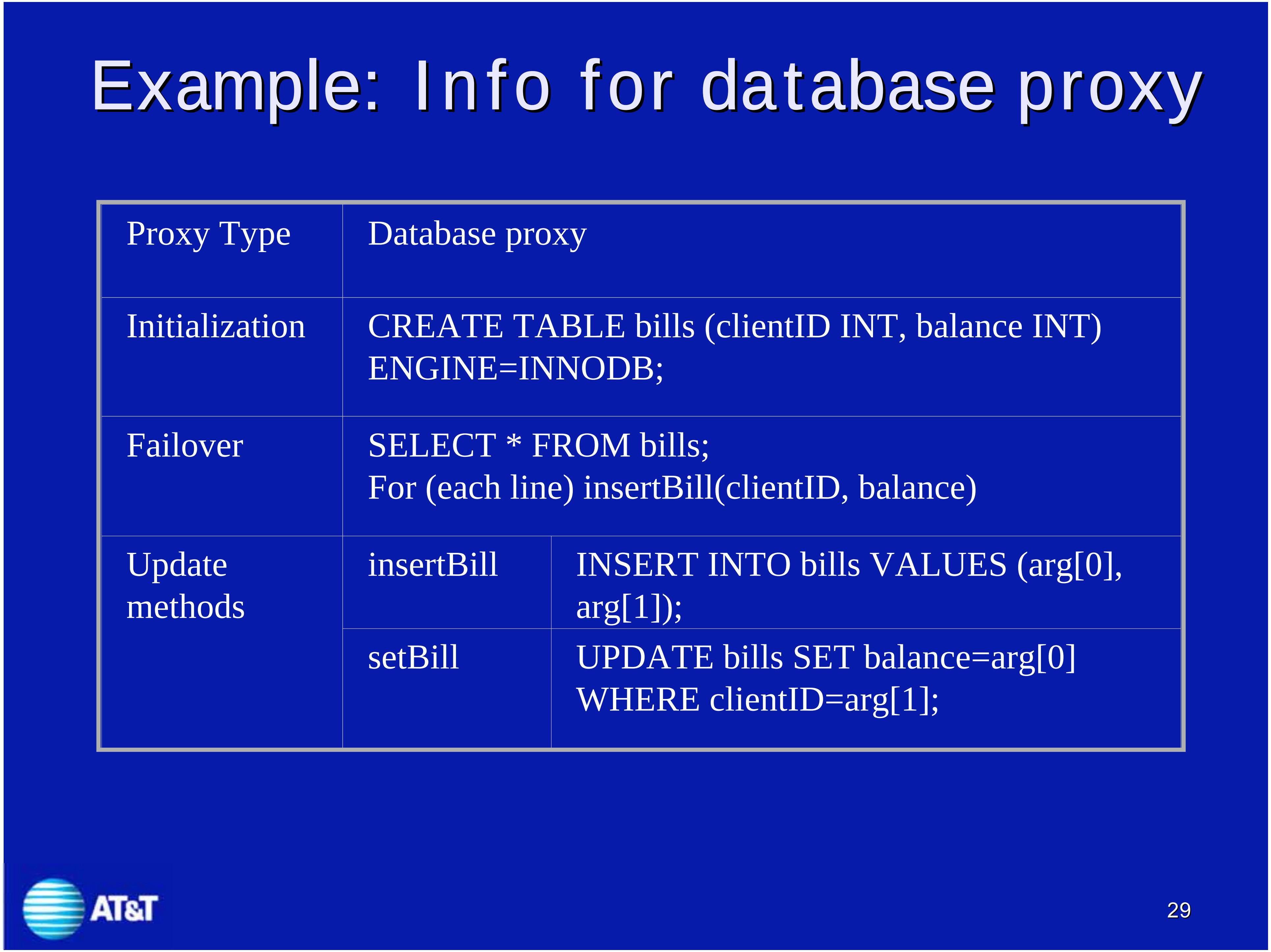
27

Configuration File

General information about the web service

- Such as the service URL, the resources the service uses...

- The information on the state update for each resource class.
- Information about transaction.

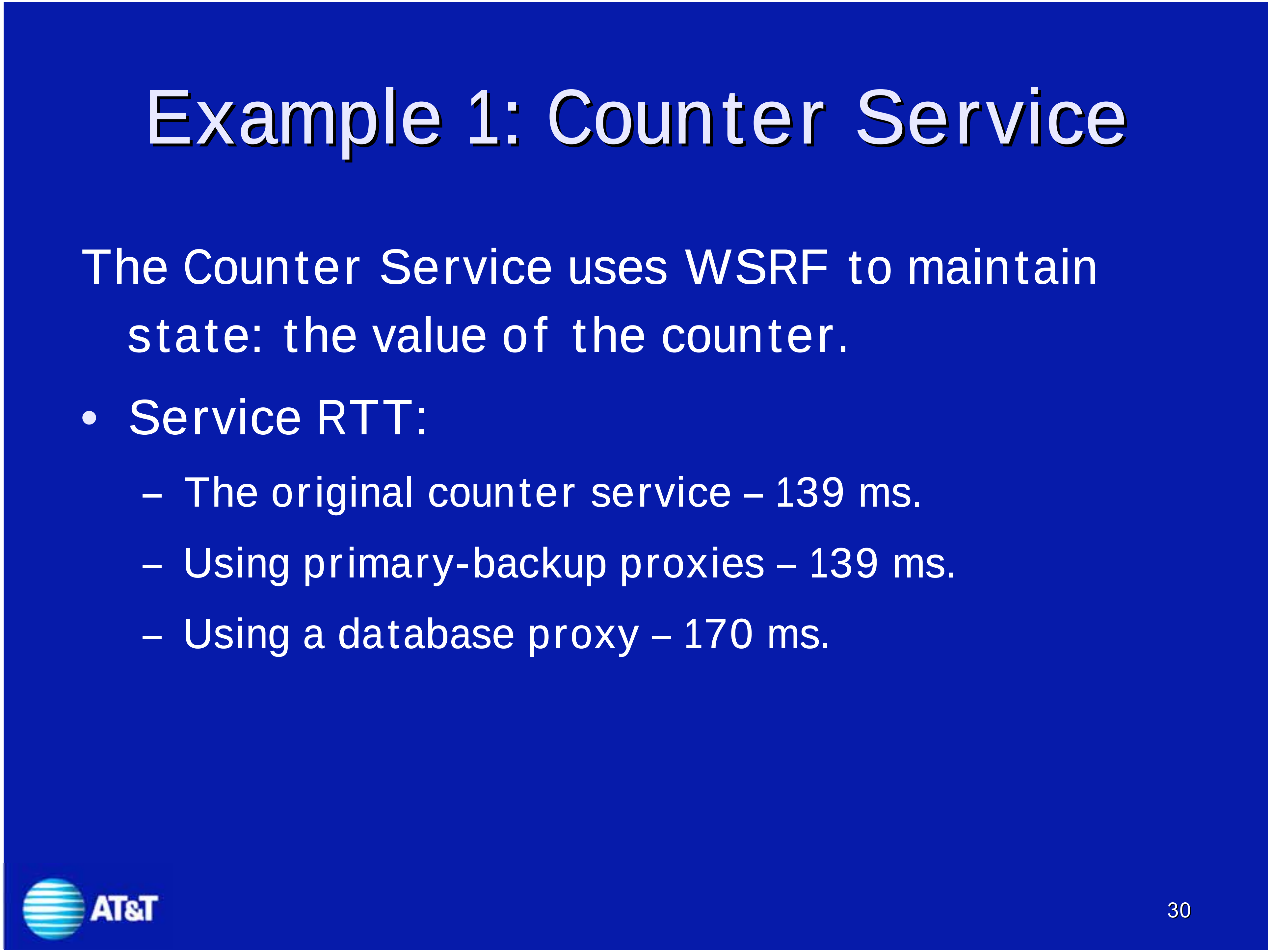


Example: Info for database proxy

Proxy Type	Database proxy	
Initialization	CREATE TABLE bills (clientID INT, balance INT) ENGINE=INNODB;	
Failover	SELECT * FROM bills; For (each line) insertBill(clientID, balance)	
Update methods	insertBill	INSERT INTO bills VALUES (arg[0], arg[1]);
	setBill	UPDATE bills SET balance=arg[0] WHERE clientID=arg[1];

AT&T

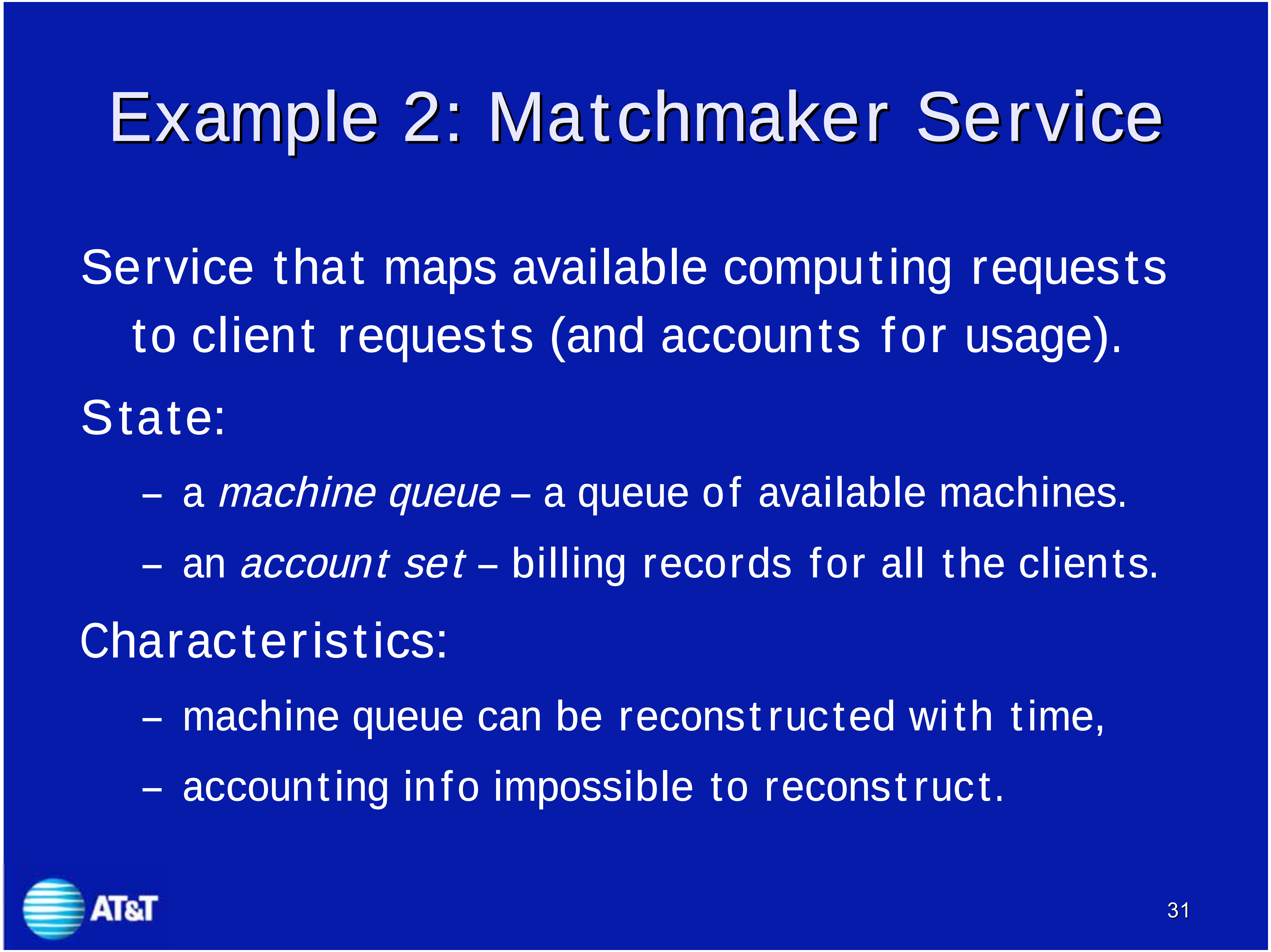
29



Example 1: Counter Service

The Counter Service uses WSRF to maintain state: the value of the counter.

- Service RTT:
  - The original counter service – 139 ms.
  - Using primary-backup proxies – 139 ms.
  - Using a database proxy – 170 ms.



Example 2: Matchmaker Service

Service that maps available computing requests to client requests (and accounts for usage).

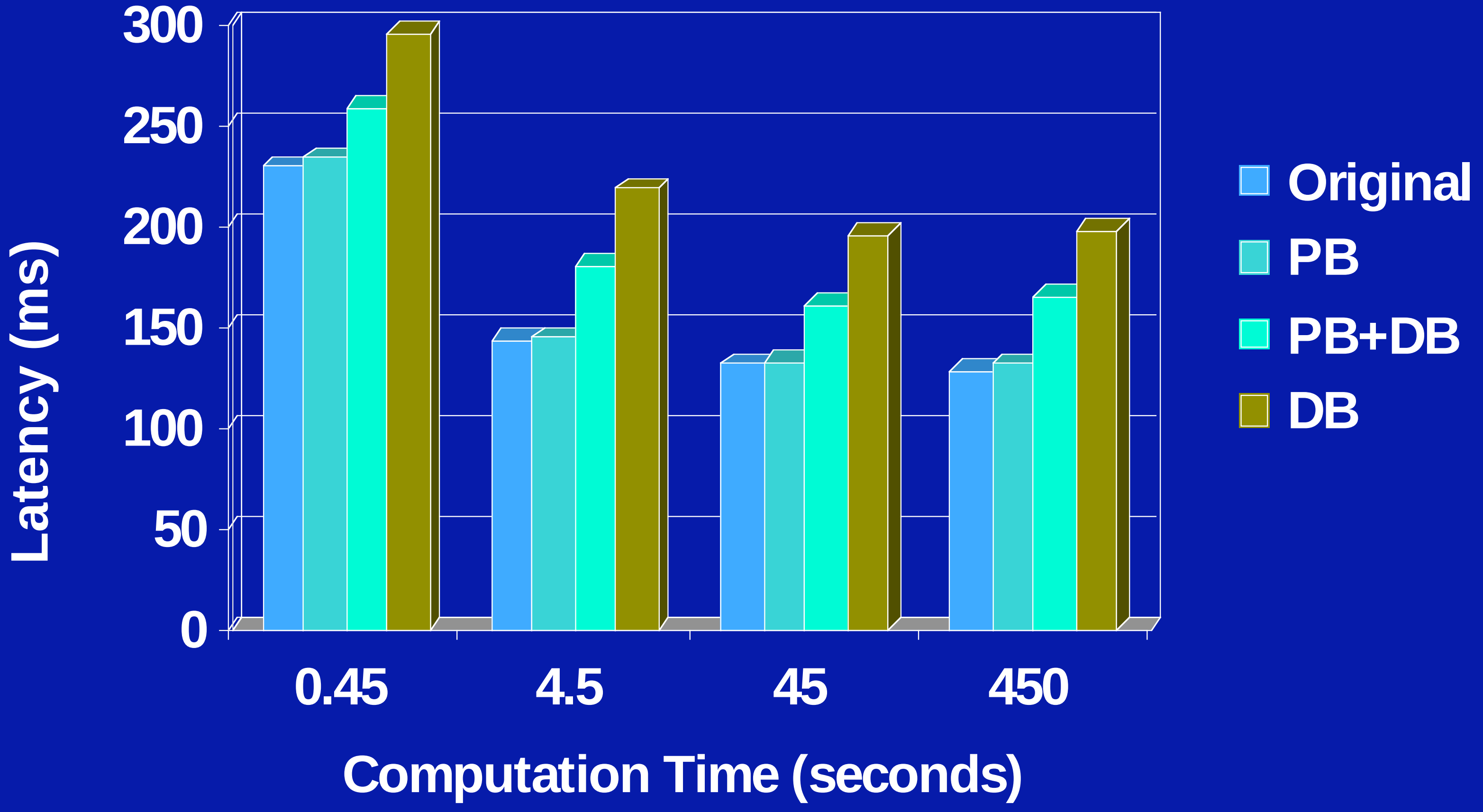
State:

- a *machine queue* – a queue of available machines.
- an *account set* – billing records for all the clients.

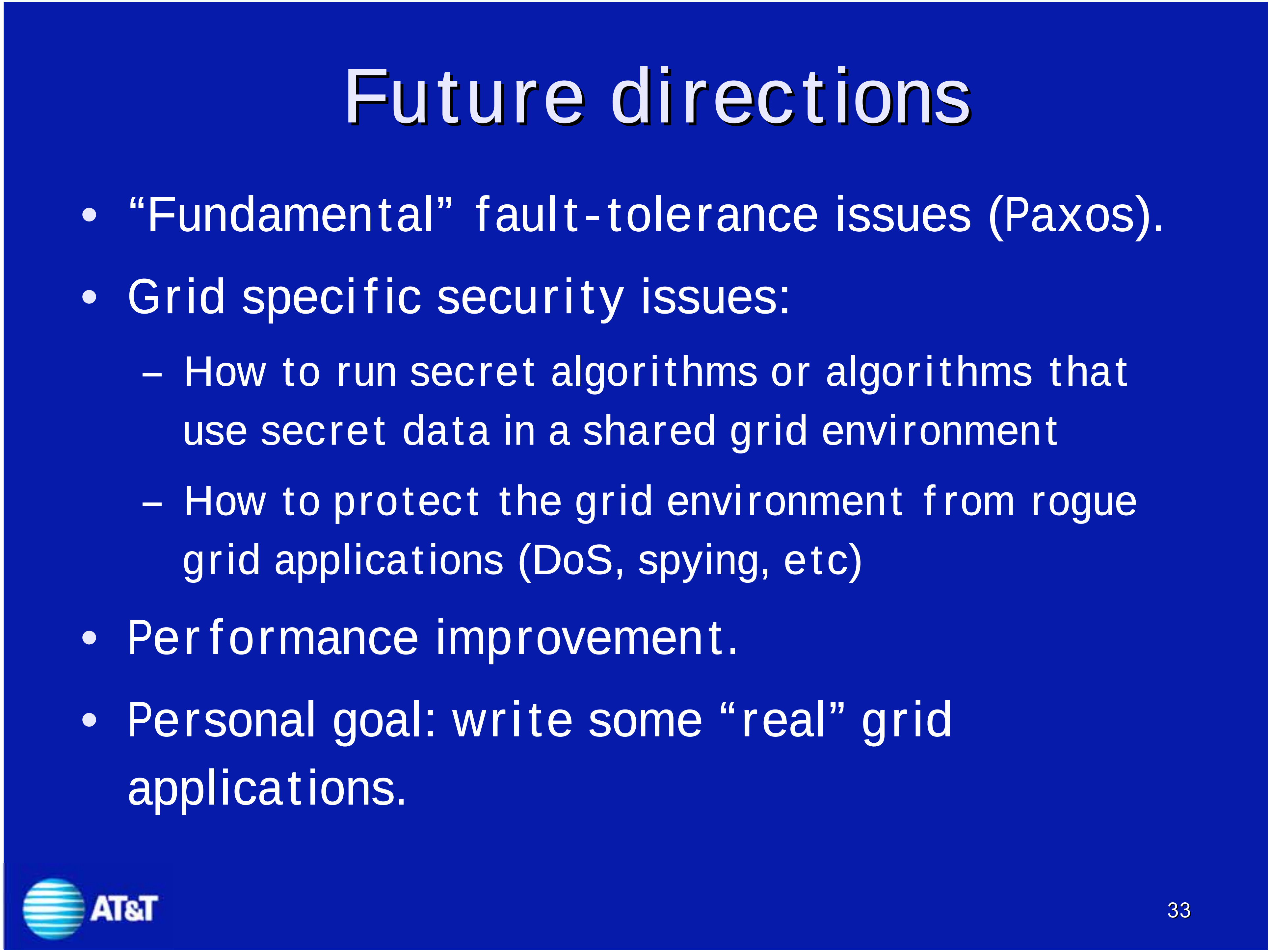
Characteristics:

- machine queue can be reconstructed with time,
- accounting info impossible to reconstruct.

Matchmaker Performance



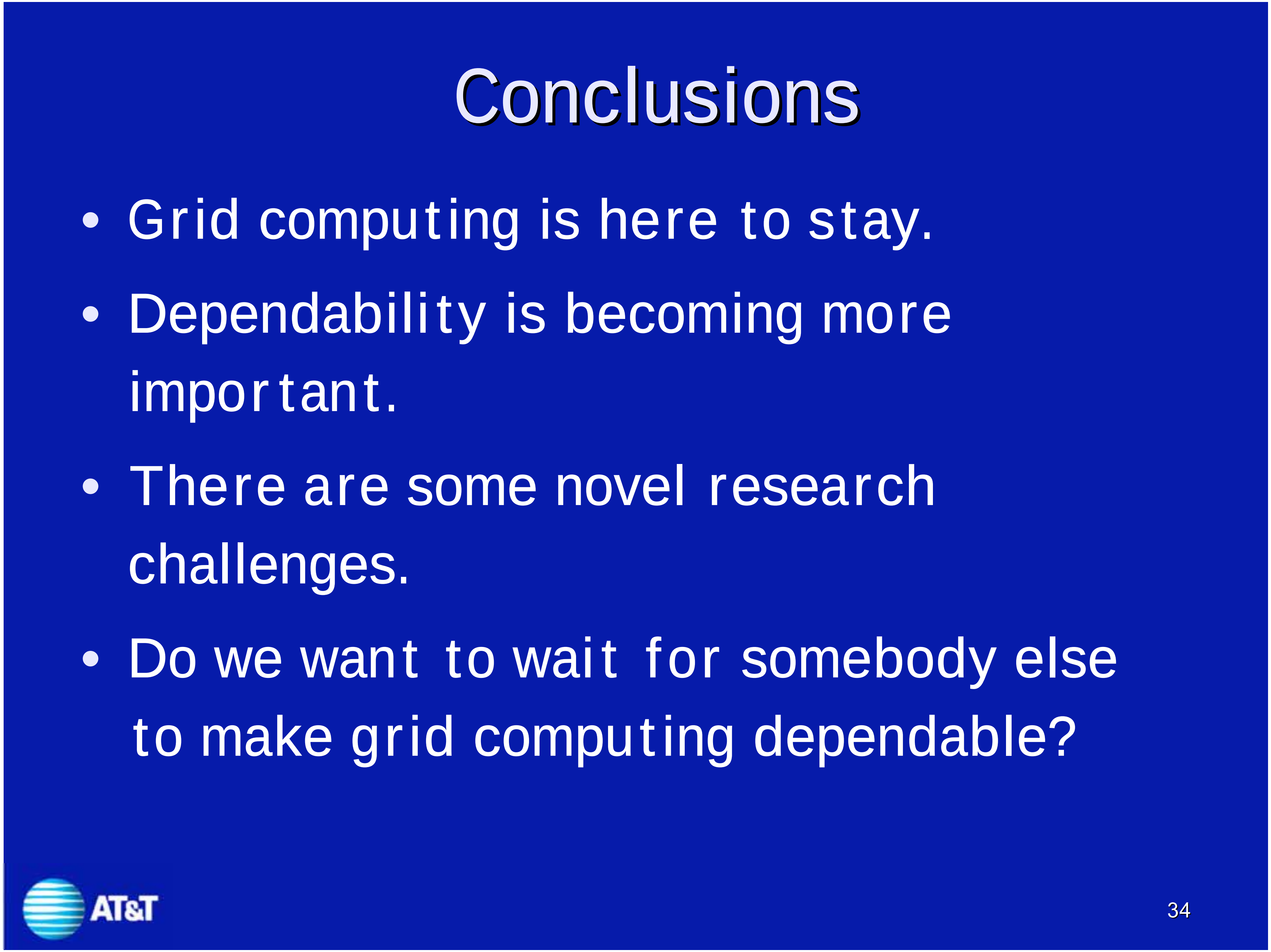
AT&T



Future directions

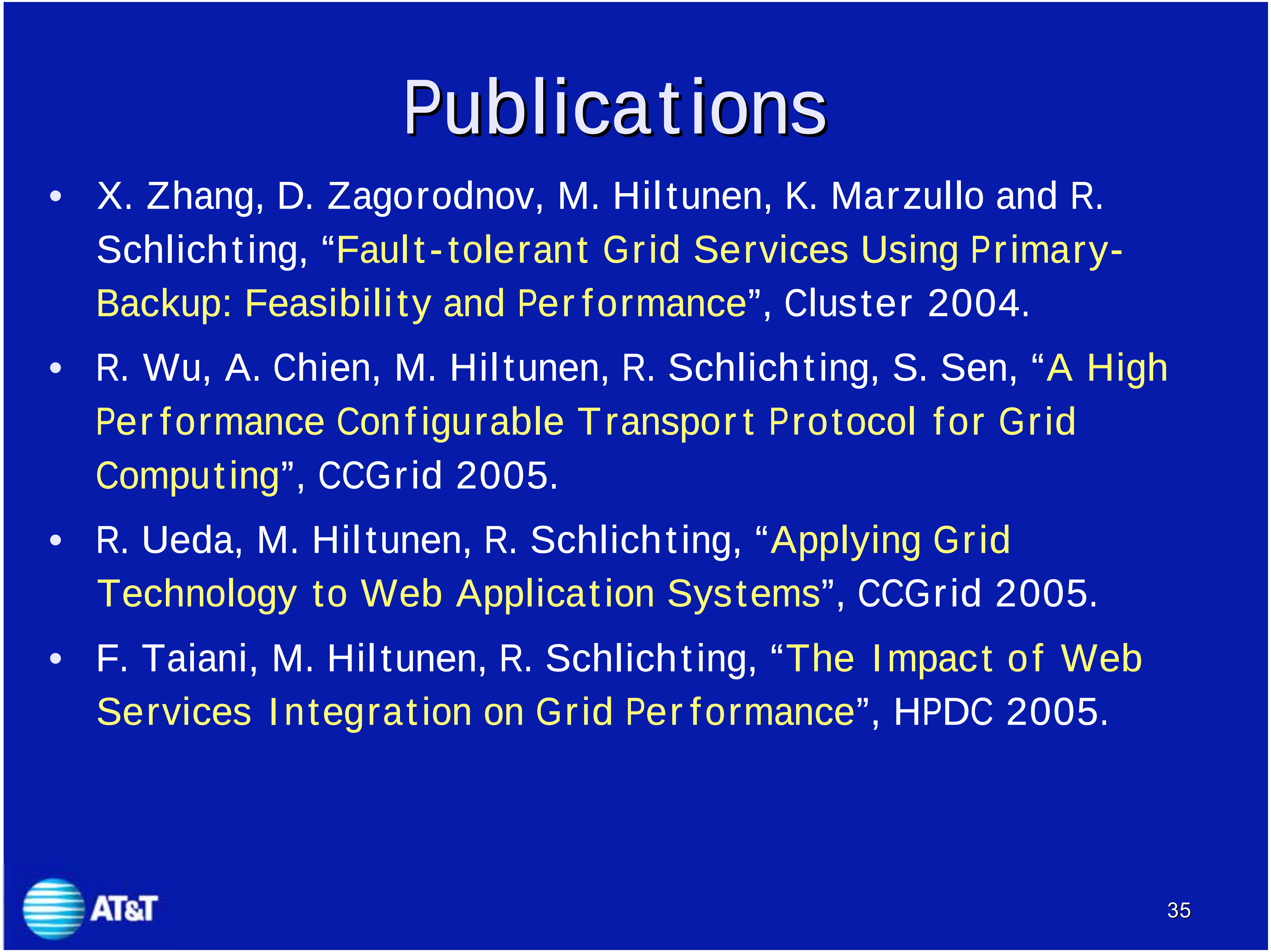
- “Fundamental” fault-tolerance issues (Paxos).
- Grid specific security issues:
  - How to run secret algorithms or algorithms that use secret data in a shared grid environment
  - How to protect the grid environment from rogue grid applications (DoS, spying, etc)
- Performance improvement.
- Personal goal: write some “real” grid applications.

33



Conclusions

- Grid computing is here to stay.
- Dependability is becoming more important.
- There are some novel research challenges.
- Do we want to wait for somebody else to make grid computing dependable?



Publications

- X. Zhang, D. Zagorodnov, M. Hiltunen, K. Marzullo and R. Schlichting, “Fault-tolerant Grid Services Using Primary-Backup: Feasibility and Performance”, Cluster 2004.
- R. Wu, A. Chien, M. Hiltunen, R. Schlichting, S. Sen, “A High Performance Configurable Transport Protocol for Grid Computing”, CCGrid 2005.
- R. Ueda, M. Hiltunen, R. Schlichting, “Applying Grid Technology to Web Application Systems”, CCGrid 2005.
- F. Taiani, M. Hiltunen, R. Schlichting, “The Impact of Web Services Integration on Grid Performance”, HPDC 2005.

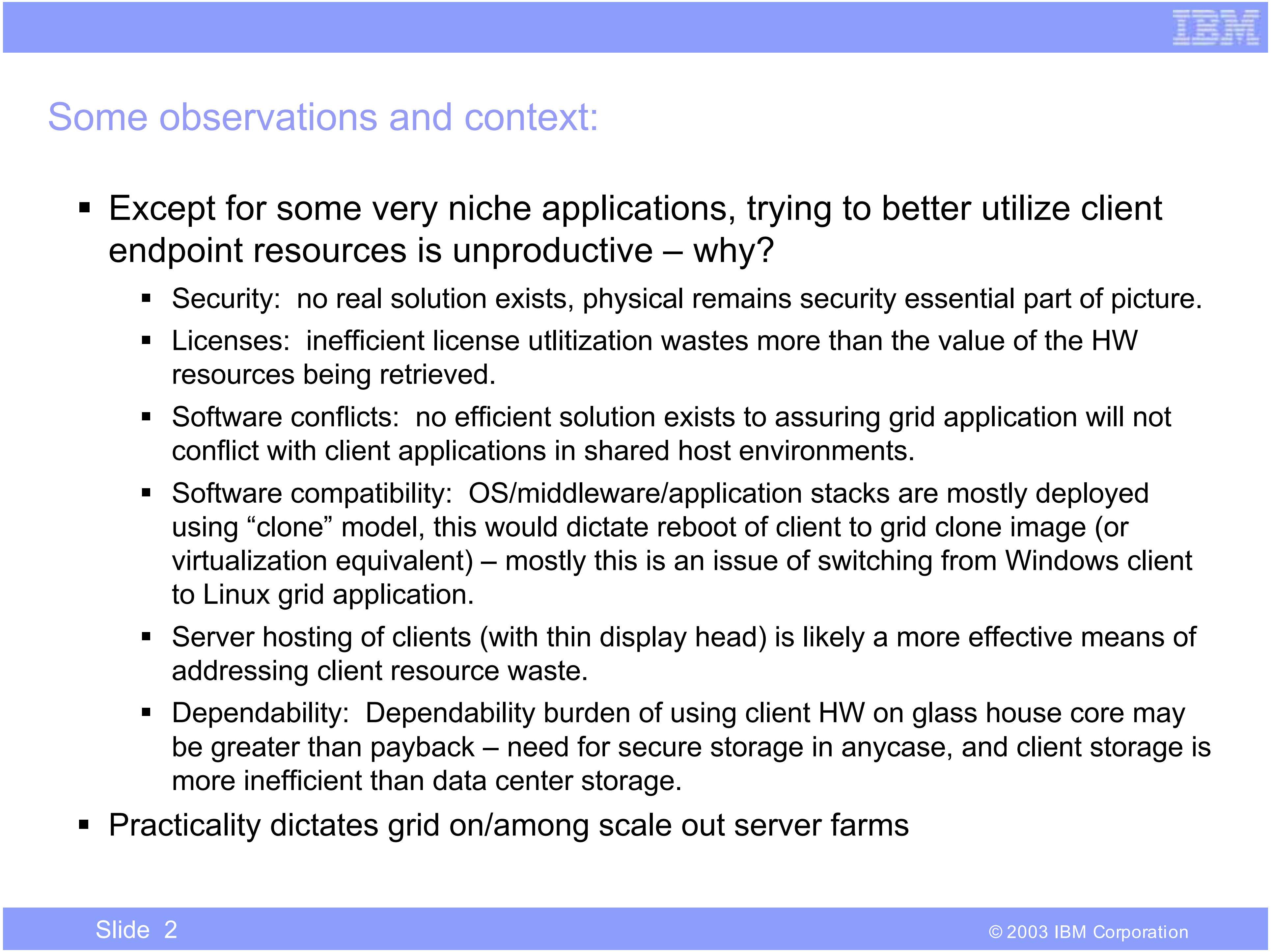
35

Grid on Blades

Basil Smith
7/2/2005

What is the problem?

- Inefficient utilization of resources (MIPS, Memory, Storage, Bandwidth)
  - Fundamentally resources are being wasted due to wide and unpredictable dynamic range in workload burdens – static or pseudo static resource allocation schemes do not work.
  - Underutilized resources in:
    - In server farms
    - At client endpoints
- Constraints
  - Security: need to run most apps with glass house class security
  - Licenses: need to get as much bang for buck for each license (this puts very real constraints on utilization of highly fragmented resources)
  - Software conflicts – hosting of grid application on a shared OS raises serious problems with conflicts and compatibility – frequently does not work at all and testing for obscure interaction is prohibitive
  - Software compatibility - applications cannot be extensively rewritten, they tend to run in context of a specific OS, middleware, and cluster environment
  - Dependability: particularly with respect to data integrity



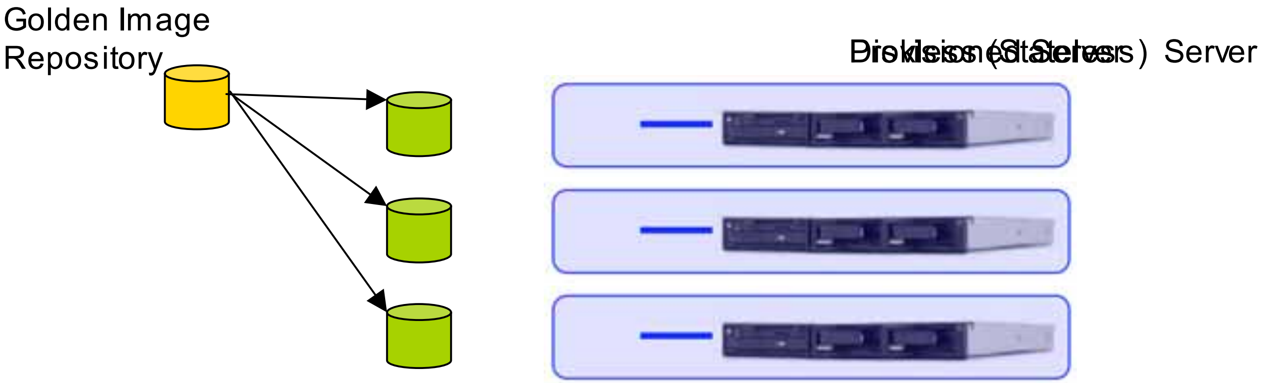
Some observations and context:

- Except for some very niche applications, trying to better utilize client endpoint resources is unproductive – why?
  - Security: no real solution exists, physical remains security essential part of picture.
  - Licenses: inefficient license utlitization wastes more than the value of the HW resources being retrieved.
  - Software conflicts: no efficient solution exists to assuring grid application will not conflict with client applications in shared host environments.
  - Software compatibility: OS/middleware/application stacks are mostly deployed using “clone” model, this would dictate reboot of client to grid clone image (or virtualization equivalent) – mostly this is an issue of switching from Windows client to Linux grid application.
  - Server hosting of clients (with thin display head) is likely a more effective means of addressing client resource waste.
  - Dependability: Dependability burden of using client HW on glass house core may be greater than payback – need for secure storage in anycase, and client storage is more inefficient than data center storage.
- Practicality dictates grid on/among scale out server farms

Slide 2

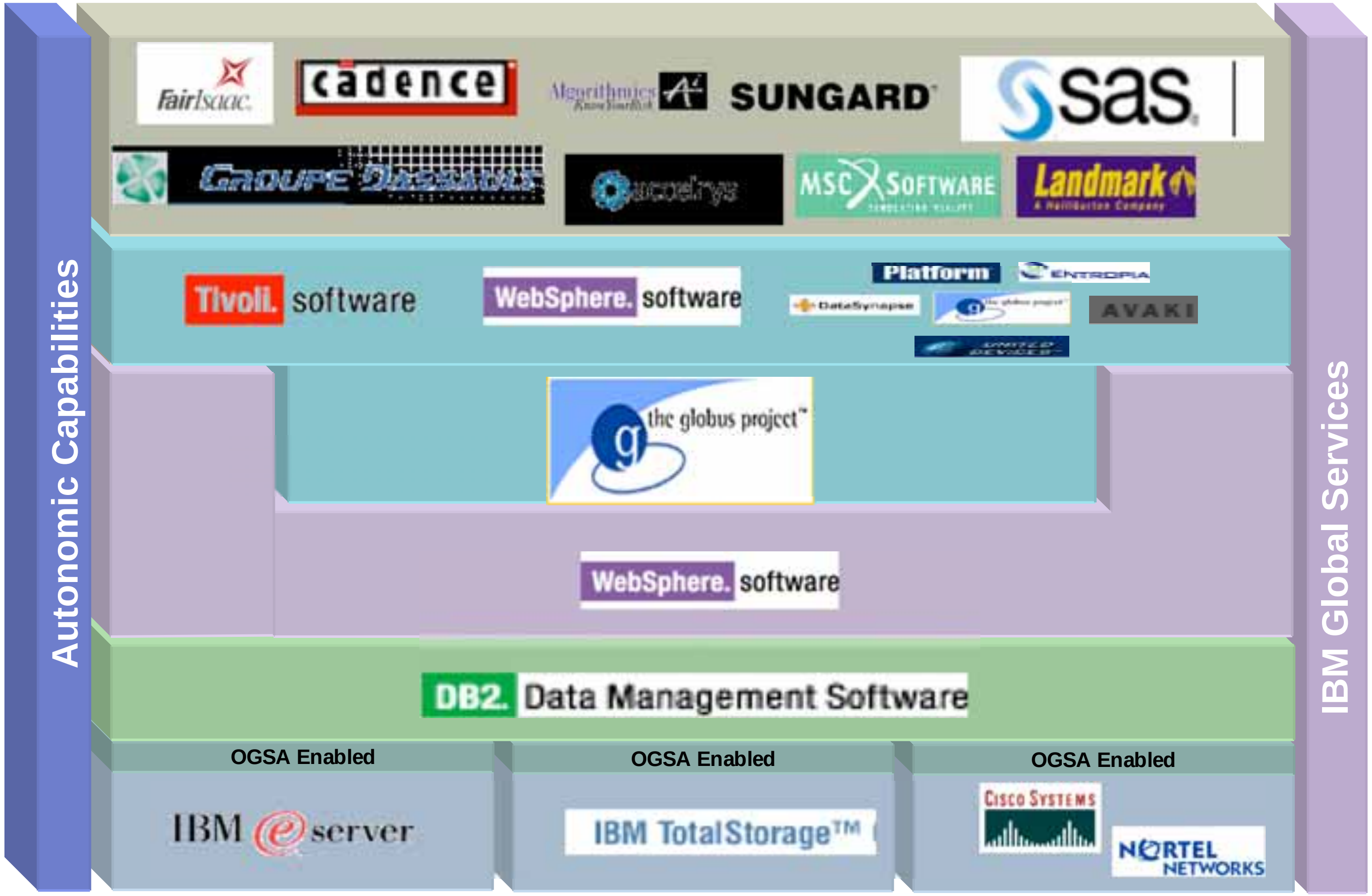
At the very bottom, what is the deployment model

- An application on a single node is deployed using “clone model”
  - Clone == boot disk image of OS/middleware/application instance, normally created from golden image, plus some customization
    - Virgin image – never been run no state beyond T0 image
      - Easily recreated from golden image
    - Dirty image – includes state changes from running image
      - May include extensive application state



Why Cloning – what's the application stack look like?

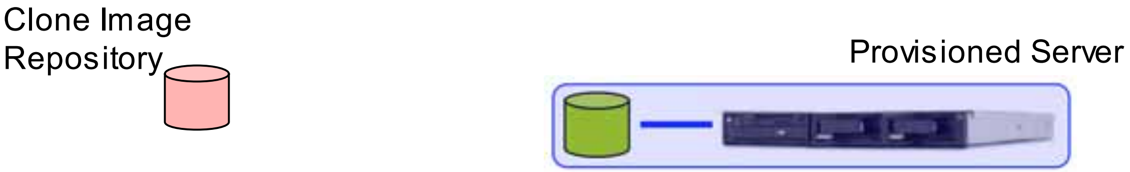
It looks like a bill board of stuff you need, and we will sell you ;-)



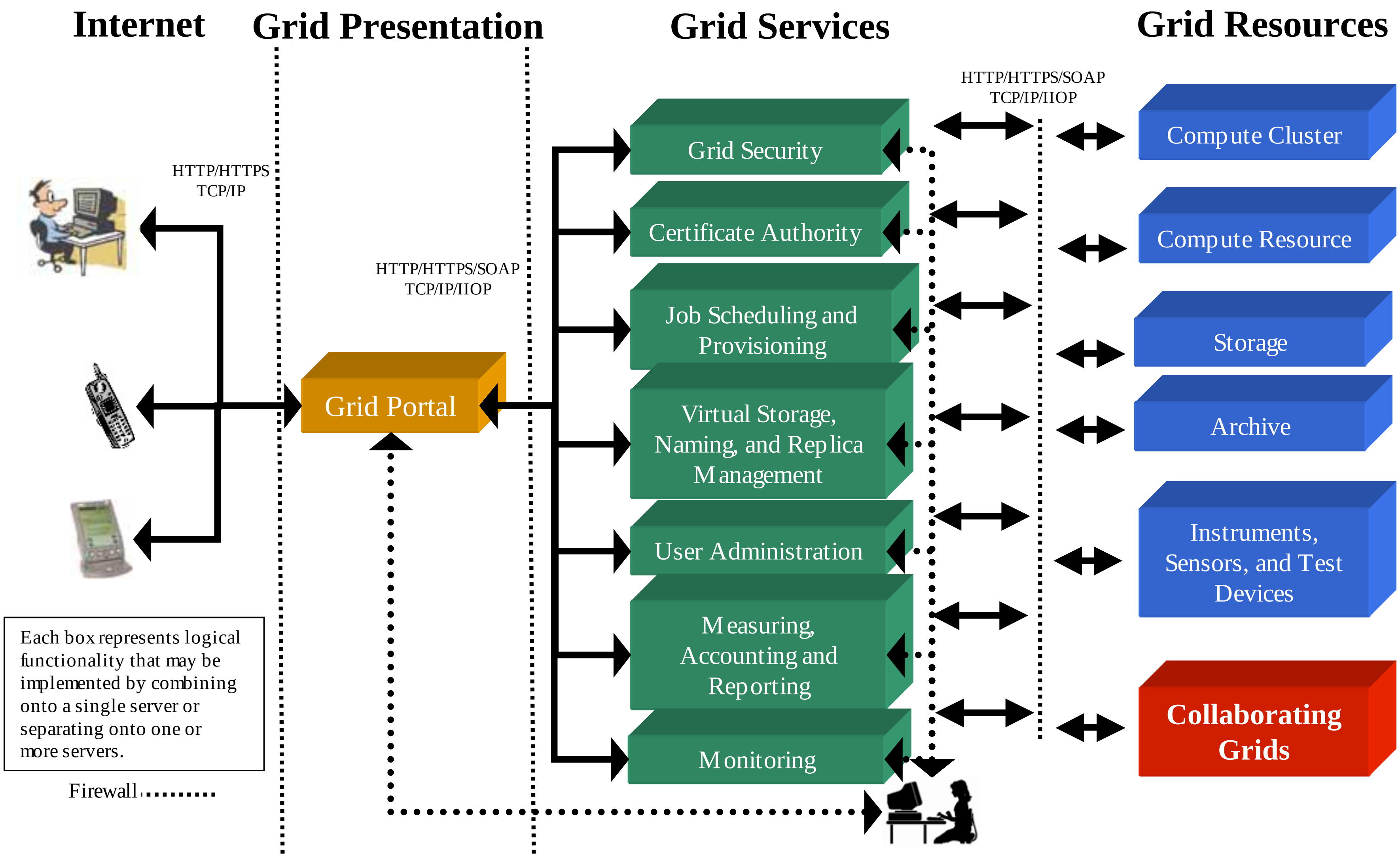
Build is tedious and release to "gold" is a lot of testing, somewhere in all of this you also might actually have to write some lines of code.

At the very bottom, retasking a server

- To retask:
  - “Hibernate” an active server (force all state to disk – a dirty clone)
  - Turn server off
  - Disconnect dirty clone of that image from server
  - Connect new clone to server
  - Boot new image

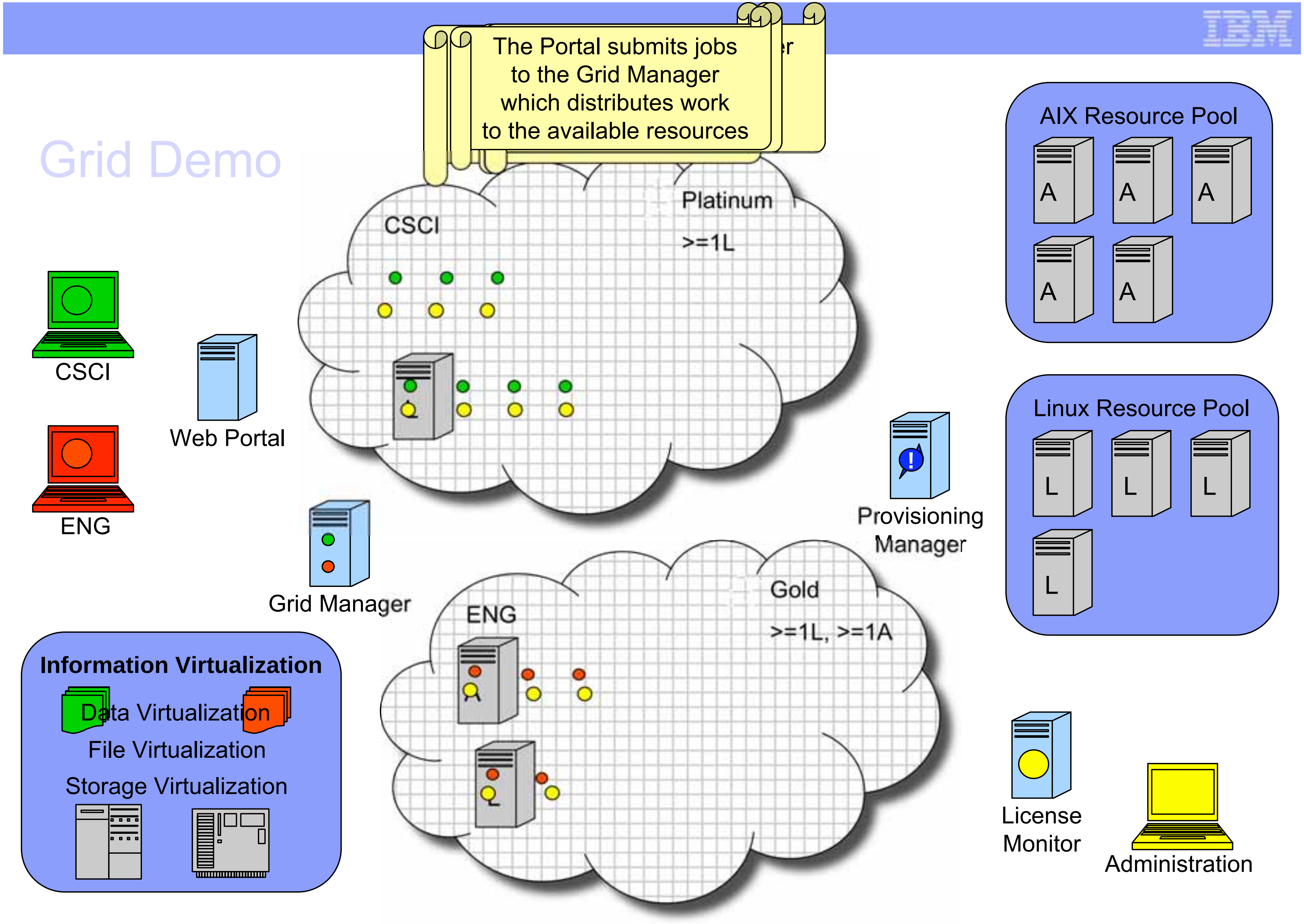


Grid Logical View



Each box represents logical functionality that may be implemented by combining onto a single server or separating onto one or more servers.

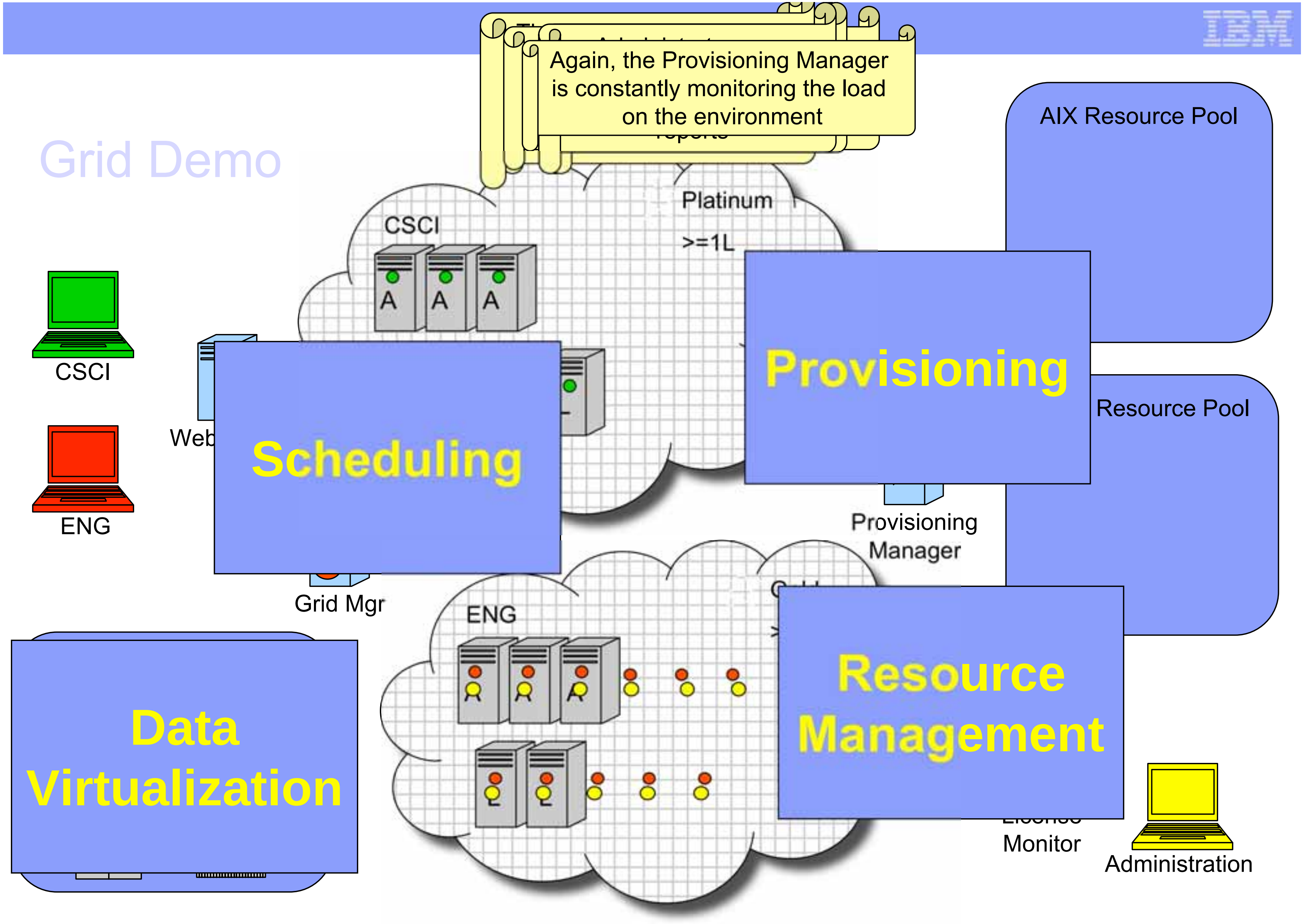
Grid Demo



The Portal submits jobs to the Grid Manager which distributes work to the available resources

Grid Demo

Again, the Provisioning Manager is constantly monitoring the load on the environment



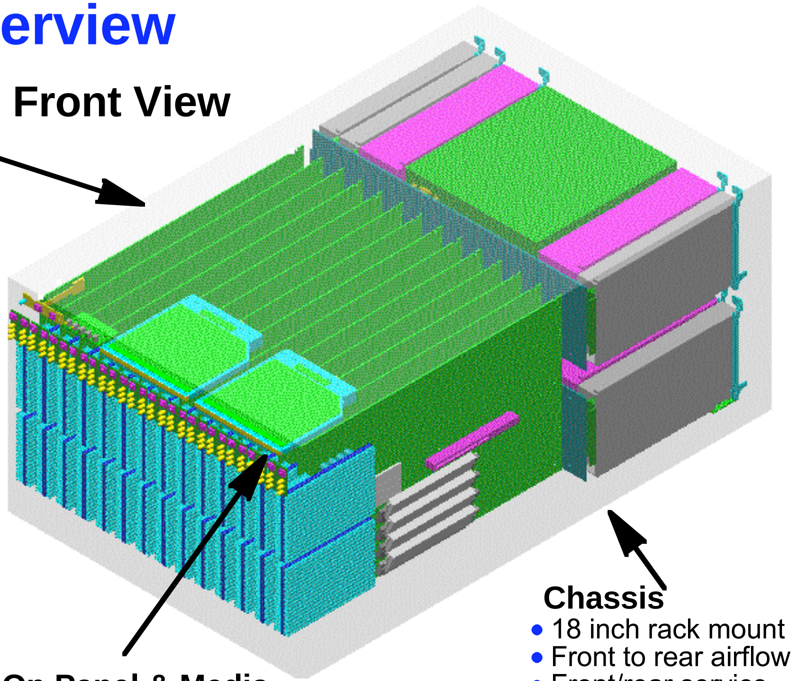
Again back to the bottom – what are these resources

eServer BladeCenter Overview

Processor Blades

- Hot swappable blades
- LEDs: Power, Alert, Info, Locate, Activity
- Buttons: Power, Reset, KVM Sel., Media Sel.
- USB, LightPath, Management, Video (HS)
- Processor Flexibility:
  - HS20 - 2-way XEON EM64T
    - 2GHz to 3.6 GHz, 800MHz FSB
    - 512MB to 8GB ECC memory
    - 2 Gb Ethernet + Opt. I/O feature card
    - Opt. to 2 SFF SCSI w/RAID0 or 1
  - HS40 - 4-way XEON MP
    - 2.0GHz to 3.0GHz, 400MHz FSB
    - 1GB to 16GB PC2100 ECC memory
    - 4 Gb Ethernet + two Opt. I/O feature card
    - Opt. to 2 SCSI disk via 'sidecar'
  - JS20 - 2-way PowerPC® 970
    - 2.2GHz, 800MHz memory
    - 512MB to 4GB ECC PC2700 memory
    - 2 Gb Ethernet + Opt. I/O feature card
    - Opt. to 2 IDE drives
- Optional - I/O Feature Cards:
  - Dual 2Gb Fibre Channel HBAs
  - Dual 1Gb Ethernet NICs (4 total)
  - 2Gb Myrinet cluster interface
  - Dual 1x InfiniBand HCAs
- Optional - dual SCSI disk 'sidecar'
  - 18.2, 36.4, 73.4, 146 or 300GB capacity
  - 10K RPM or 15K RPM speed
  - Built in mirroring, Hot swap
  - Two I/O Feature Card sockets
- Optional - dual adapter slot PCI-X 'sidecar'

Front View



Op Panel & Media

- Chassis level LEDs-
  - Power, Alert, Info,
  - Chassis 'Locate' indicator
- USB Port
- Removable storage media
  - CD & floppy disk

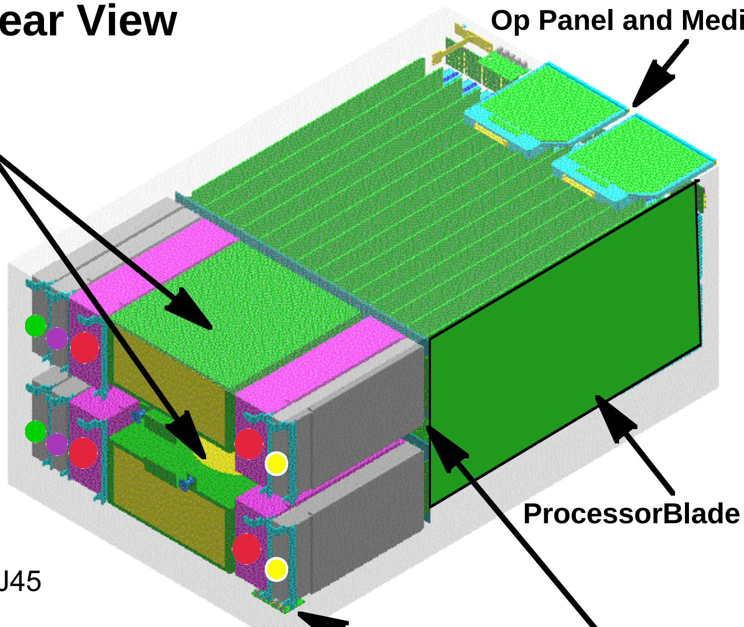
Chassis

- 18 inch rack mount
- Front to rear airflow
- Front/rear service
- Rear cabling

- "Enterprise" Rack
  - 14 CPU Blades
  - 7U high, 28" deep

- "Telco" Rack
  - 8 CPU Blades
  - 8U high, 20" deep
  - DC or AC pwr
  - NEBS ready

Again back to the bottom – what are these resources



Rear View

Blower Module (2X)
- Hot Swap, Redundant
- 300 CFM, speed controlled

🟢 **Optional Switch Module (0, 1, or 2)**
- Hot Swap, Optional Redundancy
- Input: 14 blades + 2 MM (1-3Gb + 100Mb)
  - Ethernet - (same options as below)
  - Fibre Channel - Uplink: 1/2 Gb FC SFP
    IBM SAN Switch, Brocade SAN Switch
  - OPM - Direct optical link to each blade's port
  - InfiniBand - Uplink: 12/4x IB (40Gbps total)

🟣 **Ethernet Switch Module (1 or 2)**
- Hot Swap, Optional Redundancy
- Input: 14 blades + 2 MM (Gb + 100Mb)
  - IBM Layer 2 Ethernet Switch
  - Nortel Networks L2/3, L4-7, SFP or RJ45
  - Cisco Layer 2+ Ethernet Switch
  - CPM - Direct RJ45 to each blade's port

🔴 **Power Module (2 or 4)**
- 200-240V AC (worldwide volt./freq.)
- Hot Swap, Redundant (Opt.)

🟡 **Mgmt Module (MM) (1 or 2)**
- Chassis management control point
- KVM Switches (Local and Remote)
- Hot Swap, Redundant (Opt.)

Op Panel and Media

ProcessorBlade (1-14)

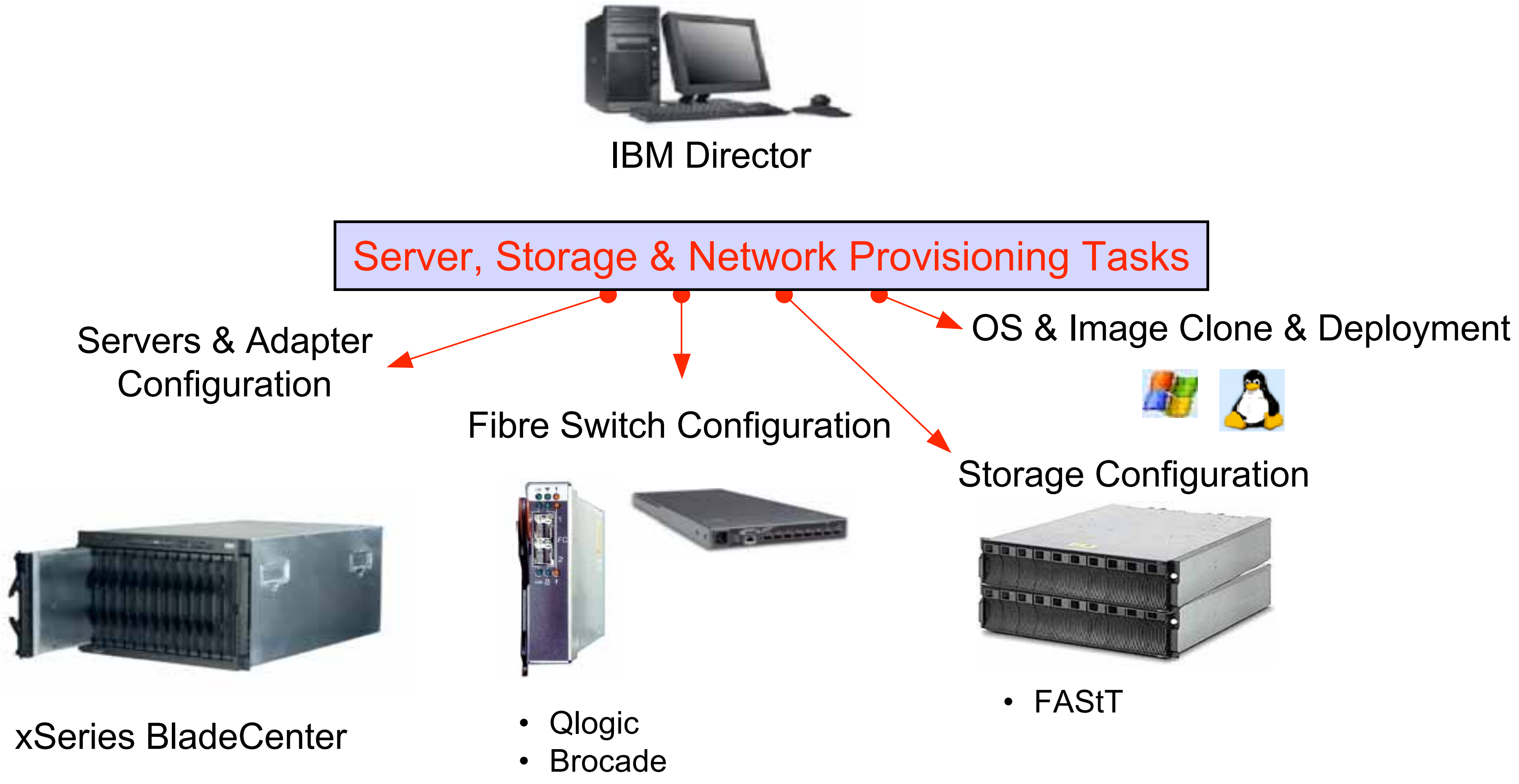
Op Panel (same LEDs)

Mid-Plane
- Redundant connections
- Point-to-Point connections
- No single point of failure

Again back to the bottom – what are these resources

Processor Blade (Dual Xeon)

Low level management to enable grid



IBM Director

Server, Storage & Network Provisioning Tasks

Servers & Adapter Configuration

Fibre Switch Configuration

Storage Configuration

OS & Image Clone & Deployment

xSeries BladeCenter

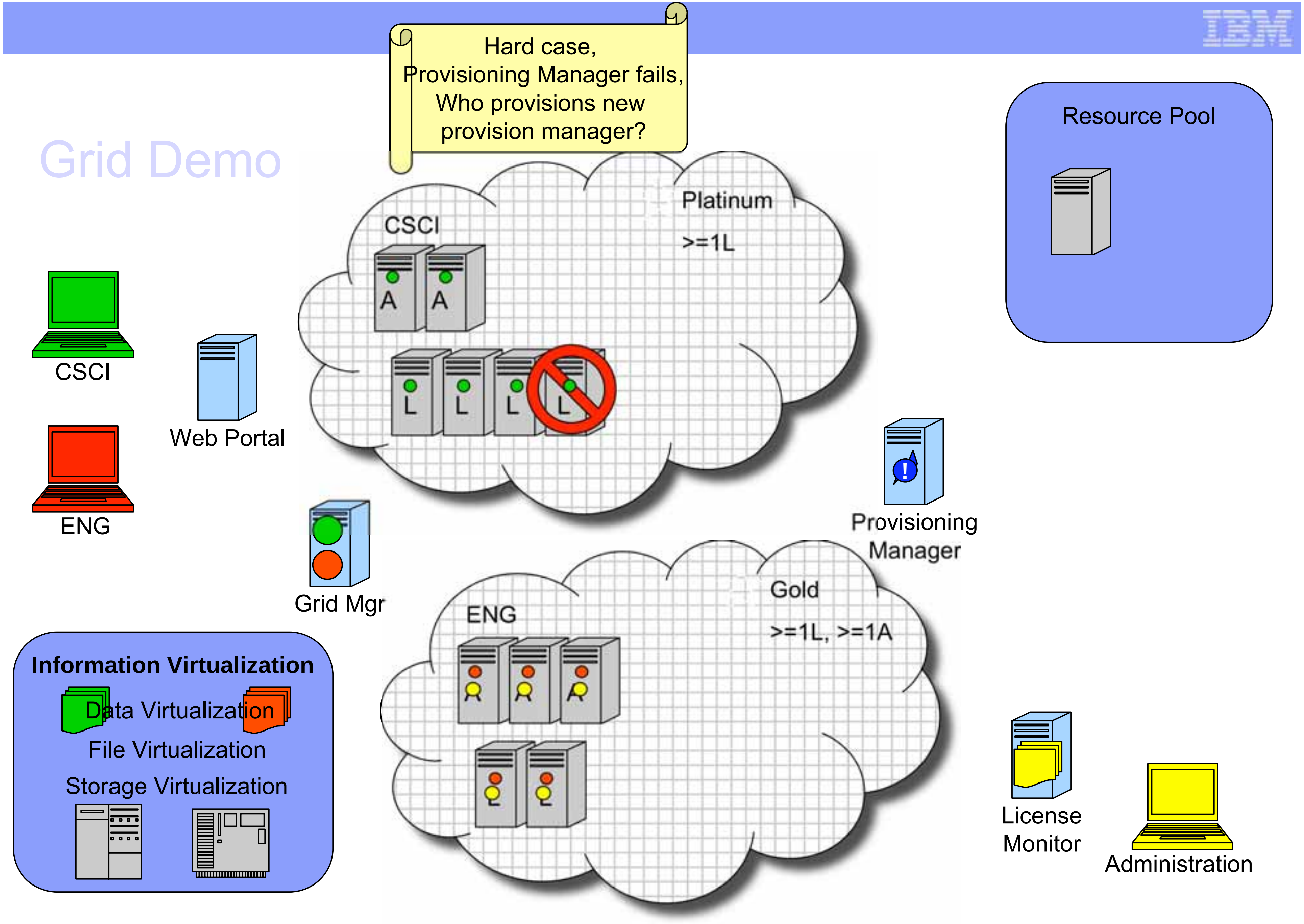
- Qlogic
- Brocade

- FAStT

Finally, the dependability challenge

- Break the problem down to known solutions
  - Classic cluster recovery for failed node in application
  - Reprovisioning of spare node to replace capacity
    - Is this with a virgin copy, checkpointed copy, or by just attaching failed image to another server and restarting
  - File and disk dependability and integrity management is critical, ultimately protecting against loss of state
    - RAID storage subsystems
    - Replicas and checkpoints (point in time copies)
    - Geographic replication (for disaster recovery)

Grid Demo

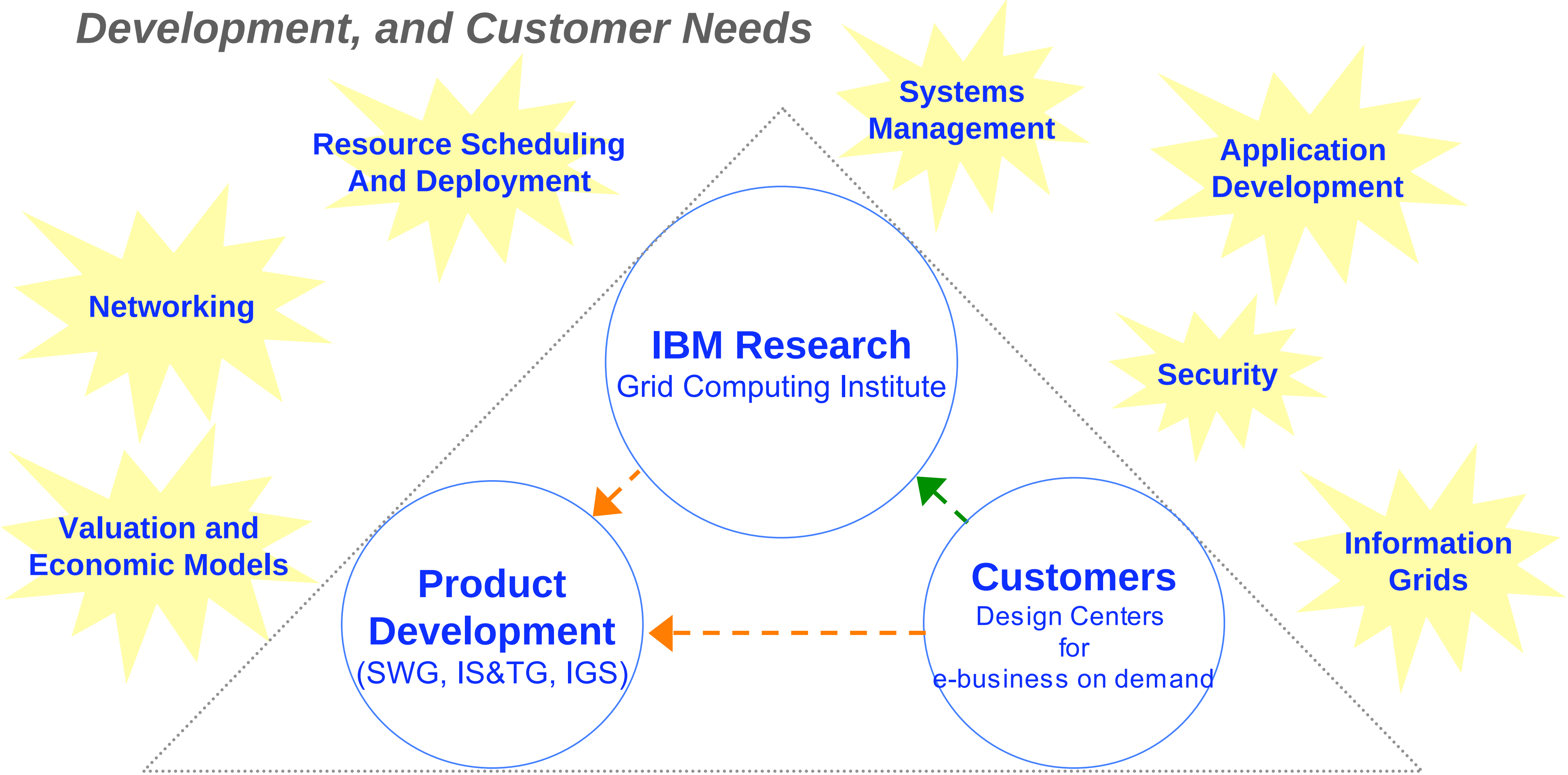

Hard case,
Provisioning Manager fails,
Who provisions new
provision manager?
Resource Pool
CSCI
Platinum
>=1L
A
A
L
L
L
L
CSCI
Web Portal
ENG
Grid Mgr
Provisioning
Manager
ENG
Gold
>=1L, >=1A
Information Virtualization
Data Virtualization
File Virtualization
Storage Virtualization
License
Monitor
Administration

The dependability challenge

- Options / candidates for availability manager
- What grid services need to be availability aware
    - Lots of problems
        - Who recovers lost licenses
        - Strategy for recoverying basic grid services.
        - Break the problem down to known solutions
        - Who keeps compatibility matrix
        - Role of virtualization
        - Whats disaster recovery procedure for storage subsystem failure

Grid Computing Institute

Aligning IBM Research with the Grid Strategy, Product Development, and Customer Needs



Resource Scheduling And Deployment

Systems Management

Application Development

Networking

Security

Valuation and Economic Models

Information Grids

IBM Research
Grid Computing Institute

Product Development
(SWG, IS&TG, IGS)

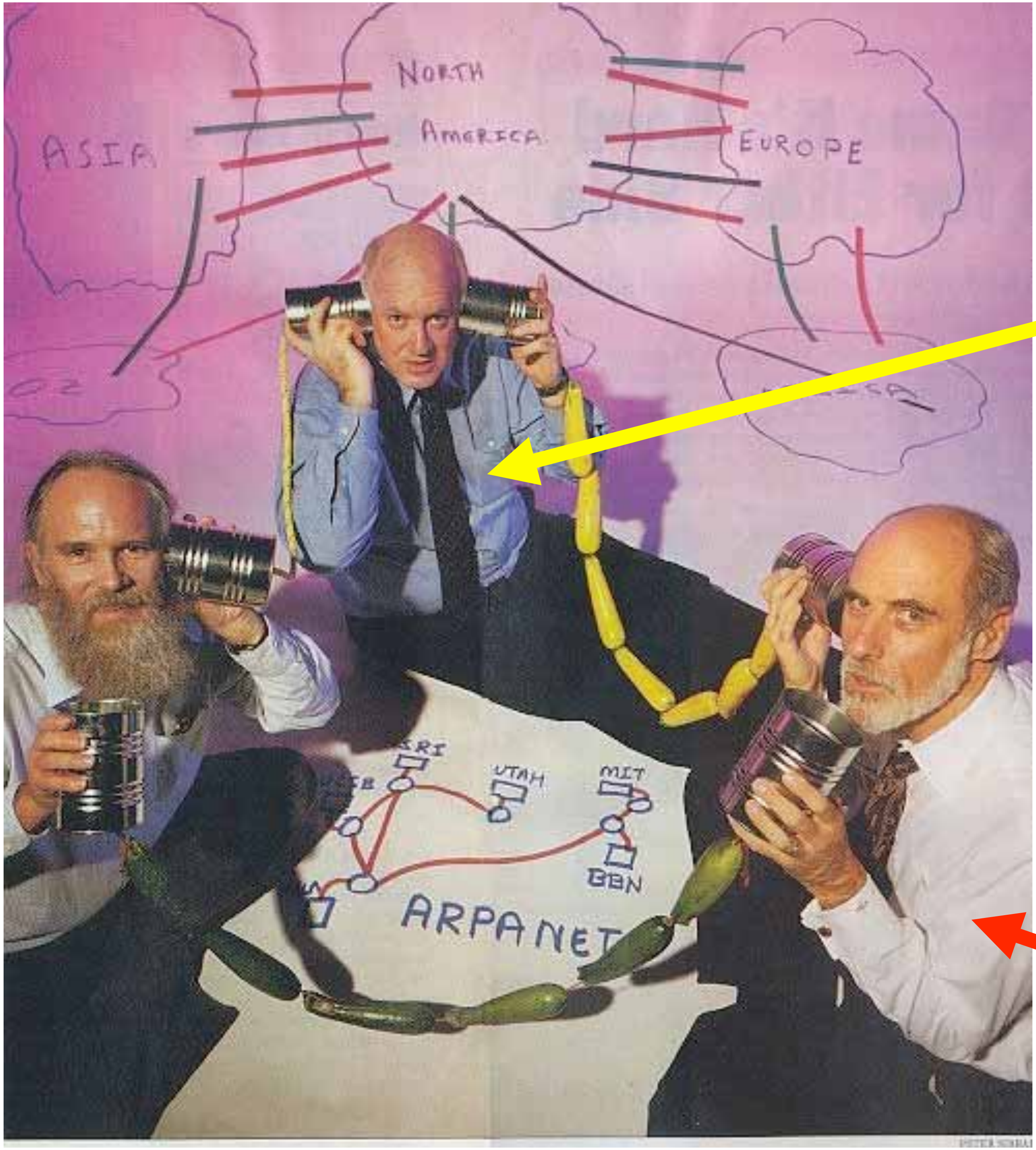
Customers
Design Centers
for
e-business on demand

Discussion:

Grid Computing Evolution and Challenges for Resilience, Performance and Scalability

Luca Simoncini

University of Pisa, Italy

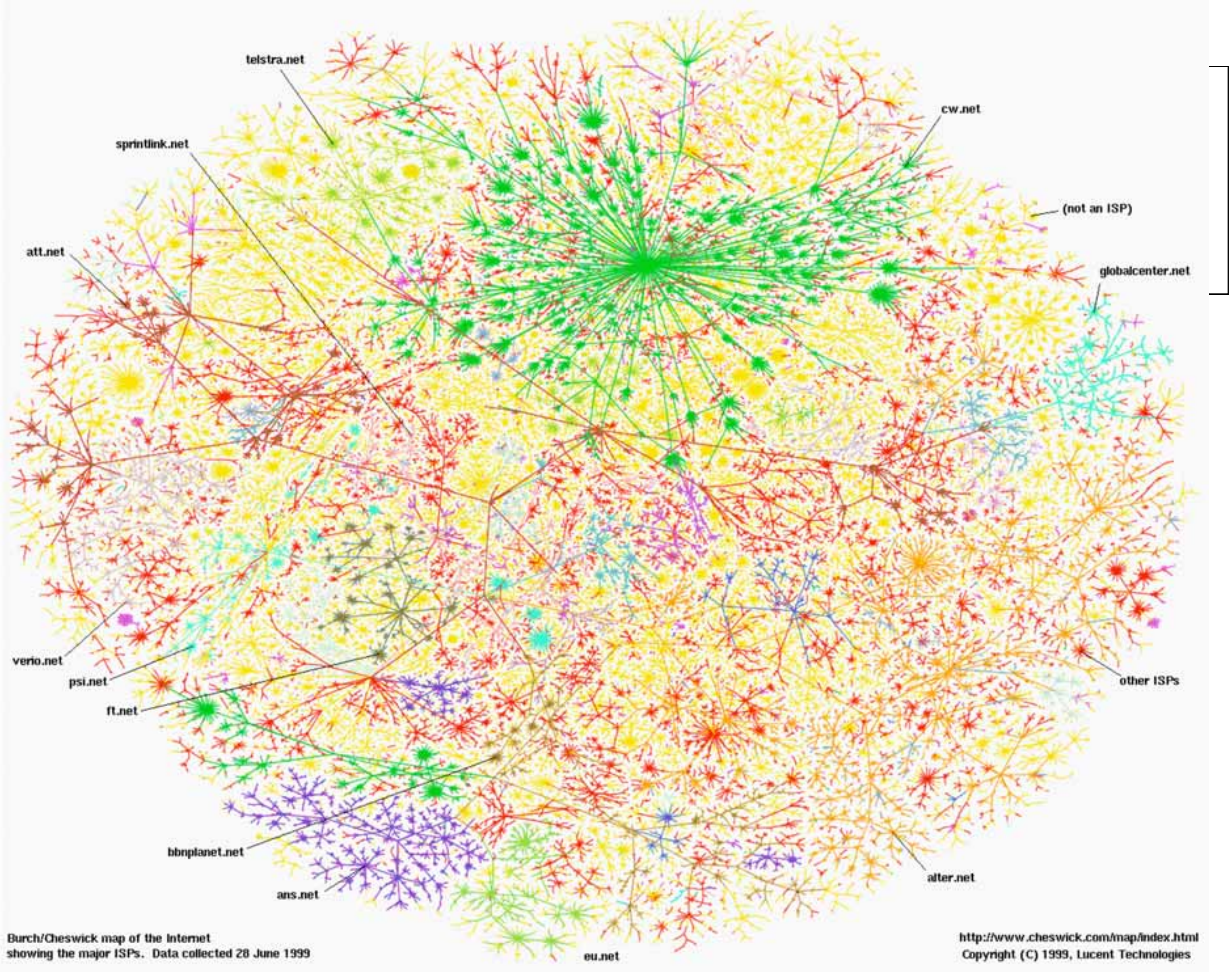


This photo was published in the August 8, 1994 issue of Newsweek and commemorates the 25th anniversary of the ARPANET. **Jon Postel**, **Steve Crocker** and I spent hours helping the photographer prepare for this shot.
Jon drew all the pictures, Steve and I strung the zucchini and the yellow squash. I think we must have collectively spent about 8 hours on this.

Note that this network can't work - there is no mouth/ear link anywhere!!!

Such was the state of networking in the primitive 1960s...

*Picture from **Vint Cerf***



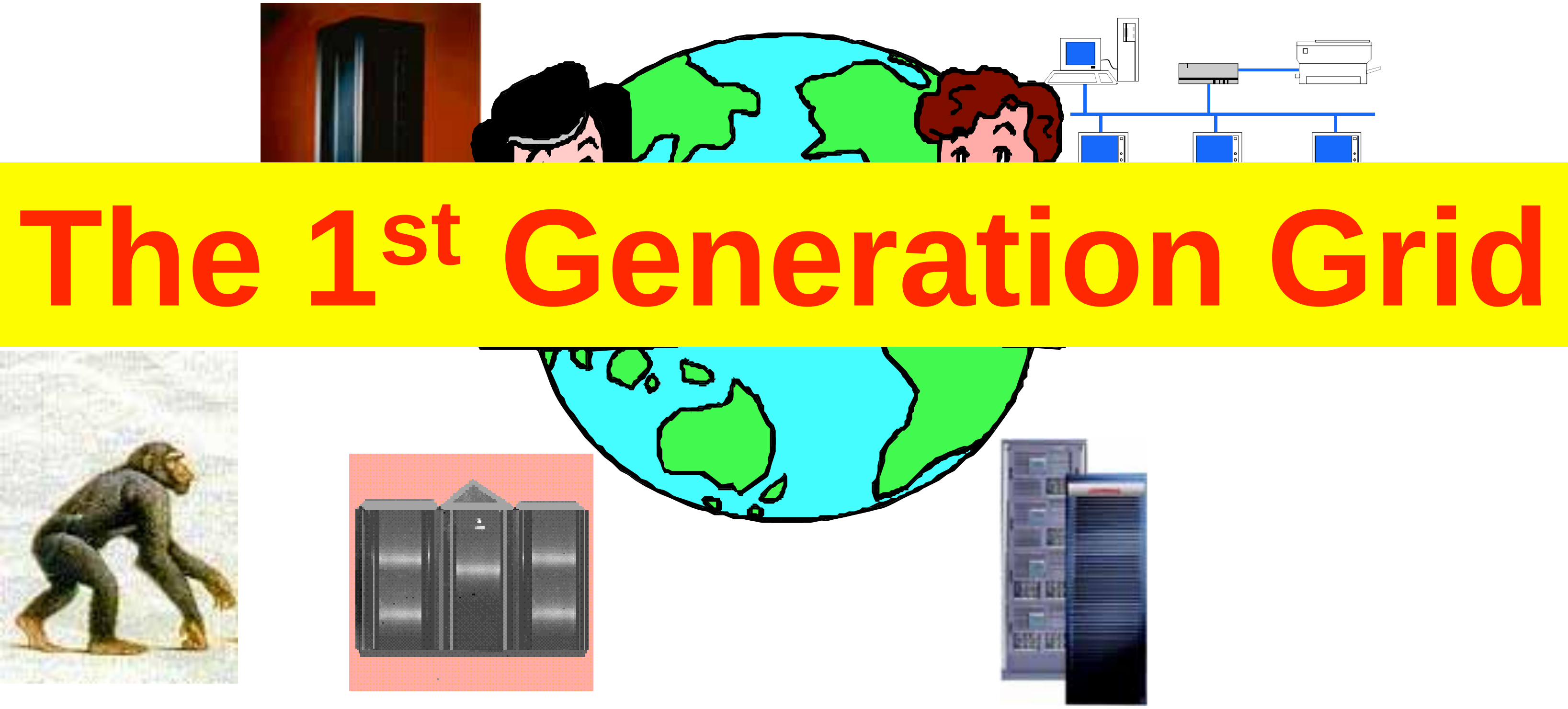
Burch/Cheswick map of the Internet
showing the major ISPs. Data collected 28 June 1999

http://www.cheswick.com/map/index.html
Copyright (C) 1999, Lucent Technologies

The term **“Grid”** means different things to different users groups and application domains.

- **Virtual organizations**. The Grid is seen as the collection of enabling technologies for building virtual organizations over the Internet.
- **Integration of resources.** The Grid is about building large-scale, distributed applications from distributed resources using a standard implementation-independent infrastructure.
- **Universal computer**. According to some (e.g., IBM-GRID25), the Grid is in effect a universal computer with memory, data storage, processing units, etc. that are distributed and are used transparently from applications.
- **Supercomputer interconnection**. The Grid is the result of interconnecting supercomputer centers together and enabling large-scale, long-running scientific computations with a very high demand regarding all kinds of computational, communication, and storage resources.
- **Distribution of computations**. Finally, there are those who see cycle-stealing applications, such as SETI@HOME, as typical Grid applications without any requirements for additional, underlying technologies.

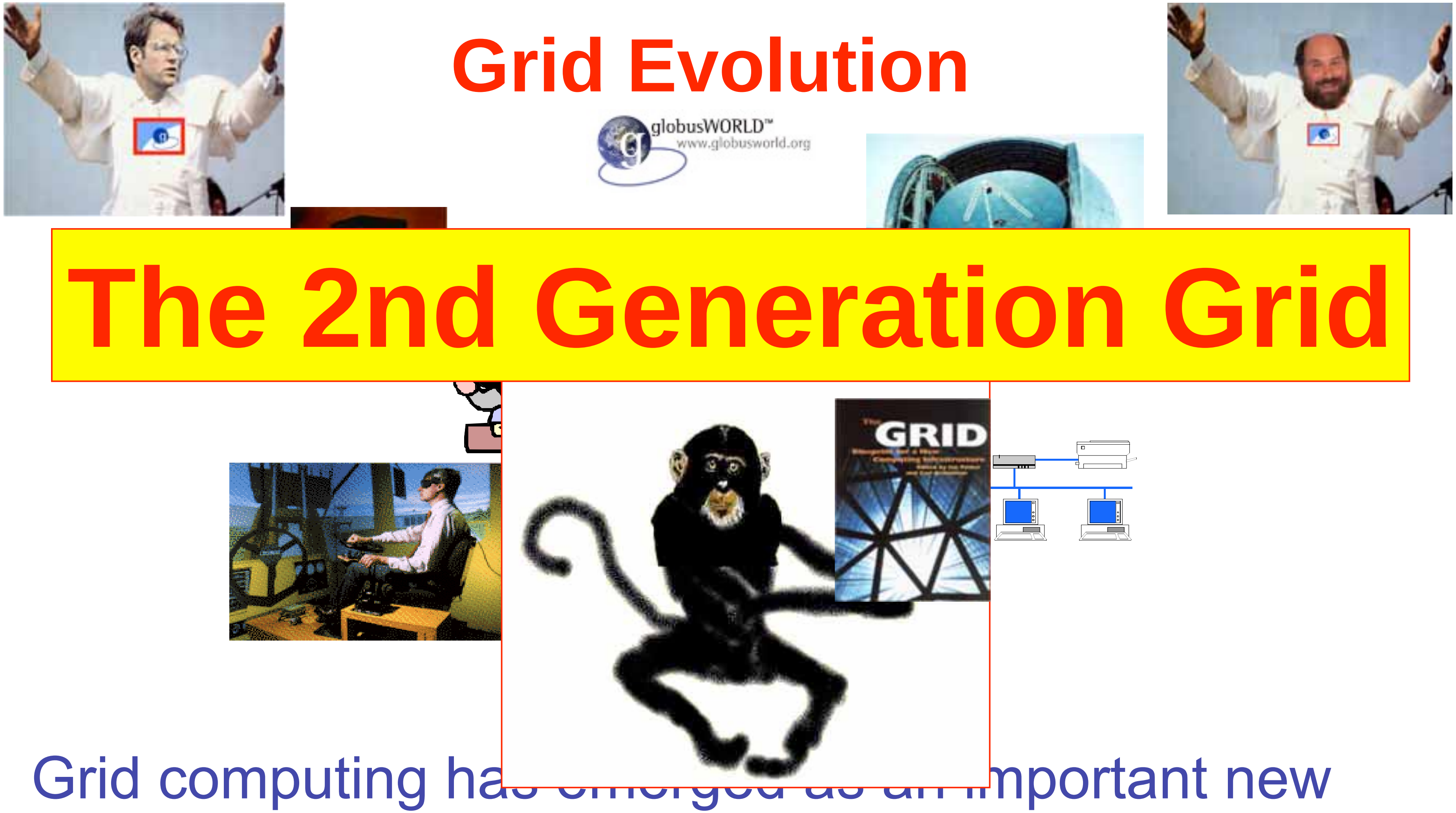
Grid Evolution - Metacomputing



The 1st Generation Grid

- Different Supercomputing Resourses
  - geographically distributed
  - used as a single **powerful** parallel machine (clear, High-Performance orientation)

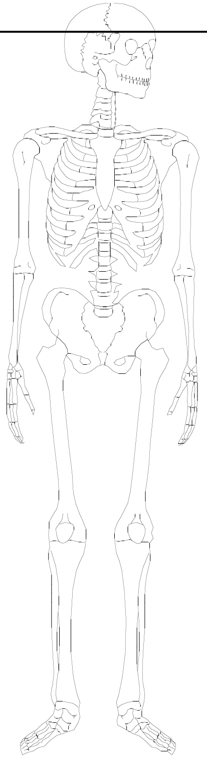
Grid Evolution

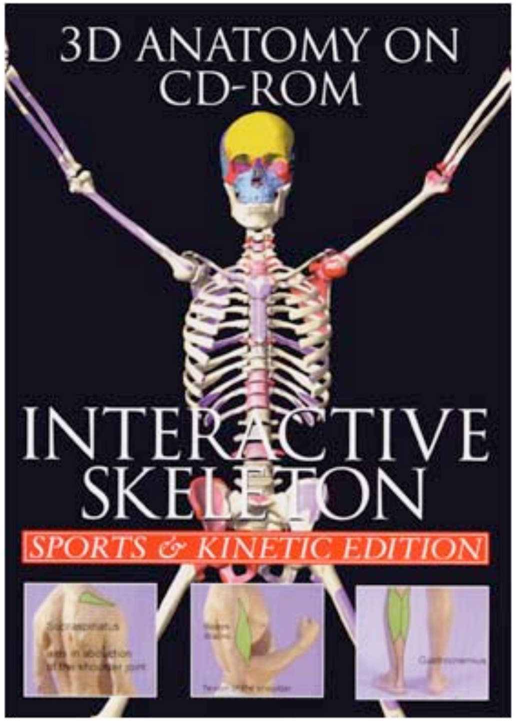


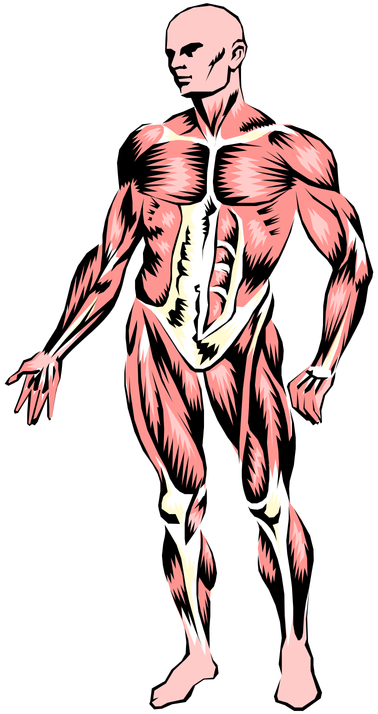
The 2nd Generation Grid

Grid computing has emerged as an important new field, distinguished from conventional distributed computing by its focus on large-scale resource sharing, innovative applications, and, in some cases, high-performance orientation.

The Anatomy of the Grid: Enabling Scalable Virtual Organizations







By Ian Foster, Carl Kesselman, and Steven Tuecke

The International Journal of High Performance Computing Applications

Volume 15, number 3, pages 200–222, Fall 2001

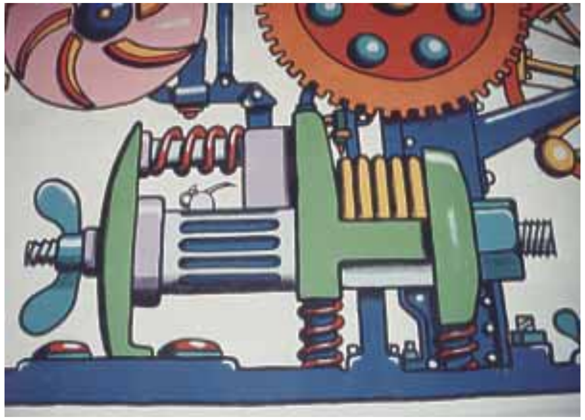
Open Question

Is the far-reaching vision offered by Grid Computing

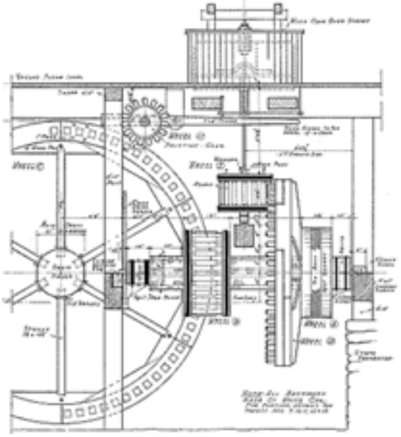
obscured by the

lack of interoperability standards

among Grid technologies ?



Interoperability



- □ Describes whether or not two components of a system that were developed with different tools or different vendor products can work together

IBM
Platform
Microsoft
The Bottom Line in Distributed & Grid Computing
AVAKI
ORACLE
NEC
hp
invent
Sun
microsystems
ENTROPIA
PC GRID COMPUTING
FUJITSU
sgi
intel

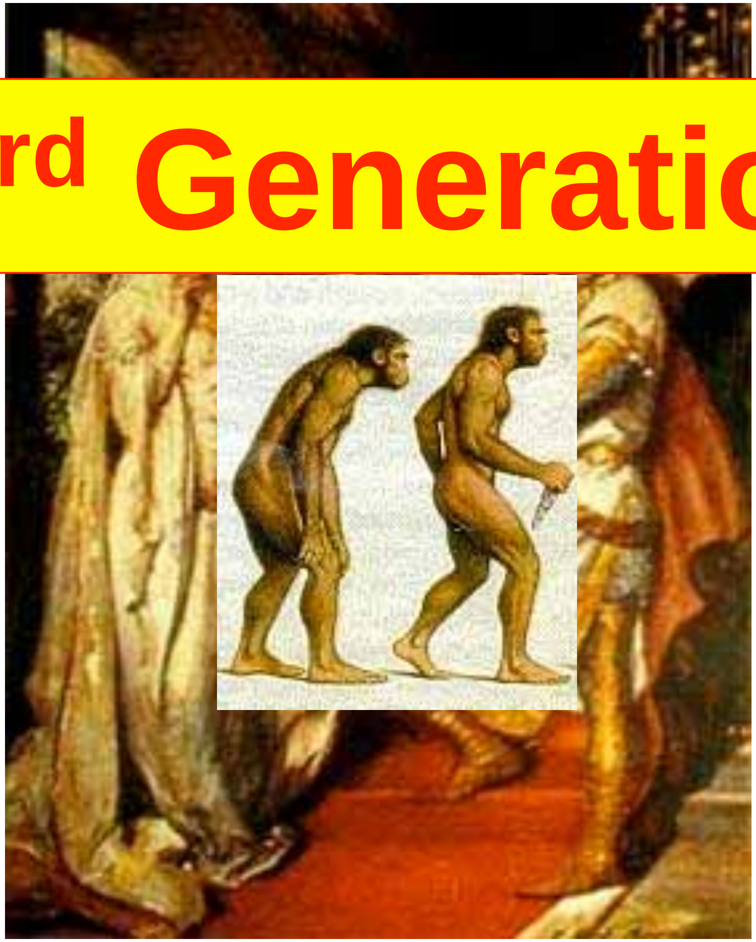
How to guarantee interoperability among Grids ?

Grid Standards & Alliances

- GGF
  - Research and Industry, use cases, architectures and specifications (OGSA, OGSI/WSRF)
- DMTF
  - Distributed Mgt. standards and models (CIM)
- OASIS
  - eBusiness & Web Services Management (WS-RF, WS-Notification, WSDM, …)
- EGA
  - Promote and grow Enterprise grid computing
- IETF
  - Internet architectures & specifications (SNMP, SMI)
- W3C
  - Web Services architectures and specifications
- SNIA
  - Advance the adoption of storage networks as complete and trusted solutions"

Grid Evolution

The 3rd Generation Grid



The marriage of the **Web technology** with the **2nd Generation Grid technology** led to new and generic Grid Services

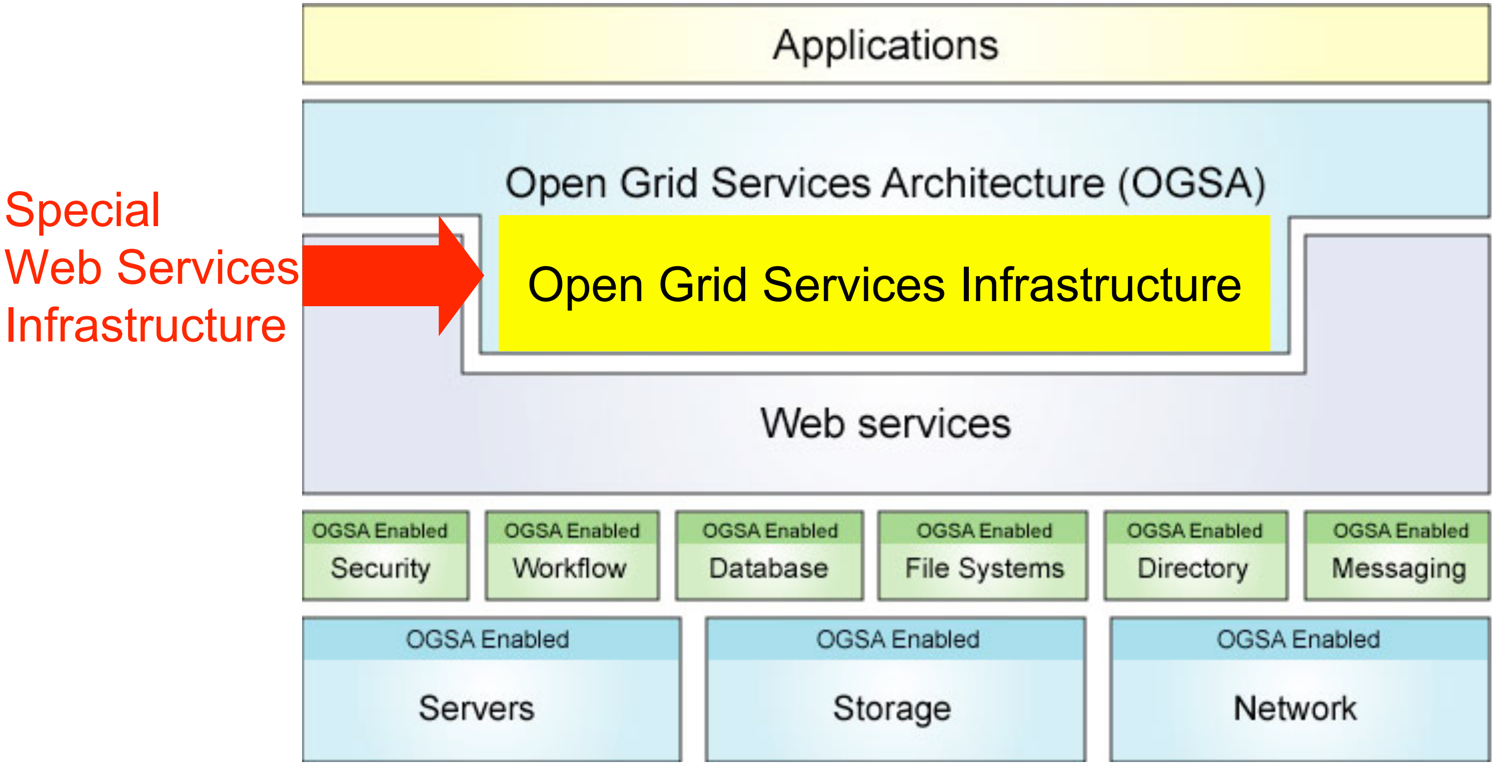
The Physiology of the Grid

An Open Grid Services Architecture for Distributed Systems Integration

I. Foster, C. Kesselman, J. Nick, S. Tuecke, January, 2002

http://www.globus.org/research/papers/ogsa.pdf

OGSA - OGSI


Special
Web Services
Infrastructure
Applications
Open Grid Services Architecture (OGSA)
Open Grid Services Infrastructure
Web services
OGSA Enabled
Security
OGSA Enabled
Workflow
OGSA Enabled
Database
OGSA Enabled
File Systems
OGSA Enabled
Directory
OGSA Enabled
Messaging
OGSA Enabled
Servers
OGSA Enabled
Storage
OGSA Enabled
Network

News

Hot News From globusWORLD™ www.globusworld.org January 20, 2004

Major Grid Services News:

The Globus Alliance and IBM in conjunction with HP announced details of the new:

WS-Resource Framework

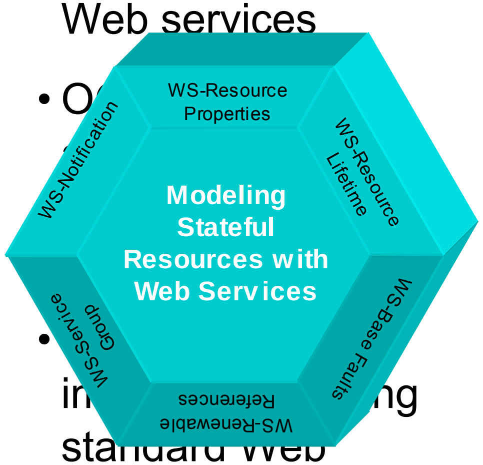
a further convergence of Grid services and Web services.

See: presentations by Daniel Sabbah of IBM and Ian Foster of the Globus Alliance for details.

How these proposals relate to OGSA

WS-Resource Framework & WS-Notification are an evolution of OGSI

- OGSA Services can be defined and implemented as Web services
- O
- in ng standard Web services development tools
- Grid applications will NOT require special Web services infrastructure



Modeling Stateful Resources with Web Services

WS-Resource Properties

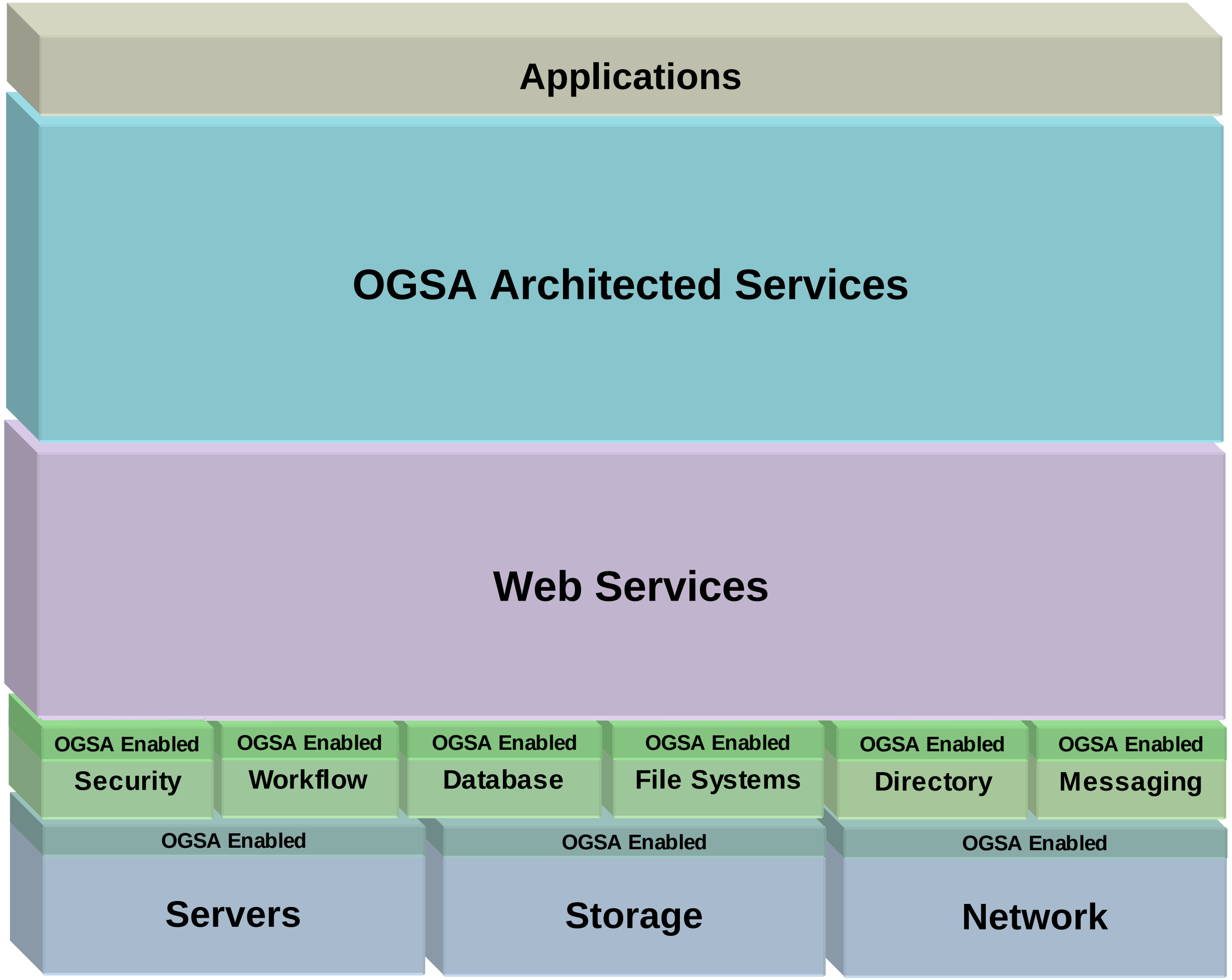
WS-Resource Lifetime

WS-Base Faults

WS-Renewable References

WS-Service Group

WS-Notification



Applications

OGSA Architected Services

Web Services

OGSA Enabled Security

OGSA Enabled Workflow

OGSA Enabled Database

OGSA Enabled File Systems

OGSA Enabled Directory

OGSA Enabled Messaging

OGSA Enabled Servers

OGSA Enabled Storage

OGSA Enabled Network

the globus consortium

January 24, 2005

About The Globus Consortium

- ❑ **The Globus Consortium - Bringing Open Source Grid Technology to the Enterprise**
  The Globus Consortium is the world's leading organization championing open source Grid technologies in the enterprise. With the support of industry leaders IBM, Intel, HP, and Sun Microsystems, the Globus Consortium draws together the vast resources of IT industry vendors, enterprise IT groups, and a vital open source developer community to advance use of the Globus Toolkit in the enterprise.

- ❑ The **Globus Toolkit** is the de facto standard for Grid infrastructure enabling IT managers to view all of their distributed computing resources around the world as a unified virtual datacenter. By giving enterprises access to computing resources as they need it, IT costs can go up and down as business demands. An open Grid infrastructure is the pre-requisite to fulfilling the promise of utility computing.

Sponsor-level members:

hp invent

IBM

intel

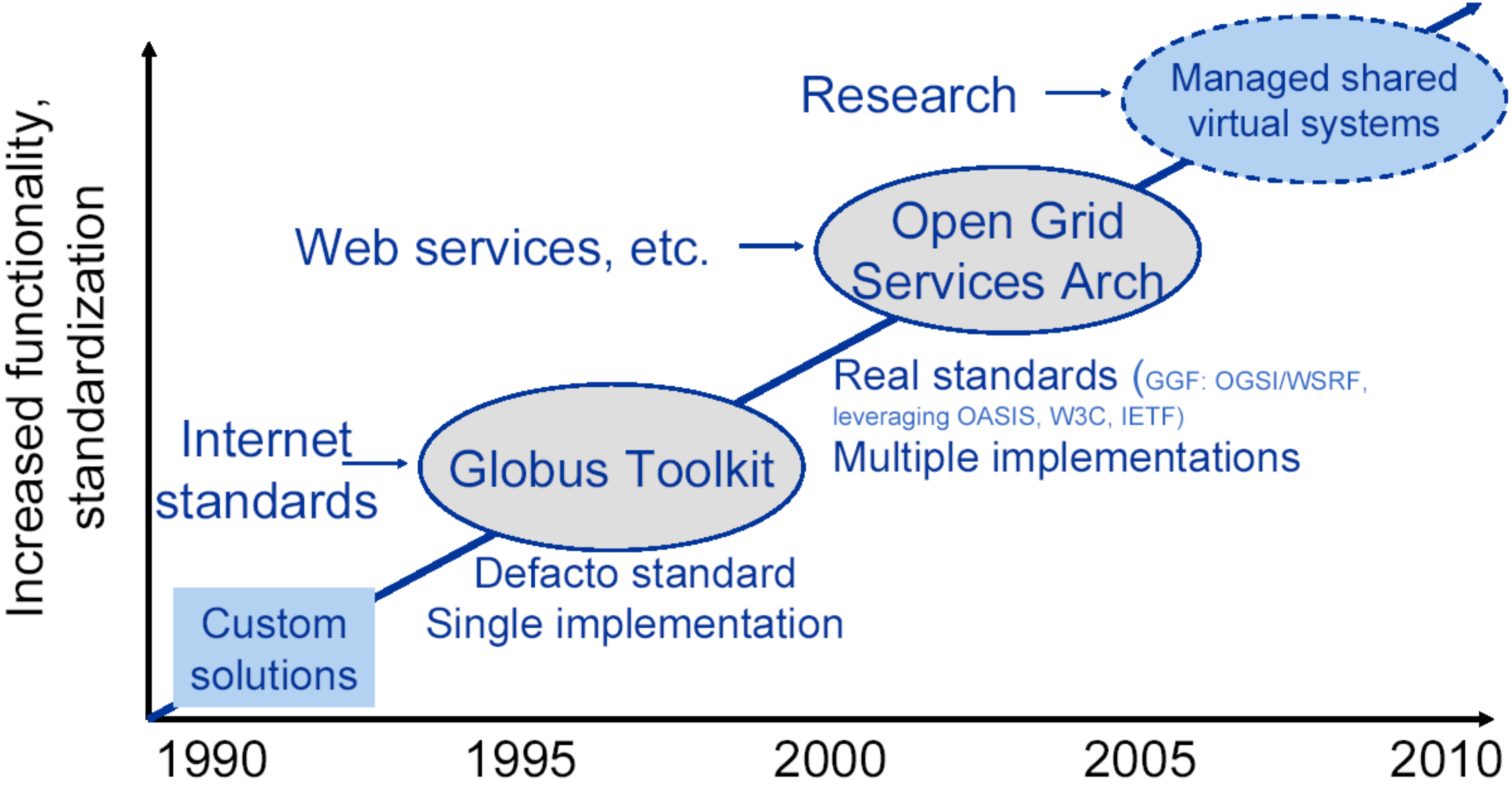
Sun microsystems

Contributor-level members:

NORTEL

Univa

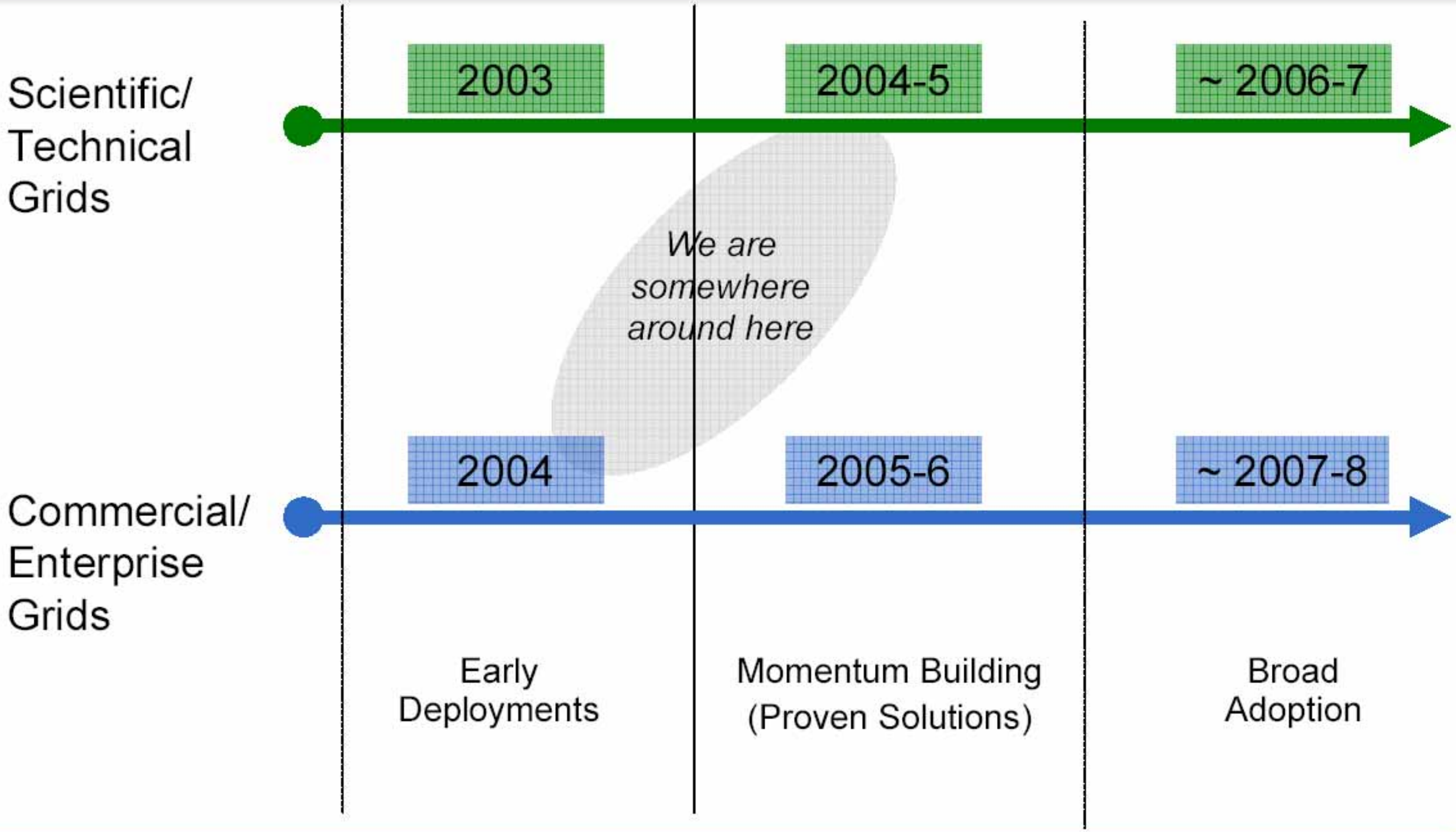
Developing Grid Standards



11

Source: Ian Foster - foster@mcs.anl.gov

Realistic Expectations

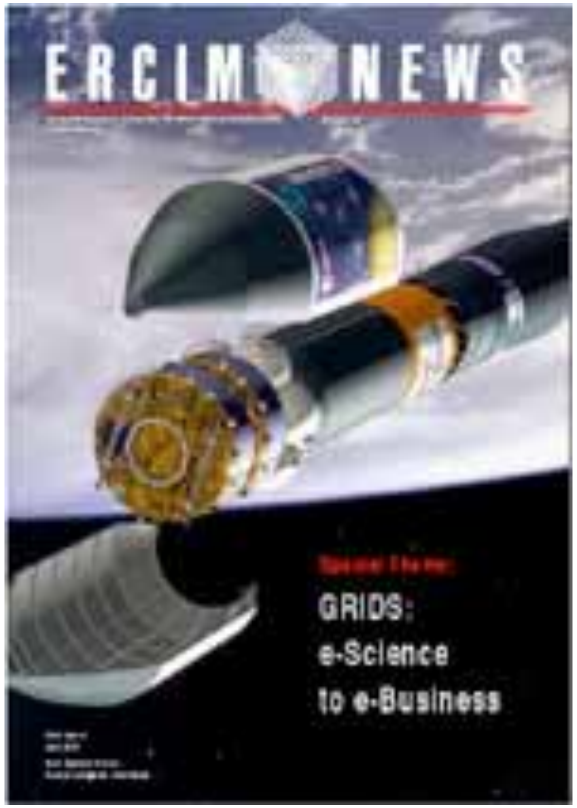


	Early Deployments	Momentum Building (Proven Solutions)	Broad Adoption
Scientific/ Technical Grids	2003	2004-5	~ 2006-7
Commercial/ Enterprise Grids	2004	2005-6	~ 2007-8

We are somewhere around here

What is boiling in the (European) pot?

ERCIM News No.59, October 2004



ERCIM News No.45, April 2001

Information Society Technologies

NGG1 and NGG2

Terms of reference

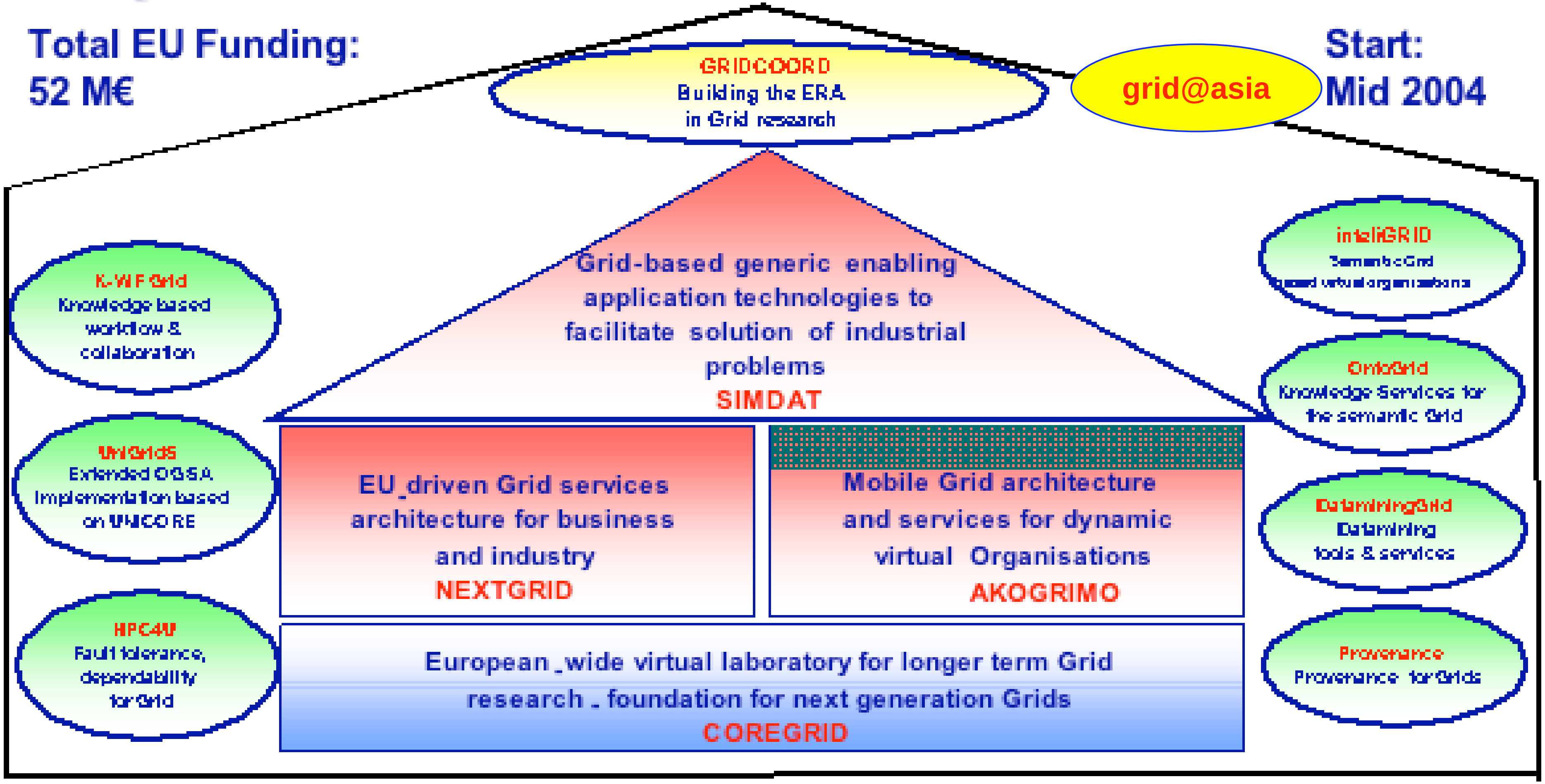
- ❑ **Identify Research Priorities**
    - ❖5 to 7 year timeframe
    - ❖Include implementation strategies
- ❑ **Propose an Implementation Roadmap**
- ❑ **Align Priorities with the European Research Agenda**
- ❑ **Network and Liaise with the Grid Community**
- ❑ **Propose actions to Improve International Collaboration**

Information Society Technologies

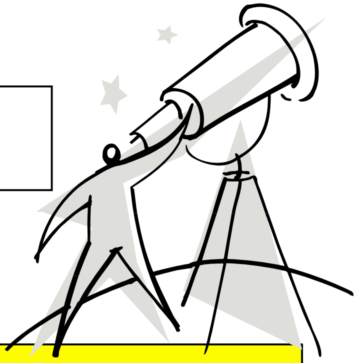
New Grid Research Projects

Total EU Funding:
52 M€

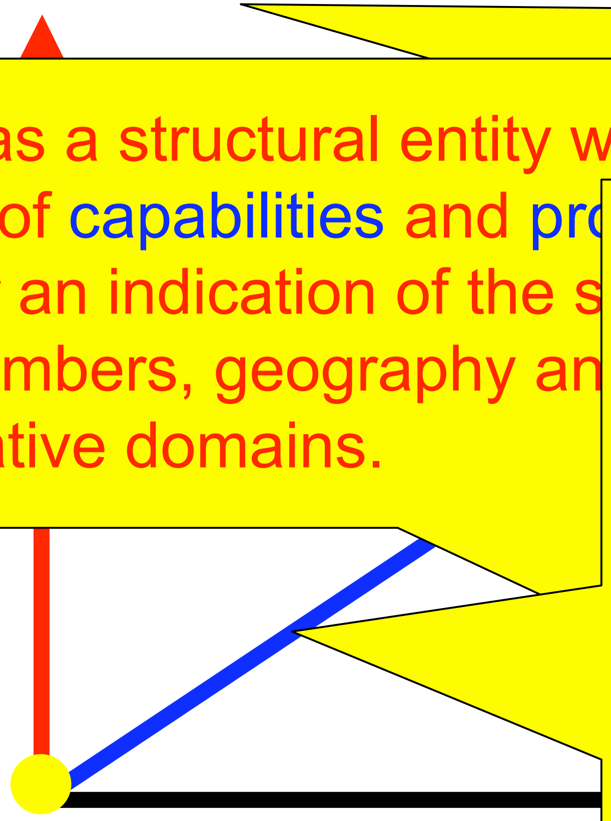
Start:
Mid 2004


GRIDCOORD
Building the ERA
in Grid research
grid@asia
Grid-based generic enabling application technologies to facilitate solution of industrial problems
SIMDAT
EU driven Grid services architecture for business and industry
NEXTGRID
Mobile Grid architecture and services for dynamic virtual Organisations
AKOGRIMO
European wide virtual laboratory for longer term Grid research - foundation for next generation Grids
COREGRID
K-WF Grid
Knowledge based workflow & collaboration
UniGridS
Extended OGSA Implementation based on UNICORE
HPC4U
Fault tolerance, dependability for Grid
inteliGRID
Semantic Grid based virtual organisations
OntoGrid
Knowledge Services for the semantic Grid
DataminingGrid
Datamining tools & services
Provenance
Provenance for Grids

NGG from 3 Different Perspectives



The end users perspective



How the Grid might be
in everyday life,
ness drives Grid

The Grid as a structural entity with a
collection of capabilities and pro
Critical for an indication of the s
term of numbers, geography an
administrative domains.

What will it be like to program the Grid? What constraints have to be observed when developing Grids?

The architectural perspective

NGG: The Wish List

- ❖ **Transparent and reliable**

- ❖ **Open to wide user and provider communities**

- ❖ **Pervasive and ubiquitous**

- ❖ **Secure and provides trust**
  - ➢ Across multiple administrative domains

- ❖ **Easy to use and to program**

- ❖ **Persistent**
  - ➢ Local and personal persistence as well as global persistence
  - ➢ Strict reproducibility

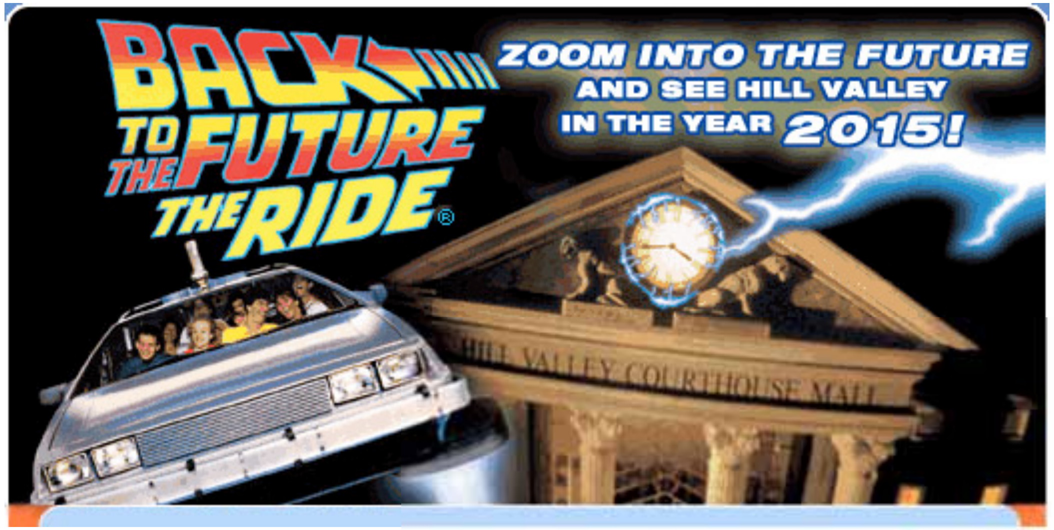
- ❖ **Person-centric**

- ❖ **Scalable and Scale Independent**
- ❖ **Easy to configure and manage**
  - – Self managing
- ❖ **Based on standards for software and protocols**

GGF OASIS IETF

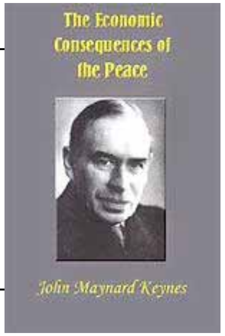
W3C WORLD WIDE WEB consortium OMG OBJECT MANAGEMENT GROUP

CCA Common Component Architecture WS-I WEB SERVICES INTEROPERABILITY ORGANIZATION



Looking into the Future

From e-Science to €-Business



❑Towards the realisation of the "invisible Grid", offering key features for A Service-oriented Knowledge Utility

- ❖ a new paradigm for software and service delivery, for the next decade.

❑**Next Generation Grids 2 - Expert Group Report**

- ❖ **http://www.cordis.lu/ist/grids/index.htm**
- ❖ **ftp://ftp.cordis.lu/pub/ist/docs/ngg2_eg_final.pdf**

Service-Oriented architecture (SOA)

Definition

http://www.service-architecture.com/web-services/articles/service-oriented_architecture_soa_definition.html

❑ A service-oriented architecture is essentially a collection of services.

❑ A service is a function that is well-defined, self-contained, and does not depend on the context or state of other services.

❑ These services communicate with each other.

❑ The communication can involve either simple data passing or it could involve two or more services coordinating some activity.

Service-Oriented architecture (SOA)

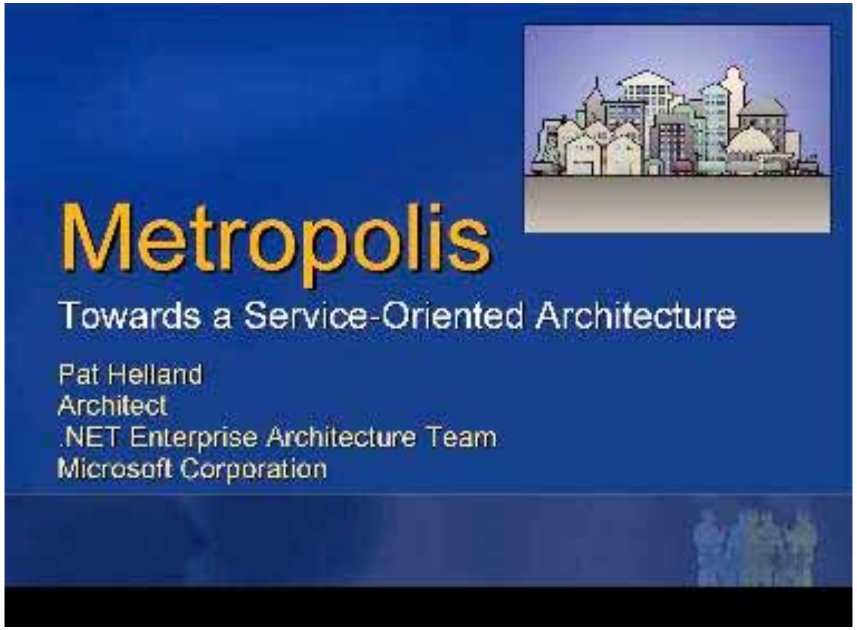
Definition

http://msdn.microsoft.com/architecture/soa/default.aspx

- ❑ The goal for Service Oriented Architecture (SOA) is a world-wide mesh of collaborating services that are published and available for invocation on a Service Bus.
- ❑ Adopting SOA is essential to delivering the business agility and IT flexibility promised by Web Services.
- ❑ These benefits are delivered not just by viewing service architecture from a technology perspective or by adopting Web Service protocols, but also by requiring the creation of a Service Oriented Environment that is based on specific key principles.

Metropolis : Envisioning the Service-Oriented Enterprise

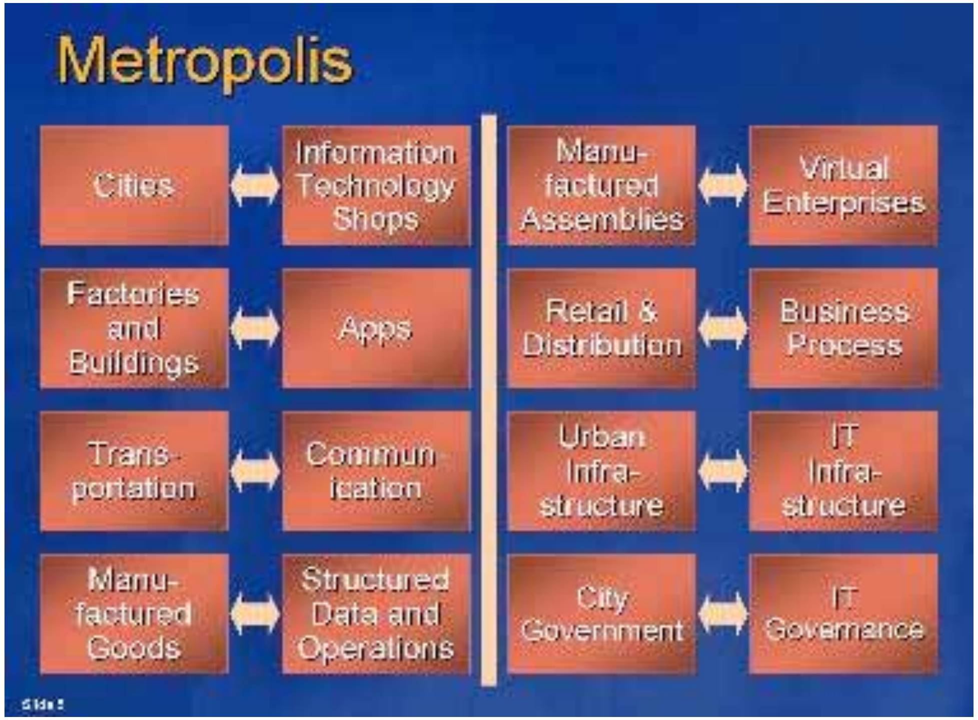
http://msdn.microsoft.com/seminar/shared/asp/view.asp?url=/architecture/media/en/metrov2_part1/manifest.xml



Metropolis

Towards a Service-Oriented Architecture

Pat Helland
Architect
.NET Enterprise Architecture Team
Microsoft Corporation

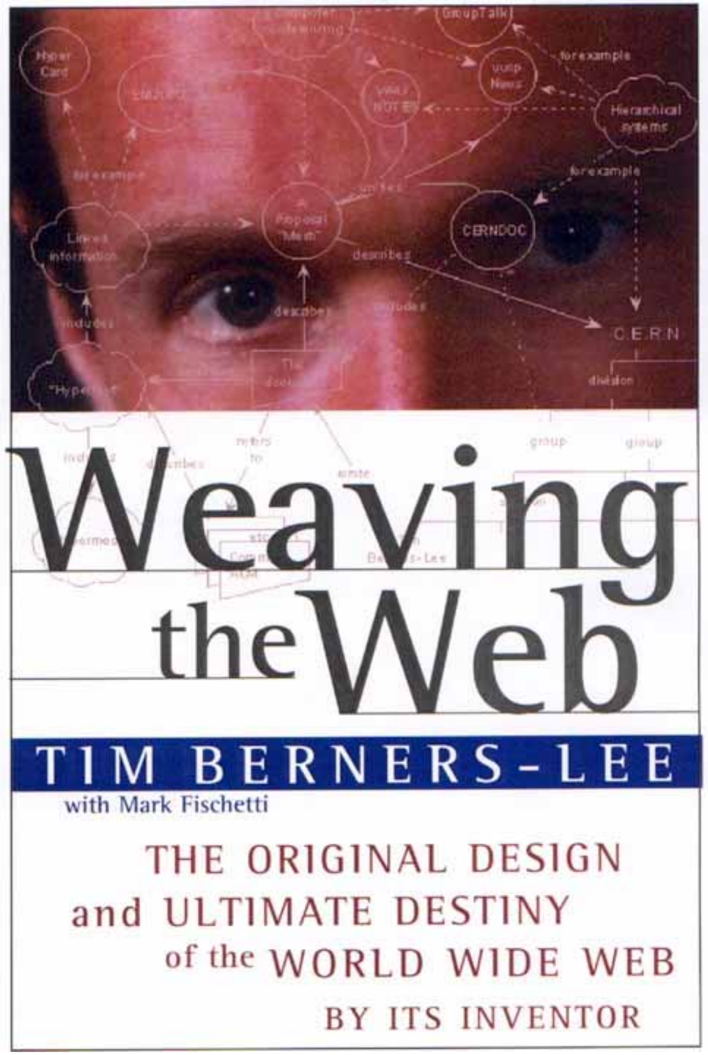


Metropolis

Cities ⇔	Information Technology Shops	Manufactured Assemblies ⇔	Virtual Enterprises
Factories and Buildings ⇔	Apps	Retail & Distribution ⇔	Business Process
Transportation ⇔	Communication	Urban Infrastructure ⇔	IT Infrastructure
Manufactured Goods ⇔	Structured Data and Operations	City Government ⇔	IT Governance

Slide 5

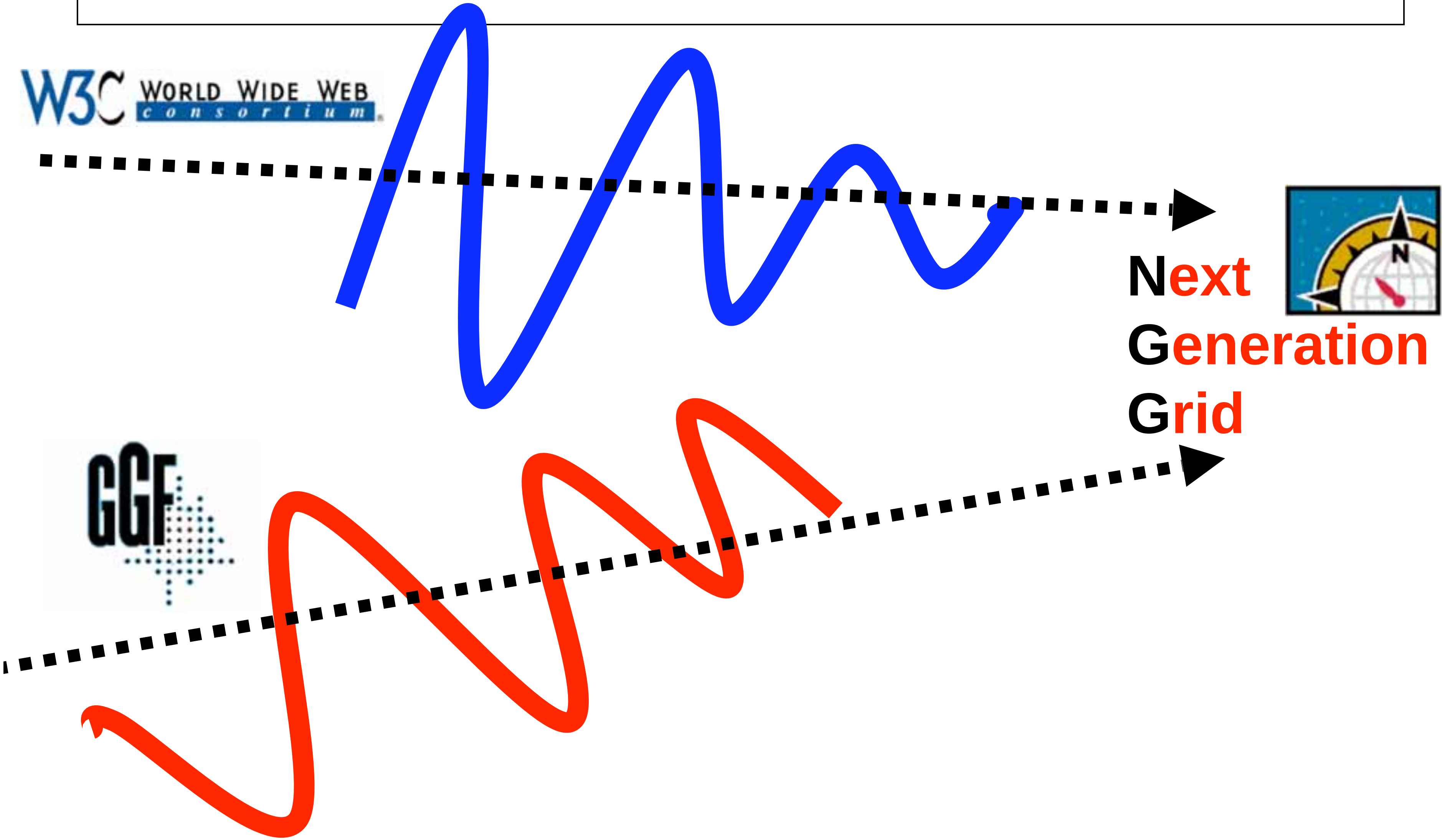
Semantic Web


Weaving the Web
TIM BERNERS-LEE
with Mark Fischetti
THE ORIGINAL DESIGN and ULTIMATE DESTINY of the WORLD WIDE WEB BY ITS INVENTOR

1999

- ❑“In the first part, the Web becomes a much more powerful means for collaboration between people …In the second part of the dream, collaborations extend to computers . ….
- ❑A ‘Semantic Web’ which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy, and our daily lives will be handled by machines talking to machines, leaving humans to provide the inspiration and intuition. . . The first step is putting data on the Web in a form that machines can naturally understand, or converting it to that form.”

Convergence of Interests



Next
Generation
Grid

Convergence is a need !

The long-awaited technologic convergence of the communications industry.

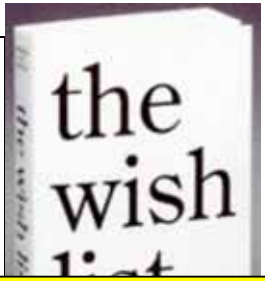
Mandatory

No Standard… ?

No Industrial/ Business Interest !

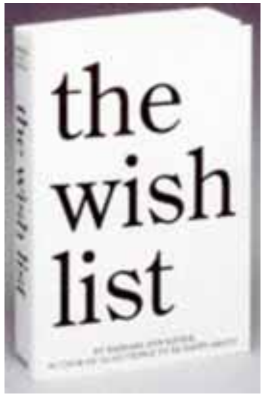
GGF
OASIS
IETF
WS-I
WEB SERVICES INTEROPERABILITY ORGANIZATION
CCA
Common Component Architecture
W3C
WORLD WIDE WEB consortium
OMG
OBJECT MANAGEMENT GROUP

Next Generation Grid Properties



The current Grid implementations DO NOT individually possess all of these properties

Future Grids NOT possessing these properties are unlikely to be of significant use and, therefore, inadequate from business perspectives



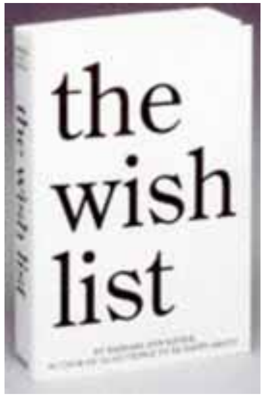
Performance and Dependability are key properties for NGG, but they are perceived as contrasting properties:

1)Long periods of grid services unavailability impact on performance

2)Techniques for resiliency may introduce overheads

Performability of grids is a holistic approach that has to include also security and business concerns

Challenges for performable grid systems and services

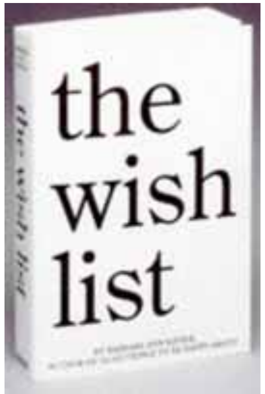


1. Standardization

- ❖ Definition of standards for metrics, models, modeling languages and formalisms
- ❖ Definition of benchmarks
- ❖ Independent approaches determine different means and tools for metrics and models
- ❖ Dominant projects that dictate standards, not necessarily have the best approach to performance and dependability

➢ Role of GGF W3C WORLD WIDE WEB consortium

and of the other standard bodies

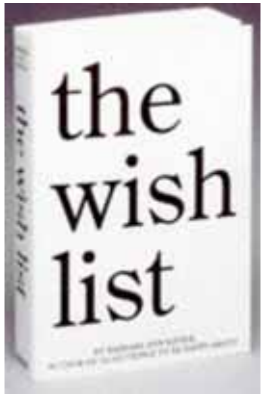


2. Virtualization

Virtualization enables a service to be offered seamlessly without awareness of what underlying services are used, their location, who provides them and if are used by others:

Hierarchy of services that can be managed as atomic entities, but introduce many problems from a modeling and measurement point of view:

- It is impossible to determine what resources are being used; different uses of the same service can be made by distinct sets of resources
- If a resources is overused, a task can be migrated to an alternative with different non-functional properties
- Different services may employ the same set of underlying services, becoming correlated and affected by common mode failures
  - this is a problem in both analysis and in design for deciding where and when using resilience techniques
- Difficult prediction of resource’s workload
  - on-line monitoring of resources but role of interdependencies
- Complexity of models of system behavior
- Little work on this issue

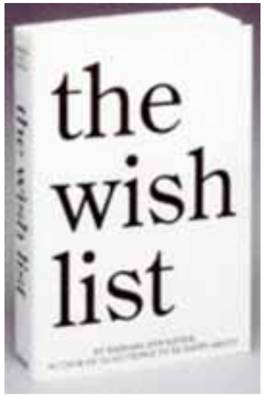


3. Measurement of complex systems

The size of grid systems, their heterogeneity and dynamicity create problems for performability analysis.

- **What to measure and where to measure**
- **Model-based evaluation of large complex systems will have to cope with large state spaces**
- **Simulation will have unacceptable run times**
- **Analytical models of complex systems, if available, are very costly to solve**

- **Need of techniques for efficient solutions of large models and for finding simple approximations**
- **Production of trustworthy approximations and verifiable techniques for model simplification**

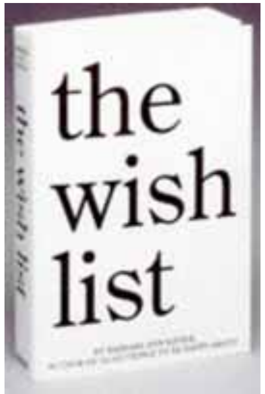


4. Resource management

Effective management of resources is a key part for providing QoS to customers; managing performability requires up to date knowledge of the state of the system operation:

- ➢ **Being entirely up to date is unreasonable**
- ➢ **Performance may be increased if the choice of where directing a particular request is based on the best information available**
- ➢ **Predictive mechanisms:**
  - **efficient decomposition techniques**
  - **accurate approximations**
  - **scenario specific heuristics**

- ➢ **Identification of quasi-optimal policies and their evaluation**
- ➢ **Application oriented easily usable mechanisms**



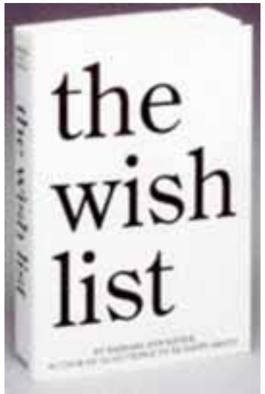
5. Realistic parameterization of systems

Performability models are only as good as the data that is used to populate them. If performance or availability is predicted on a conservative estimate for user demand then the system may have too little capacity and a far poorer expected performability

It is important to have accurate information *on demand* and for proposed models to be accurately verified against real data

Quite apart some work on grid scheduling, still much is to be done for:

- providing the right level of information across a wide range of systems in an accurate and timely manner
- providing new applications with accurate historical data from similar applications to be able to make accurate performability predictions



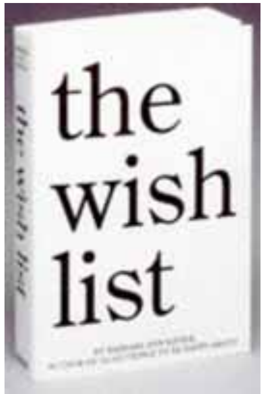
6. Business metrics

- Real metrics of interest are financial
- Increasing performability introduces costs

there is a need for a trade-off

- Grid systems are not simply a technical solution, but rather a different way of organizing business
- The core model is going to be a business process model and the technical models are going to be add-ons to this
- Need of understanding of charging models and their impact on user behavior

The relationship between charging and performability is very complex



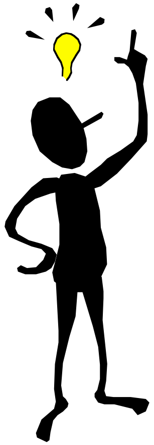
7. Performance and security

- Grid systems involve sharing of large set of personal data some of which very valuable
- Protection of data is a key issue
- Making open systems secure is difficult and can introduce large unwanted overheads
- Some users may privilege performance over security and decide to turn off security measure
- Even if security developers do not consider performability as orthogonal to security, for sure, it is a secondary consideration for them.

Much work has to be done:

- to define acceptable trade-offs between security and performability
- to identify accurate even if approximate measures of security

More Research is needed…



- introduction of performability services
- understanding, integration of all these viewpoints and their absorption into standards

More international cooperation is needed….

Session 1.2

Practice and Experiments

Moderator and Rapporteur

Jaynarayan H. Lala, Raytheon Company, Arlington, VA, USA

Grid Computing

Customer Interest, Expectation, and Requirement for Grid in Dependability Context

48th Meeting of IFIP Working Group 10.4

Takanori Seki, Distinguished Engineer
Technical Sales Support, IBM Japan

30 (4*7.5)

Contents

- Customer Expectation to Grid
- Roadblocks for Grid Implementation
- Grid with Reasonable Dependability

Customer Expectation to Grid

- Many customers expects Grid as
  - Platform for a wider variety of applications
    - Small enterprise HPC market
    - Transactional and e-business applications
  - Transparent adoption to applications
    - Less/no application modification
    - Transparent migration from current assets
  - Grid benefits >> Current tech implementation
    - Faster execution, higher throughput, lower IT costs etc.
    - Substantial benefits needed for new platform
      - Faster and cheaper implementation with open computing

Customer Expectation to Grid

- Many customers expects Grid as
  - Reasonable dependability environment
    - Availability with high availability or disaster recovery
    - Policy-based service level or expected service level
      - Only run in batch window/expected response time
      - Allocate resource for you anytime
    - Simple maintenance ability like single system
      - No more complexity
    - Secure like dedicated resources
  - Comparison to current platform
    - If not equal or better, good excuse not to adopt
  - Do not care standards yet
    - Within enterprise

Roadblocks for Grid Implementation

- IT Silo
  - Application platform dependence
    - Fairly connected with OS/database/middleware
  - Application-specific system management
    - System monitoring/operation
    - High availability and disaster recovery
- Non-IT Silo
  - Financial
    - IT budget allocated to each end user (Business owner, not IT dept.)
  - Organizational
    - No incentive to share as culture
- Enterprise IT optimization initiative needed
  - CEO/CIO high priority issue
  - Enterprise Architecture/IT Governance

Grid with Reasonable Dependability

- Grid as Enterprise-wide initiative
  - Not only tech. but total IT governance initiative
  - Restoration of the mainframe-idea but virtual
    - Total system management/IT resource optimization
    - User does not care the infrastructure, but application only
- Great benefits
  - With reasonable dependability
  - Open computing had great benefits but reasonable dependability
    - Quick implementation, cheap HW/SW, rich/interactive GUI
- Approach could be
  - As a part of enterprise optimization direction
  - Packaged solution even only for a single application
  - Almost middleware supports Grid (cross organization feature)
    - Open Standard maturity needed

ON DEMAND BUSINESS™

Japanese Business Grid Project Objectives & Key Technical Issues

IFIP Conference, July 2005

Nobutoshi Sagawa (Hitachi Ltd)
Toshiyuki Nakata (NEC Corporation)
Hiro Kishimoto (Fujitsu Ltd)

Thanks to all the teams in the BUSINESS GRID COMPUTING PROJECT

Contents

- Objectives
- Business Grid Middleware
- Demonstration Screen Shots
- Relevant Standardization Efforts
- Things to Do

Contents

- Objectives
- Business Grid Middleware
- Demonstration Screen Shots
- Relevant Standardization Efforts
- Things to Do

Business Grid Consortium

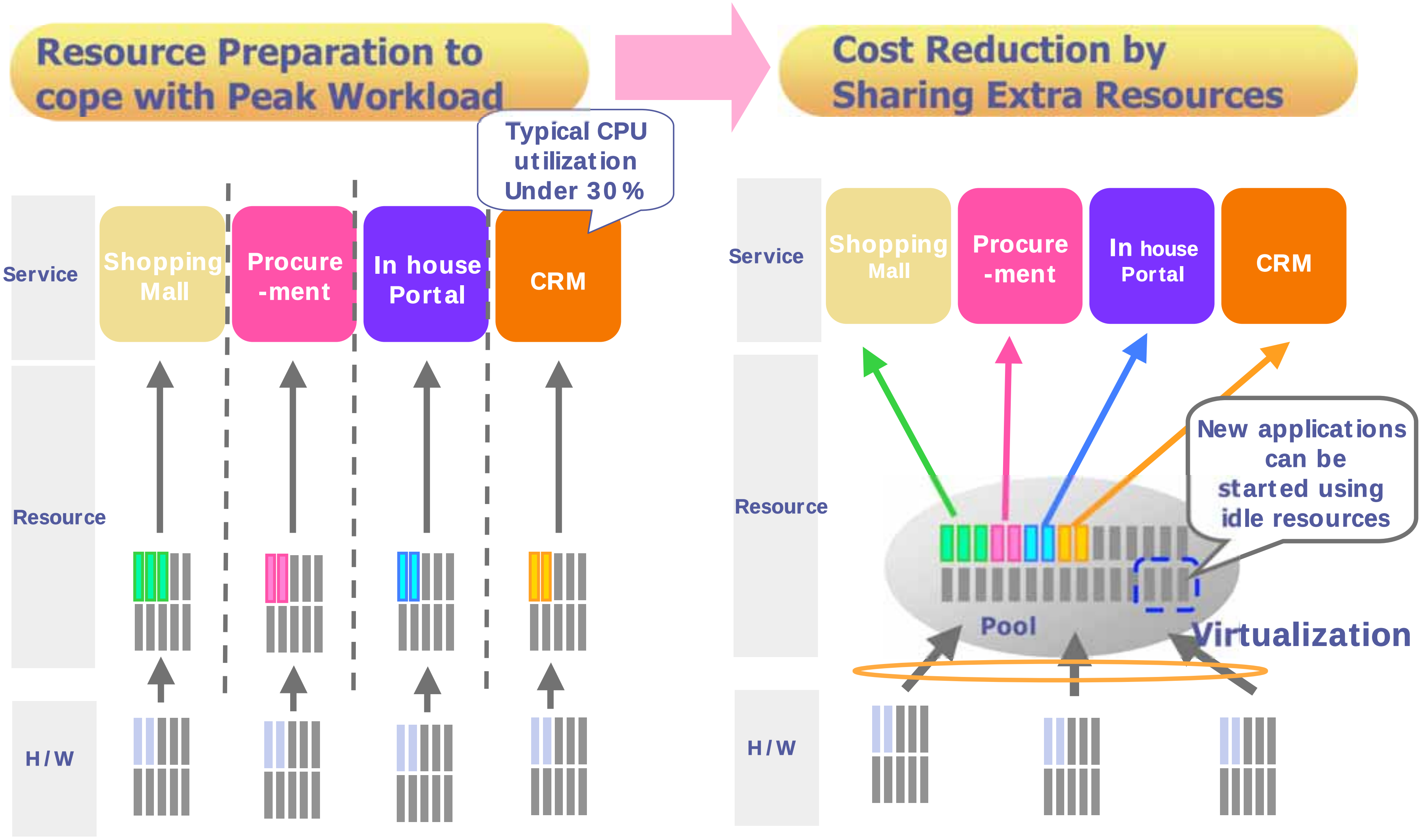
- Mission: Develop Business Grid middleware
  - Next generation business application infrastructure
  - Contribute to international standardization
- Three year project: 2003 - 2005
- Industry Members: Fujitsu, Hitachi, and NEC
- Collaboration with Grid Technology Research Center
  - Agency of Industrial Science and Technology (AIST)
- Matching funds from the METI
  - About half of the funding is from METI
- Coordinated by IPA (Information-technology Promotion Agency, Japan)
- Distribute resulting components as high-quality open-source
- Two main objectives:
  - Objective 1: Reduce IT Infrastructure Costs
  - Objective 2: Support Business Continuity

METI: Ministry of Economy, Trade, and Industry

Objective 1: Reduce IT Infrastructure Costs

- Better utilization of IT resources
  - Optimal and dynamic resource allocation
  - Share available resources
- Integrated management of heterogeneous environment
- Automate System Management
  - Simplify the job of system administrators
  - Reduce human errors
- Lower overhead of trying out new business
  - Set up new services at low initial cost
  - And scale them up easily if successful
- Enable resource sharing among multiple organizations

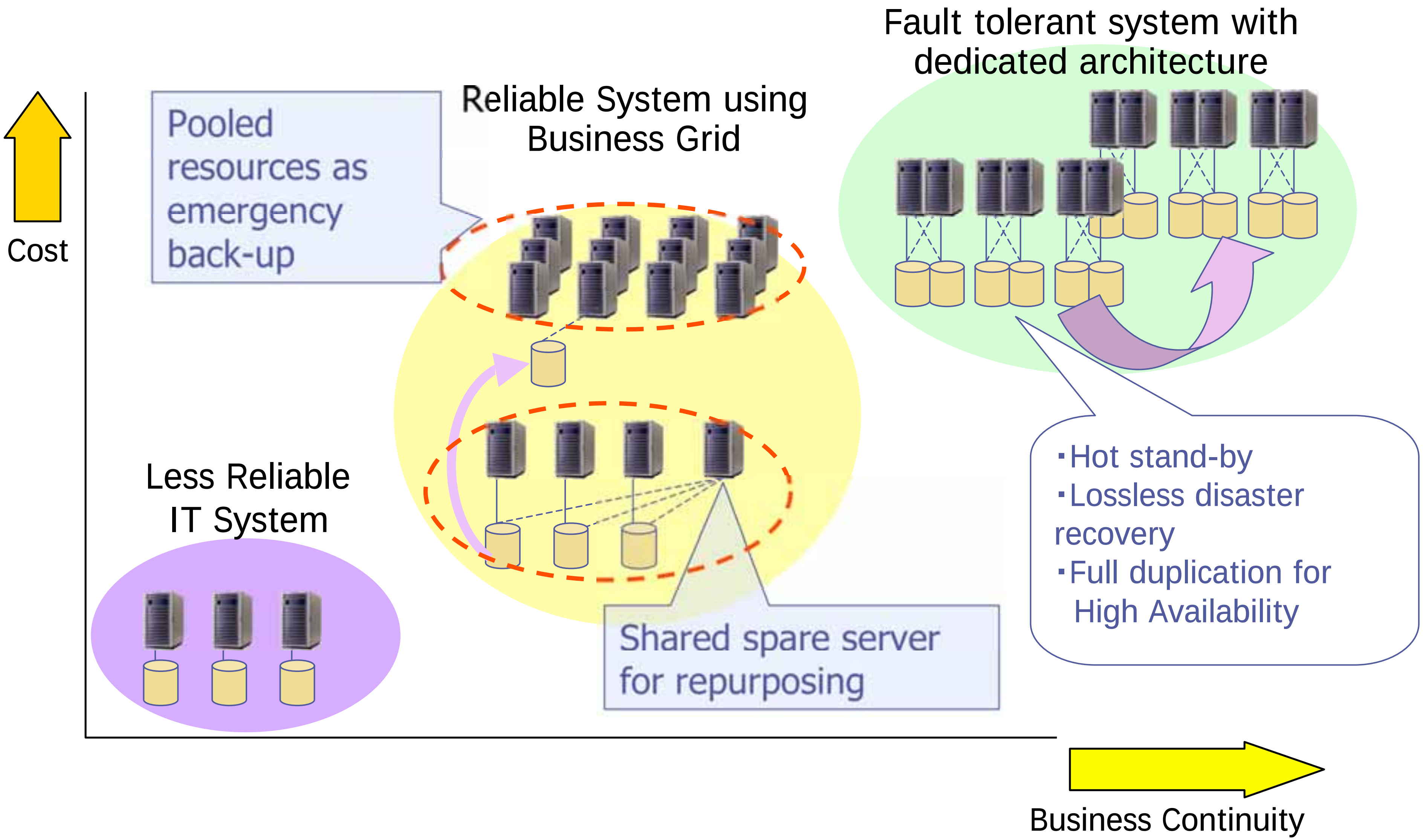
Objective 1: Reduce IT Infrastructure Costs


Resource Preparation to cope with Peak Workload
Typical CPU utilization Under 30%
Service
Shopping Mall
Procure -ment
In house Portal
CRM
Resource
H/W
Cost Reduction by Sharing Extra Resources
Service
Shopping Mall
Procure -ment
In house Portal
CRM
Resource
New applications can be started using idle resources
Pool
Virtualization
H/W

Objective 2: Support Business Continuity

- Robust IT environment
  - Respond to unexpected load spikes
- Reliable IT environment
  - Standards-based support for disaster recovery at reasonable cost
    - Database replication
    - Failover to remote site

Objective 2: Support Business Continuity

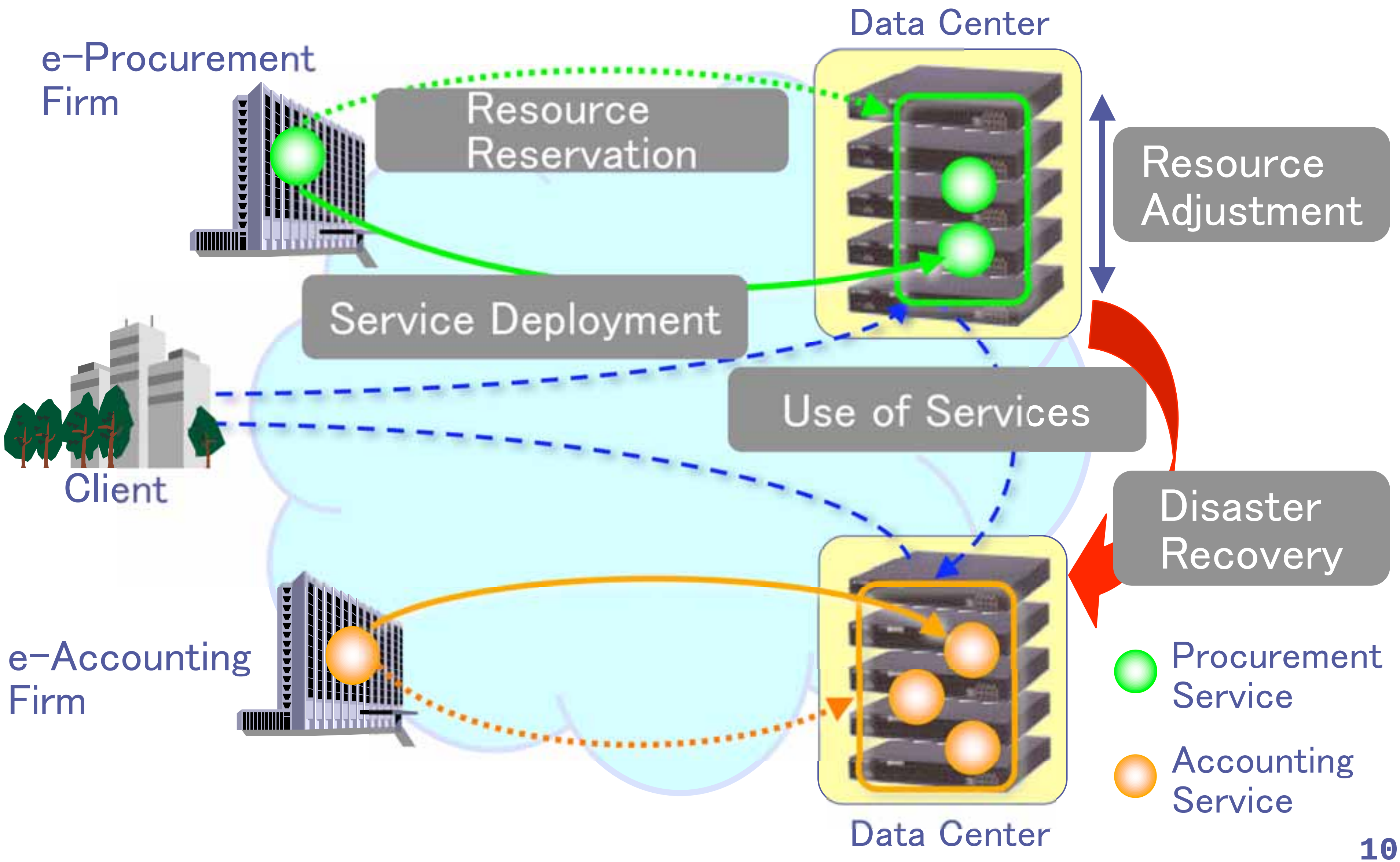

Fault tolerant system with dedicated architecture
Reliable System using Business Grid
Pooled resources as emergency back-up
Cost
Less Reliable IT System
Shared spare server for repurposing
Business Continuity

- Hot stand-by
- Lossless disaster recovery
- Full duplication for High Availability

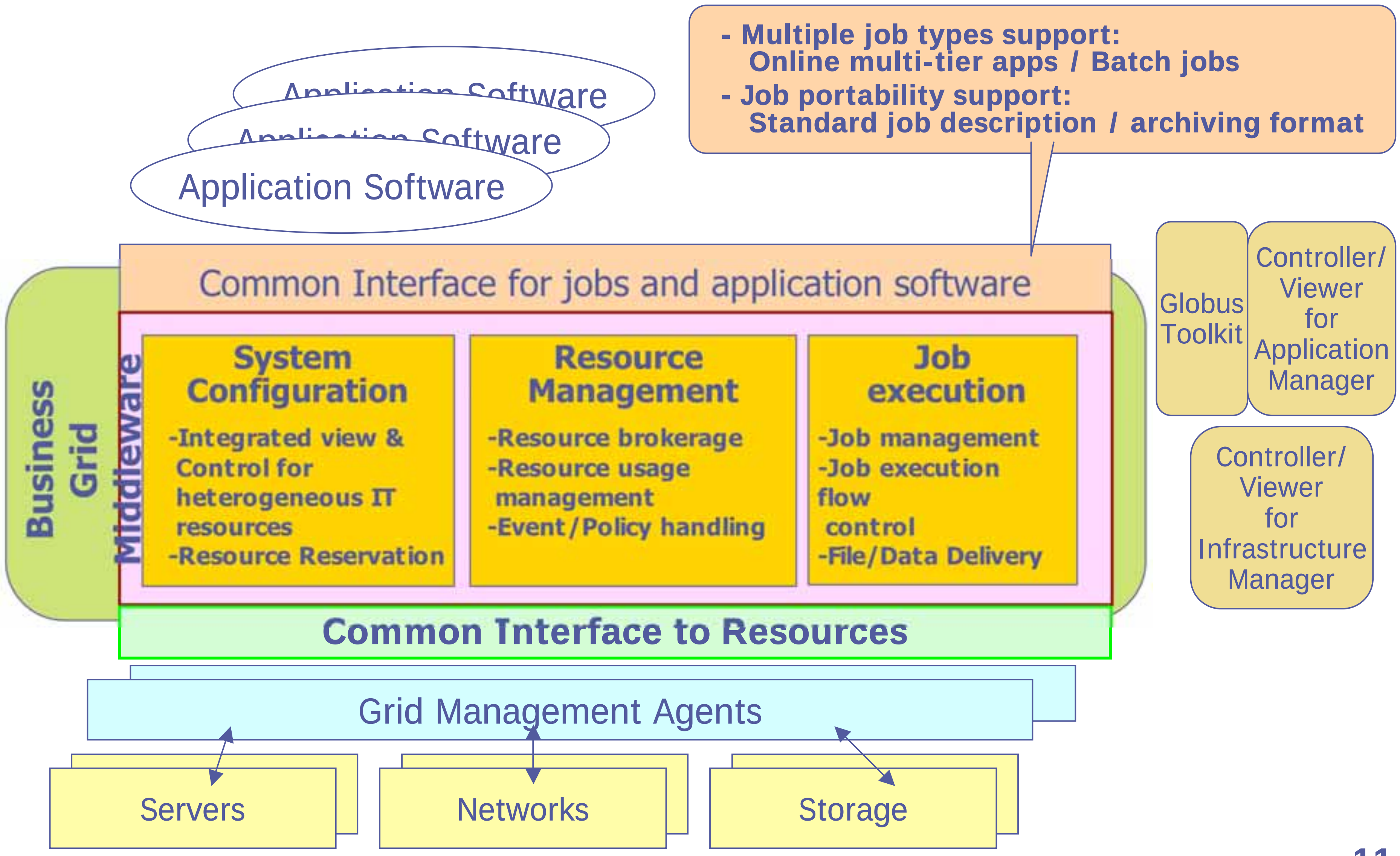
Contents

- Objectives
- Business Grid Middleware
- Demonstration Screen Shots
- Relevant Standardization Efforts
- Things to Do

What Needs to be Supported by the Middleware



Business Grid Key Components

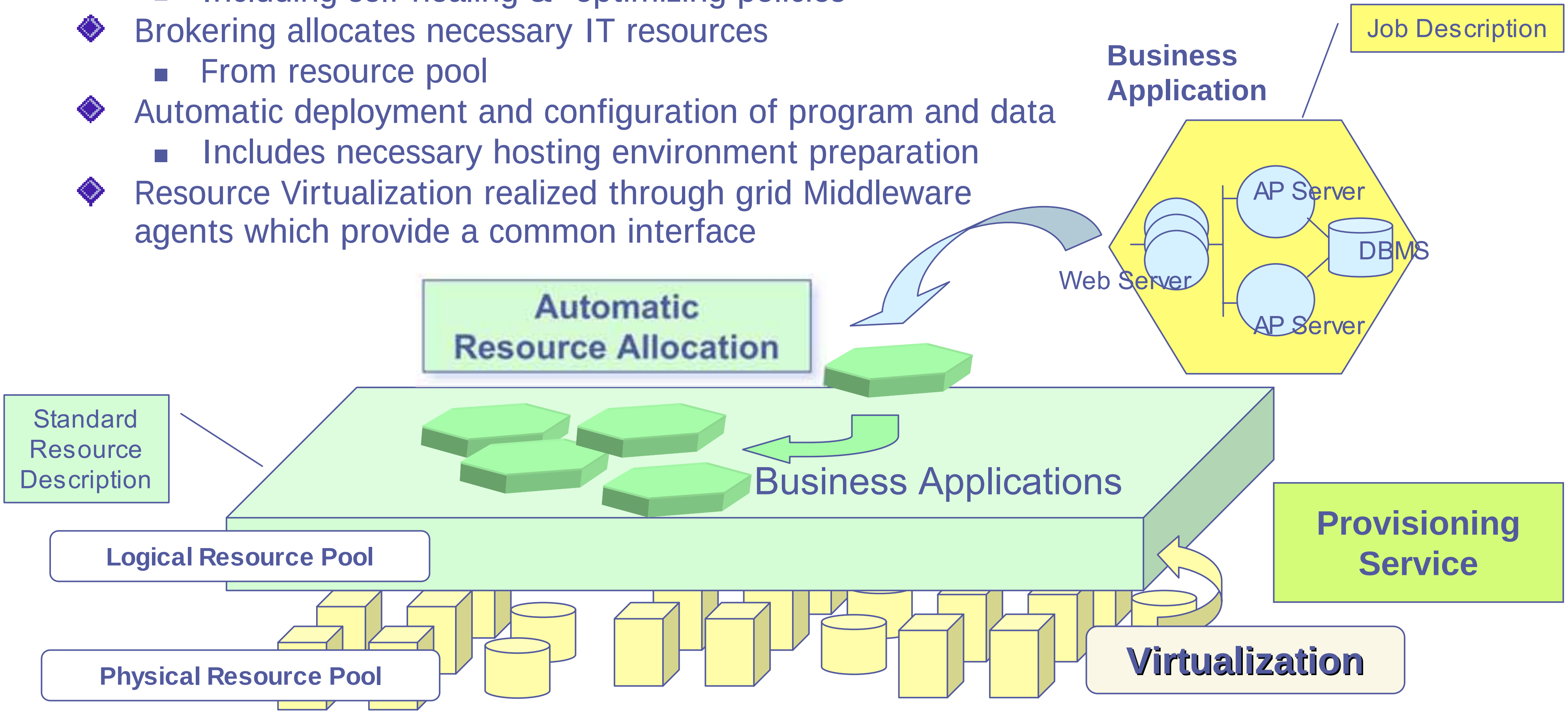


- Multiple job types support:
  Online multi-tier apps / Batch jobs
- Job portability support:
  Standard job description / archiving format

11

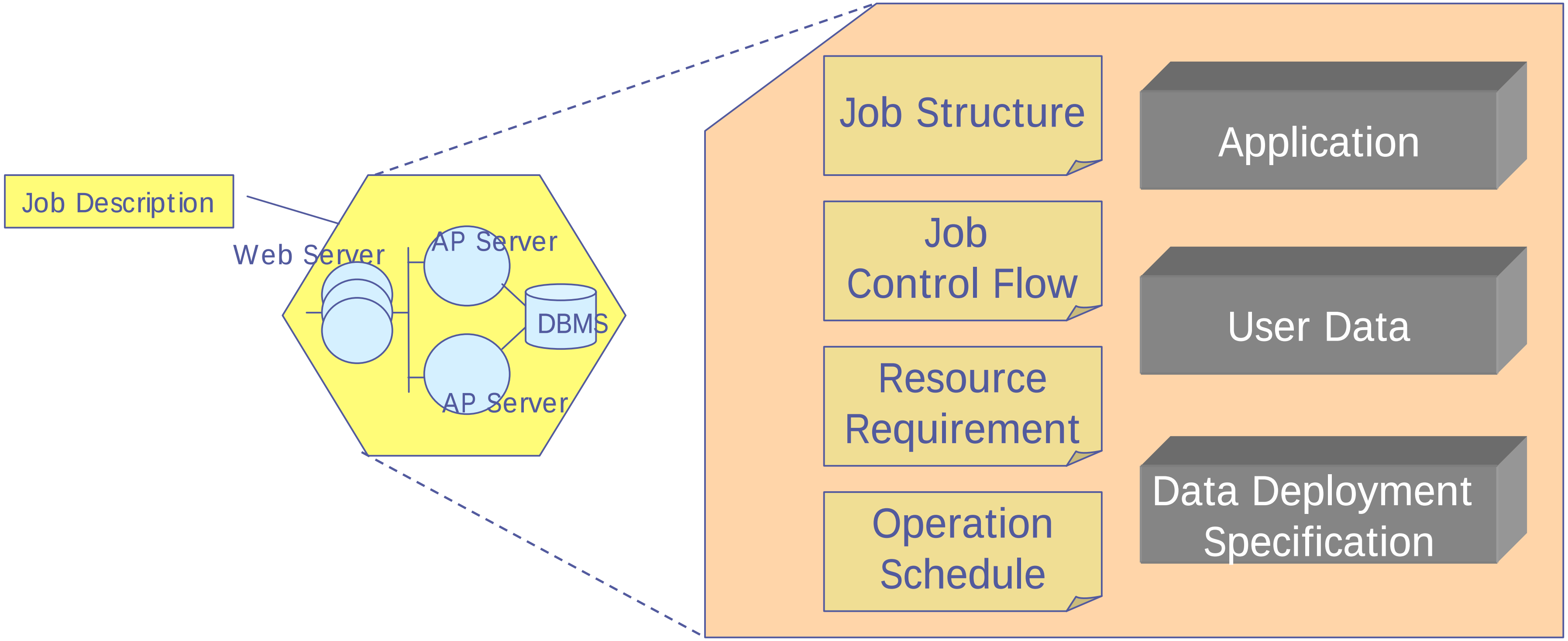
Big Picture - how it works -

- Job Submission
  - Standard job description and application contents service (using WS-Agreement protocol)
  - Including self-healing & -optimizing policies
- Brokering allocates necessary IT resources
  - From resource pool
- Automatic deployment and configuration of program and data
  - Includes necessary hosting environment preparation
- Resource Virtualization realized through grid Middleware agents which provide a common interface



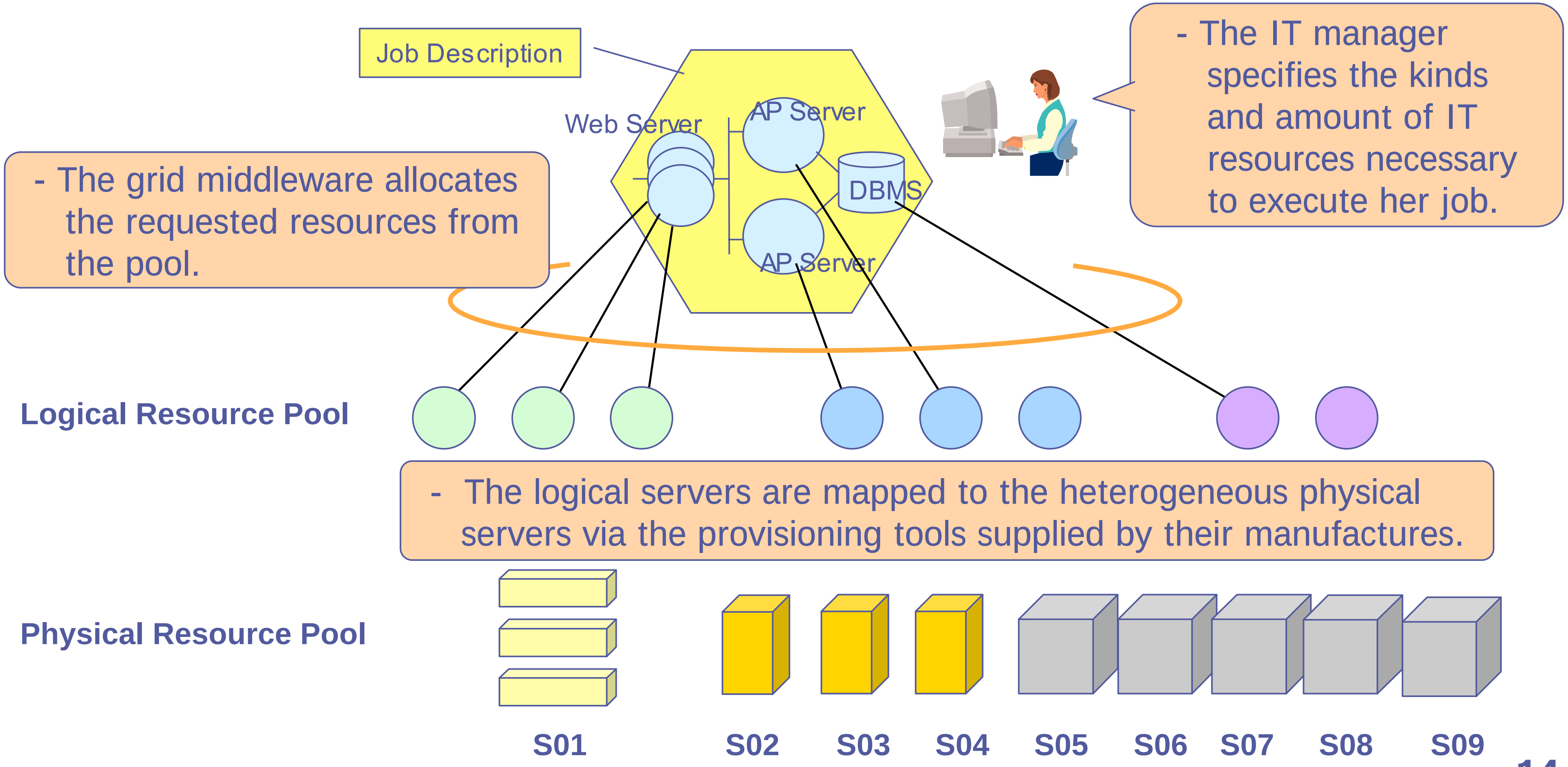
Job Description

- The job description in BizGrid not only archives the relevant execution modules, but also maintains all necessary information in one package, in order to manage the entire lifecycle of the operation.
- The description contains the specification of job structure (e.g. 3-tier Web App). It enables mapping between the job and virtualized resources, automatic deployment of execution modules and autonomic control of the resource allocation.



Optimum Resource Allocation

- The grid middleware finds and allocates the optimum amount and kinds of resources from the virtualized pool, enabling increased resource utilization with minimum human intervention.



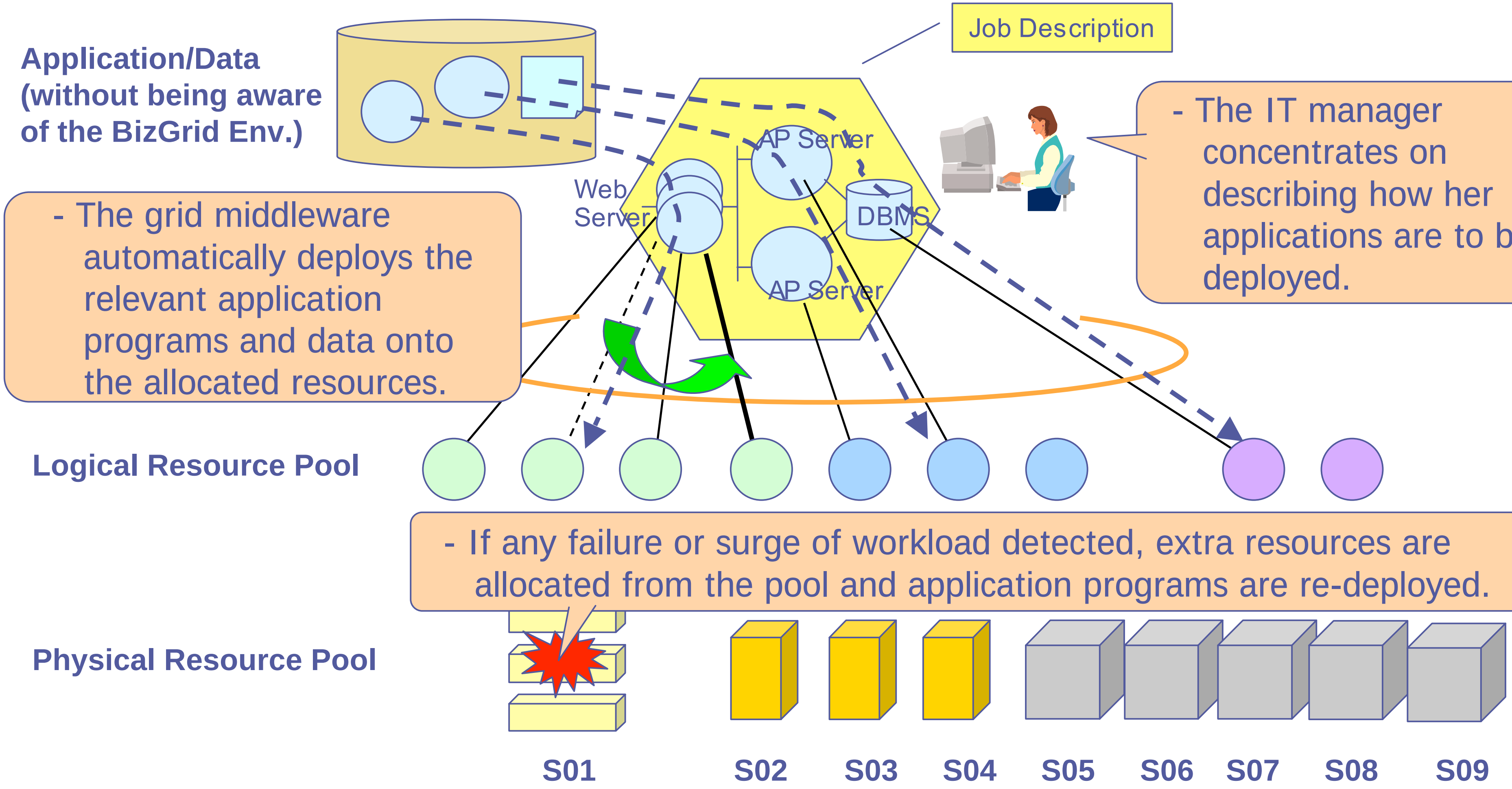
- The IT manager specifies the kinds and amount of IT resources necessary to execute her job.

- The grid middleware allocates the requested resources from the pool.

- The logical servers are mapped to the heterogeneous physical servers via the provisioning tools supplied by their manufactures.

Dynamic Deployment of Business Application

- Based upon the job description, the relevant application programs and data are automatically deployed onto the allocated resources in a consistent manner.
- The application programs need not be aware of the Business Grid Interface.

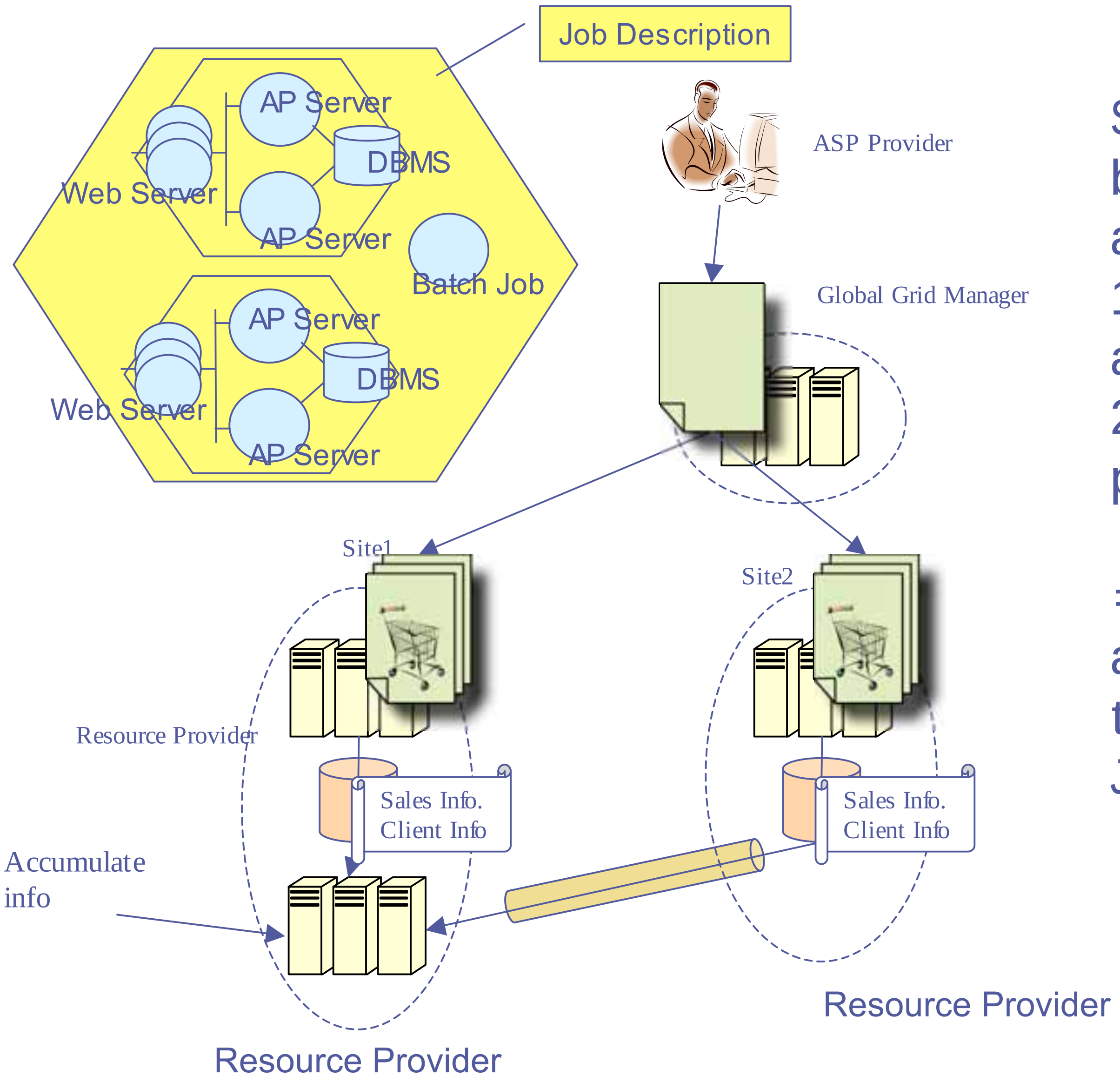


- The grid middleware automatically deploys the relevant application programs and data onto the allocated resources.

- The IT manager concentrates on describing how her applications are to be deployed.

- If any failure or surge of workload detected, extra resources are allocated from the pool and application programs are re-deployed.

Realization of Wide-Area Business Grid

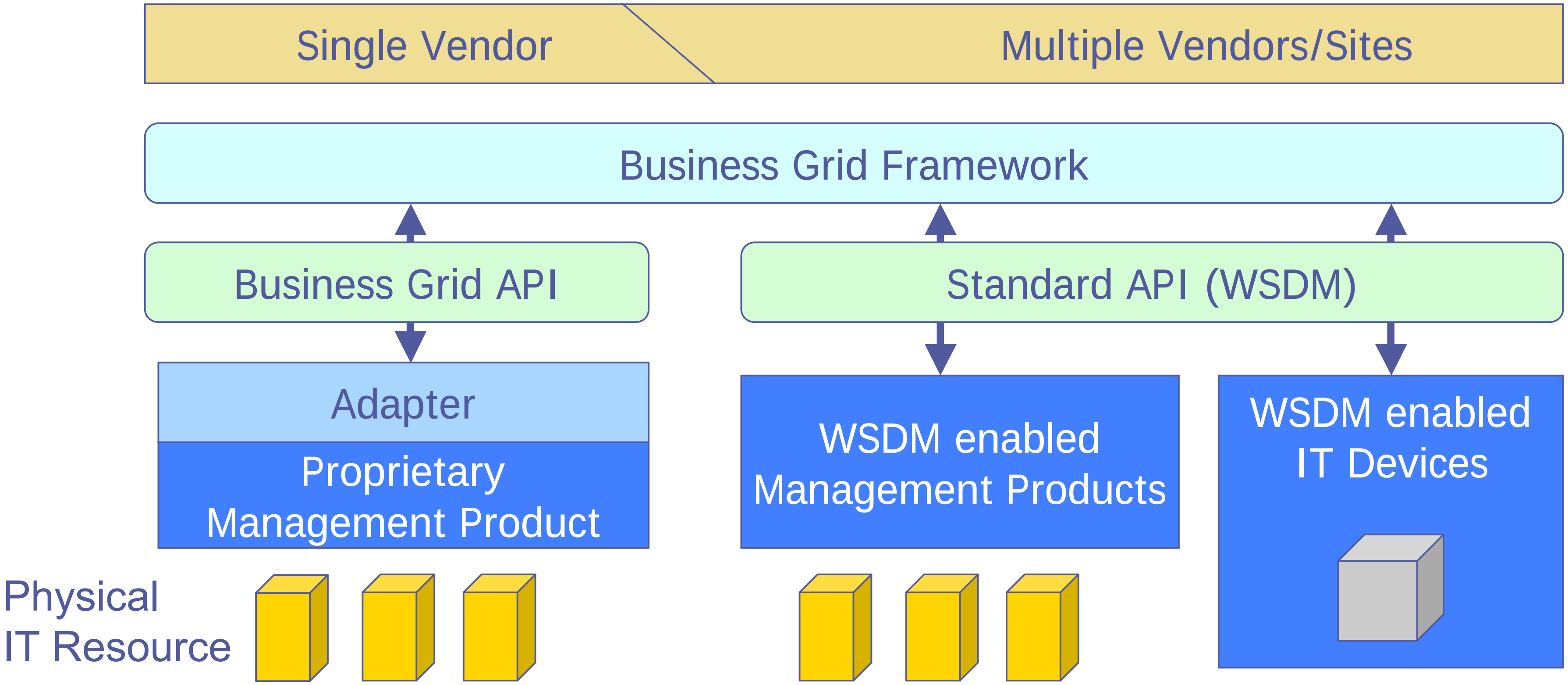


Share IT resources based on the contract / agreement among :
1)Distributed Centers in an Enterprise,
2)Among Trusted partner Data Centers

=> Make it possible for an ASP Provider (client) to dispatch a Complex Job from an entry point

Resource Virtualization

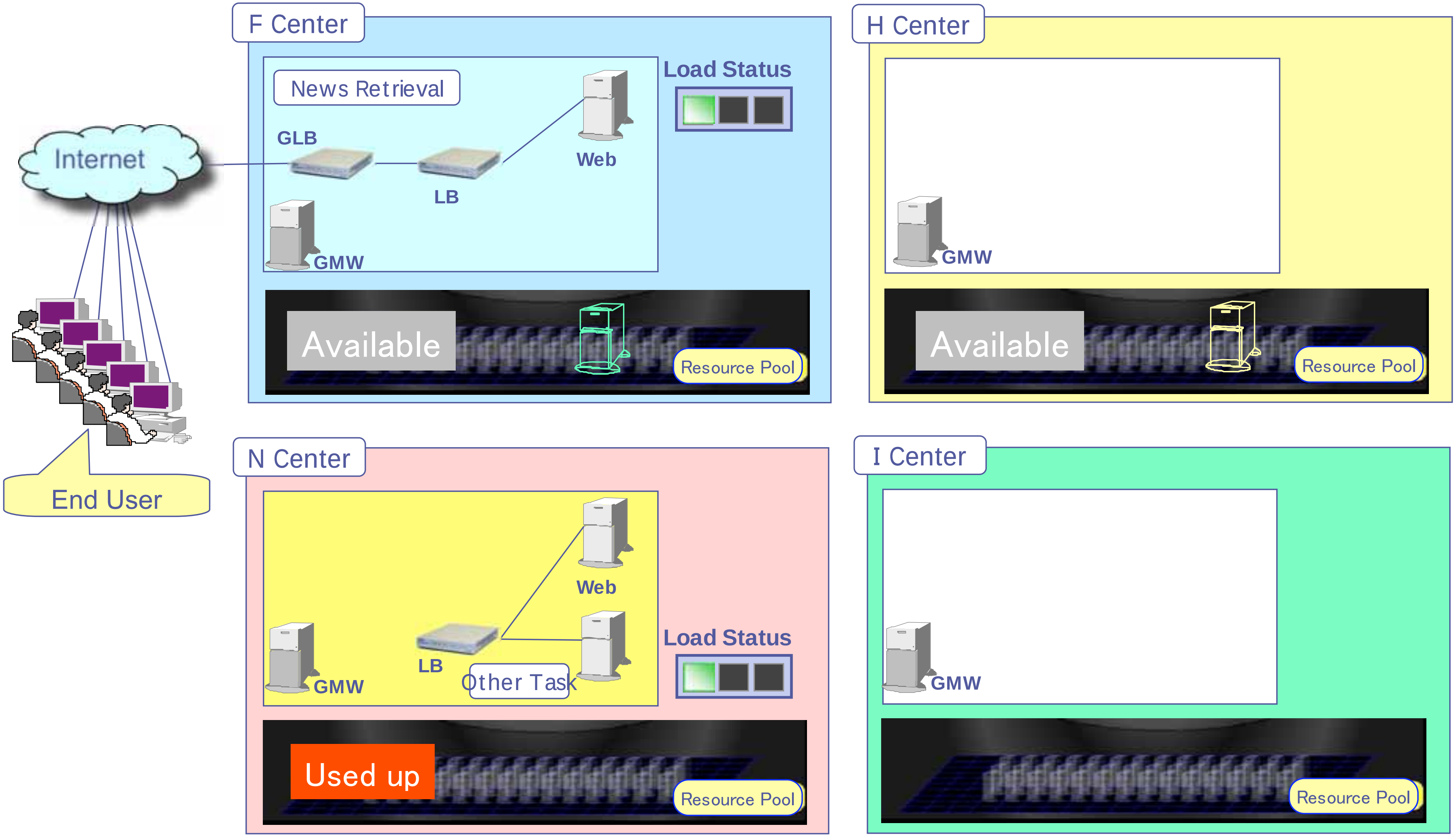
- Currently, BizGrid adopts its own API to describe and control IT devices.
- Efforts are being made to adopt the standardized API (e.g. WSDM) so that WSDM enabled management products and IT devices can also be managed by the business grid framework in a seamless way.



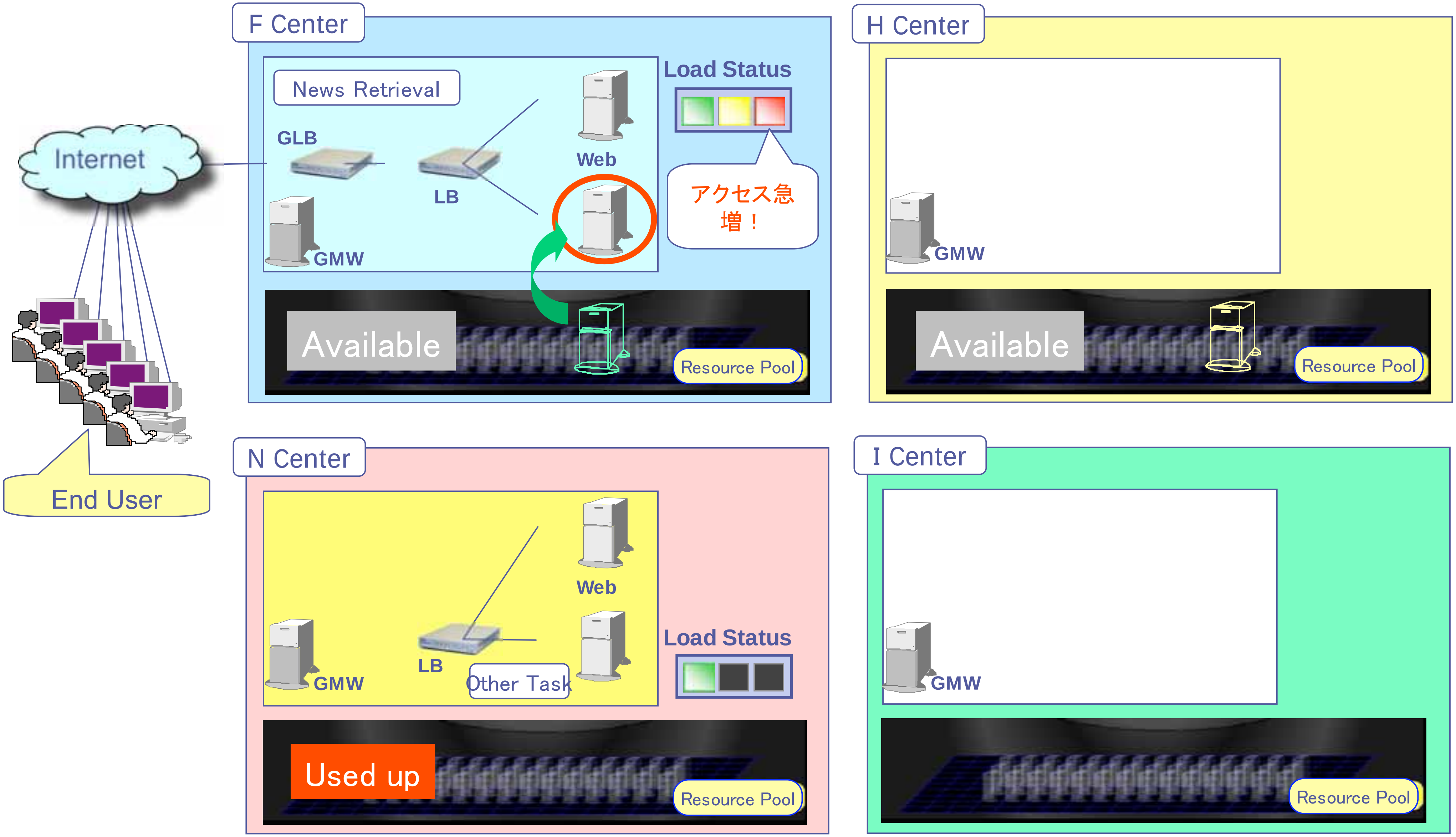
Contents

- Objectives
- Business Grid Middleware
- Demonstration Screen Shots
- Early Adopters
- Relevant Standardization Efforts
- Things to Do

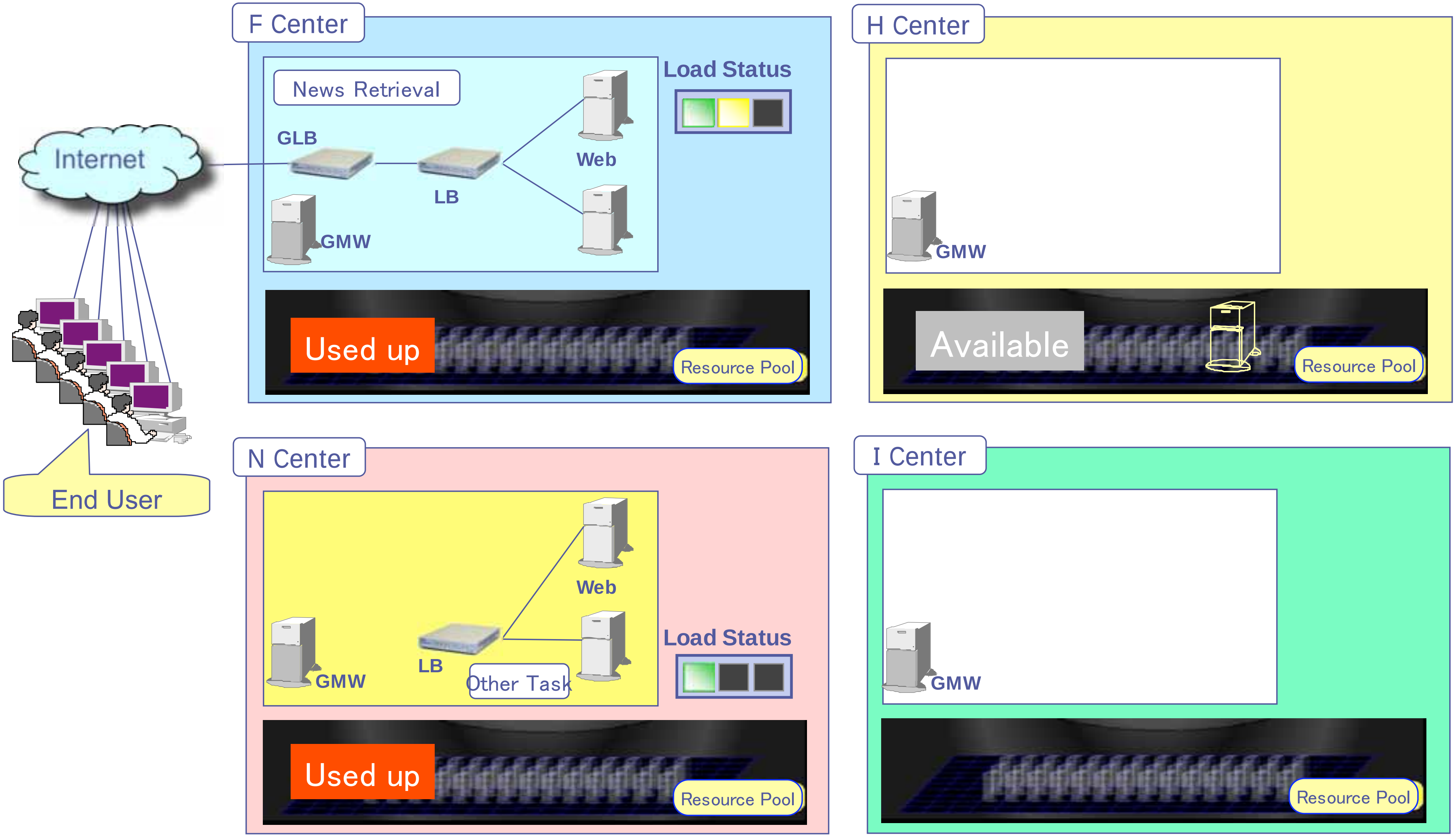
Demonstration (Scenario 1)



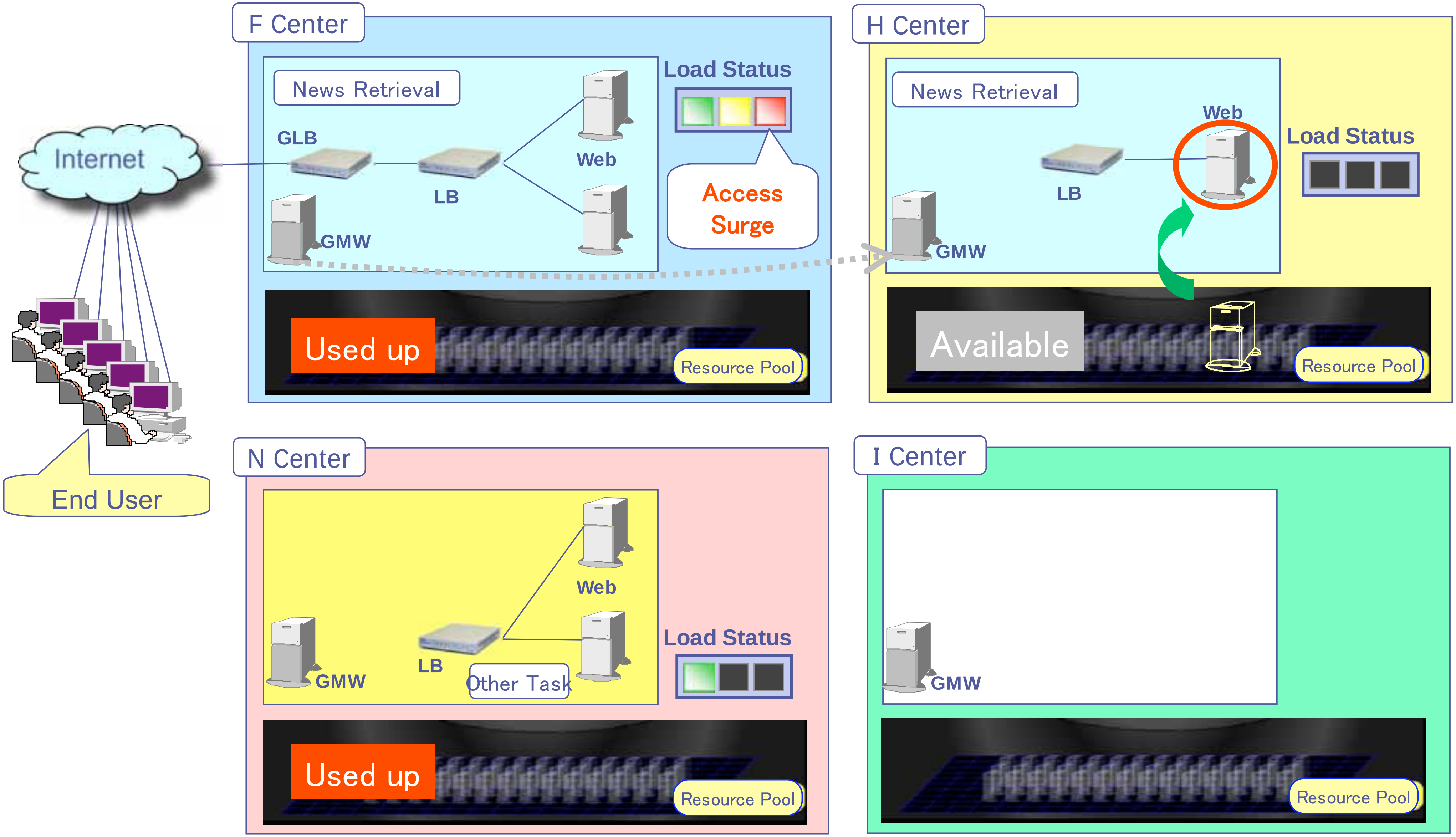
Demonstration (Scenario 2-1)



Demonstration (Scenario 2-2)

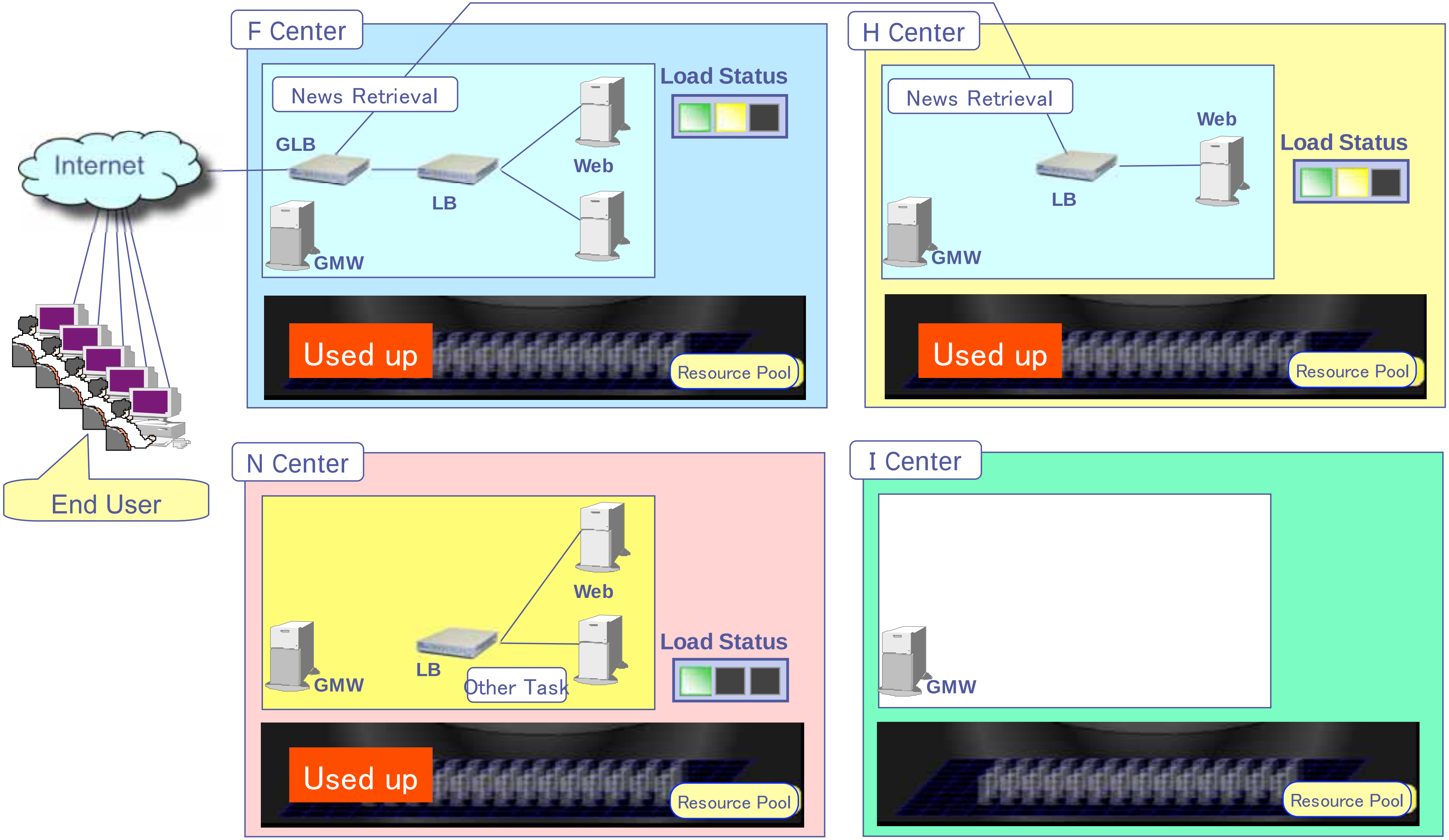


Demonstration (Scenario 3-1)



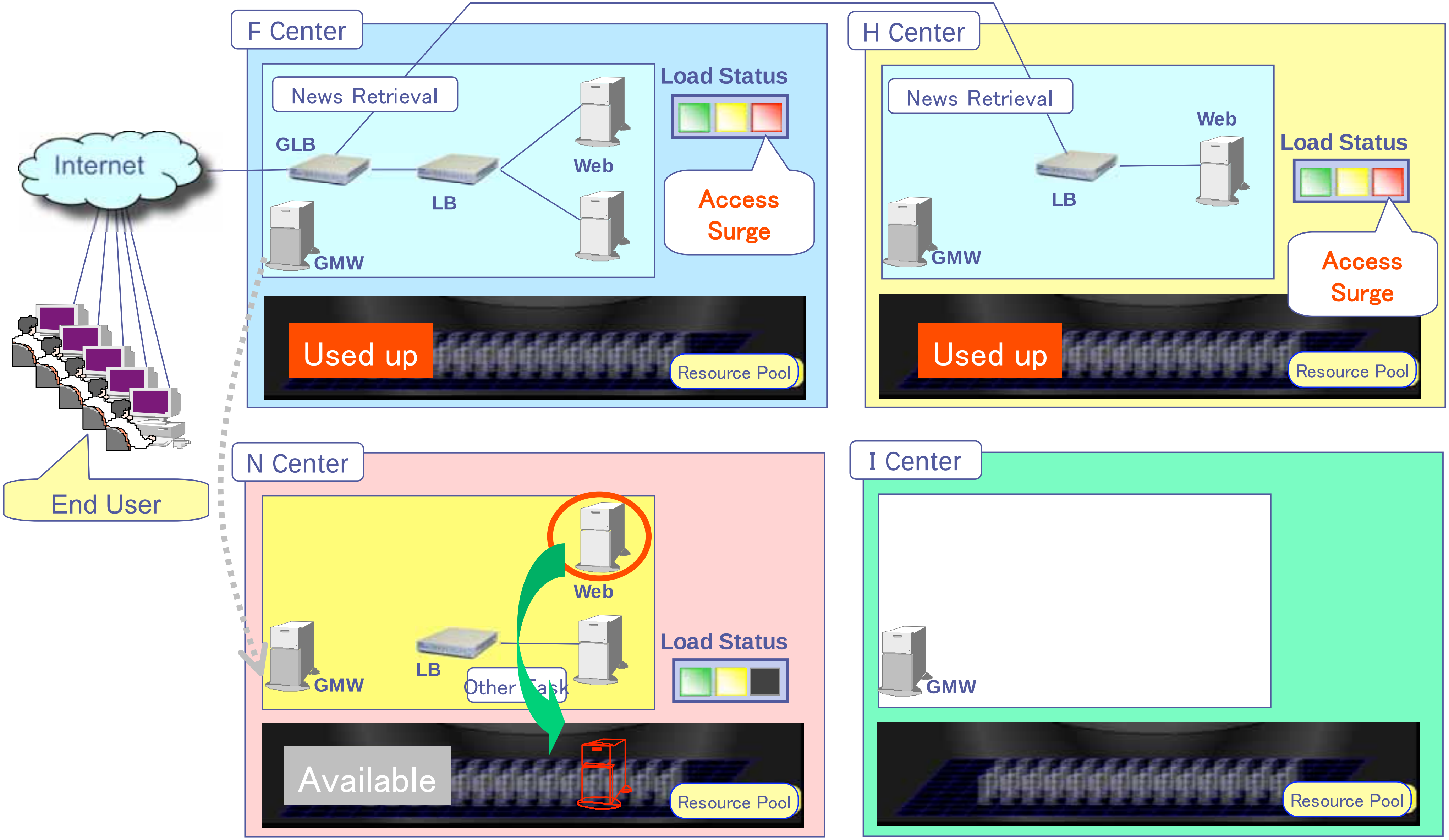
22

Demonstration (Scenario 3-2)

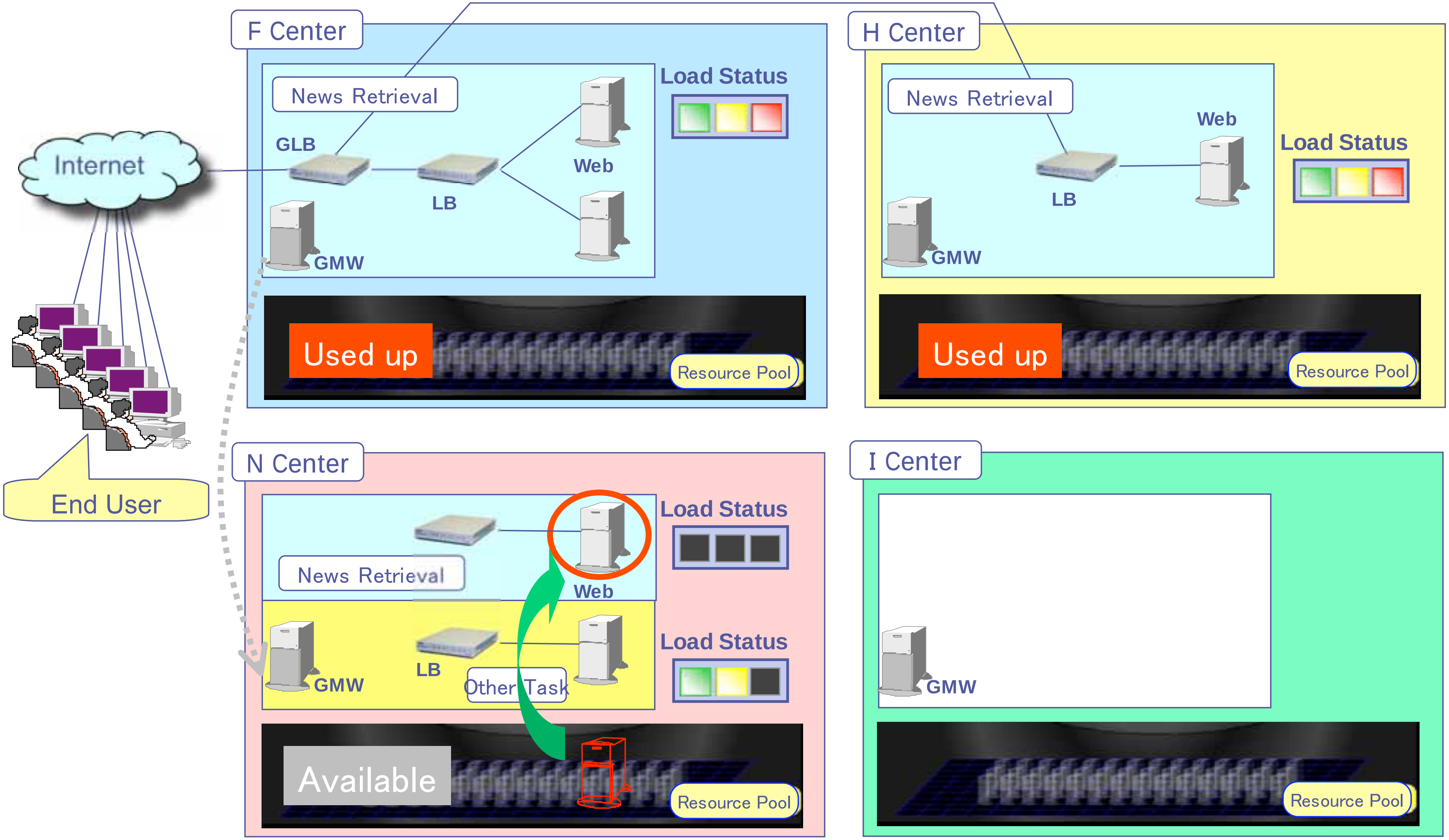


23

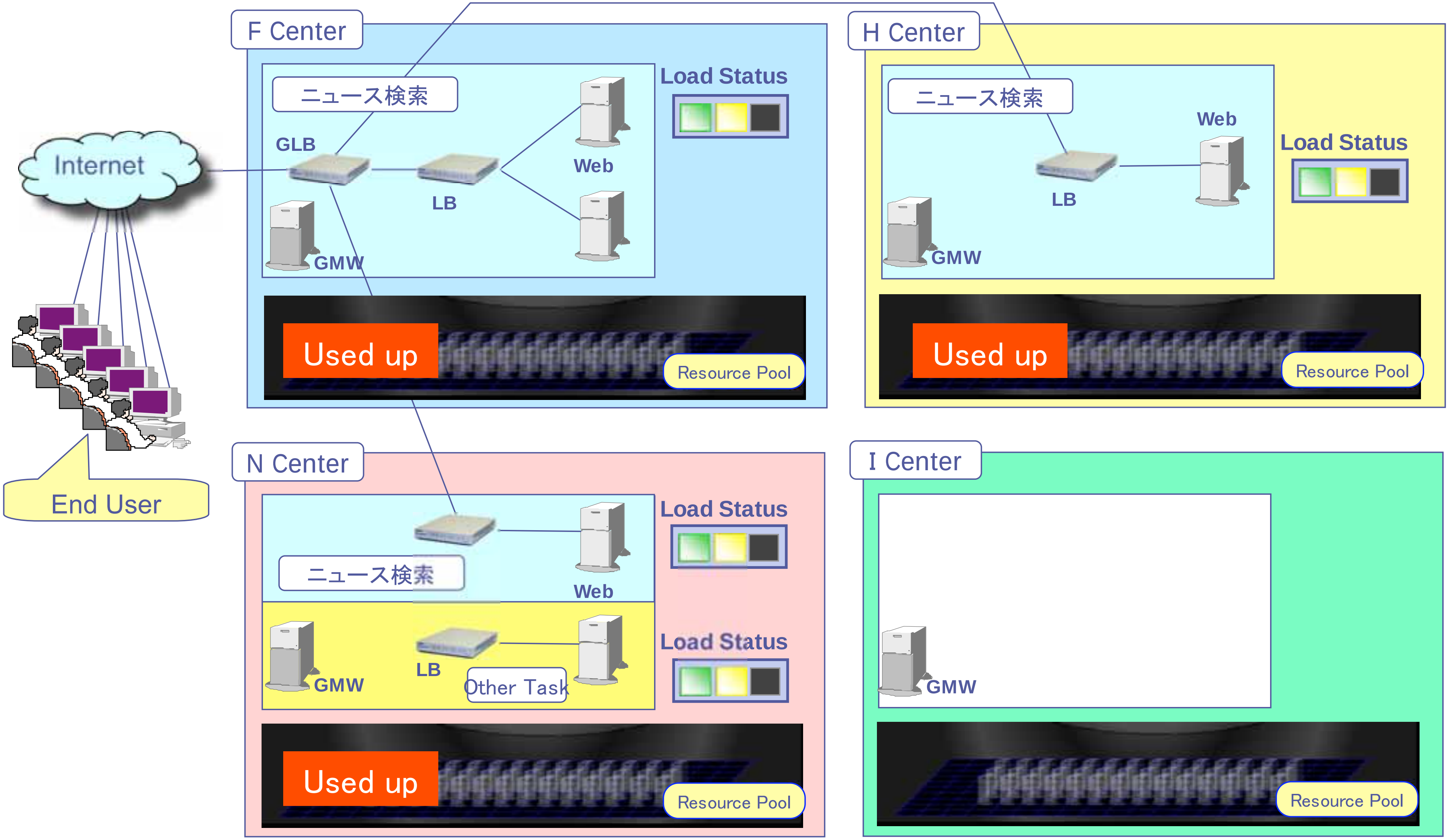
Demonstration (Scenario 4-1)



Demonstration (Scenario 4-2)



Demonstration (Scenario 4-3）



Contents

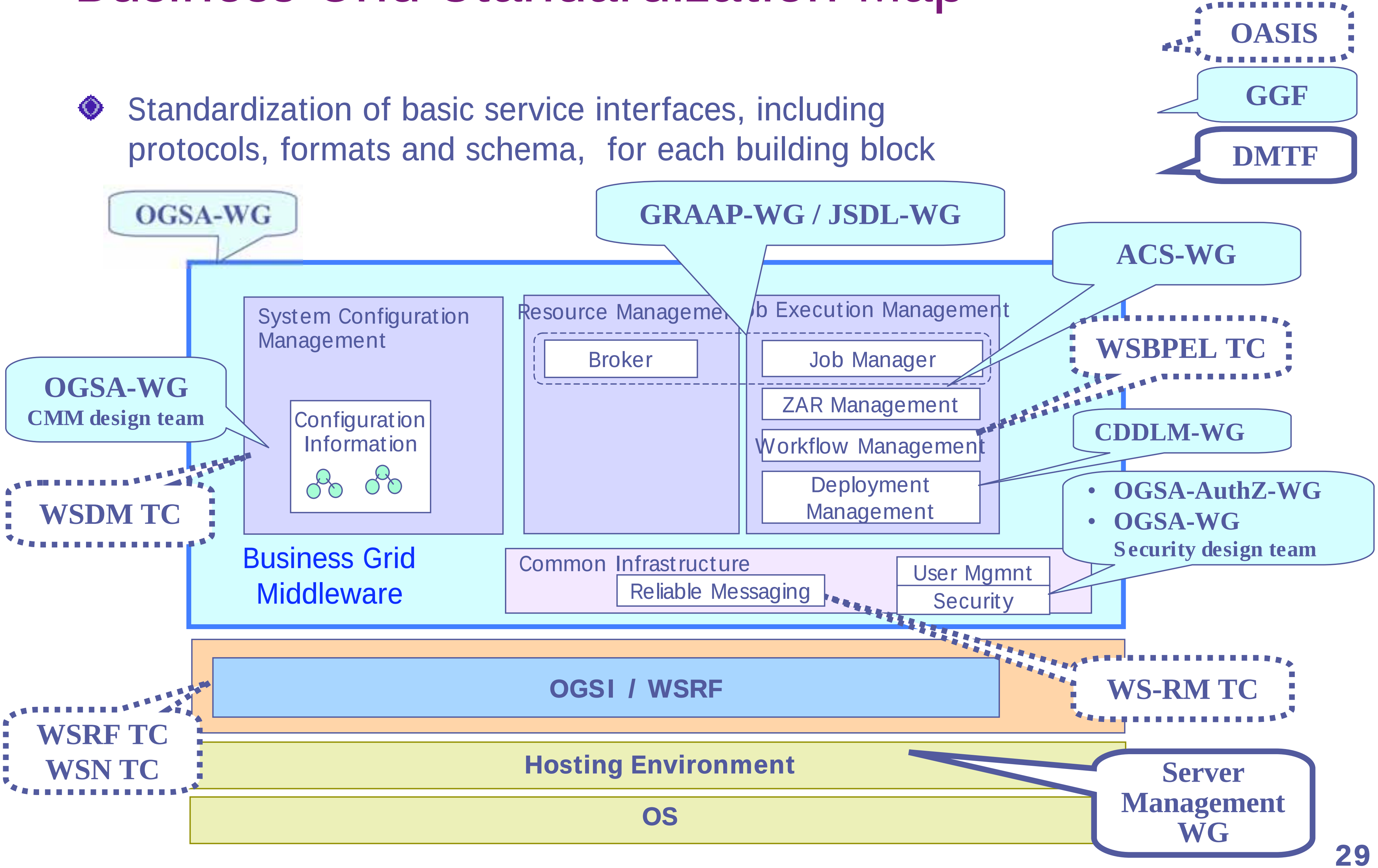
- Objectives
- Business Grid Middleware
- Demonstration Screen Shots
- Relevant Standardization Efforts
- Things to Do

Relevant Standardization Bodies

- GGF
  - OGSA-WG (architecture, roadmap, WG factory, resource management)
  - ACS-WG (application archiving format and archiver API)
  - JSDL-WG, GRAAP-WG (job portability)
  - CDDLM-WG (configuration, deployment, lifecycle management)
- OASIS
  - WSDM TC
  - WSRM TC
  - WSBPEL TC
  - WSRF TC, WSN TC
- DMTF
  - Server Management WG
  - Utility Computing WG

Business Grid Standardization Map

- Standardization of basic service interfaces, including protocols, formats and schema, for each building block



Contents

- Objectives
- Business Grid Middleware
- Demonstration Screen Shots
- Relevant Standardization Efforts
- Things to Do

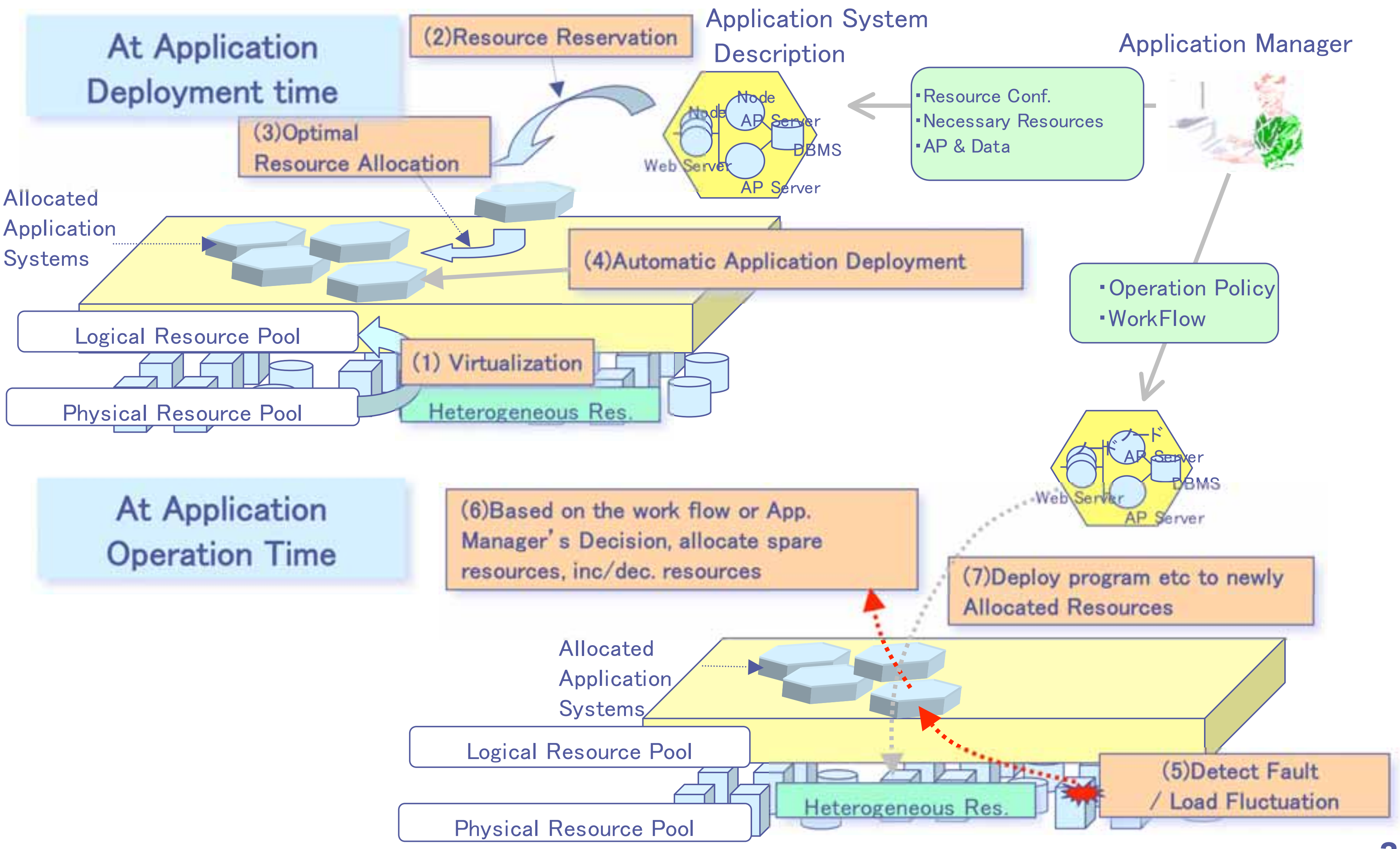
Project Status / Things to do

- Two thirds of the project have finished.
- Initial version of the business grid middleware has been developed and basic features are tried out.
- Features developed so far include:
  - Monitoring and registering underlying IT resources (both hardware and software)
  - Submitting and controlling e-Business applications
  - Allocating IT resources required by the application
  - Deploying and configuring e-Business application
  - Primitive functions for enabling policy based self-managing functionality
  - Controlling multiple data centers i.e. Local/global two layered grid
  - Autonomic and more dynamic control of the resources
- Features to be developed this fiscal year (-03/2006) will include:
  - Adoption of emerging standards from GGF, OASIS, DMTF and other standardization bodies
  - Field test in collaboration with a number of real industry users

Thank you!

Back-up Slides

Functions of Business Grid Middleware



At Application Deployment time

(2)Resource Reservation

(3)Optimal Resource Allocation

Application System Description

Application Manager

- Resource Conf.
- Necessary Resources
- AP & Data

Allocated Application Systems

(4)Automatic Application Deployment

- Operation Policy
- WorkFlow

Logical Resource Pool

(1) Virtualization

Physical Resource Pool

Heterogeneous Res.

At Application Operation Time

(6)Based on the work flow or App. Manager' s Decision, allocate spare resources, inc/dec. resources

(7)Deploy program etc to newly Allocated Resources

Allocated Application Systems

Logical Resource Pool

Physical Resource Pool

Heterogeneous Res.

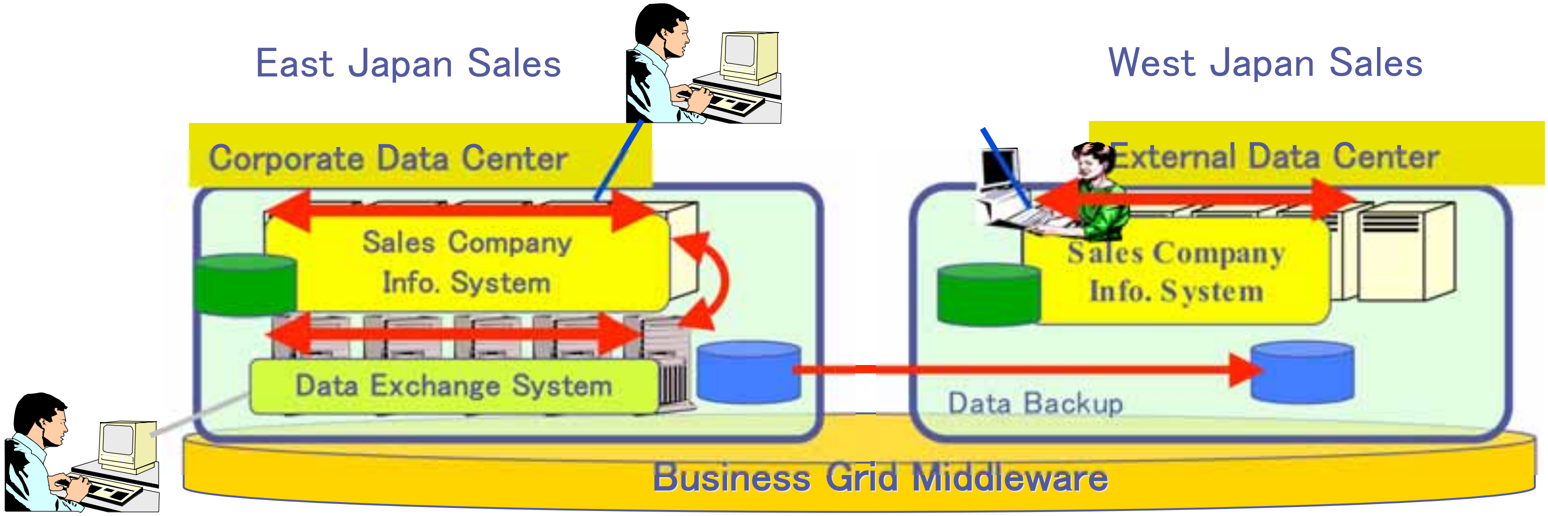
(5)Detect Fault / Load Fluctuation

34

Outline and Goal of the Trial System

Keep the Investment Cost / Management Cost of the Core Information System low and at the same time improve business continuity in case of disaster by allocating the system to multiple sites

A) Wide-Area Load Balancing
B) Disaster Recovery
C) Effective System Management

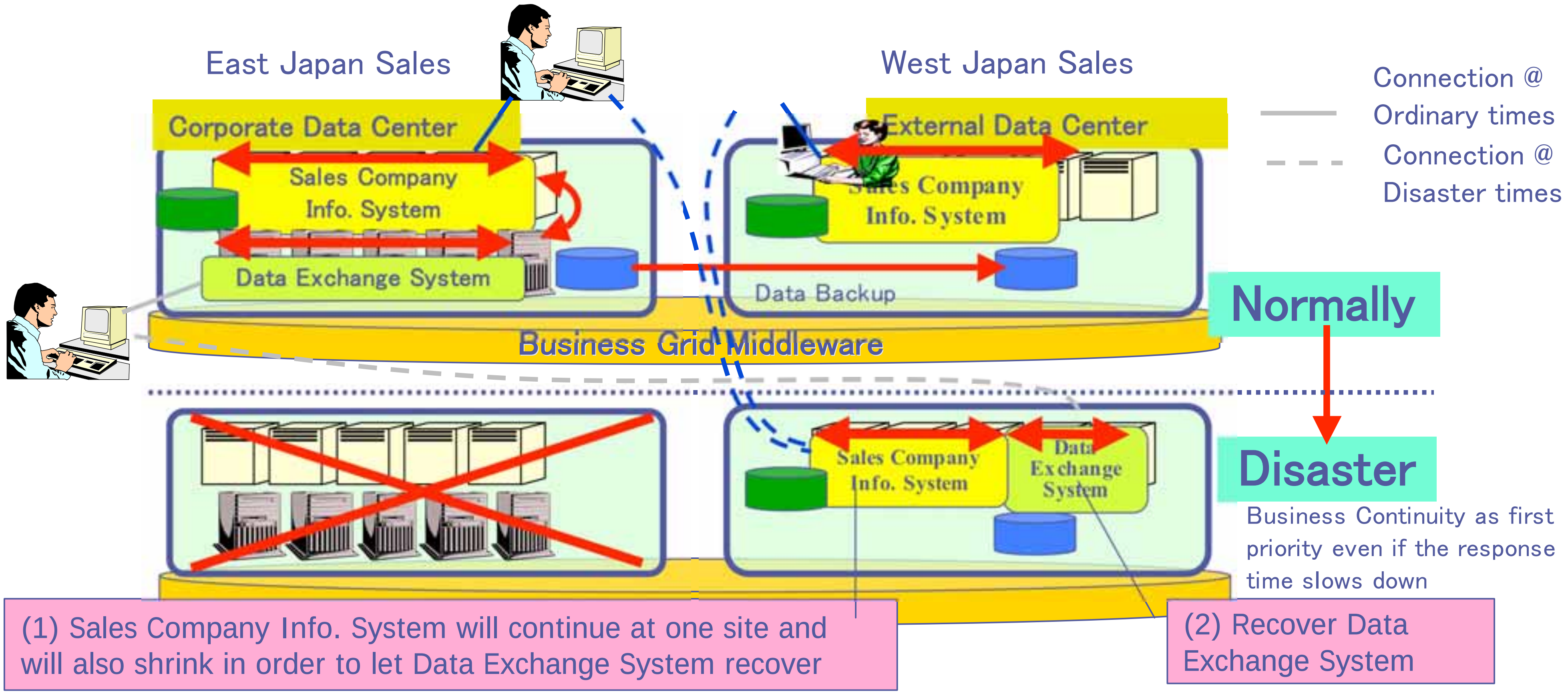


A) Wide–Area Load Balancing: At Normal times, use the IT resource effectively for disaster recovery as spare system for high load business apps and guarantee the quick response

Outline and Goal of the Trial System

B) Disaster Recovery: In case of disaster at the data center for the corporate site, let the application continue at the external data center

C) Effective System Management: Change the configuration automatically and optimize business app management among multiple data centers.



East Japan Sales

West Japan Sales

Corporate Data Center

External Data Center

Sales Company Info. System

Data Exchange System

Data Backup

Business Grid Middleware

Connection @ Ordinary times

Connection @ Disaster times

Normally

Disaster

Business Continuity as first priority even if the response time slows down

(1) Sales Company Info. System will continue at one site and will also shrink in order to let Data Exchange System recover

(2) Recover Data Exchange System

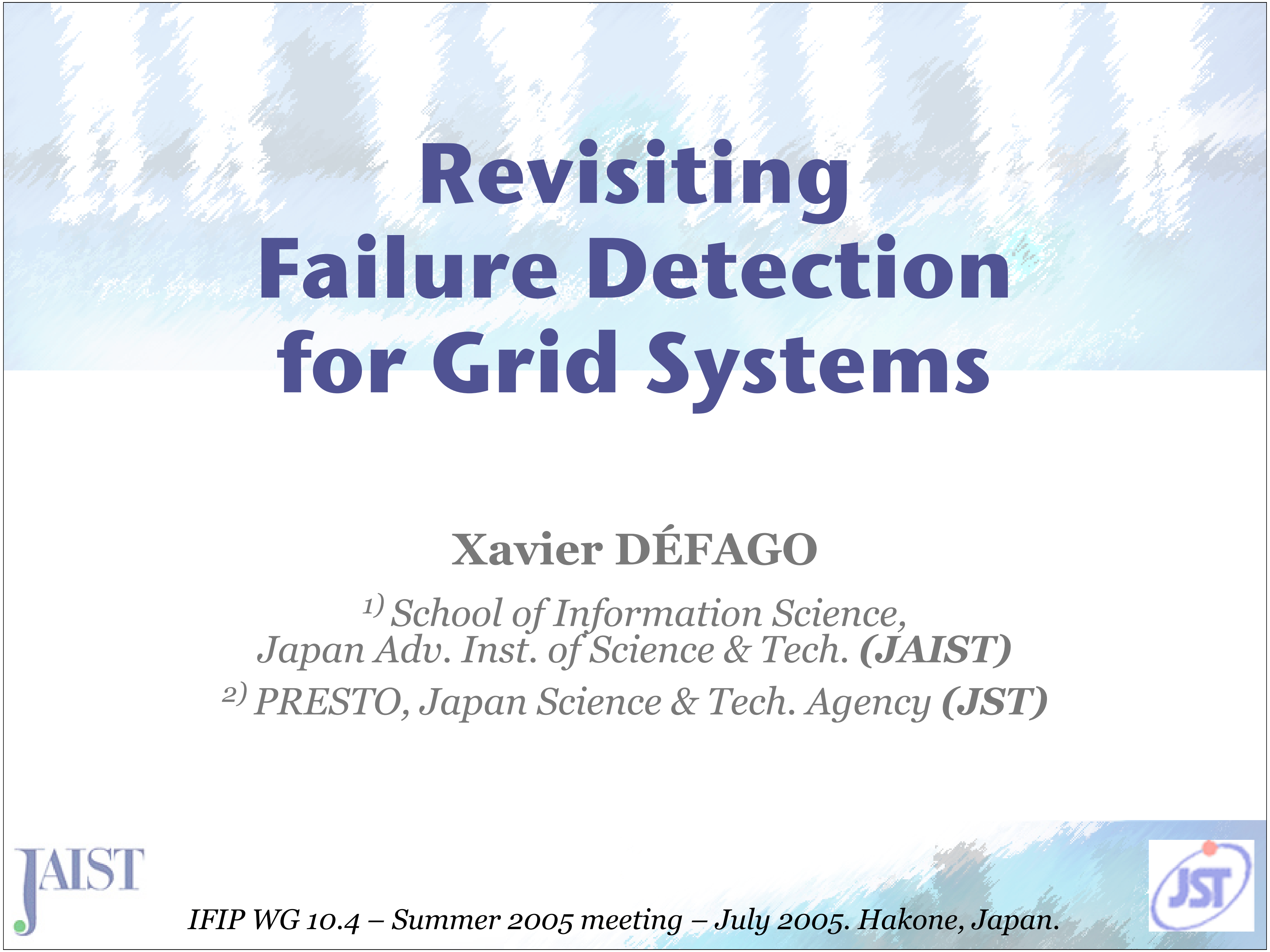
36

Session 1.3

Fault Tolerance in Grid Computing

Moderator and Rapporteur

Richard D. Schlichting, AT&T Labs - Research, Florham Park, NJ, USA



Revisiting Failure Detection for Grid Systems

Xavier DÉFAGO

[1] School of Information Science, Japan Adv. Inst. of Science & Tech. ***(JAIST)***

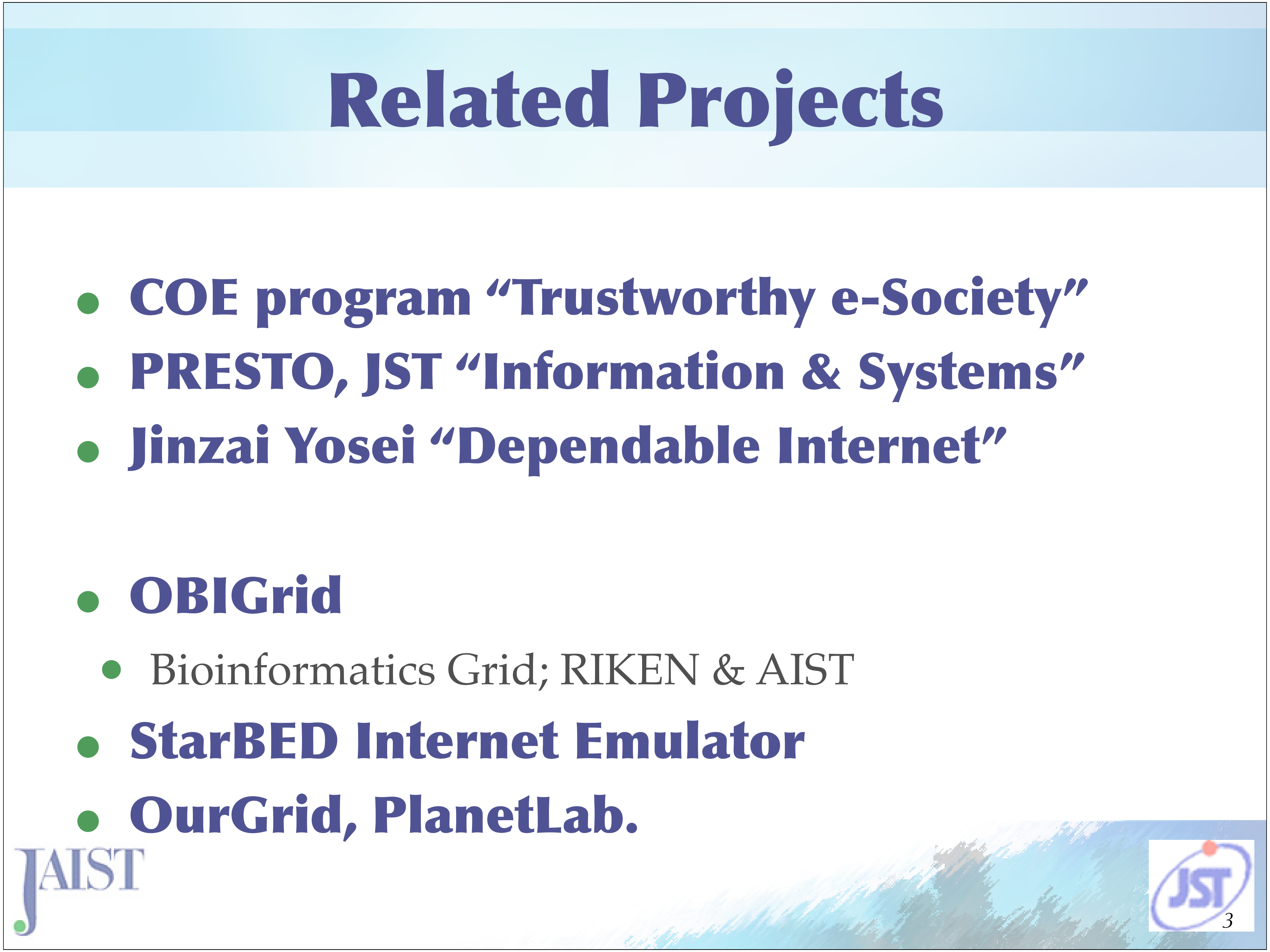
[2] PRESTO, Japan Science & Tech. Agency ***(JST)***

IFIP WG 10.4 – Summer 2005 meeting – July 2005. Hakone, Japan.

Acknowledgements

- **Naohiro HAYASHIBARA**
  - now at Tokyo Denki University
- **Péter URBÁN**
- **Rami YARED**
- **Takuya KATAYAMA**

- **… and many people through enlightening discussions**

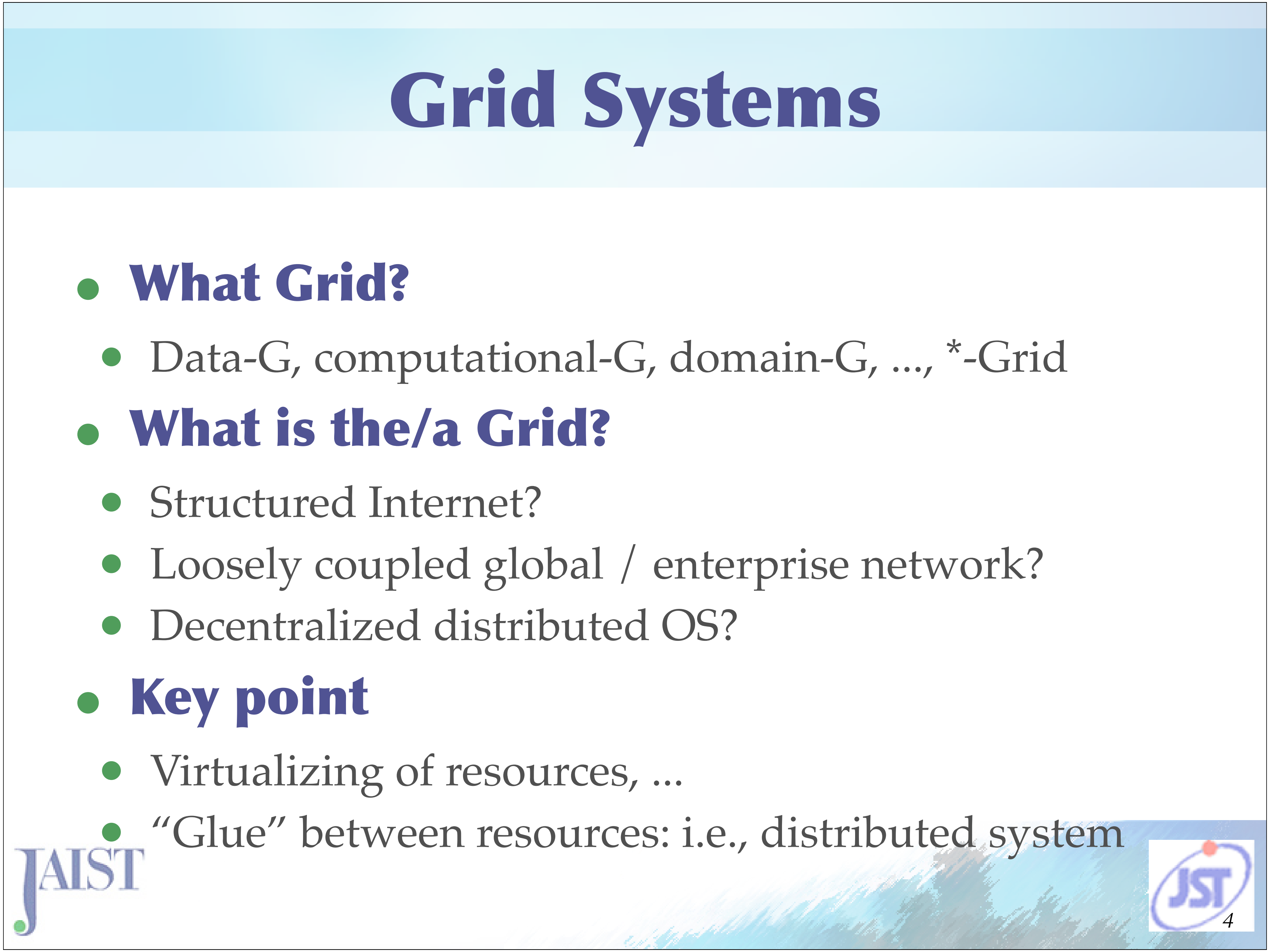


Related Projects

- **COE program “Trustworthy e-Society”**
- **PRESTO, JST “Information & Systems”**
- **Jinzai Yosei “Dependable Internet”**

- **OBIGrid**
  - Bioinformatics Grid; RIKEN & AIST
- **StarBED Internet Emulator**
- **OurGrid, PlanetLab.**

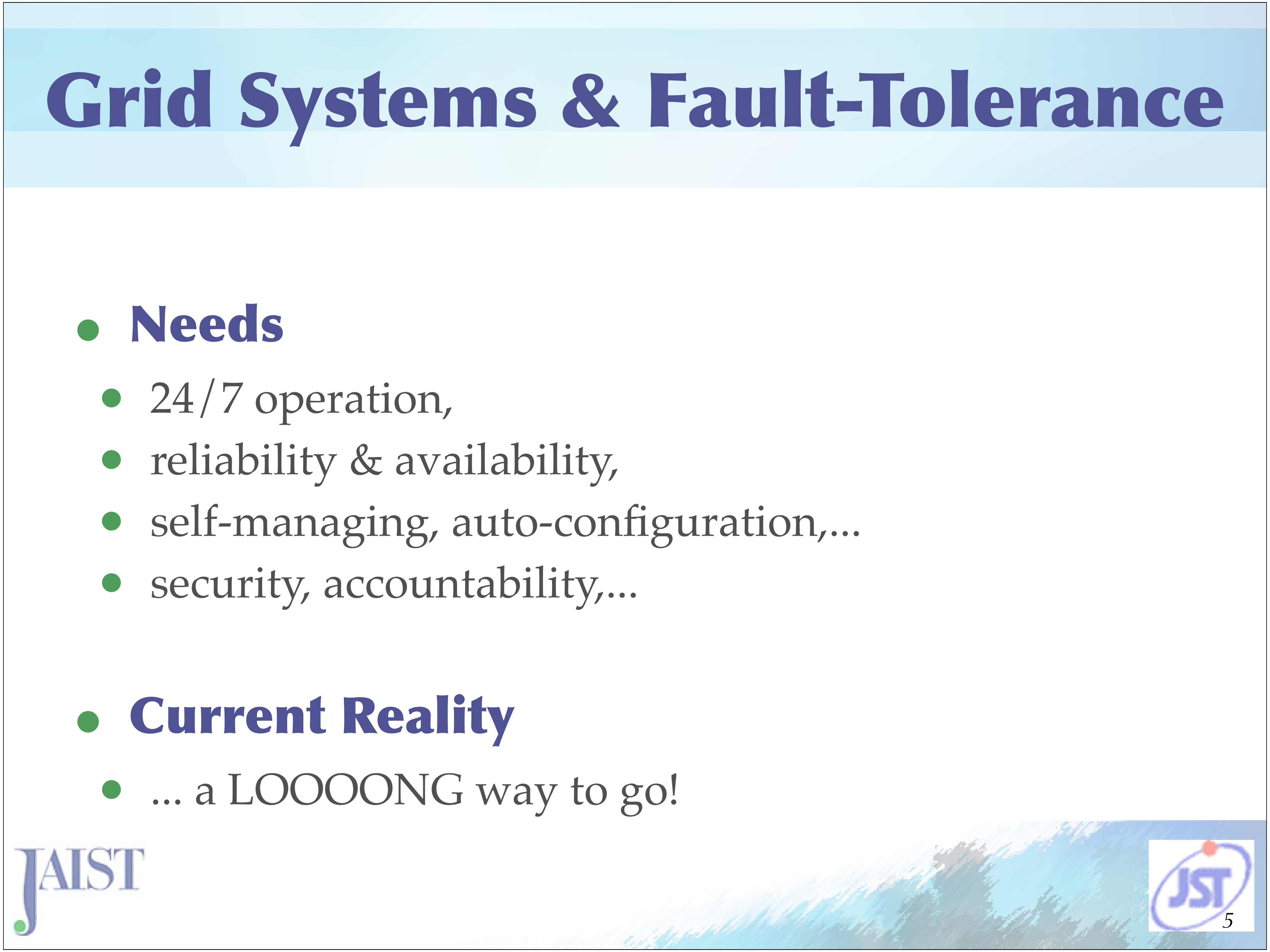
3



Grid Systems

- **What Grid?**
  - Data-G, computational-G, domain-G, ..., *-Grid
- **What is the/a Grid?**
  - Structured Internet?
  - Loosely coupled global / enterprise network?
  - Decentralized distributed OS?
- **Key point**
  - Virtualizing of resources, ...
  - “Glue” between resources: i.e., distributed system

4



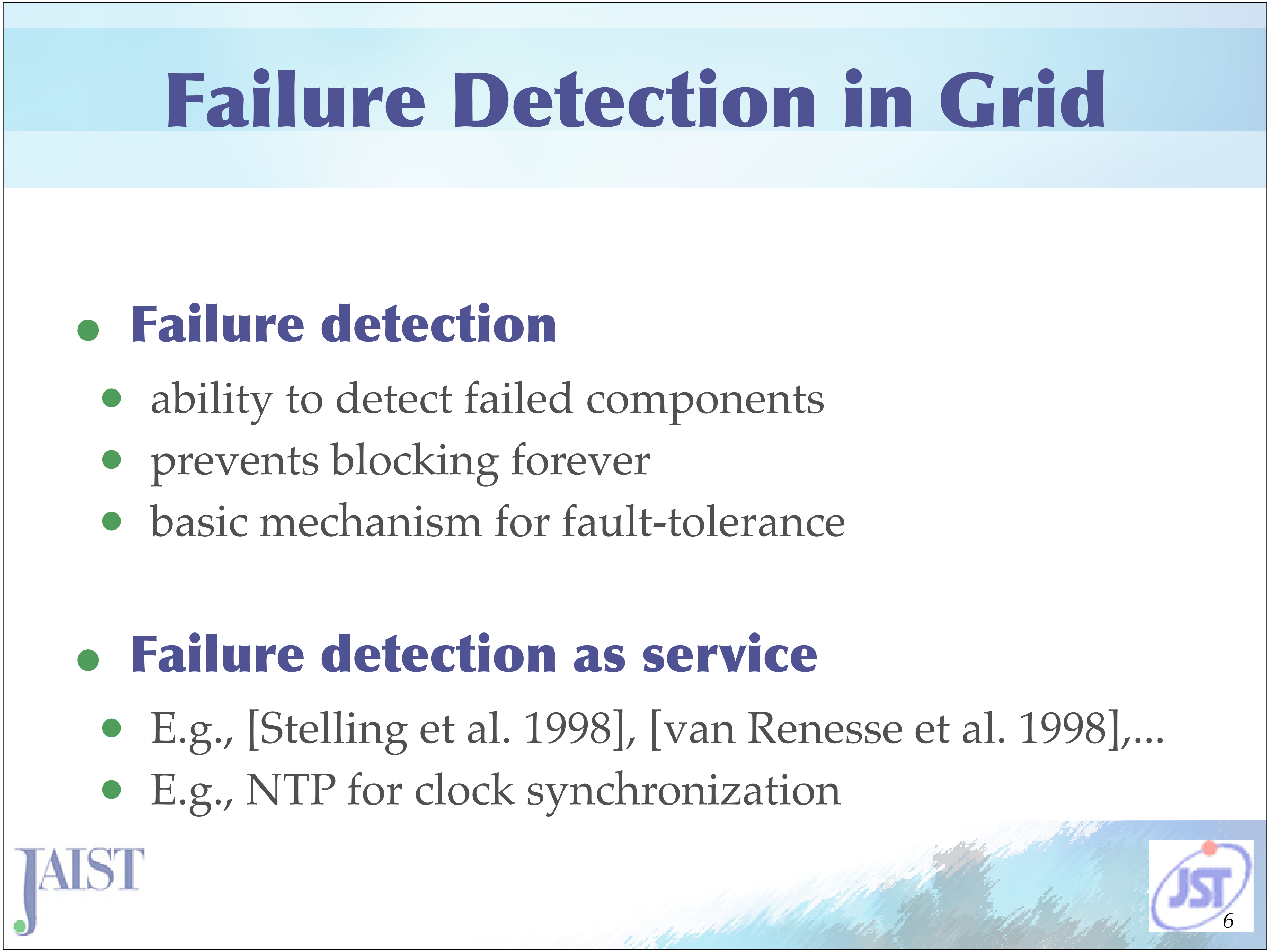
Grid Systems & Fault-Tolerance

- **Needs**
  - 24/7 operation,
  - reliability & availability,
  - self-managing, auto-configuration,...
  - security, accountability,...
- **Current Reality**
  - ... a LOOOONG way to go!

JAIST

JST

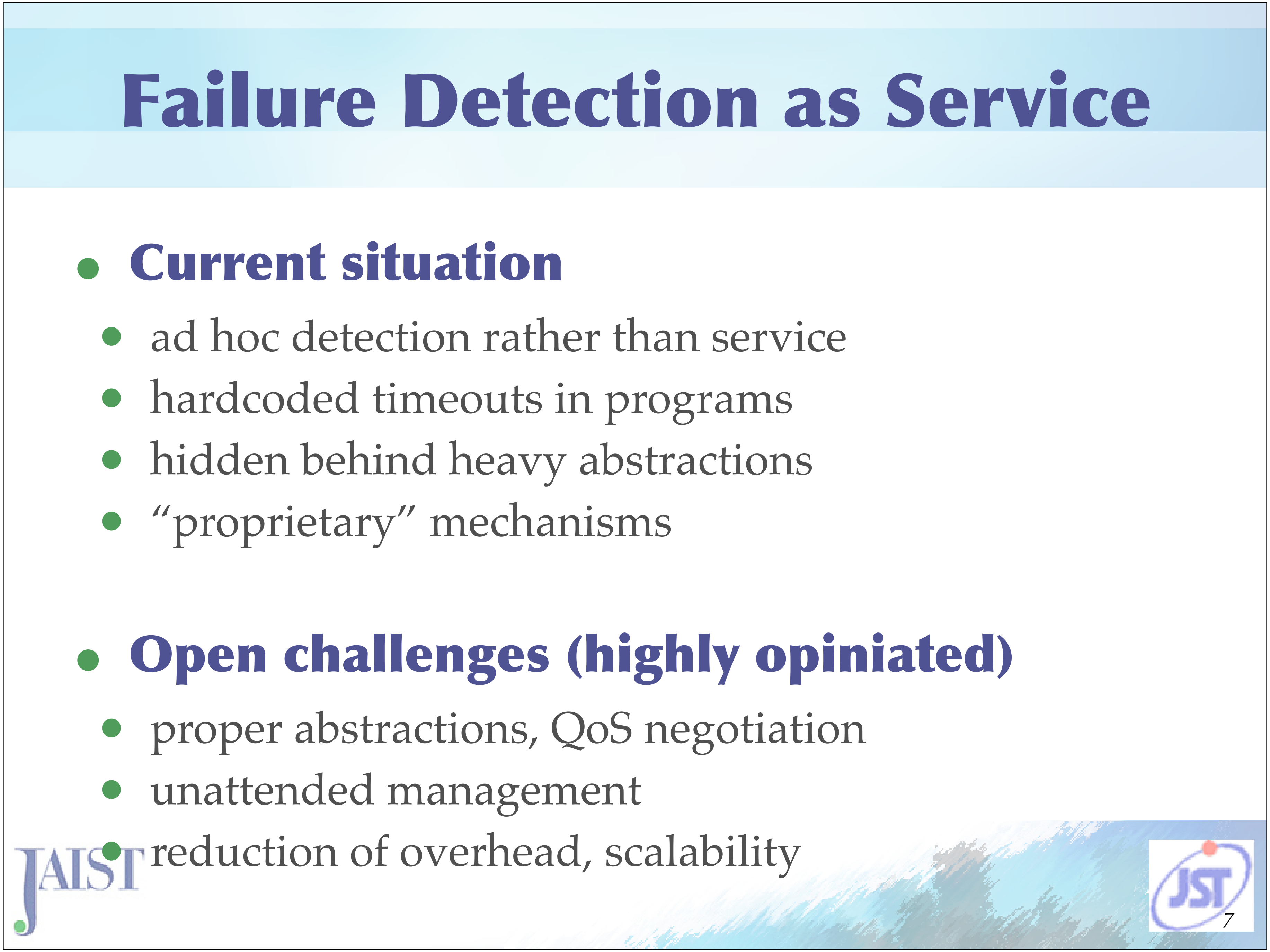
5



Failure Detection in Grid

- **Failure detection**
  - ability to detect failed components
  - prevents blocking forever
  - basic mechanism for fault-tolerance
- **Failure detection as service**
  - E.g., [Stelling et al. 1998], [van Renesse et al. 1998],...
  - E.g., NTP for clock synchronization

6

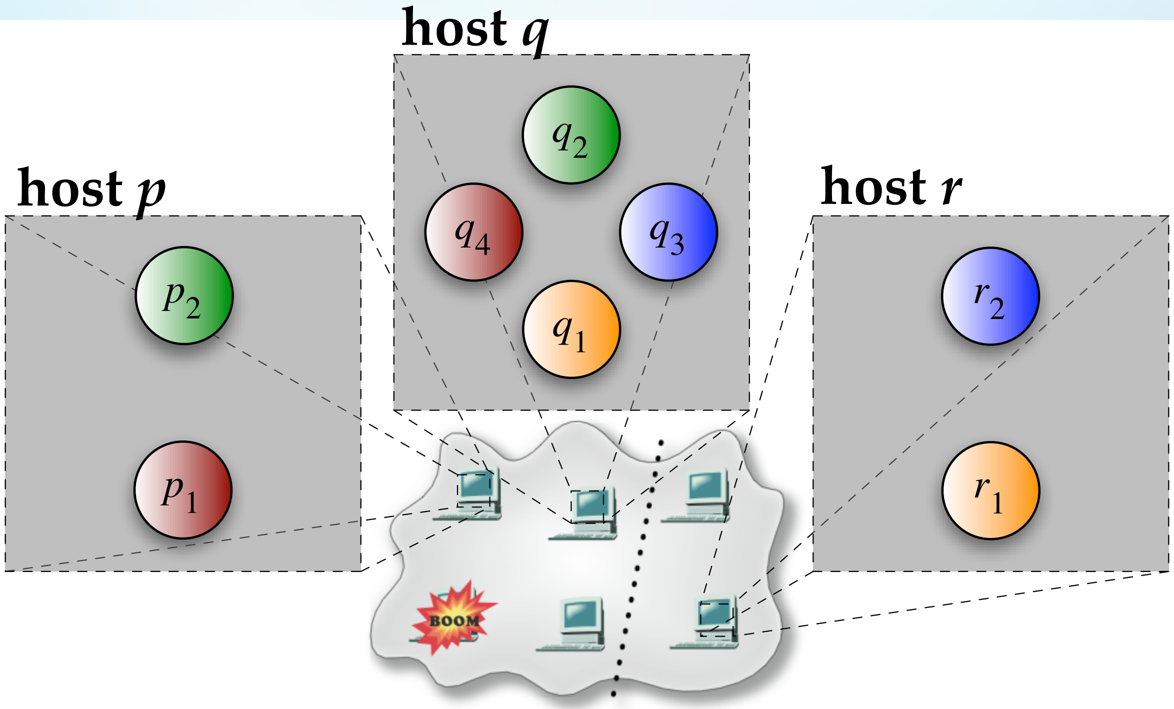


Failure Detection as Service

- **Current situation**
  - ad hoc detection rather than service
  - hardcoded timeouts in programs
  - hidden behind heavy abstractions
  - “proprietary” mechanisms
- **Open challenges (highly opiniated)**
  - proper abstractions, QoS negotiation
  - unattended management
  - reduction of overhead, scalability

7

Simult. Indep. Requirements



applications

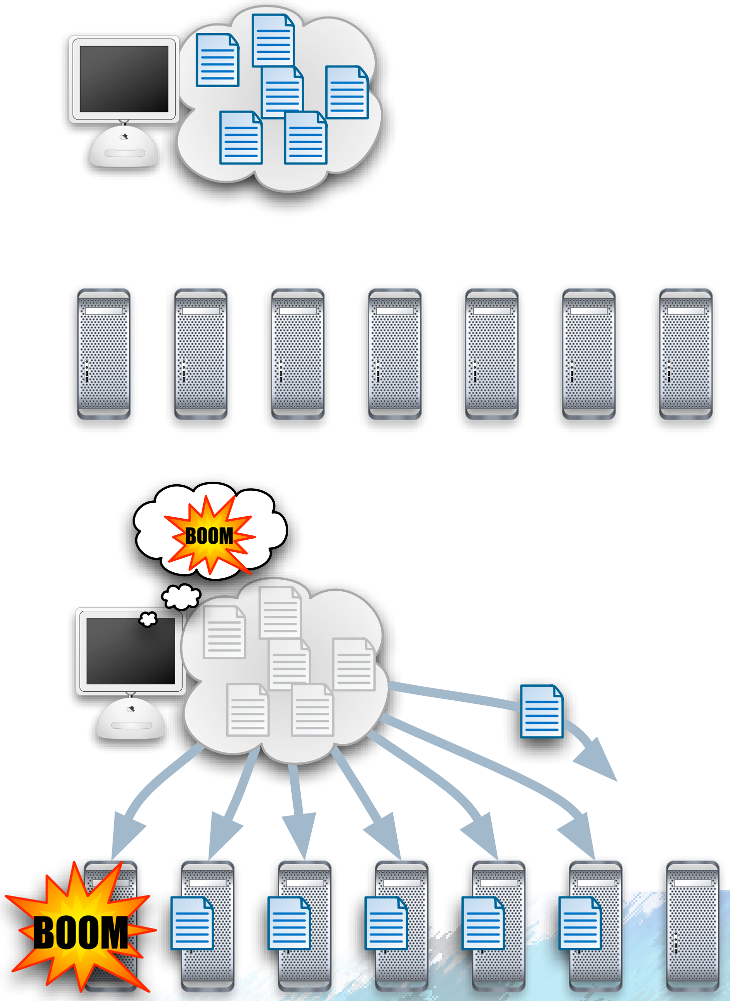
- ● ● ● A_α
- ● ● ● A_β
- ● ● ● A_γ
- ● ● ● A_θ

- **Large-scale systems**
  - Many distributed applications simultaneously
  - Different requirements

8

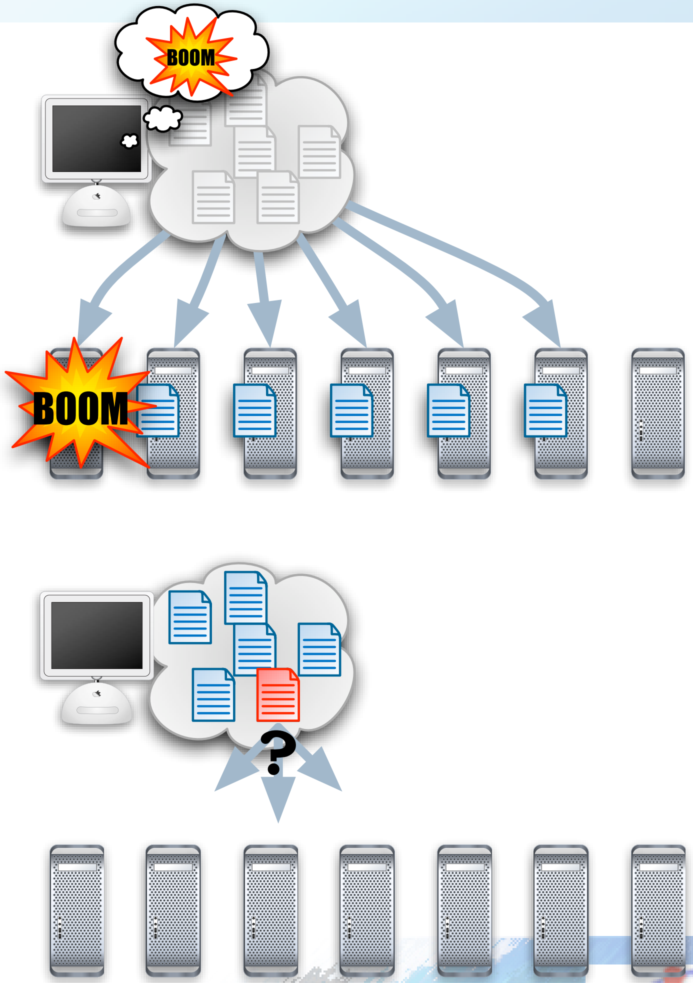
Example / Motivation

- **Simple case**
  - “Bag-of-Tasks” computations
  - Dispatch tasks
  - Wait for results
- **Environment**
  - **Partial failures**
  - Heterogeneous
  - Unpredictable comm.



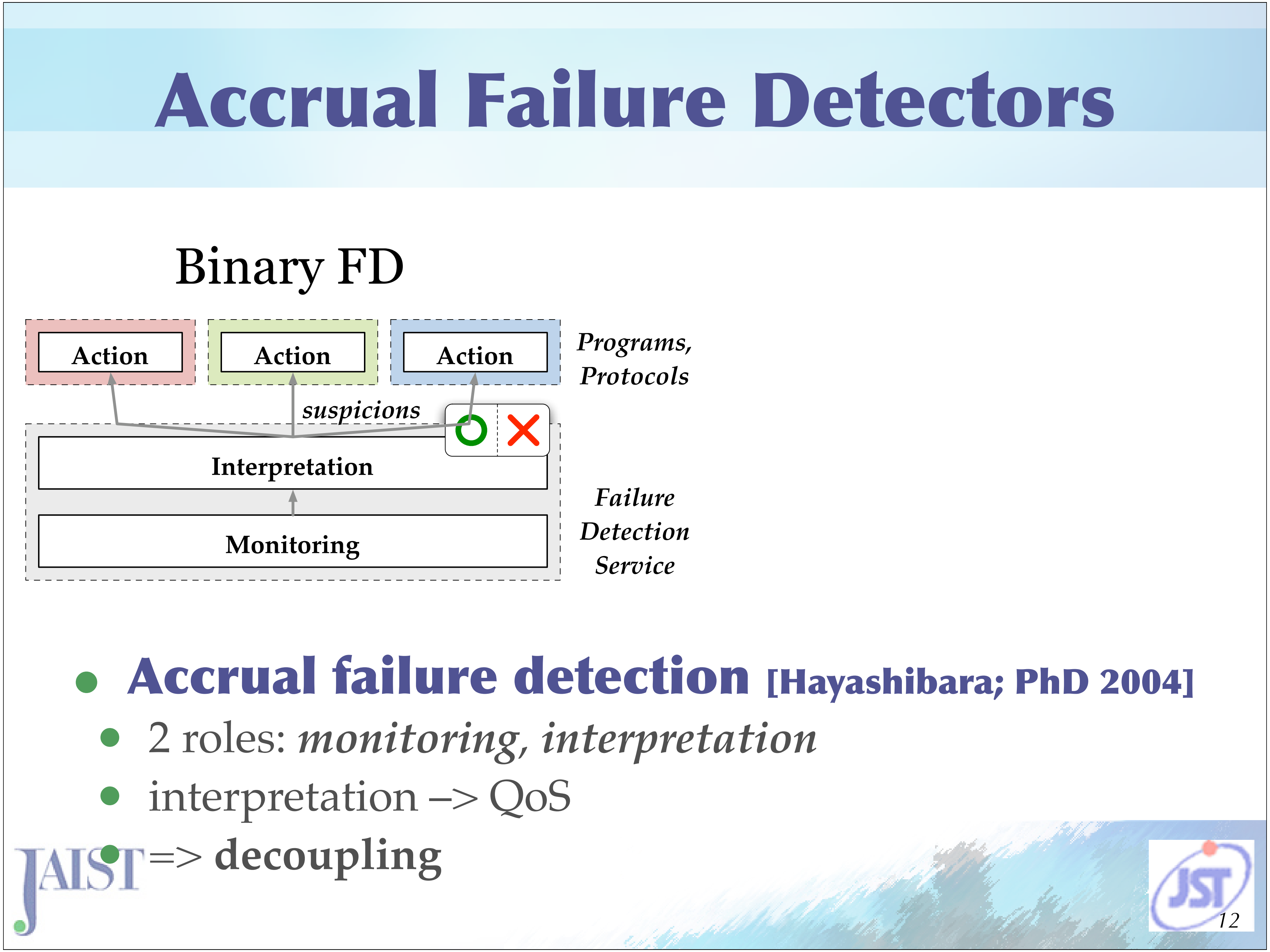
Usage Patterns

- **Case 1:**
  - Cost varies with time:
    - amount work completed
    - available resources



10

Abstractions



Accrual Failure Detectors

Binary FD

Action Action Action

Programs, Protocols

suspicions

○ ✗

Interpretation

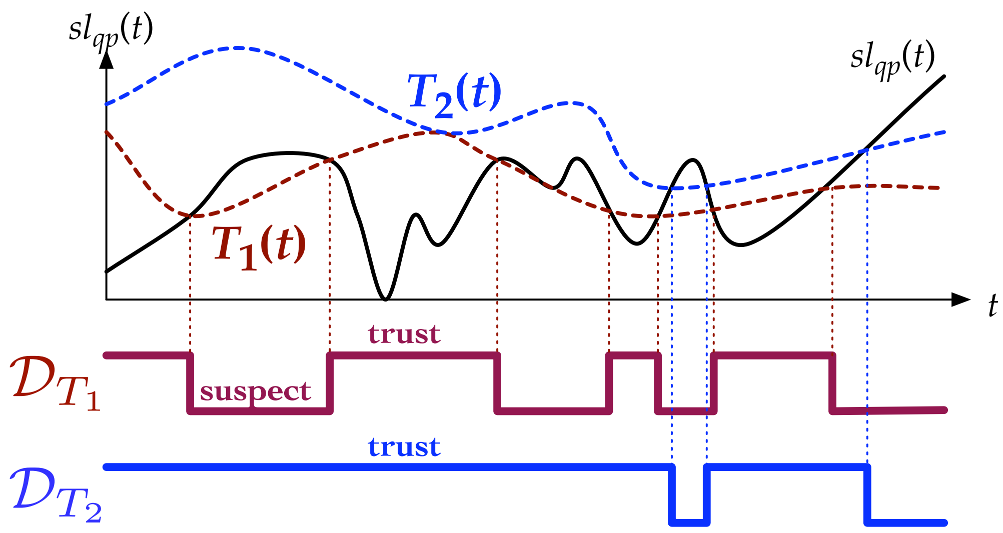
Monitoring

Failure Detection Service

- **Accrual failure detection** **[Hayashibara; PhD 2004]**
  - 2 roles: *monitoring, interpretation*
  - interpretation –> QoS
  - => **decoupling**

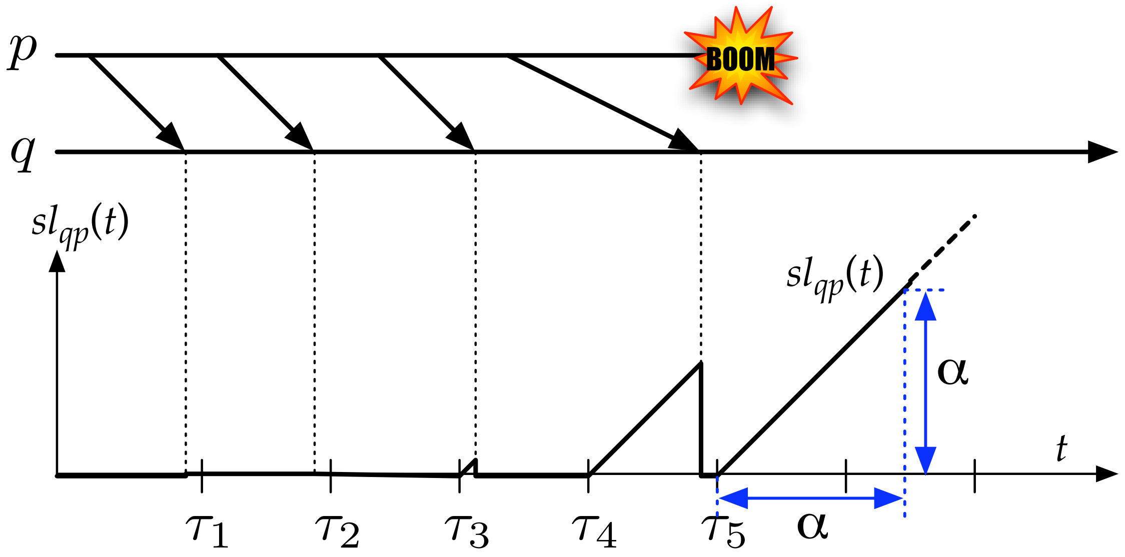
12

Accrual Failure Detectors



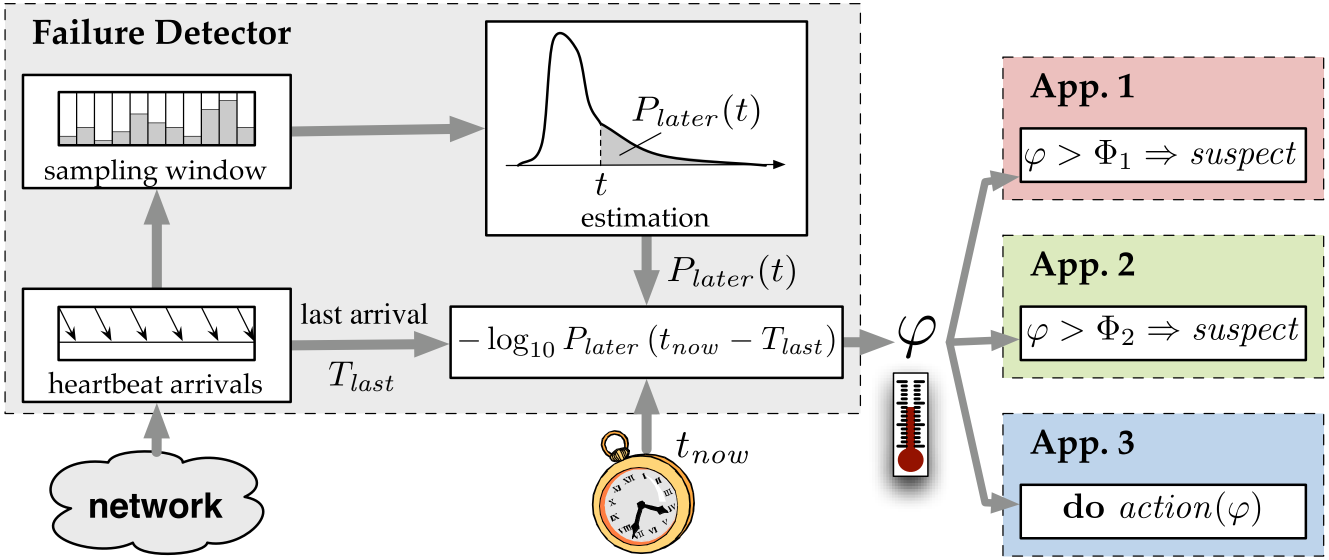
- **Accrual FD abstraction** **[Défago et al.; DSN 2005]**
  - combine different QoS
  - properties; relation w/FD theory

Chen FD as Accrual



- **Chen-based adaptation** **[Chen et al.; TC 2002]**
  - After freshness point, increase with time
  - **Reset** when receive heartbeat
  - Safety margin α set with threshold

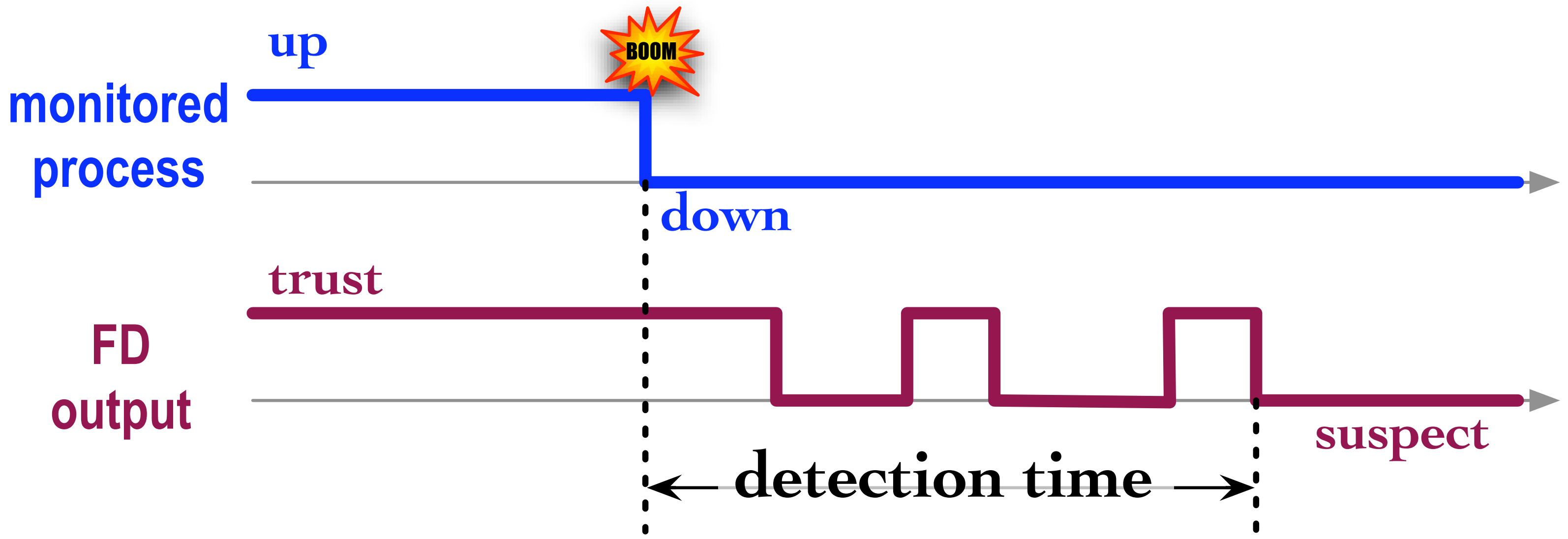
φ Accrual FD



- **φ failure detector** **[Hayashibara et al.; SRDS 2004]**
  - Heartbeat based, estimate arrival distribution

QoS of Failure Detectors

QoS of Failure Detectors

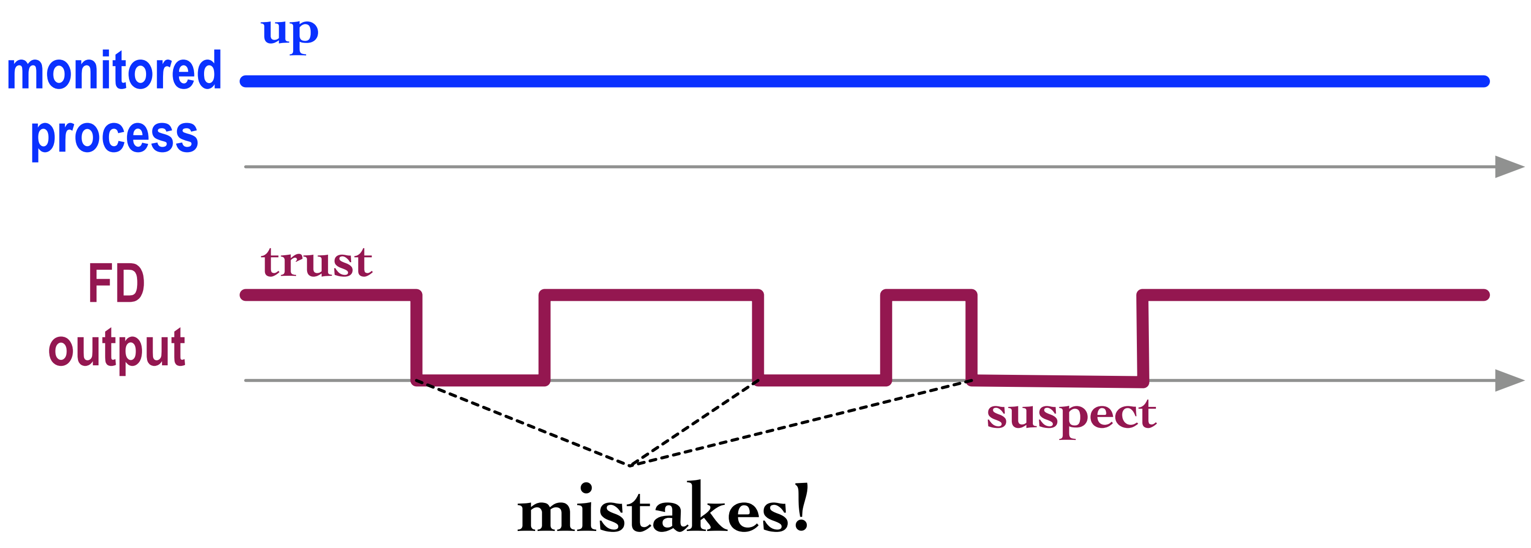


- **Metrics**

  when *p* faulty:

  - Detection time

QoS of Failure Detectors

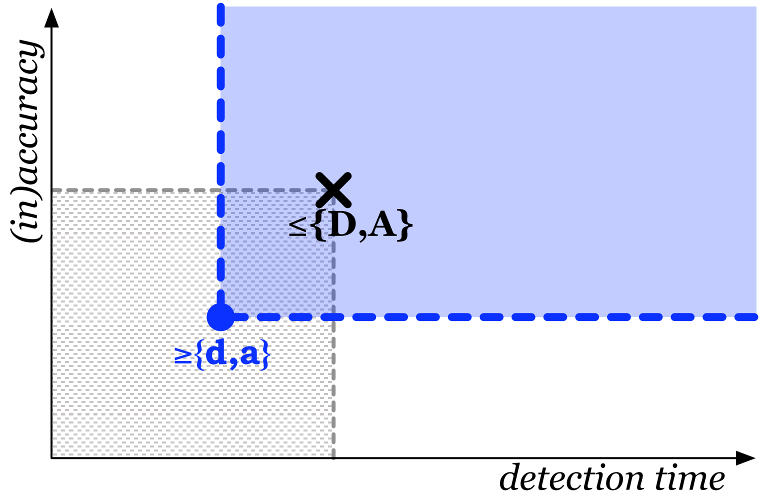


- **Metrics (accuracy)**

  when *p* correct:
  - average mistake rate
  - query accuracy prob.
  - good period duration

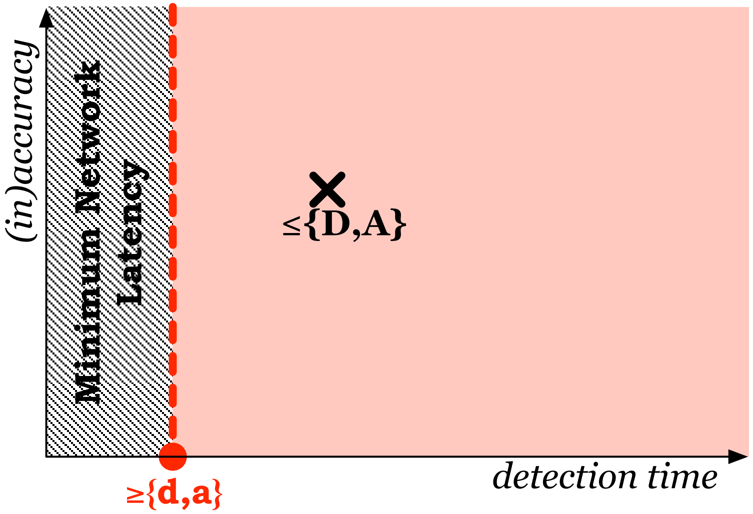
18

Requirements vs. Guarantees

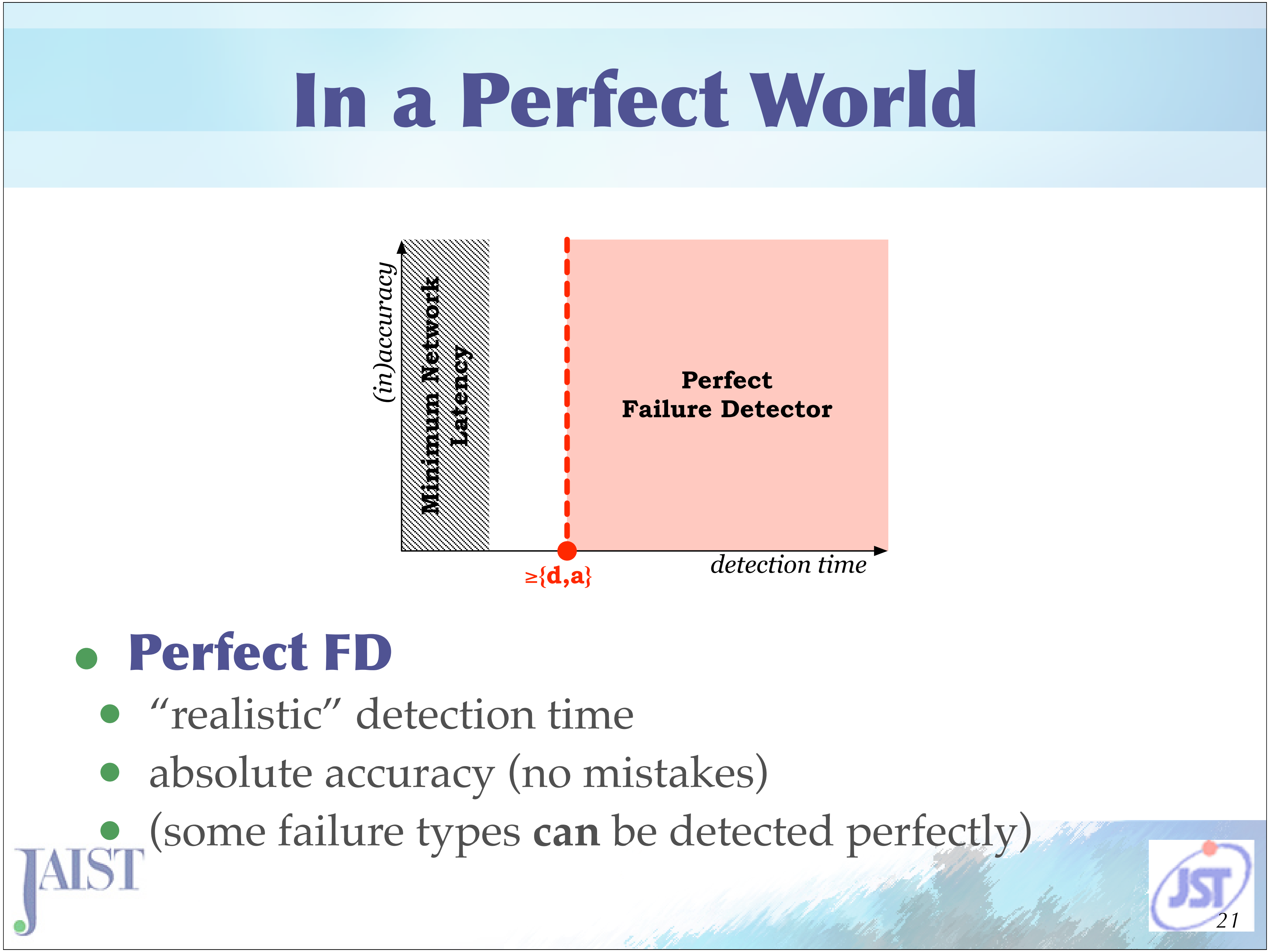


- **Application requirements**
  - ≤{D,A} : max. detect. time, max. mistakes
- **FD QoS**
  - ≥{d,a} : effect. detection time, effect. mistakes

In a Perfect World



- **Ideal**
  - FD limited by min. network latency
  - “acceptable” network/system load

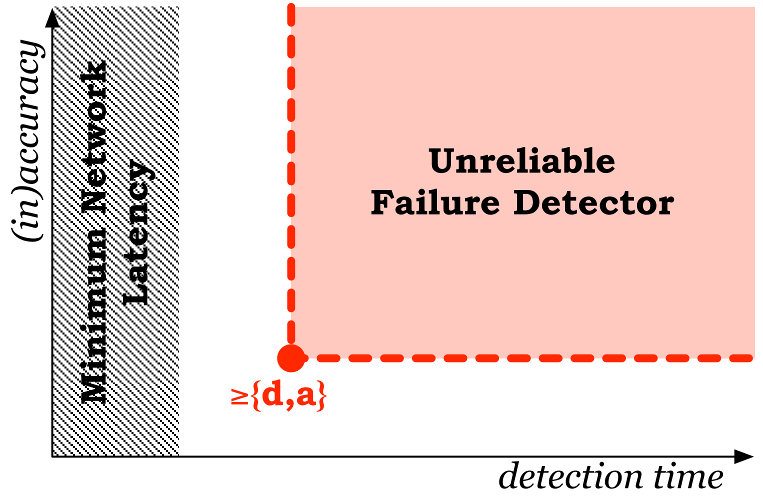


In a Perfect World

- **Perfect FD**
  - “realistic” detection time
  - absolute accuracy (no mistakes)
  - (some failure types **can** be detected perfectly)

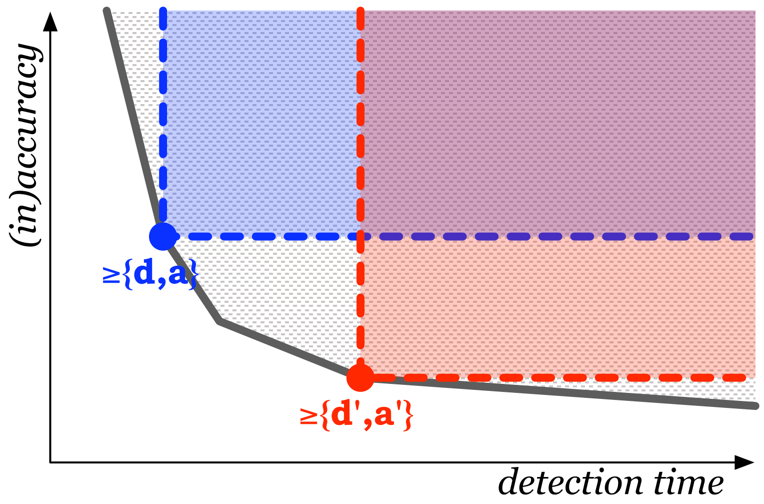
21

In a Less Perfect World



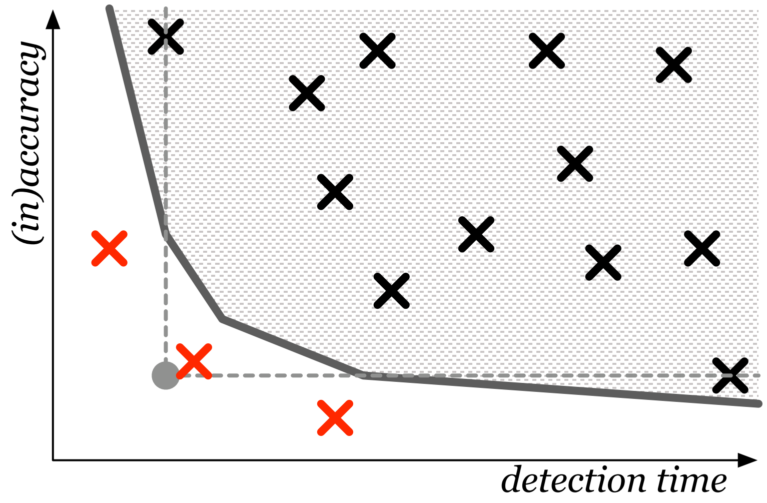
- **Unreliable FDs**
  - “realistic” detection latency
  - imperfect accuracy

Parametric Failure Detector



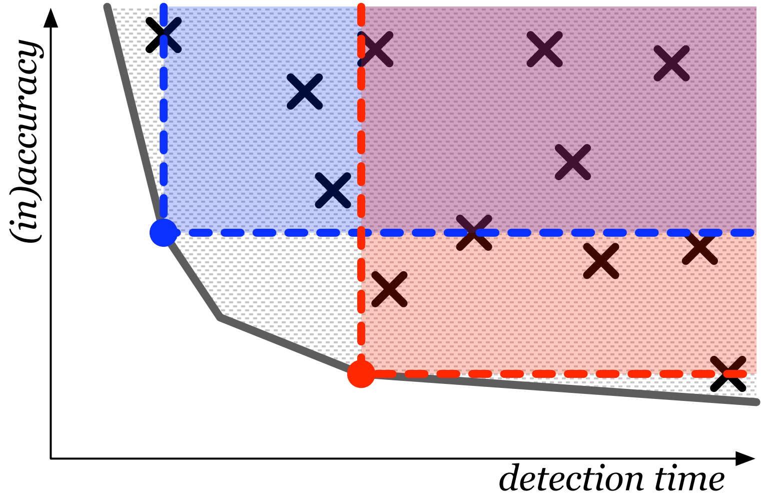
- **Parametric FDs**
  - Parameter value defines FD best QoS
  - E.g., Chen FD,...
  - Tradeoff: accuracy <-> detection latency

QoS Coverage



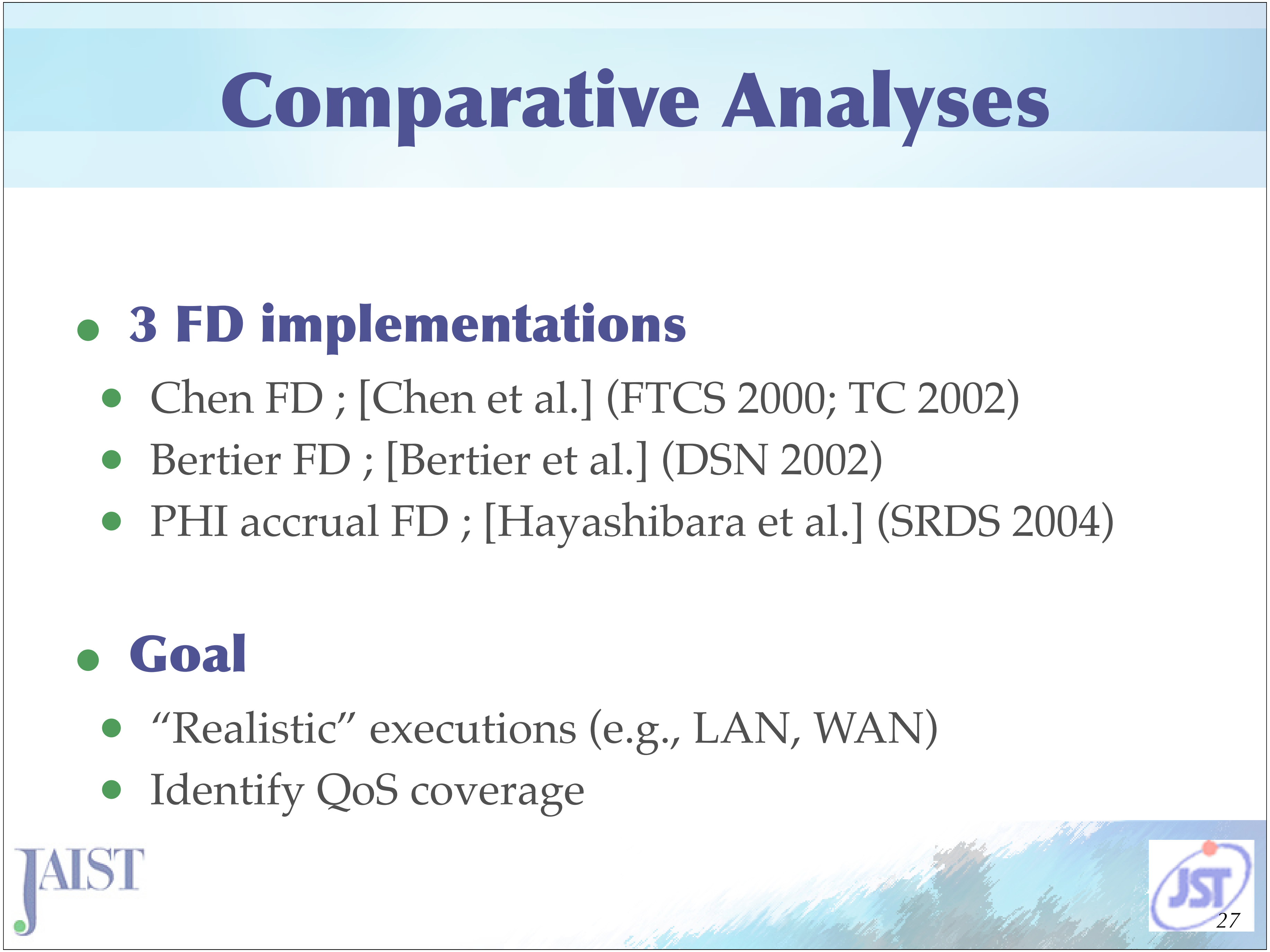
- **Coverage of FD**
  - FD **could** be tuned to support app. req.
  - Measure of FD

Dynamic QoS Coverage



- **Approximate coverage**
  - Instantiate several QoS sets
  - Find minimal set; minimal change

Experimentation

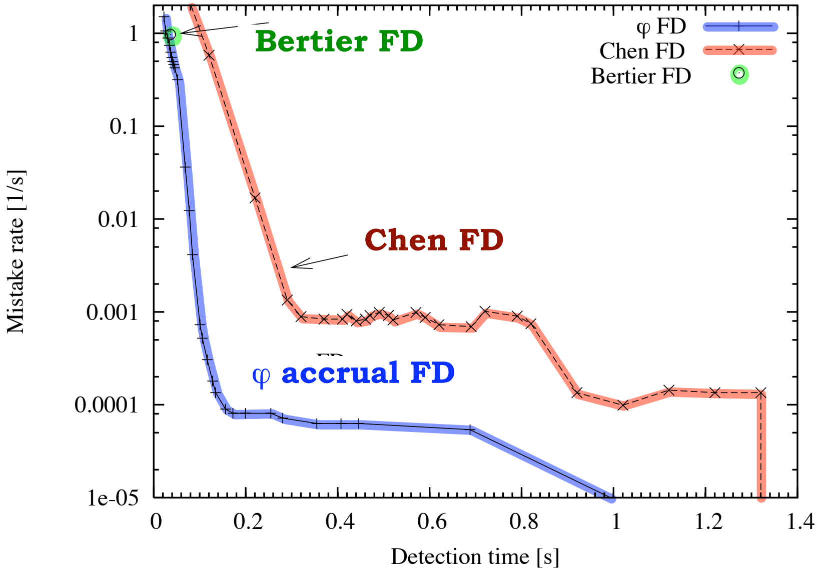


Comparative Analyses

- **3 FD implementations**
  - Chen FD ; [Chen et al.] (FTCS 2000; TC 2002)
  - Bertier FD ; [Bertier et al.] (DSN 2002)
  - PHI accrual FD ; [Hayashibara et al.] (SRDS 2004)
- **Goal**
  - “Realistic” executions (e.g., LAN, WAN)
  - Identify QoS coverage

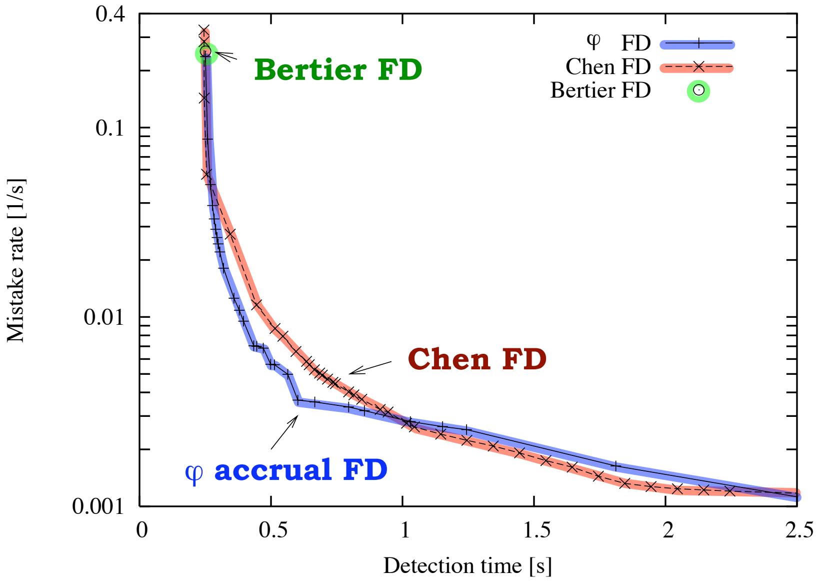
Experimentation: LAN

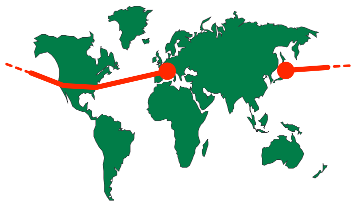
- **LAN**
  - single FastEther hub
- **Parameters**
  - HB interval: 20 ms
  - Duration: 5½ hour
  - Total HB: 1’000’000
  - no loss



Experimentation: WAN

- **WAN**
  - JAIST (JP) – EPFL (CH)
- **Parameters**
  - HB interval: 100 ms
  - Duration: 1 week
  - Total HB: ~ 6’000’000



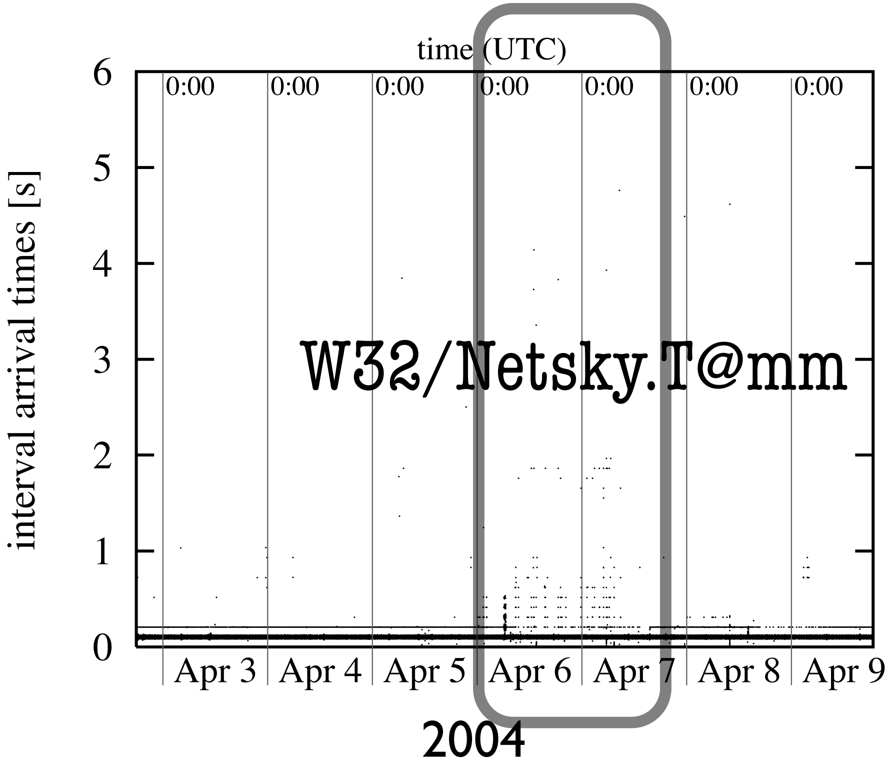


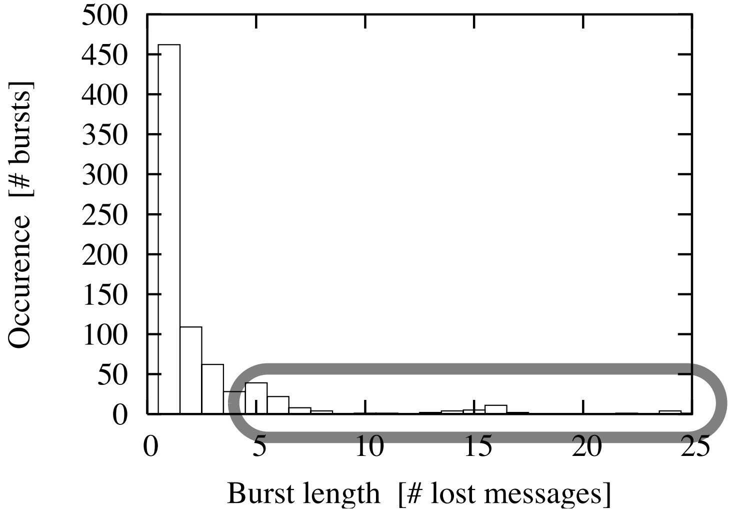
JAIST

JST

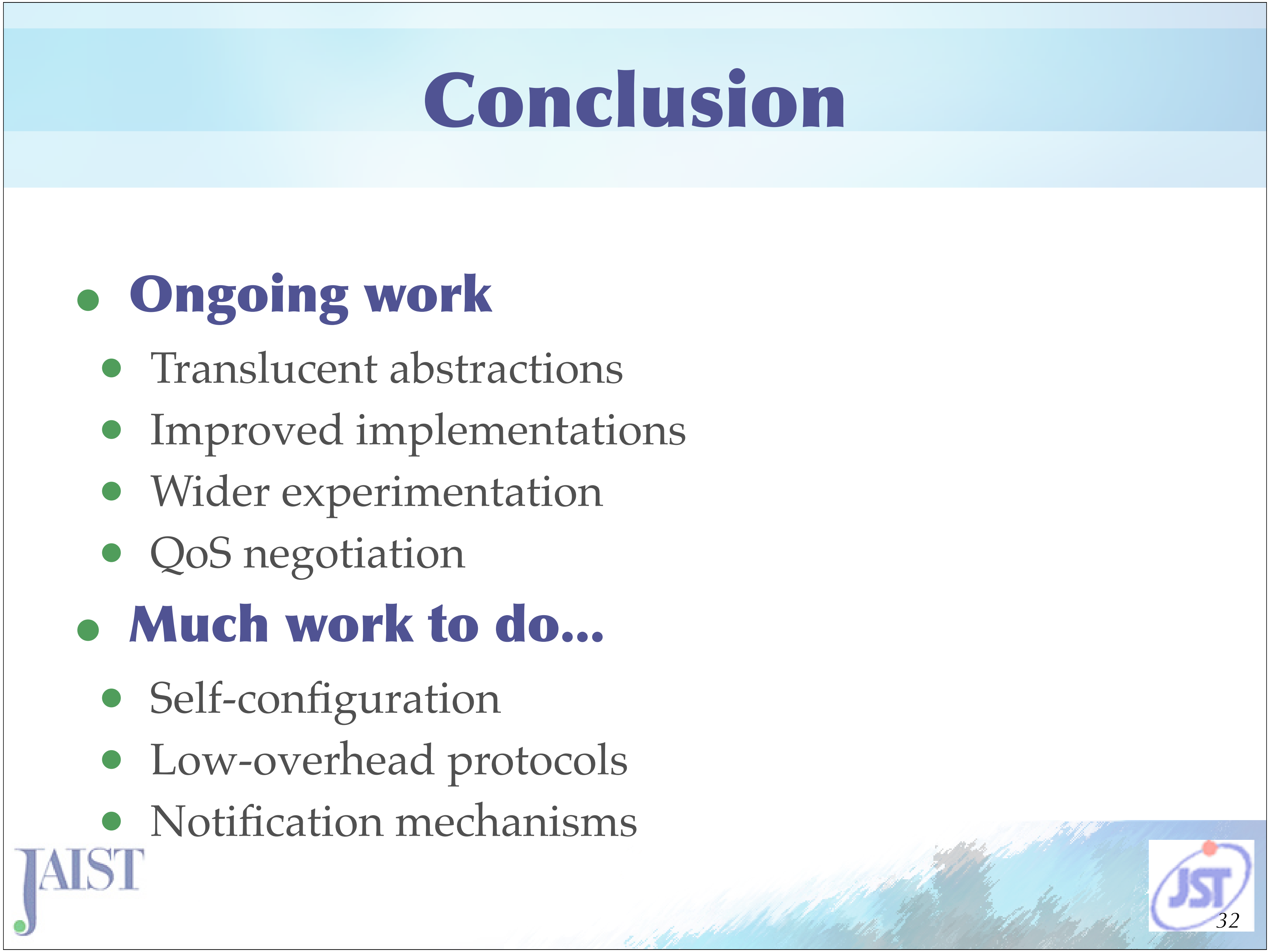
29

Experimentation: WAN





Wrapping Up



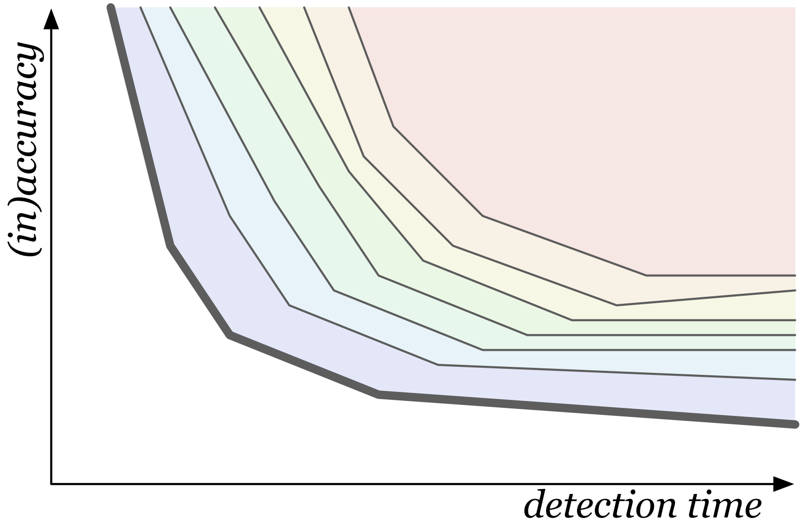
Conclusion

- **Ongoing work**
  - Translucent abstractions
  - Improved implementations
  - Wider experimentation
  - QoS negotiation
- **Much work to do...**
  - Self-configuration
  - Low-overhead protocols
  - Notification mechanisms

32

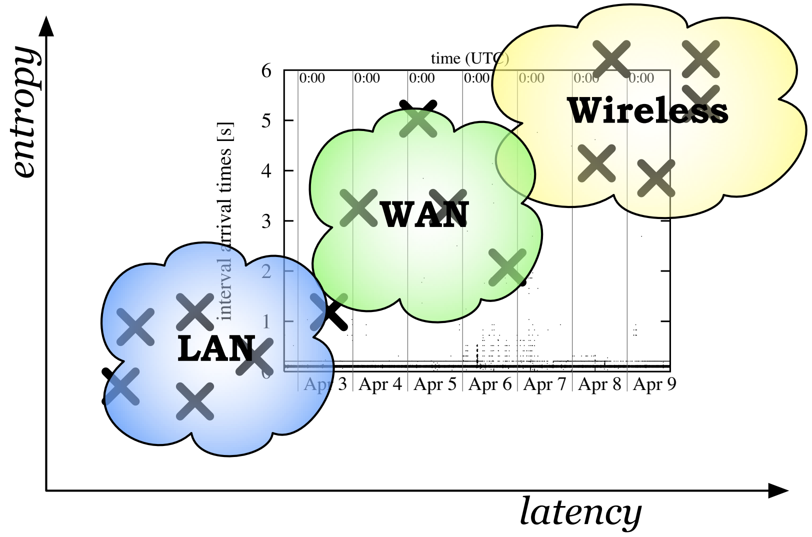
Future Directions

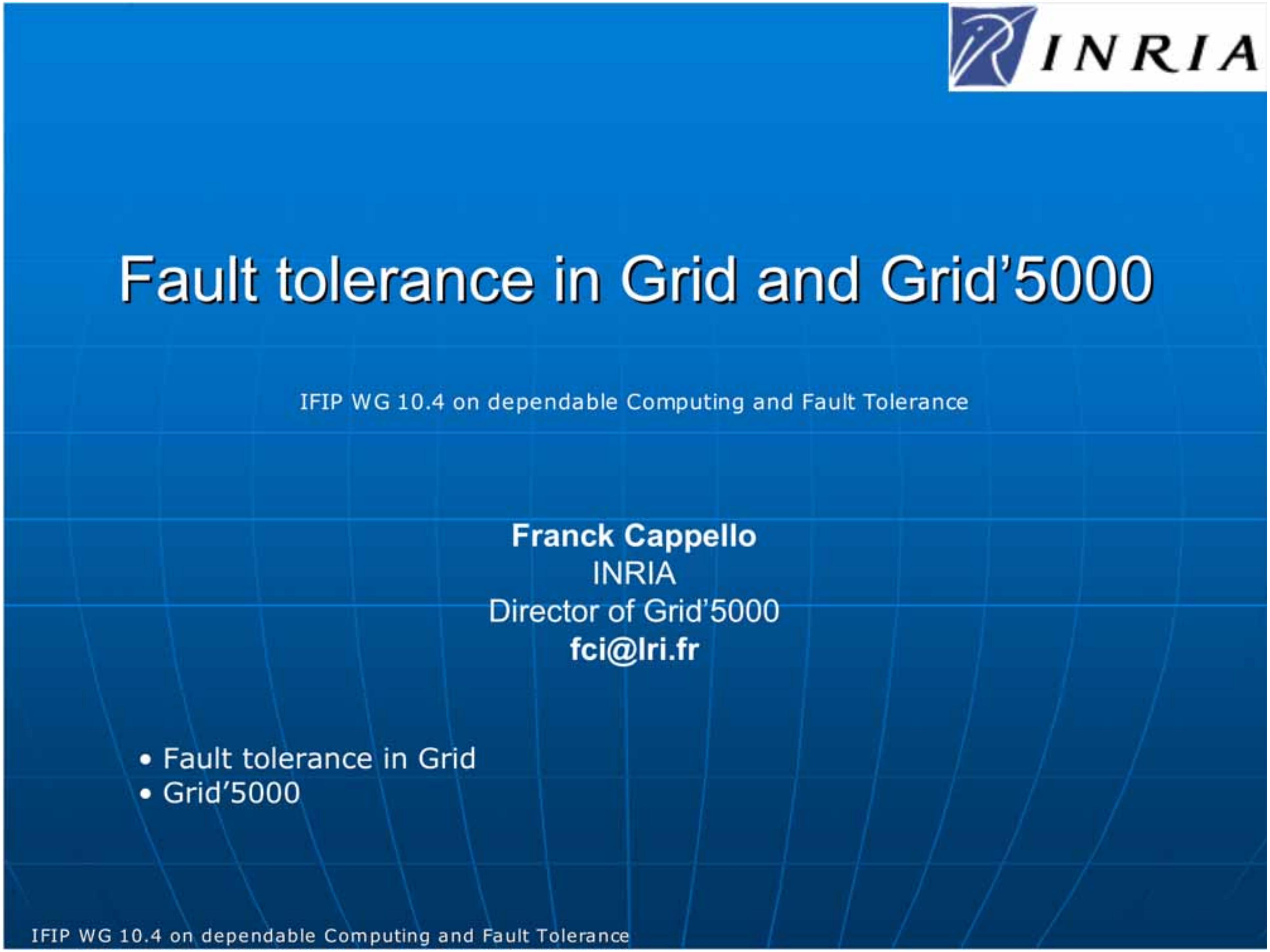
- **QoS Coverage**
  - stricter definition
  - gradients (uncertainty)
- **QoS negotiation**
  - dynamic (re-)negotiation
  - prob./best-effort negotiation
  - fail-safe enforcement



Future Directions

- **Other environments**
  - E.g., wireless, dial-up,...
- **Characterize traffic**
  - metrics
  - clustering
  - “benchmarking” sets





INRIA

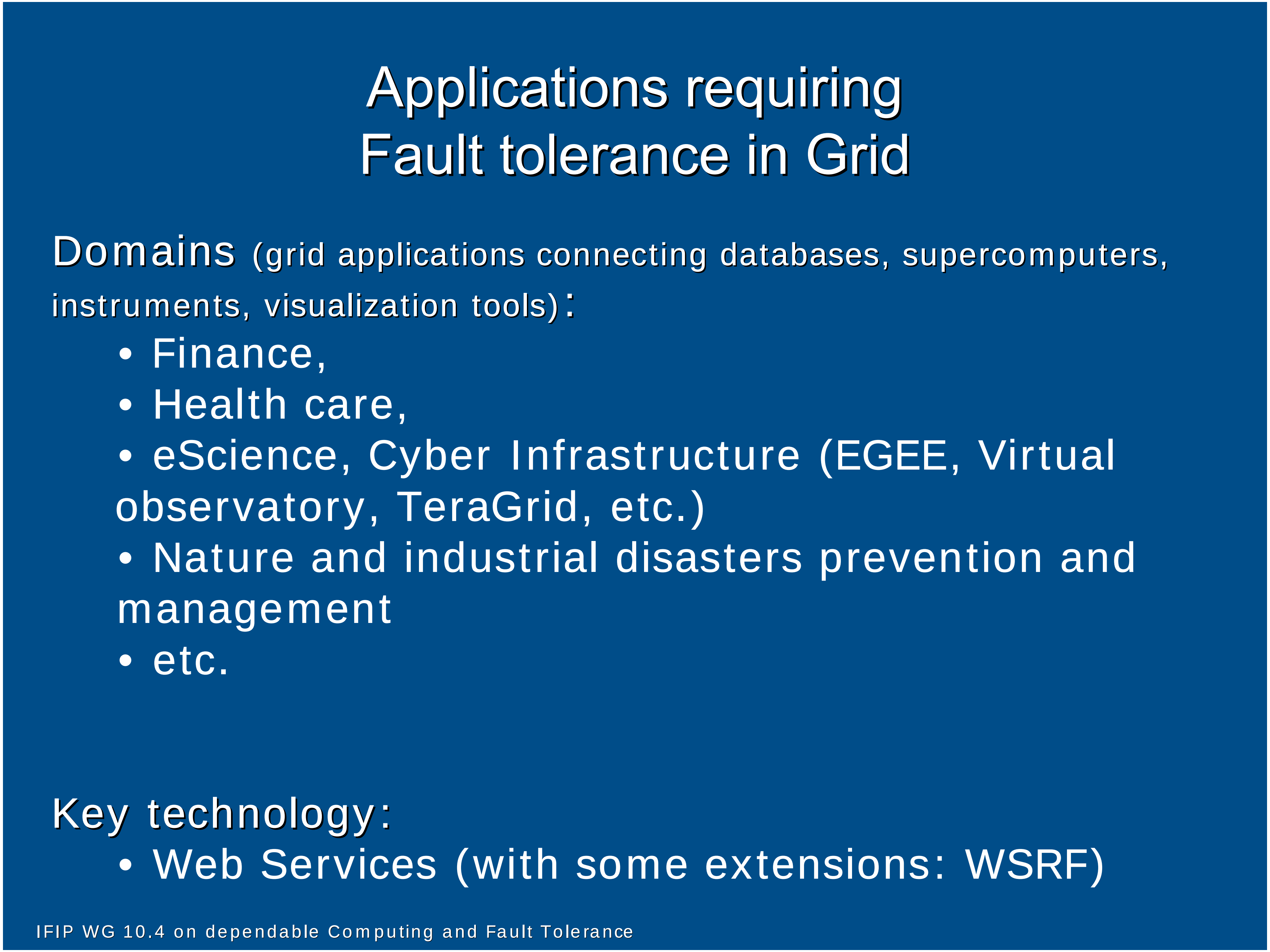
Fault tolerance in Grid and Grid’5000

IFIP WG 10.4 on dependable Computing and Fault Tolerance

Franck Cappello
INRIA
Director of Grid’5000
fci@lri.fr

- Fault tolerance in Grid
- Grid’5000

IFIP WG 10.4 on dependable Computing and Fault Tolerance



Applications requiring Fault tolerance in Grid

Domains (grid applications connecting databases, supercomputers, instruments, visualization tools):

- Finance,
- Health care,
- eScience, Cyber Infrastructure (EGEE, Virtual observatory, TeraGrid, etc.)
- Nature and industrial disasters prevention and management
- etc.

Key technology:

- Web Services (with some extensions: WSRF)

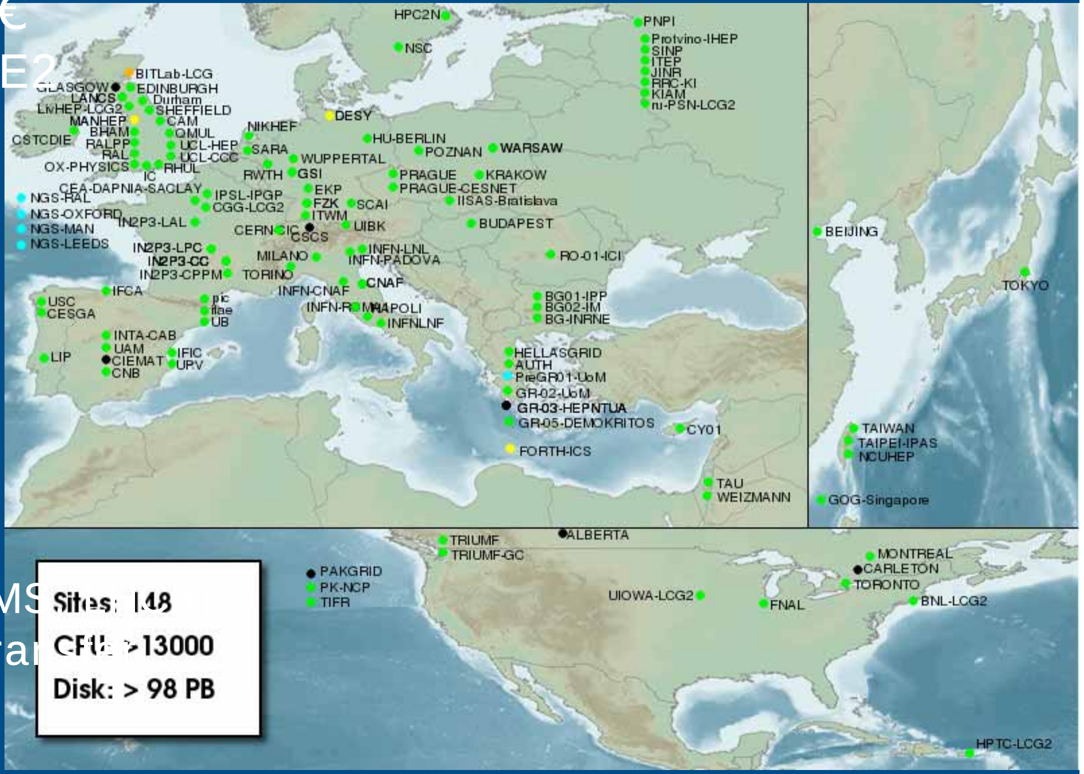
The EGEE project (Enabling Grid for E-SciencE)

- Building and Maintaining a large scale computing infrastructure
- Provide support for Scientists using it.

Size:

Users: 3000
Institutes: 70
Countries: 27
Sites: 148
CPU: > 13000
Disk > 98 PB

Duration: 2 years
Cost: 32M€
Next: EGEE2

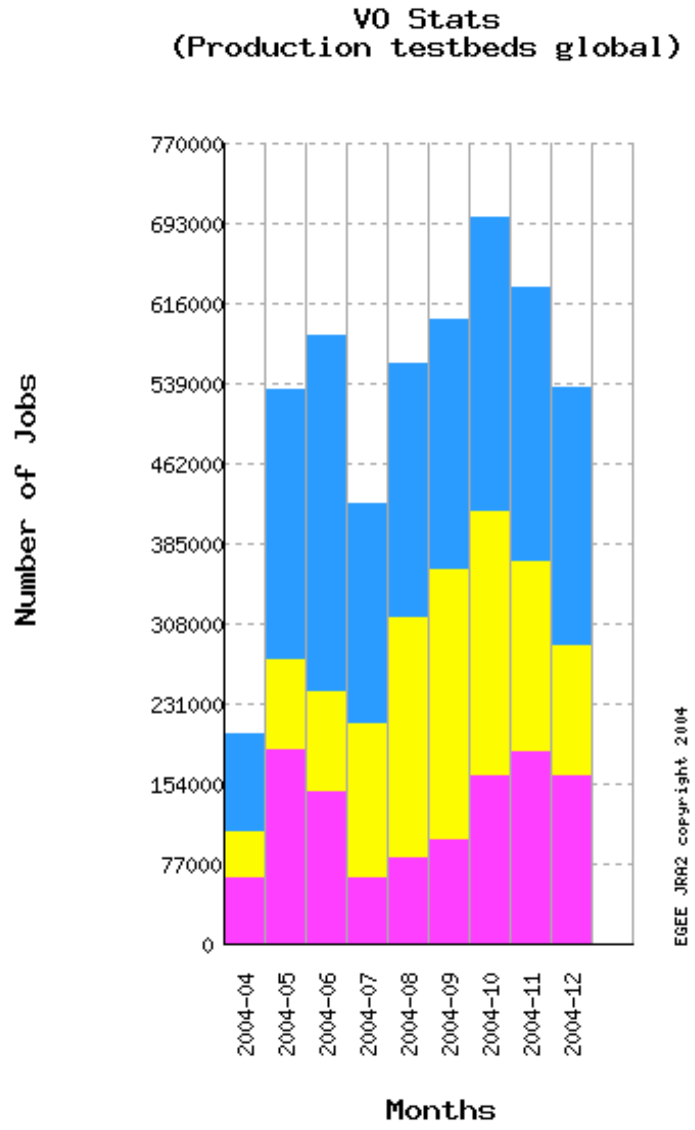


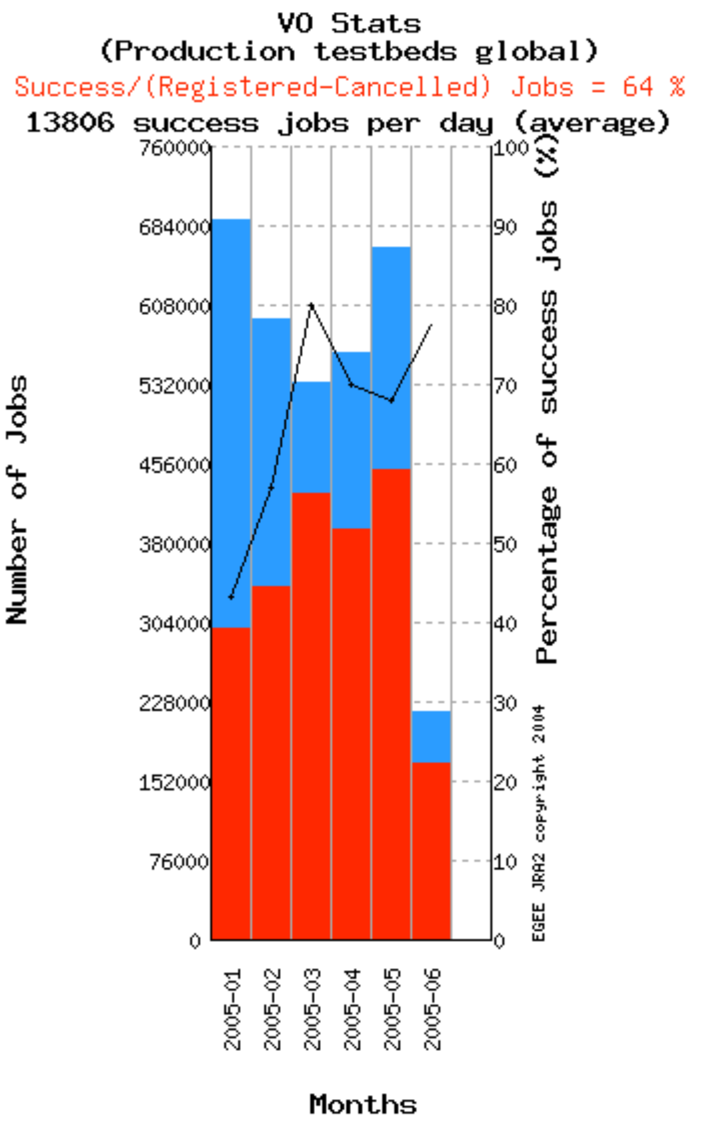
Pilot applications:
LHC experiment (Alice, Atlas, CMS
→ Scale, high bandwidth data tra

Biomedical experiments:
→ Security, Ease of use, distributed data base

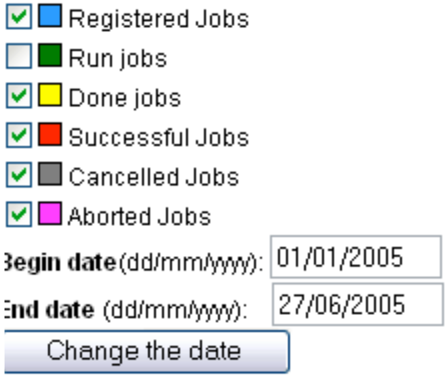
IFIP WG 10.4 on dependable Computing and Fault Tolerance

Job Statistics in EGEE (Enabling Grid for E-SciencE)





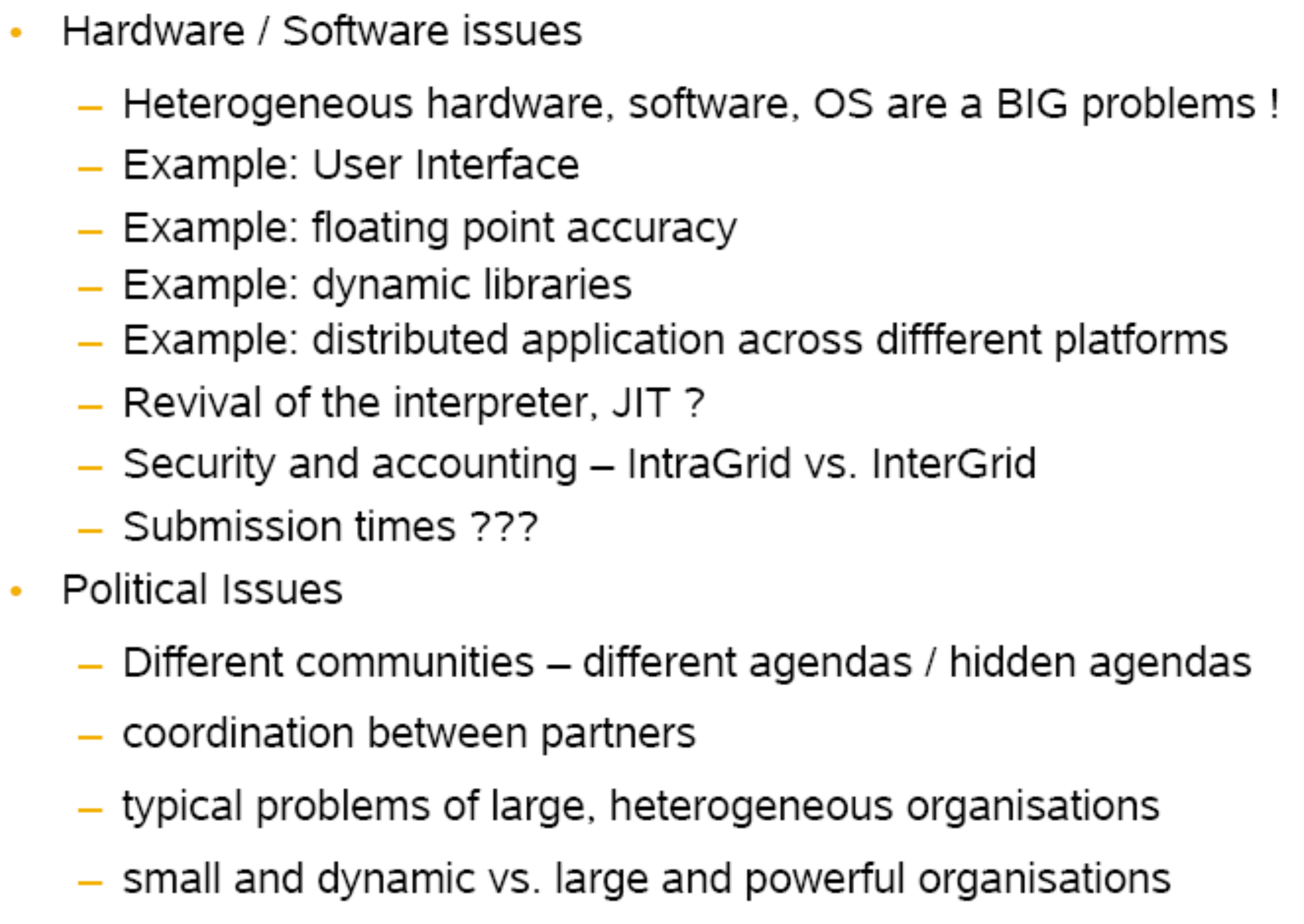
- Successful jobs are jobs whose execution is terminated (done jobs) AND whose done status is OK.
- Registered jobs are registered by the User Interface.
- Cancelled jobs are jobs indicated as cancelled by user.


☑ 🟦 Registered Jobs
□ 🟩 Run jobs
☑ 🟨 Done jobs
☑ 🟥 Successful Jobs
☑ ⬛ Cancelled Jobs
☑ 🟪 Aborted Jobs

3egin date(dd/mm/yyyy): 01/01/2005

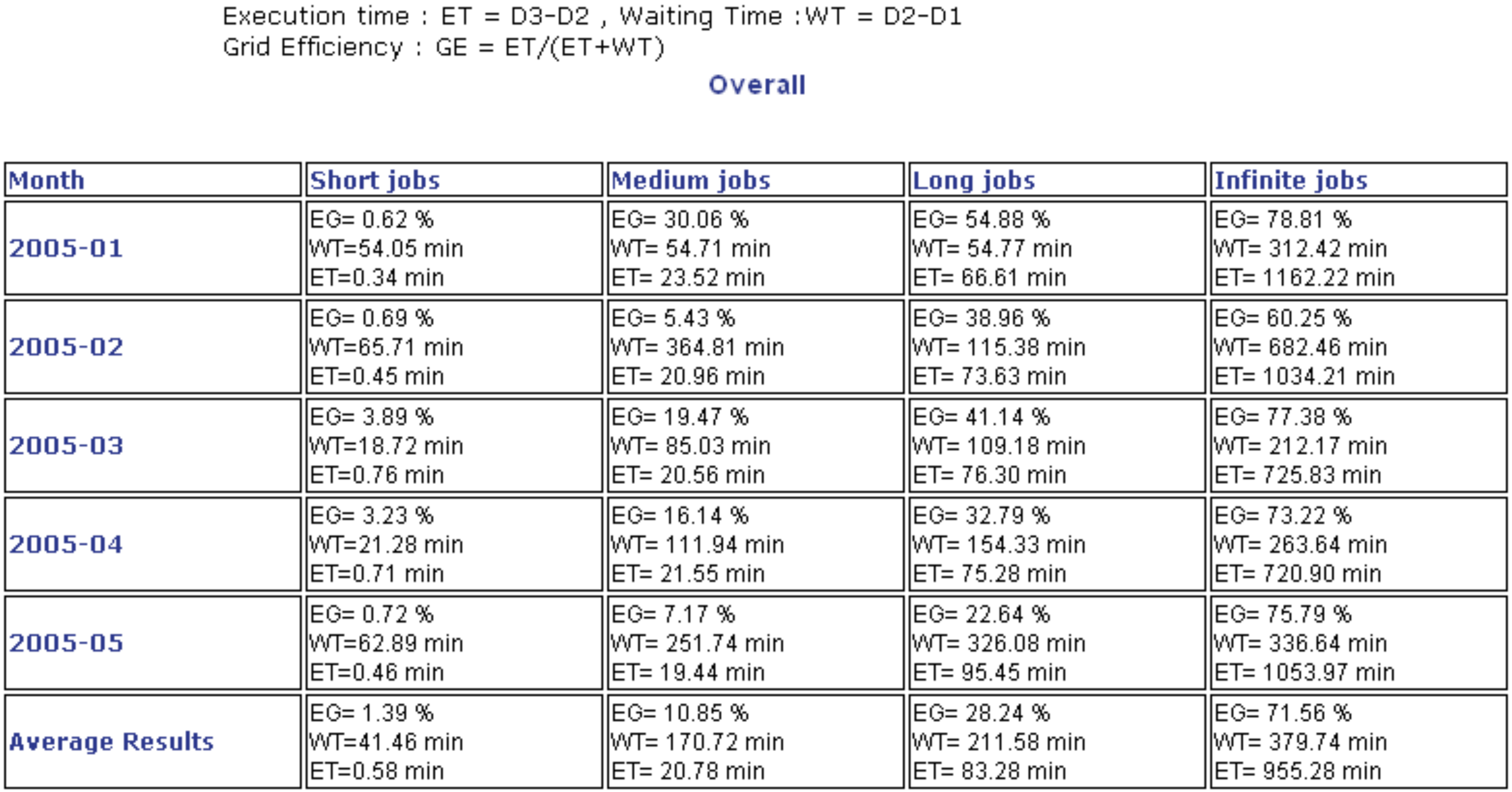
End date (dd/mm/yyyy): 27/06/2005

EGEE issues and problems



- Hardware / Software issues
  - Heterogeneous hardware, software, OS are a BIG problems !
  - Example: User Interface
  - Example: floating point accuracy
  - Example: dynamic libraries
  - Example: distributed application across diffferent platforms
  - Revival of the interpreter, JIT ?
  - Security and accounting – IntraGrid vs. InterGrid
  - Submission times ???
- Political Issues
  - Different communities – different agendas / hidden agendas
  - coordination between partners
  - typical problems of large, heterogeneous organisations
  - small and dynamic vs. large and powerful organisations

Job Efficiency in EGEE



Execution time : ET = D3-D2 , Waiting Time :WT = D2-D1
Grid Efficiency : GE = ET/(ET+WT)

Overall

Month	Short jobs	Medium jobs	Long jobs	Infinite jobs

2005-01	EG= 0.62 %
	WT=54.05 min
	ET=0.34 min
		EG= 30.06 %
		WT= 54.71 min
		ET= 23.52 min
			EG= 54.88 %
			WT= 54.77 min
			ET= 66.61 min
				EG= 78.81 %
				WT= 312.42 min
				ET= 1162.22 min

2005-02	EG= 0.69 %
	WT=65.71 min
	ET=0.45 min
		EG= 5.43 %
		WT= 364.81 min
		ET= 20.96 min
			EG= 38.96 %
			WT= 115.38 min
			ET= 73.63 min
				EG= 60.25 %
				WT= 682.46 min
				ET= 1034.21 min

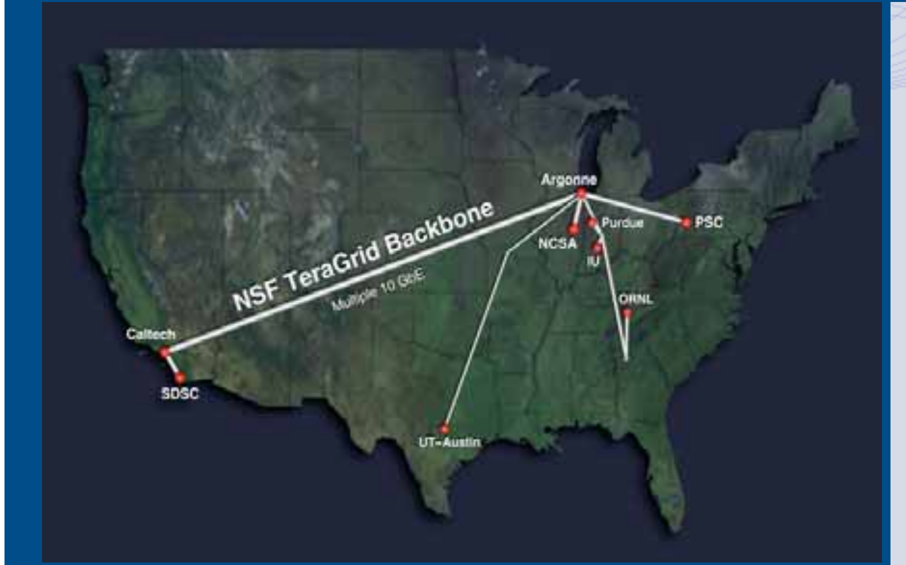
2005-03	EG= 3.89 %
	WT=18.72 min
	ET=0.76 min
		EG= 19.47 %
		WT= 85.03 min
		ET= 20.56 min
			EG= 41.14 %
			WT= 109.18 min
			ET= 76.30 min
				EG= 77.38 %
				WT= 212.17 min
				ET= 725.83 min

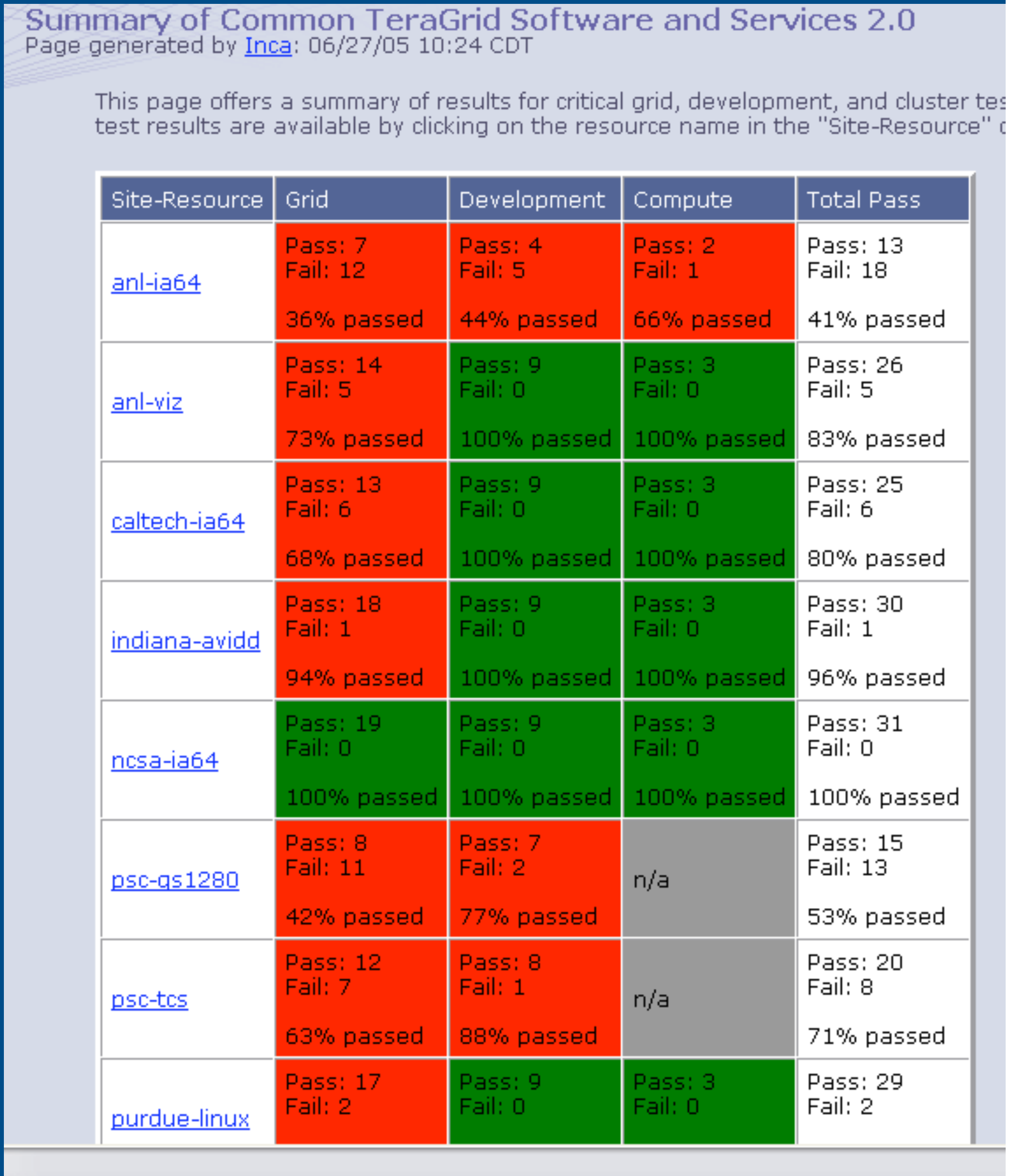
2005-04	EG= 3.23 %
	WT=21.28 min
	ET=0.71 min
		EG= 16.14 %
		WT= 111.94 min
		ET= 21.55 min
			EG= 32.79 %
			WT= 154.33 min
			ET= 75.28 min
				EG= 73.22 %
				WT= 263.64 min
				ET= 720.90 min

2005-05	EG= 0.72 %
	WT=62.89 min
	ET=0.46 min
		EG= 7.17 %
		WT= 251.74 min
		ET= 19.44 min
			EG= 22.64 %
			WT= 326.08 min
			ET= 95.45 min
				EG= 75.79 %
				WT= 336.64 min
				ET= 1053.97 min

Average Results	EG= 1.39 %
	WT=41.46 min
	ET=0.58 min
		EG= 10.85 %
		WT= 170.72 min
		ET= 20.78 min
			EG= 28.24 %
			WT= 211.58 min
			ET= 83.28 min
				EG= 71.56 %
				WT= 379.74 min
				ET= 955.28 min

Software Status in TERA GRID 1/2





Summary of Common TeraGrid Software and Services 2.0

Page generated by Inca: 06/27/05 10:24 CDT

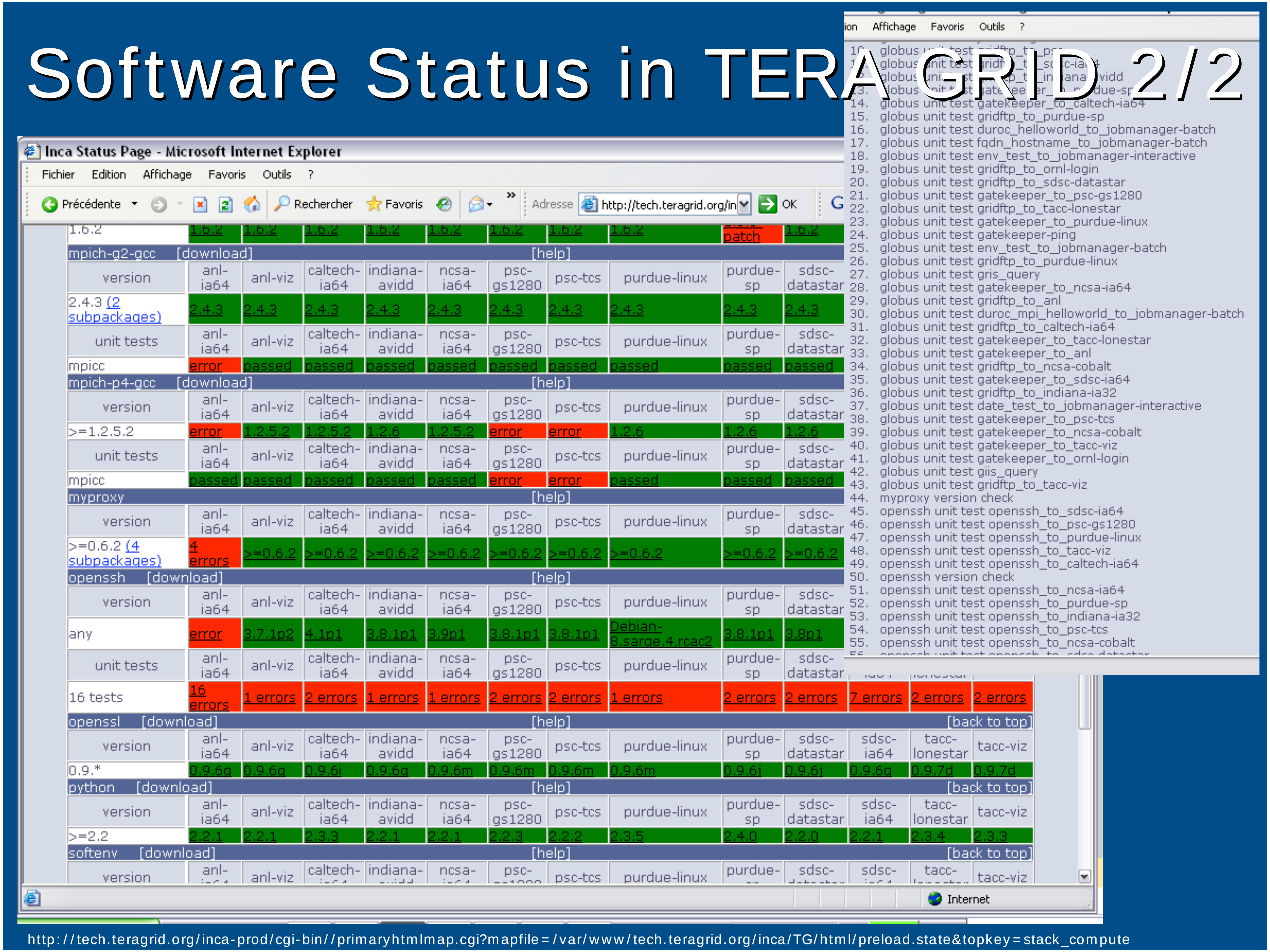
This page offers a summary of results for critical grid, development, and cluster tes
test results are available by clicking on the resource name in the "Site-Resource" c

Site-Resource	Grid	Development	Compute	Total Pass
anl-ia64	Pass: 7 Fail: 12 36% passed	Pass: 4 Fail: 5 44% passed	Pass: 2 Fail: 1 66% passed	Pass: 13 Fail: 18 41% passed
anl-viz	Pass: 14 Fail: 5 73% passed	Pass: 9 Fail: 0 100% passed	Pass: 3 Fail: 0 100% passed	Pass: 26 Fail: 5 83% passed
caltech-ia64	Pass: 13 Fail: 6 68% passed	Pass: 9 Fail: 0 100% passed	Pass: 3 Fail: 0 100% passed	Pass: 25 Fail: 6 80% passed
indiana-avidd	Pass: 18 Fail: 1 94% passed	Pass: 9 Fail: 0 100% passed	Pass: 3 Fail: 0 100% passed	Pass: 30 Fail: 1 96% passed
ncsa-ia64	Pass: 19 Fail: 0 100% passed	Pass: 9 Fail: 0 100% passed	Pass: 3 Fail: 0 100% passed	Pass: 31 Fail: 0 100% passed
psc-qs1280	Pass: 8 Fail: 11 42% passed	Pass: 7 Fail: 2 77% passed	n/a	Pass: 15 Fail: 13 53% passed
psc-tcs	Pass: 12 Fail: 7 63% passed	Pass: 8 Fail: 1 88% passed	n/a	Pass: 20 Fail: 8 71% passed
purdue-linux	Pass: 17 Fail: 2	Pass: 9 Fail: 0	Pass: 3 Fail: 0	Pass: 29 Fail: 2

TeraGrid:
-integrated, persistent computational resource.
-Deployment completed in September 2004,
-40 teraflops of computing power
-nearly 2 petabytes of storage,
-interconnections at 10-30 gigabits/sec.
(via a dedicated national network.)

http://tech.teragrid.org/inca-prod/cgi-bin//primaryhtmlmap.cgi?mapfile=/var/www/tech.teragrid.org/inca/TG/html/preload.state&topkey=exec

Software Status in TERA GRID 2/2


Inca Status Page - Microsoft Internet Explorer
Fichier Edition Affichage Favoris Outils ?
Précédente
Rechercher
Favoris
Adresse
http://tech.teragrid.org/in
OK
Internet
ion Affichage Favoris Outils ?

1.6.2	1.6.2	1.6.2	1.6.2	1.6.2	1.6.2	1.6.2	1.6.2	1.6.2	patch	1.6.2

mpich-g2-gcc [download] [help]

version	anl-ia64	anl-viz	caltech-ia64	indiana-avidd	ncsa-ia64	psc-gs1280	psc-tcs	purdue-linux	purdue-sp	sdsc-datastar
2.4.3 (2 subpackages)	2.4.3	2.4.3	2.4.3	2.4.3	2.4.3	2.4.3	2.4.3	2.4.3	2.4.3	2.4.3
unit tests	anl-ia64	anl-viz	caltech-ia64	indiana-avidd	ncsa-ia64	psc-gs1280	psc-tcs	purdue-linux	purdue-sp	sdsc-datastar
mpicc	error	passed	passed	passed	passed	passed	passed	passed	passed	passed

mpich-p4-gcc [download] [help]

version	anl-ia64	anl-viz	caltech-ia64	indiana-avidd	ncsa-ia64	psc-gs1280	psc-tcs	purdue-linux	purdue-sp	sdsc-datastar
>=1.2.5.2	error	1.2.5.2	1.2.5.2	1.2.6	1.2.5.2	error	error	1.2.6	1.2.6	1.2.6
unit tests	anl-ia64	anl-viz	caltech-ia64	indiana-avidd	ncsa-ia64	psc-gs1280	psc-tcs	purdue-linux	purdue-sp	sdsc-datastar
mpicc	passed	passed	passed	passed	passed	error	error	passed	passed	passed

myproxy [help]

version	anl-ia64	anl-viz	caltech-ia64	indiana-avidd	ncsa-ia64	psc-gs1280	psc-tcs	purdue-linux	purdue-sp	sdsc-datastar
>=0.6.2 (4 subpackages)	4 errors	>=0.6.2	>=0.6.2	>=0.6.2	>=0.6.2	>=0.6.2	>=0.6.2	>=0.6.2	>=0.6.2	>=0.6.2

openssh [download] [help]

version	anl-ia64	anl-viz	caltech-ia64	indiana-avidd	ncsa-ia64	psc-gs1280	psc-tcs	purdue-linux	purdue-sp	sdsc-datastar			
any	error	3.7.1p2	4.1p1	3.8.1p1	3.9p1	3.8.1p1	3.8.1p1	Debian-8.sarge.4.rcac2	3.8.1p1	3.8p1			
unit tests	anl-ia64	anl-viz	caltech-ia64	indiana-avidd	ncsa-ia64	psc-gs1280	psc-tcs	purdue-linux	purdue-sp	sdsc-datastar			
16 tests	16 errors	1 errors	2 errors	1 errors	1 errors	2 errors	2 errors	1 errors	2 errors	2 errors	7 errors	2 errors	2 errors

openssl [download] [help] [back to top]

version	anl-ia64	anl-viz	caltech-ia64	indiana-avidd	ncsa-ia64	psc-gs1280	psc-tcs	purdue-linux	purdue-sp	sdsc-datastar	sdsc-ia64	tacc-lonestar	tacc-viz
0.9.*	0.9.6g	0.9.6g	0.9.6i	0.9.6g	0.9.6m	0.9.6m	0.9.6m	0.9.6m	0.9.6j	0.9.6j	0.9.6g	0.9.7d	0.9.7d

python [download] [help] [back to top]

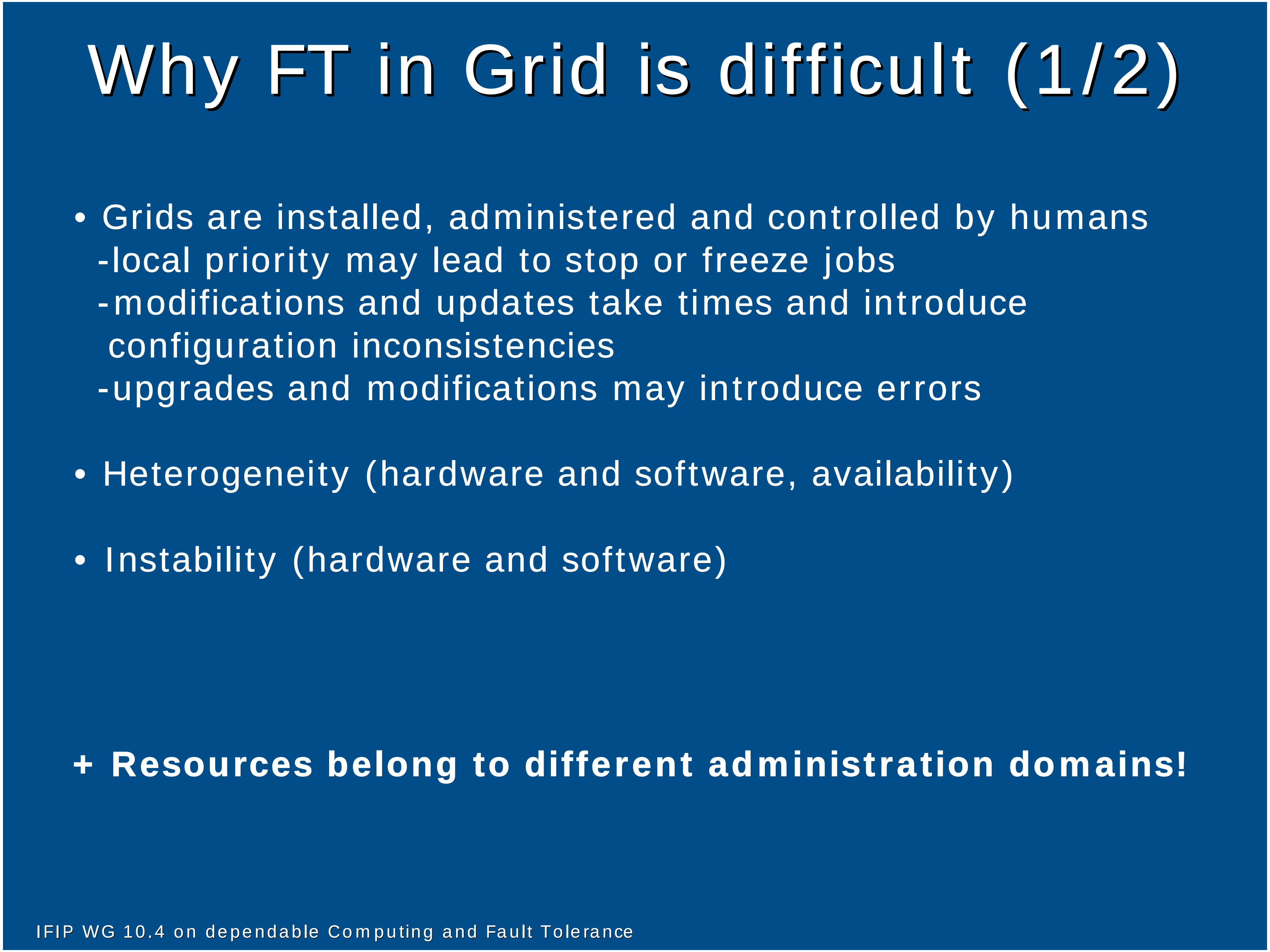
version	anl-ia64	anl-viz	caltech-ia64	indiana-avidd	ncsa-ia64	psc-gs1280	psc-tcs	purdue-linux	purdue-sp	sdsc-datastar	sdsc-ia64	tacc-lonestar	tacc-viz
>=2.2	2.2.1	2.2.1	2.3.3	2.2.1	2.2.1	2.2.3	2.2.2	2.3.5	2.4.0	2.2.0	2.2.1	2.3.4	2.3.3

softenv [download] [help] [back to top]

version	anl-ia64	anl-viz	caltech-ia64	indiana-avidd	ncsa-ia64	psc-gs1280	psc-tcs	purdue-linux	purdue-sp	sdsc-datastar	sdsc-ia64	tacc-lonestar	tacc-viz

10. globus unit test gridftp_to_psc
11. globus unit test gridftp_to_sdsc-ia64
12. globus unit test gridftp_to_indiana-avidd
13. globus unit test gatekeeper_to_purdue-sp
14. globus unit test gatekeeper_to_caltech-ia64
15. globus unit test gridftp_to_purdue-sp
16. globus unit test duroc_helloworld_to_jobmanager-batch
17. globus unit test fqdn_hostname_to_jobmanager-batch
18. globus unit test env_test_to_jobmanager-interactive
19. globus unit test gridftp_to_ornl-login
20. globus unit test gridftp_to_sdsc-datastar
21. globus unit test gatekeeper_to_psc-gs1280
22. globus unit test gridftp_to_tacc-lonestar
23. globus unit test gatekeeper_to_purdue-linux
24. globus unit test gatekeeper-ping
25. globus unit test env_test_to_jobmanager-batch
26. globus unit test gridftp_to_purdue-linux
27. globus unit test gris_query
28. globus unit test gatekeeper_to_ncsa-ia64
29. globus unit test gridftp_to_anl
30. globus unit test duroc_mpi_helloworld_to_jobmanager-batch
31. globus unit test gridftp_to_caltech-ia64
32. globus unit test gatekeeper_to_tacc-lonestar
33. globus unit test gatekeeper_to_anl
34. globus unit test gridftp_to_ncsa-cobalt
35. globus unit test gatekeeper_to_sdsc-ia64
36. globus unit test gridftp_to_indiana-ia32
37. globus unit test date_test_to_jobmanager-interactive
38. globus unit test gatekeeper_to_psc-tcs
39. globus unit test gatekeeper_to_ncsa-cobalt
40. globus unit test gatekeeper_to_tacc-viz
41. globus unit test gatekeeper_to_ornl-login
42. globus unit test giis_query
43. globus unit test gridftp_to_tacc-viz
44. myproxy version check
45. openssh unit test openssh_to_sdsc-ia64
46. openssh unit test openssh_to_psc-gs1280
47. openssh unit test openssh_to_purdue-linux
48. openssh unit test openssh_to_tacc-viz
49. openssh unit test openssh_to_caltech-ia64
50. openssh version check
51. openssh unit test openssh_to_ncsa-ia64
52. openssh unit test openssh_to_purdue-sp
53. openssh unit test openssh_to_indiana-ia32
54. openssh unit test openssh_to_psc-tcs
55. openssh unit test openssh_to_ncsa-cobalt
56. openssh unit test openssh_to_sdsc-datastar

http://tech.teragrid.org/inca-prod/cgi-bin//primaryhtmlmap.cgi?mapfile=/var/www/tech.teragrid.org/inca/TG/html/preload.state&topkey=stack_compute

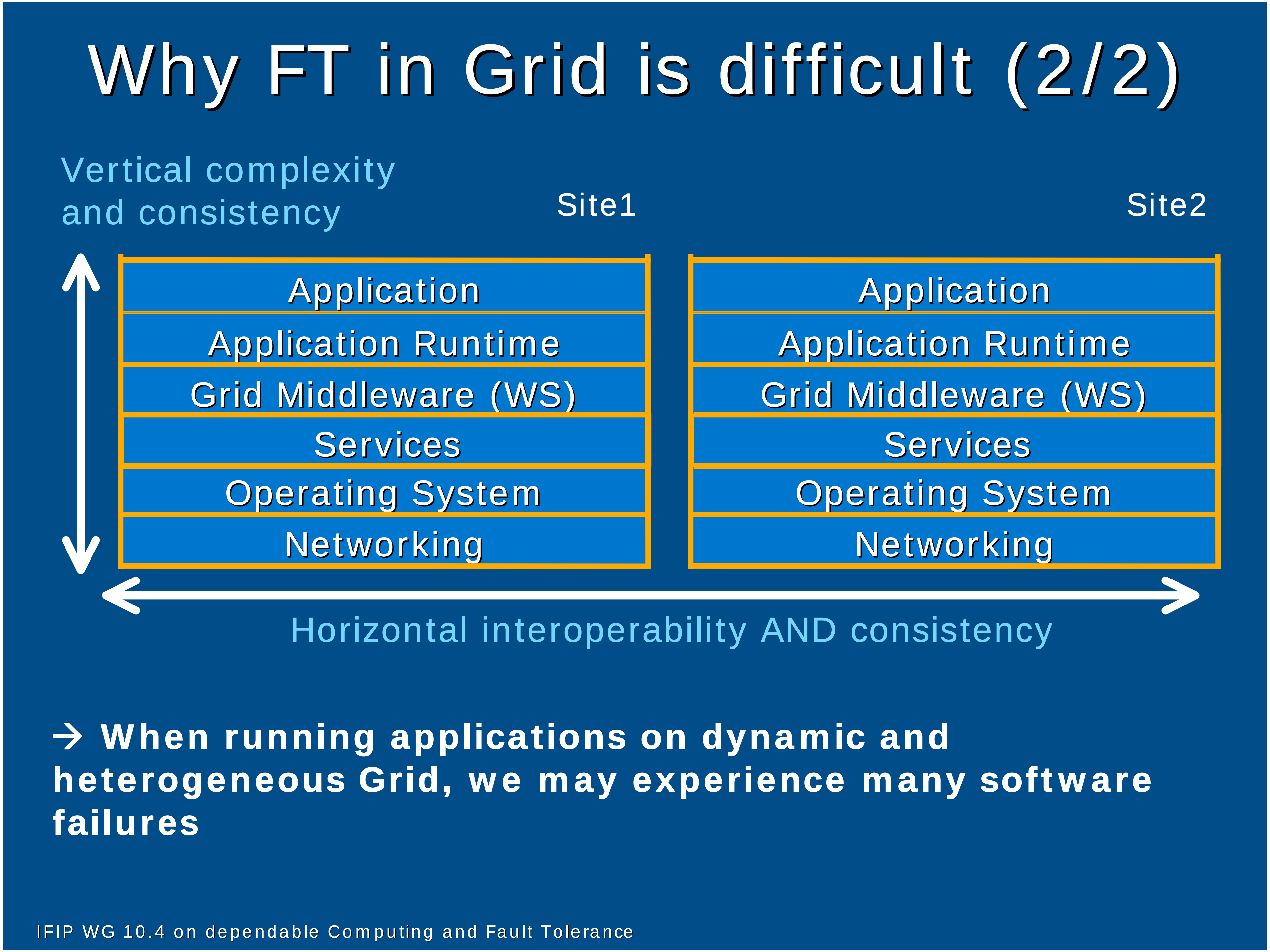


Why FT in Grid is difficult (1/2)

- Grids are installed, administered and controlled by humans
  - -local priority may lead to stop or freeze jobs
  - -modifications and updates take times and introduce configuration inconsistencies
  - -upgrades and modifications may introduce errors
- Heterogeneity (hardware and software, availability)
- Instability (hardware and software)

+ Resources belong to different administration domains!

IFIP WG 10.4 on dependable Computing and Fault Tolerance



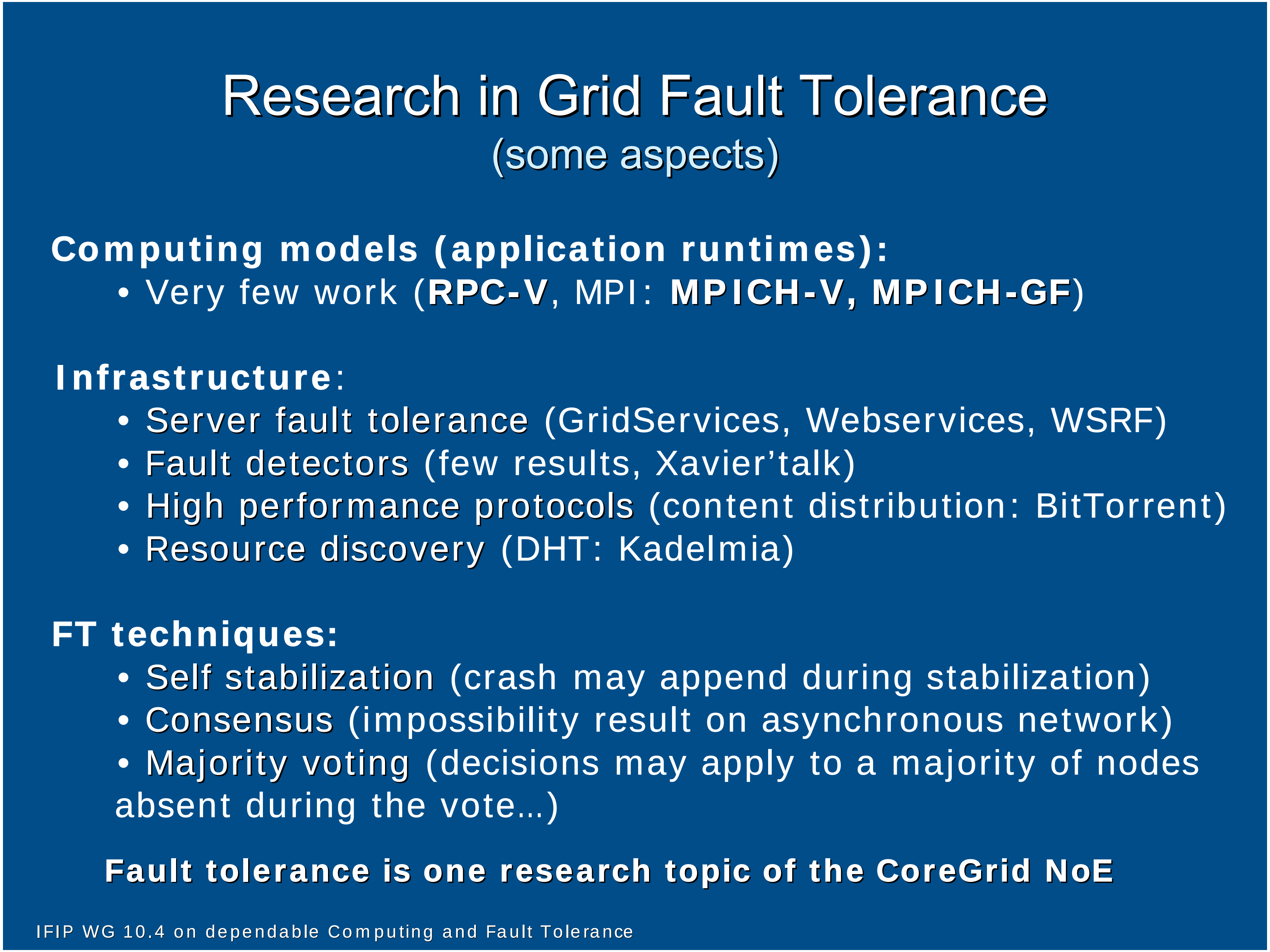
Why FT in Grid is difficult (2/2)

Vertical complexity
and consistency

Site1	Site2
Application	Application
Application Runtime	Application Runtime
Grid Middleware (WS)	Grid Middleware (WS)
Services	Services
Operating System	Operating System
Networking	Networking

Horizontal interoperability AND consistency

→ **When running applications on dynamic and heterogeneous Grid, we may experience many software failures**



Research in Grid Fault Tolerance

(some aspects)

Computing models (application runtimes):

- Very few work (**RPC-V**, MPI: **MPICH-V, MPICH-GF**)

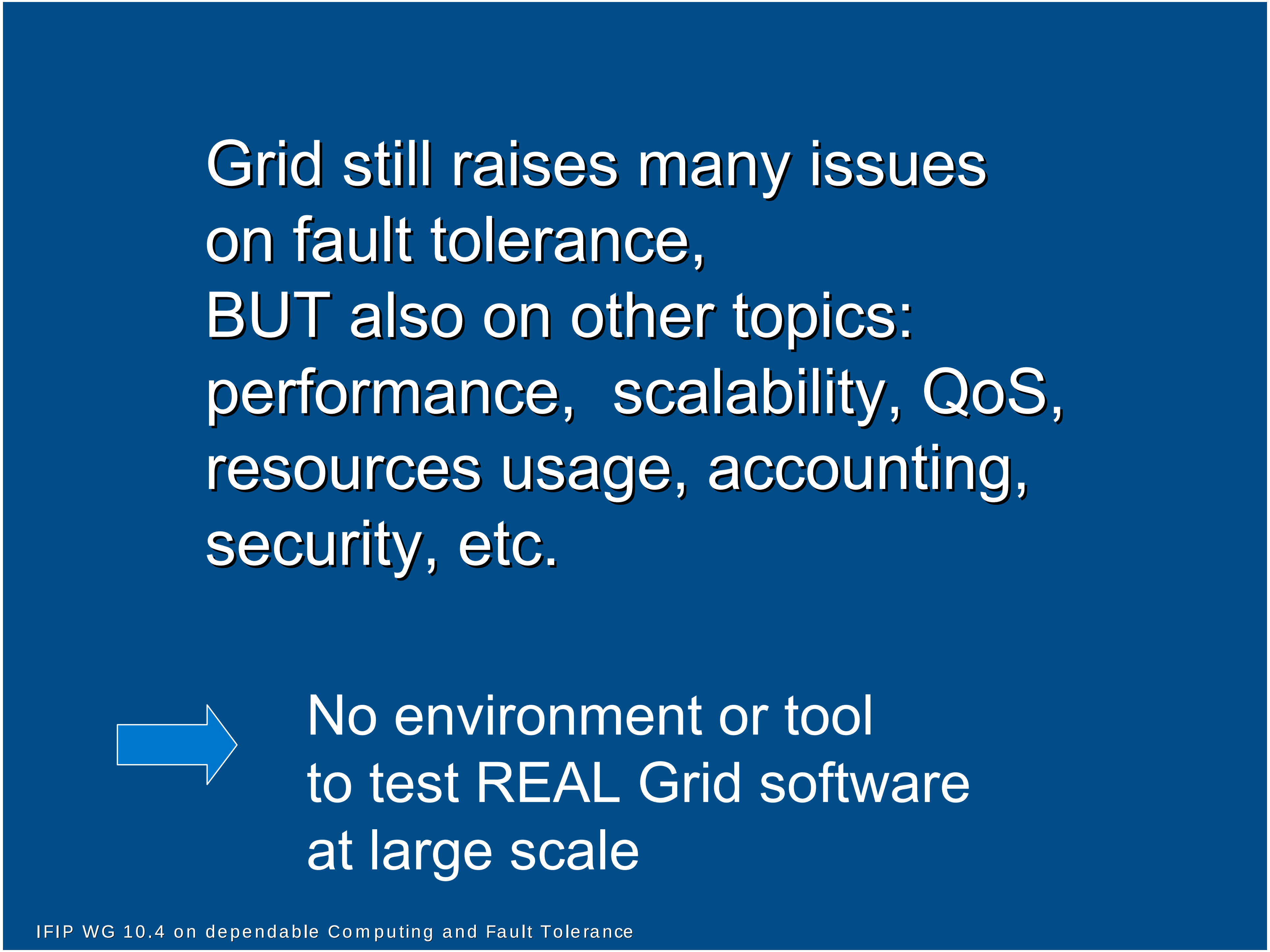
Infrastructure:

- Server fault tolerance (GridServices, Webservices, WSRF)
- Fault detectors (few results, Xavier'talk)
- High performance protocols (content distribution: BitTorrent)
- Resource discovery (DHT: Kadelmia)

FT techniques:

- Self stabilization (crash may append during stabilization)
- Consensus (impossibility result on asynchronous network)
- Majority voting (decisions may apply to a majority of nodes absent during the vote...)

Fault tolerance is one research topic of the CoreGrid NoE



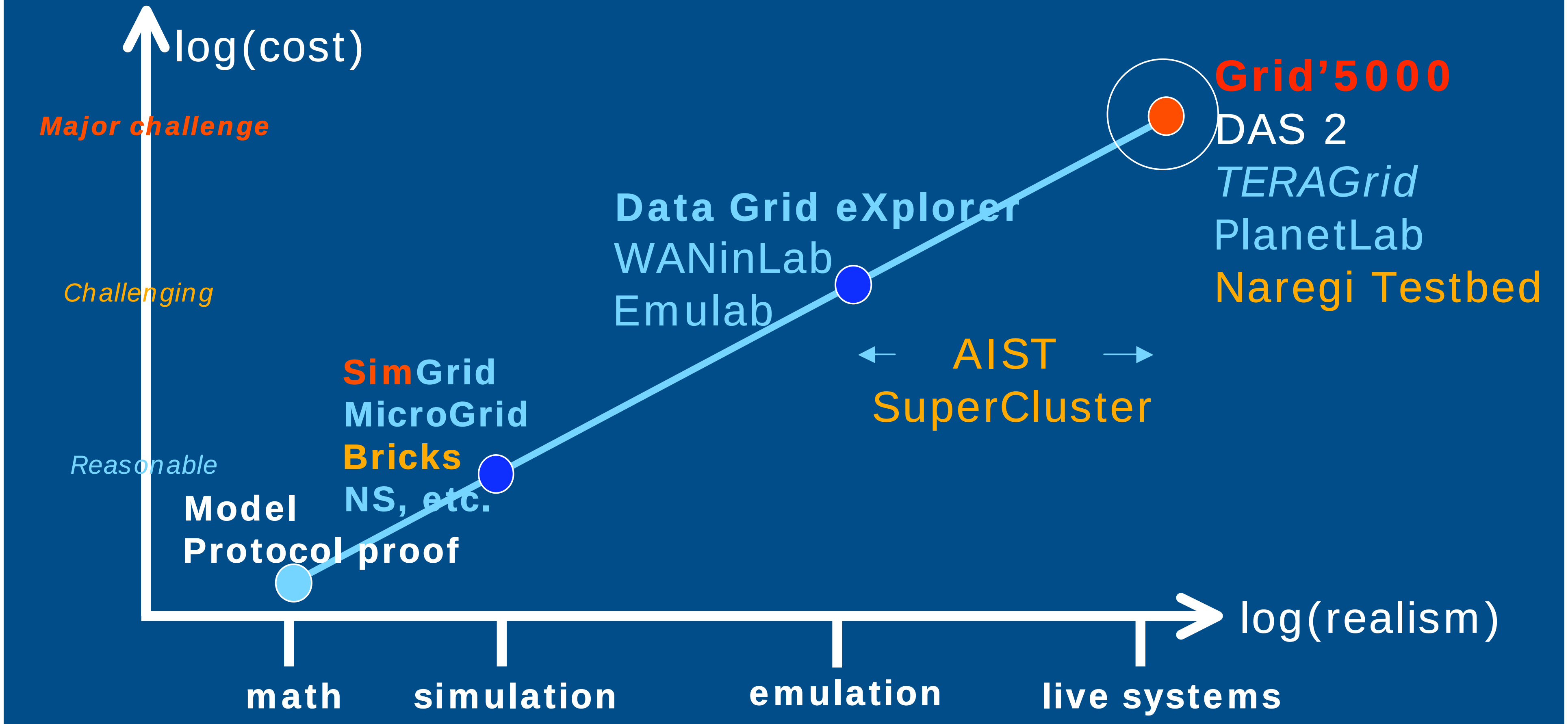
Grid still raises many issues
on fault tolerance,
BUT also on other topics:
performance, scalability, QoS,
resources usage, accounting,
security, etc.

⇨ No environment or tool
to test REAL Grid software
at large scale

We need Grid experimental tools

In the first ½ of 2003, the design and development of two Grid experimental platforms was decided:

→ Grid'5000 as a real life system



log(cost)

Major challenge

Challenging

Reasonable

Model
Protocol proof

SimGrid
MicroGrid
Bricks
NS, etc.

Data Grid eXplorer
WANinLab
Emulab

← AIST →
SuperCluster

Grid'5000
DAS 2
TERAGrid
PlanetLab
Naregi Testbed

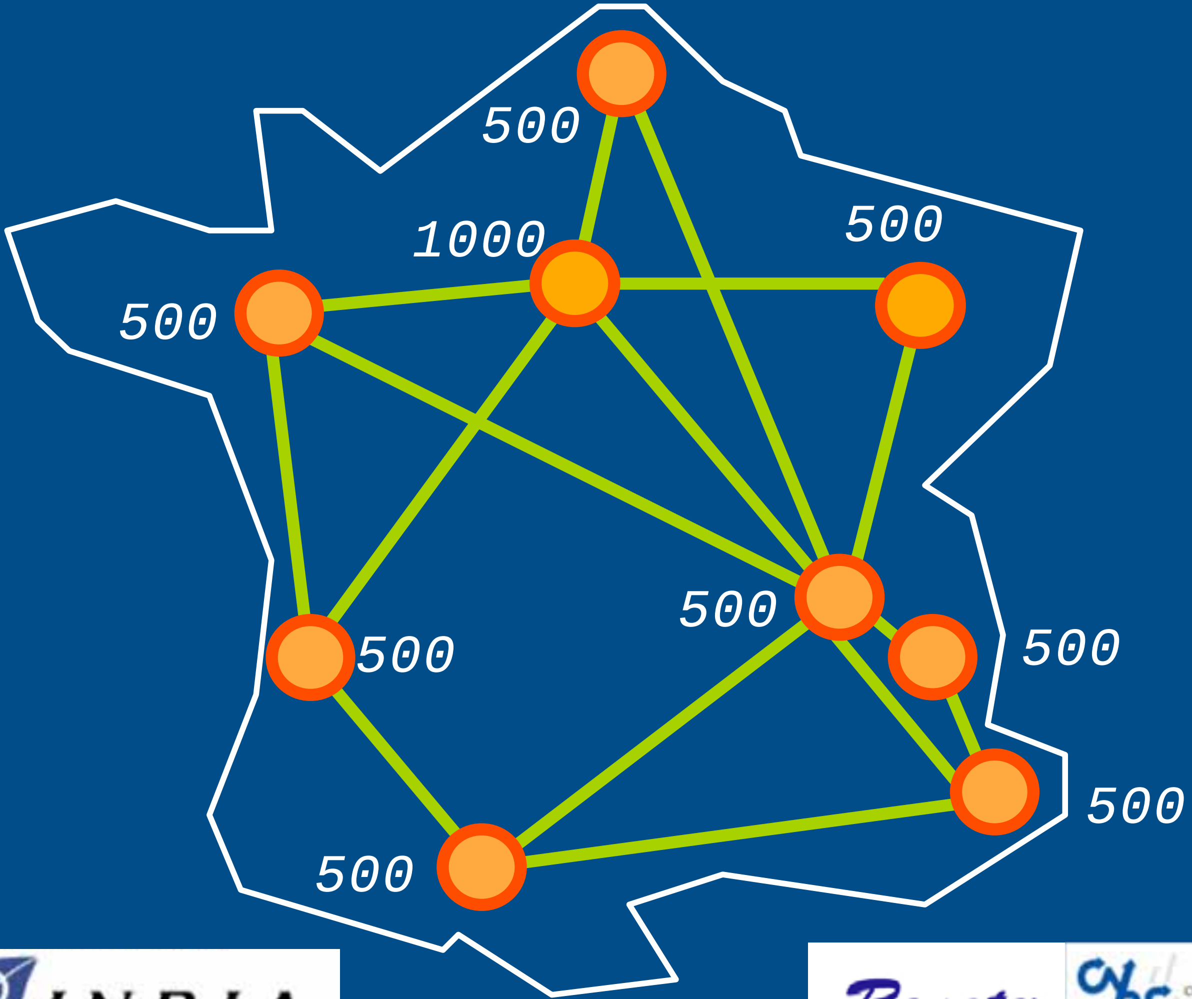
log(realism)

math simulation emulation live systems

IFIP WG 10.4 on dependable Computing and Fault Tolerance

Grid'5000

The largest research Instrument to study Grid issues



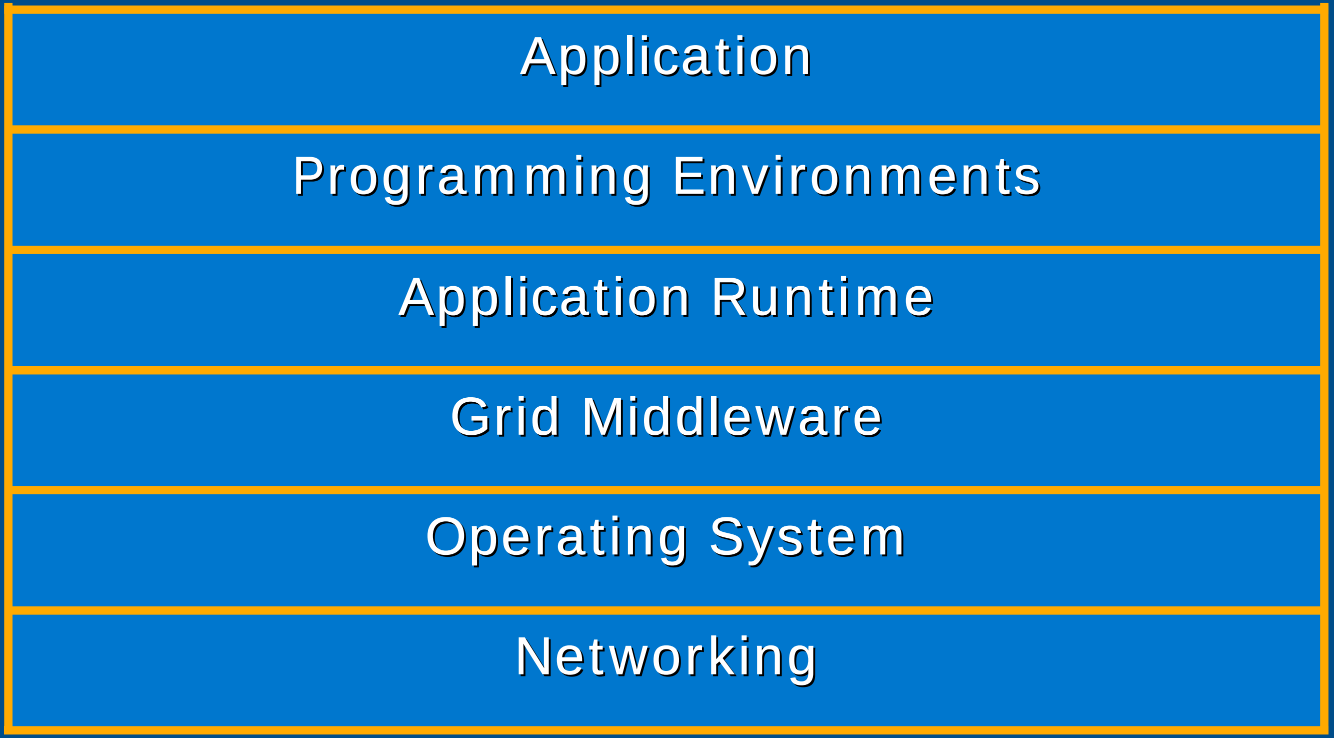
━ RENATER

Grid’5000 foundations: Collection of experiments to be done

- Networking
  - End host communication layer (interference with local communications)
  - High performance long distance protocols (improved TCP)
  - High Speed Network Emulation
- Middleware / OS
  - Scheduling / data distribution in Grid
  - Fault tolerance in Grid
  - Resource management
  - Grid SSI OS and Grid I/O
  - Desktop Grid/P2P systems
- Programming
  - Component programming for the Grid (Java, Corba)
  - GRID-RPC
  - GRID-MPI
  - Code Coupling
- Applications
  - Multi-parametric applications (Climate modeling/Functional Genomic)
  - Large scale experimentation of distributed applications (Electromagnetism, multi-material fluid mechanics, parallel optimization algorithms, CFD, astrophysics
  - Medical images, Collaborating tools in virtual 3D environment

Grid'5000 goal:

Experimenting fault tolerance and many other topics on all layers of the Grid software stack



Application

Programming Environments

Application Runtime

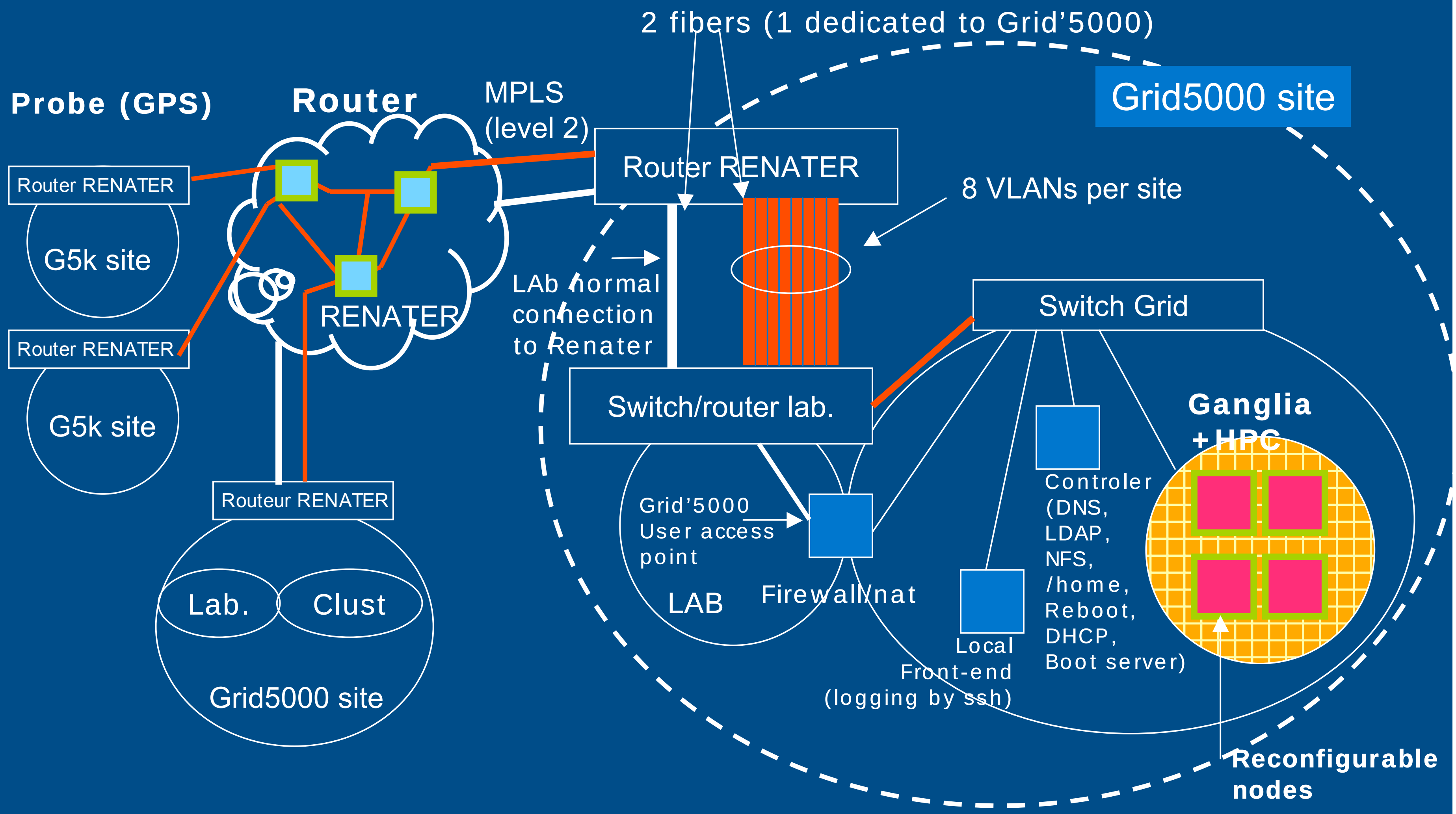
Grid Middleware

Operating System

Networking

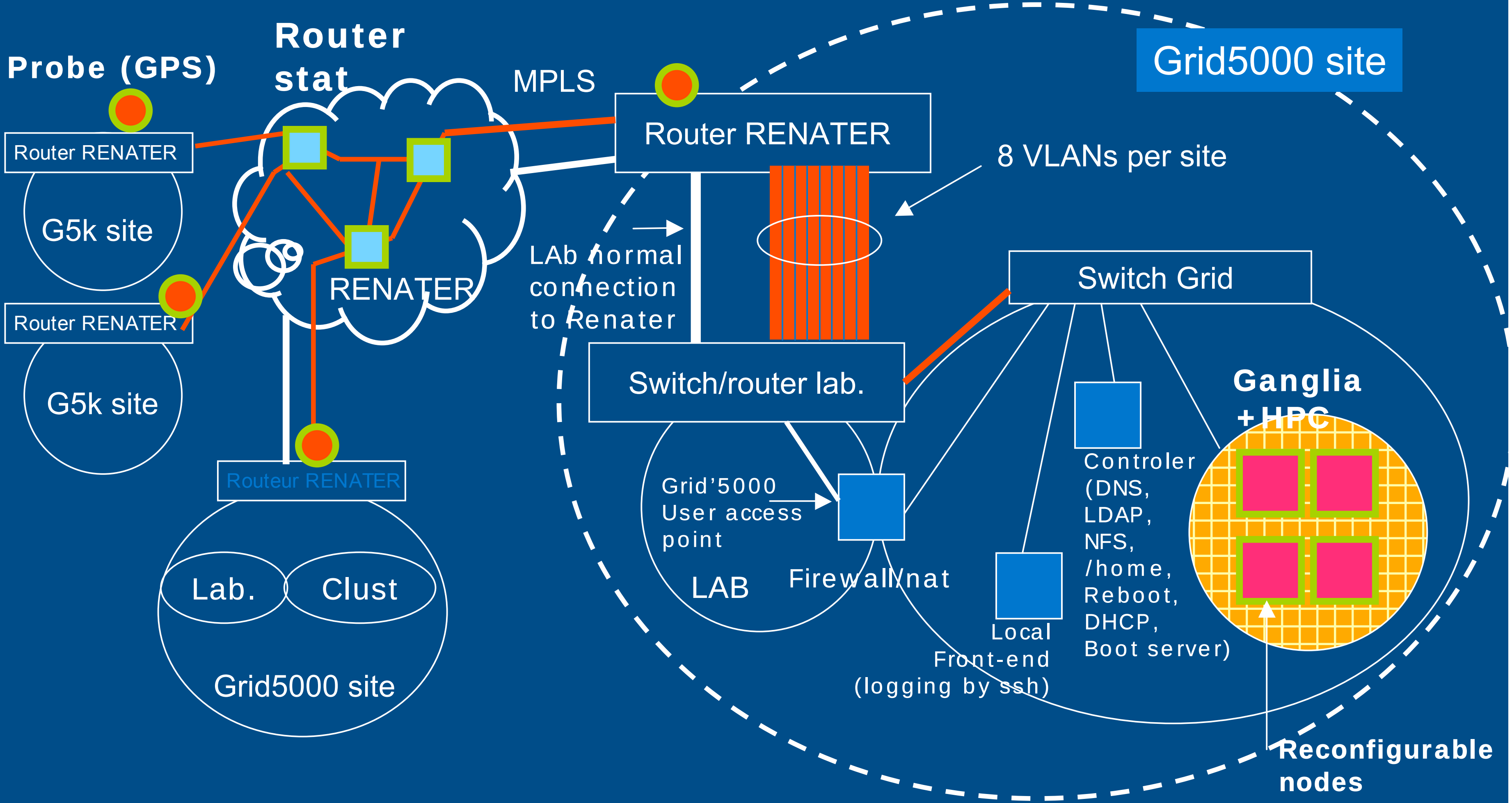
→ A highly reconfigurable, controllable and monitorable experimental platform

Confinement / isolation



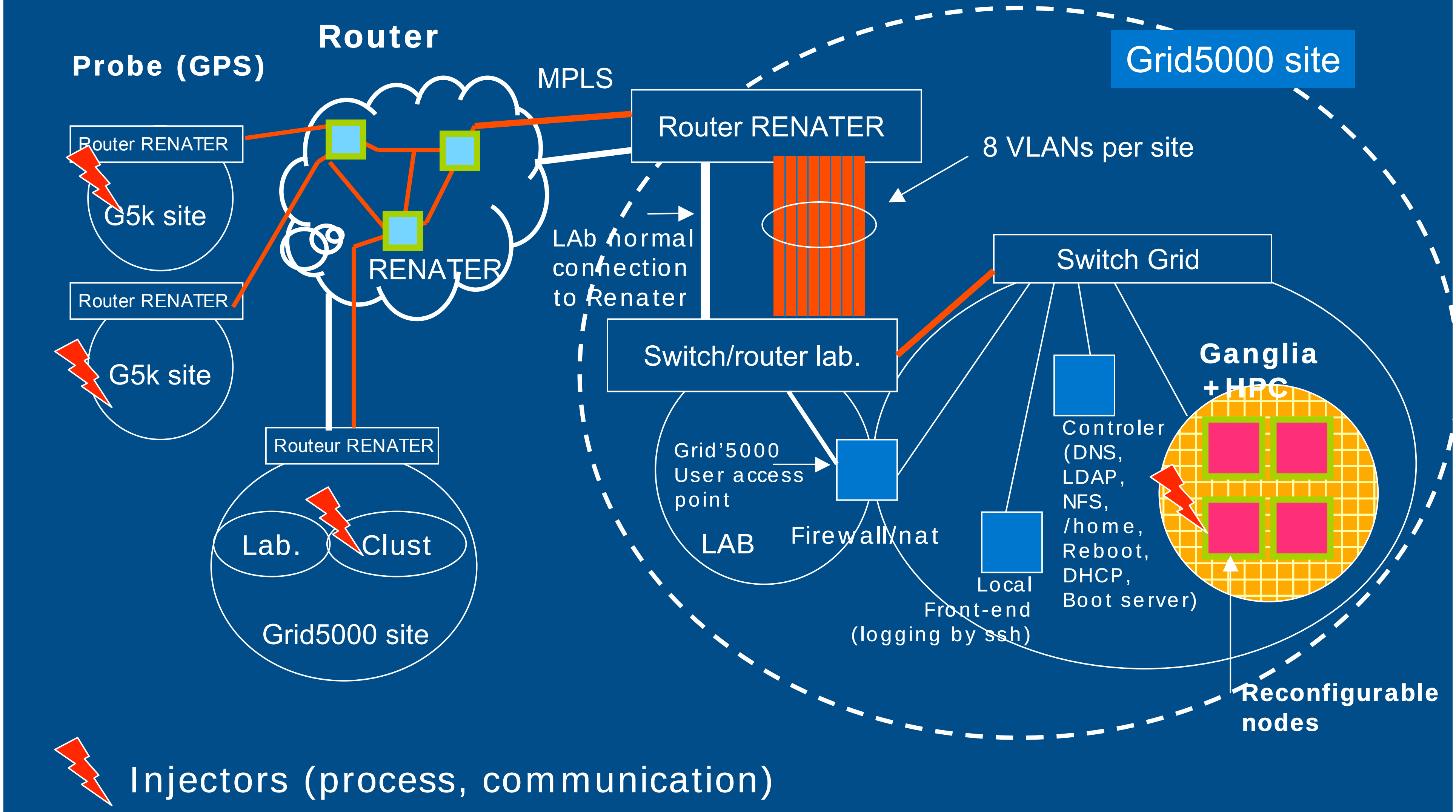
IFIP WG 10.4 on dependable Computing and Fault Tolerance

Observation & Monitoring



IFIP WG 10.4 on dependable Computing and Fault Tolerance

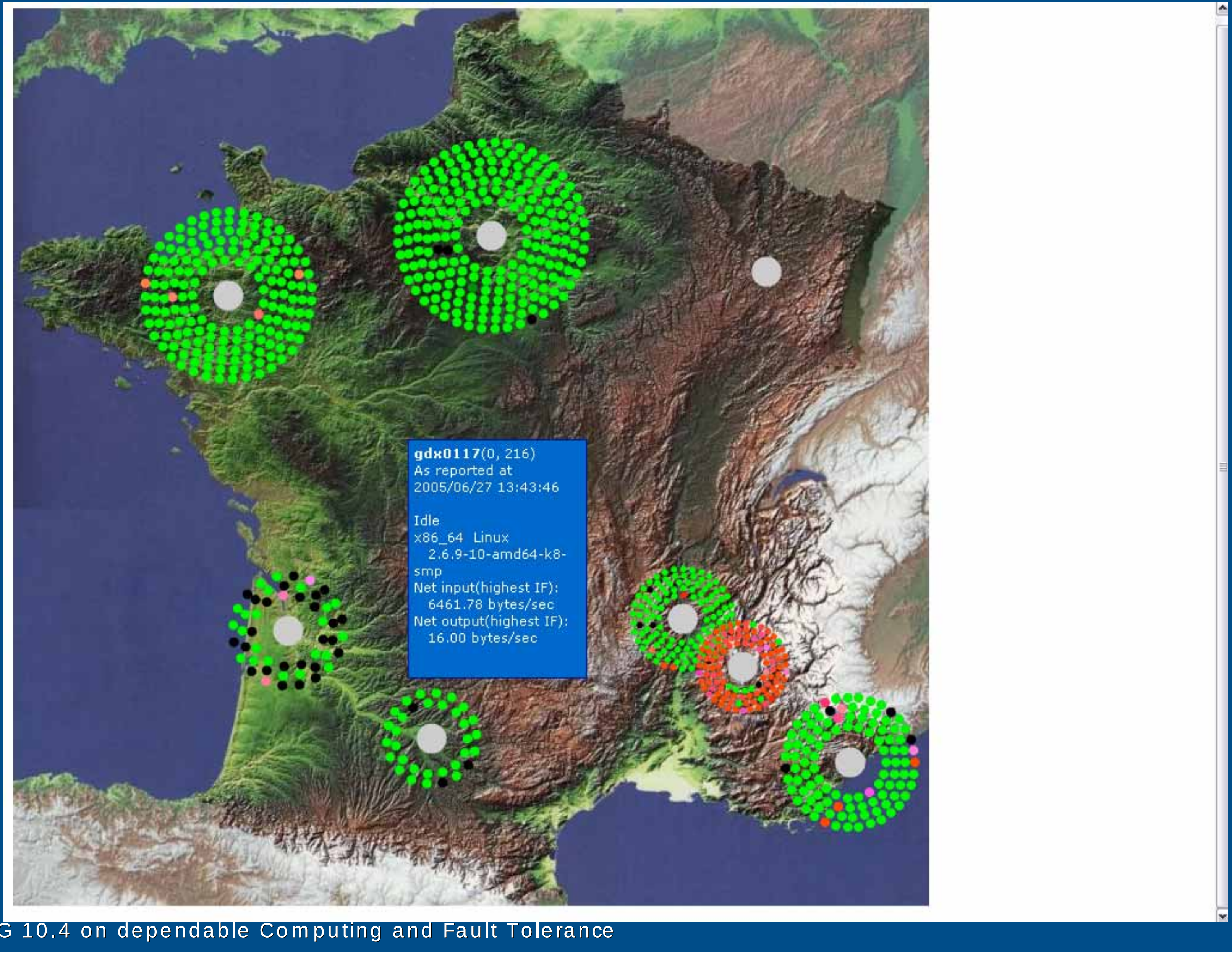
Workload/Traffic & Fault injection



Injectors (process, communication)

Rennes
Lyon
Sophia
Grenoble
Bordeaux
Toulouse
Orsay

Grid'5000 Global Observer



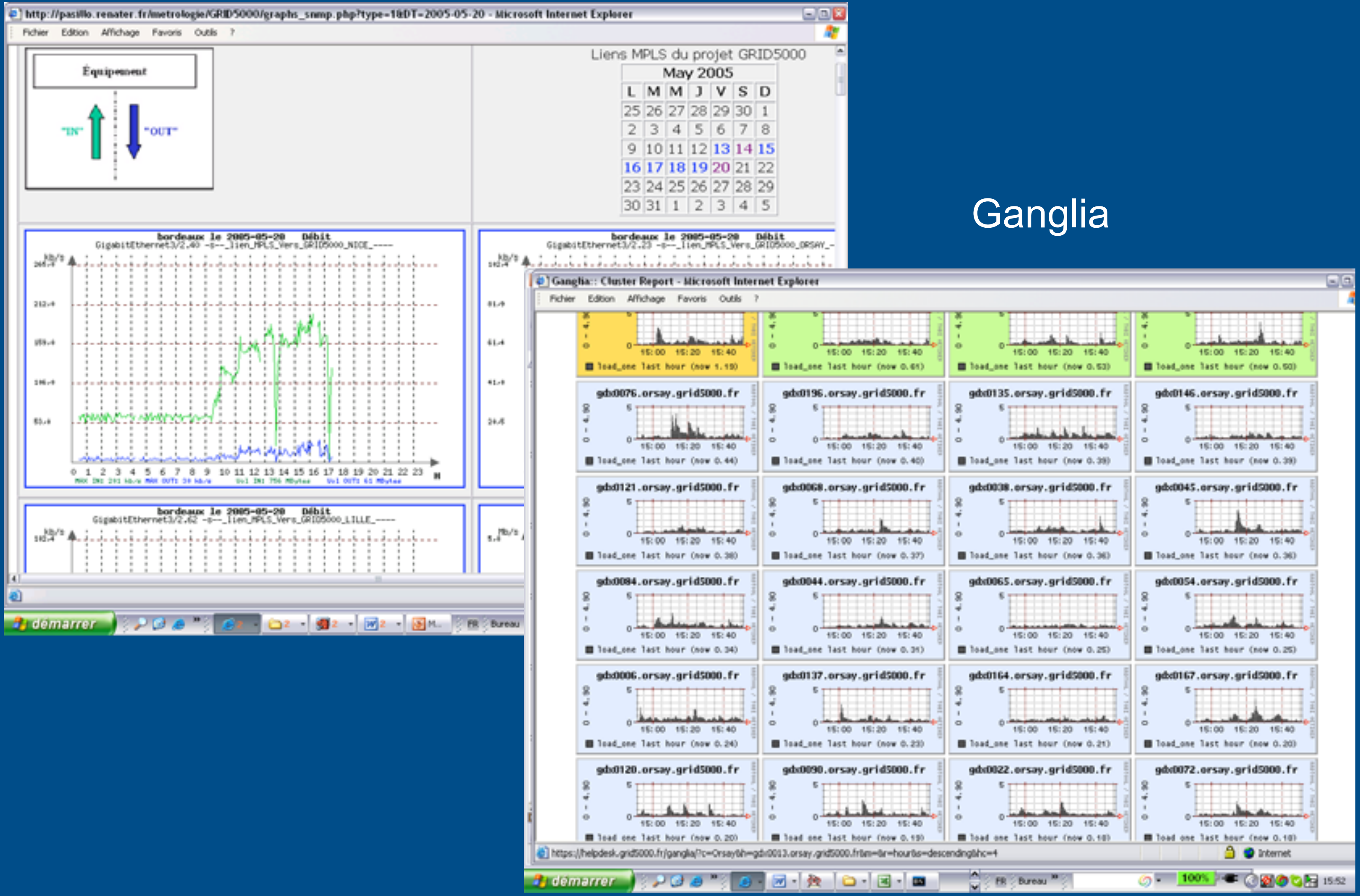
gdx0117(0, 216)
As reported at
2005/06/27 13:43:46

Idle
x86_64 Linux
  2.6.9-10-amd64-k8-
smp
Net input(highest IF):
  6461.78 bytes/sec
Net output(highest IF):
  16.00 bytes/sec

Grid'5000 Monitoring tools

Network traffic

Ganglia



http://pasillo.renater.fr/metrologie/GRID5000/graphs_snmp.php?type=1&DT=2005-05-20 - Microsoft Internet Explorer

Liens MPLS du projet GRID5000

May 2005						
L	M	M	J	V	S	D
25	26	27	28	29	30	1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31	1	2	3	4	5

bordeaux le 2005-05-20 Débit GigabitEthernet3/2.40 -=-_lien_MPLS_Vers_GRID5000_NICE_----

bordeaux le 2005-05-20 Débit GigabitEthernet3/2.23 -=-_lien_MPLS_Vers_GRID5000_ORSAY_-

bordeaux le 2005-05-20 Débit GigabitEthernet3/2.62 -=-_lien_MPLS_Vers_GRID5000_LILLE_----

Ganglia:: Cluster Report - Microsoft Internet Explorer

gdx0076.orsay.grid5000.fr load_one last hour (now 0.44)

gdx0196.orsay.grid5000.fr load_one last hour (now 0.40)

gdx0135.orsay.grid5000.fr load_one last hour (now 0.39)

gdx0146.orsay.grid5000.fr load_one last hour (now 0.39)

gdx0121.orsay.grid5000.fr load_one last hour (now 0.38)

gdx0068.orsay.grid5000.fr load_one last hour (now 0.37)

gdx0038.orsay.grid5000.fr load_one last hour (now 0.36)

gdx0045.orsay.grid5000.fr load_one last hour (now 0.36)

gdx0084.orsay.grid5000.fr load_one last hour (now 0.34)

gdx0044.orsay.grid5000.fr load_one last hour (now 0.31)

gdx0065.orsay.grid5000.fr load_one last hour (now 0.25)

gdx0054.orsay.grid5000.fr load_one last hour (now 0.25)

gdx0006.orsay.grid5000.fr load_one last hour (now 0.24)

gdx0137.orsay.grid5000.fr load_one last hour (now 0.23)

gdx0164.orsay.grid5000.fr load_one last hour (now 0.21)

gdx0167.orsay.grid5000.fr load_one last hour (now 0.20)

gdx0120.orsay.grid5000.fr load_one last hour (now 0.20)

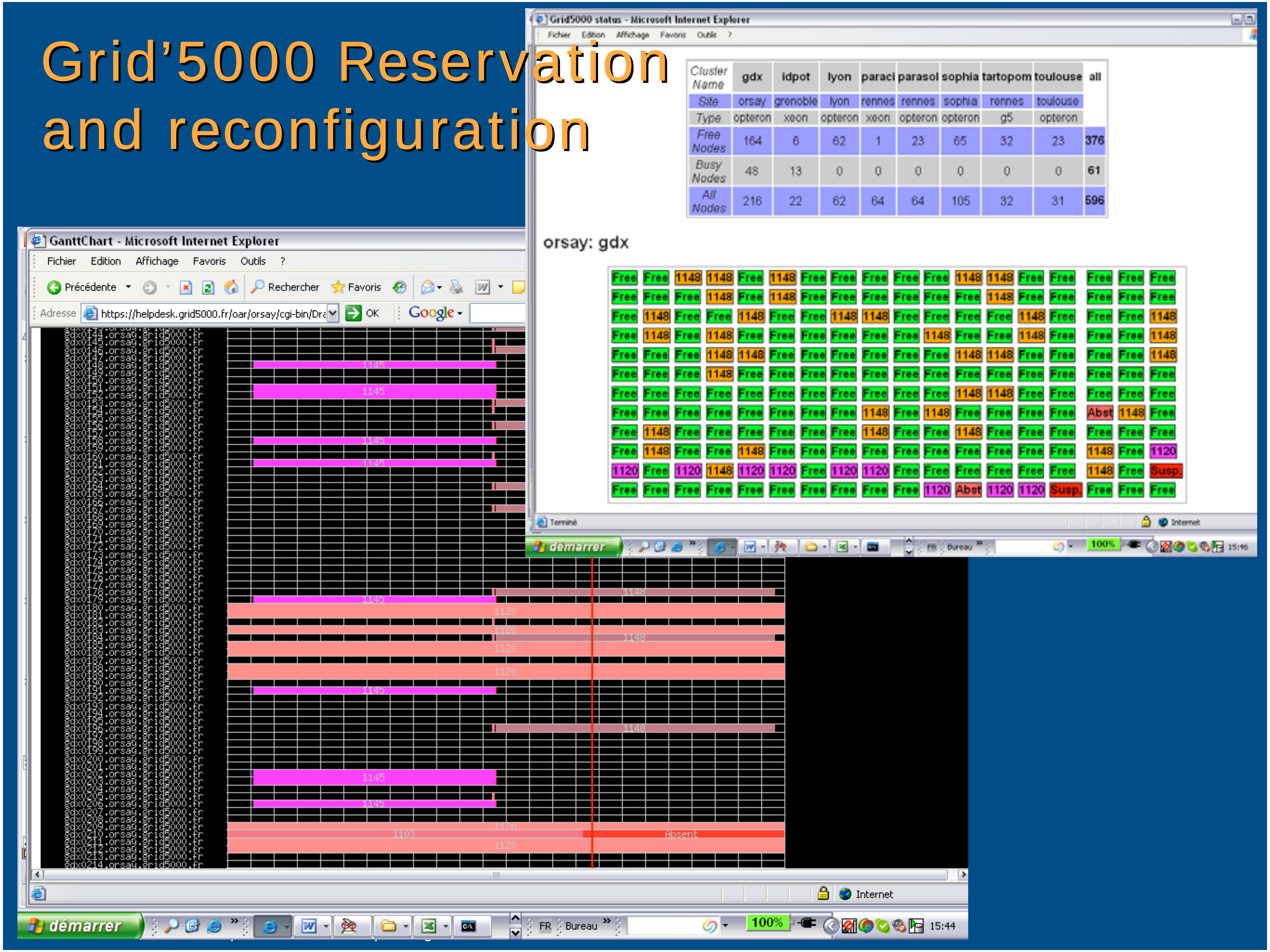
gdx0090.orsay.grid5000.fr load_one last hour (now 0.19)

gdx0022.orsay.grid5000.fr load_one last hour (now 0.18)

gdx0072.orsay.grid5000.fr load_one last hour (now 0.18)

https://helpdesk.grid5000.fr/ganglia/?c=Orsay&h=gdx0013.orsay.grid5000.fr&m=&r=hour&s=descending&hc=4

Grid’5000 Reservation and reconfiguration


Grid5000 status - Microsoft Internet Explorer
Fichier Edition Affichage Favoris Outils ?
Terminé
Internet
démarrer
FR Bureau
100%
15:46
GanttChart - Microsoft Internet Explorer
Fichier Edition Affichage Favoris Outils ?
Précédente
Rechercher
Favoris
Adresse
https://helpdesk.grid5000.fr/oar/orsay/cgi-bin/Dra
OK
Google
Internet
démarrer
FR Bureau
100%
15:44
1145
1148
1120
1103
Absent

Cluster Name	gdx	idpot	lyon	paraci	parasol	sophia	tartopom	toulouse	all
Site	orsay	grenoble	lyon	rennes	rennes	sophia	rennes	toulouse	
Type	opteron	xeon	opteron	xeon	opteron	opteron	g5	opteron	
Free Nodes	164	6	62	1	23	65	32	23	376
Busy Nodes	48	13	0	0	0	0	0	0	61
All Nodes	216	22	62	64	64	105	32	31	596

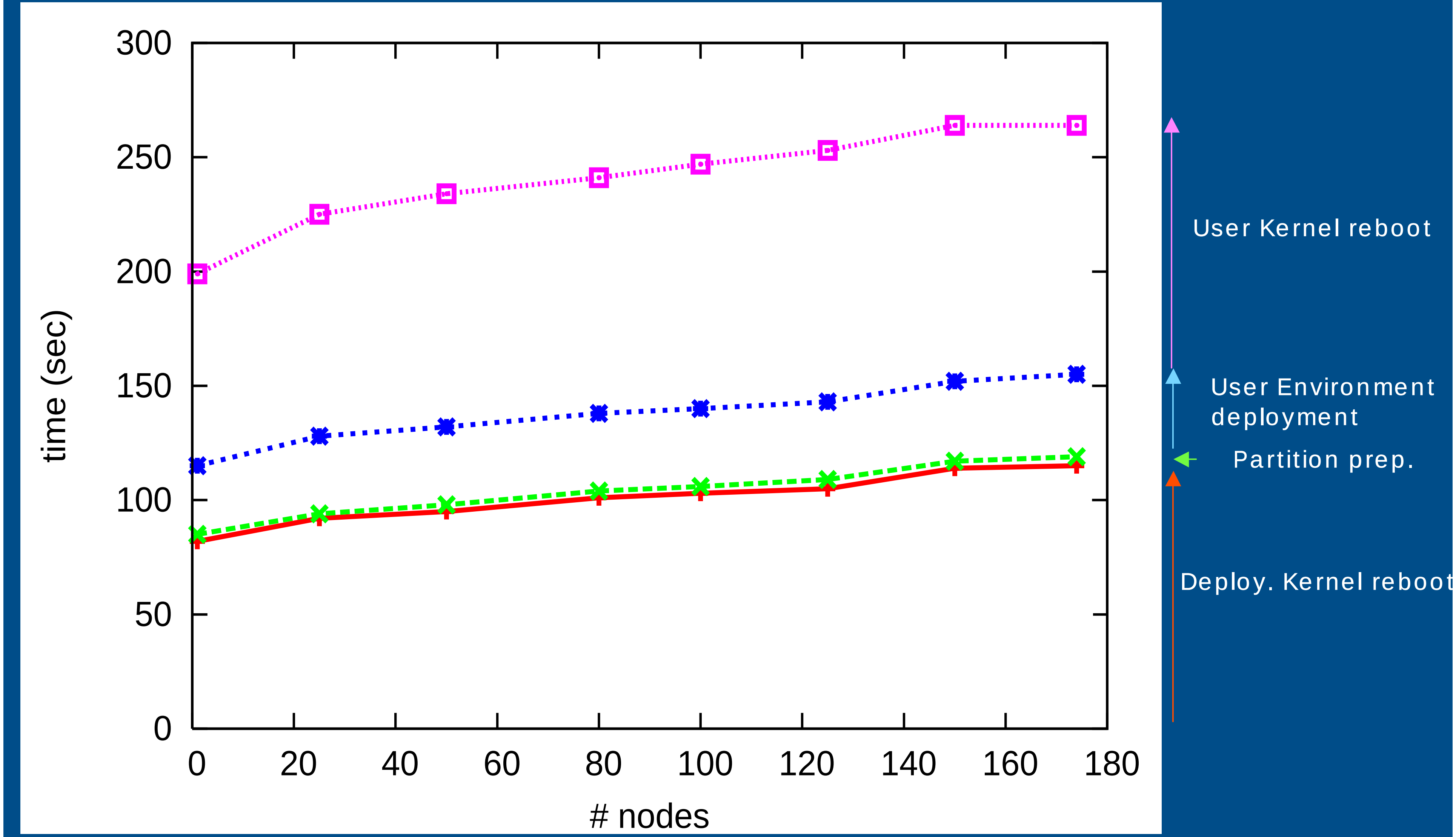
orsay: gdx

Free	Free	1148	1148	Free	1148	Free	Free	Free	Free	Free	1148	1148	Free	Free	Free	Free	Free
Free	Free	Free	1148	Free	1148	Free	Free	Free	Free	Free	Free	1148	Free	Free	Free	Free	Free
Free	1148	Free	Free	1148	Free	Free	1148	1148	Free	Free	Free	Free	1148	Free	Free	Free	1148
Free	1148	Free	1148	Free	Free	Free	Free	Free	Free	1148	Free	Free	1148	Free	Free	Free	1148
Free	Free	Free	1148	1148	Free	Free	Free	Free	Free	Free	1148	1148	Free	Free	Free	Free	1148
Free	Free	Free	1148	Free	Free	Free	Free	Free	Free	Free	Free	Free	Free	Free	Free	Free	Free
Free	Free	Free	Free	Free	Free	Free	Free	Free	Free	Free	1148	1148	Free	Free	Free	Free	Free
Free	Free	Free	Free	Free	Free	Free	Free	1148	Free	1148	Free	Free	Free	Free	Abst	1148	Free
Free	1148	Free	Free	Free	Free	Free	Free	1148	Free	Free	1148	Free	Free	Free	Free	Free	Free
Free	1148	Free	Free	1148	Free	Free	Free	Free	Free	Free	Free	Free	Free	Free	1148	Free	1120
1120	Free	1120	1148	1120	1120	Free	1120	1120	Free	Free	Free	Free	Free	Free	1148	Free	Susp.
Free	Free	Free	Free	Free	Free	Free	Free	Free	Free	1120	Abst	1120	1120	Susp.	Free	Free	Free

```
gdx0144.orsay.grid5000.fr
gdx0145.orsay.grid5000.fr
gdx0146.orsay.grid5000.fr
gdx0147.orsay.grid5000.fr
gdx0148.orsay.grid5000.fr
gdx0149.orsay.grid5000.fr
gdx0150.orsay.grid5000.fr
gdx0151.orsay.grid5000.fr
gdx0152.orsay.grid5000.fr
gdx0153.orsay.grid5000.fr
gdx0154.orsay.grid5000.fr
gdx0155.orsay.grid5000.fr
gdx0156.orsay.grid5000.fr
gdx0157.orsay.grid5000.fr
gdx0158.orsay.grid5000.fr
gdx0159.orsay.grid5000.fr
gdx0160.orsay.grid5000.fr
gdx0161.orsay.grid5000.fr
gdx0162.orsay.grid5000.fr
gdx0163.orsay.grid5000.fr
gdx0164.orsay.grid5000.fr
gdx0165.orsay.grid5000.fr
gdx0166.orsay.grid5000.fr
gdx0167.orsay.grid5000.fr
gdx0168.orsay.grid5000.fr
gdx0169.orsay.grid5000.fr
gdx0170.orsay.grid5000.fr
gdx0171.orsay.grid5000.fr
gdx0172.orsay.grid5000.fr
gdx0173.orsay.grid5000.fr
gdx0174.orsay.grid5000.fr
gdx0175.orsay.grid5000.fr
gdx0176.orsay.grid5000.fr
gdx0177.orsay.grid5000.fr
gdx0178.orsay.grid5000.fr
gdx0179.orsay.grid5000.fr
gdx0180.orsay.grid5000.fr
gdx0181.orsay.grid5000.fr
gdx0182.orsay.grid5000.fr
gdx0183.orsay.grid5000.fr
gdx0184.orsay.grid5000.fr
gdx0185.orsay.grid5000.fr
gdx0186.orsay.grid5000.fr
gdx0187.orsay.grid5000.fr
gdx0188.orsay.grid5000.fr
gdx0189.orsay.grid5000.fr
gdx0190.orsay.grid5000.fr
gdx0191.orsay.grid5000.fr
gdx0192.orsay.grid5000.fr
gdx0193.orsay.grid5000.fr
gdx0194.orsay.grid5000.fr
gdx0195.orsay.grid5000.fr
gdx0196.orsay.grid5000.fr
gdx0197.orsay.grid5000.fr
gdx0198.orsay.grid5000.fr
gdx0199.orsay.grid5000.fr
gdx0200.orsay.grid5000.fr
gdx0201.orsay.grid5000.fr
gdx0202.orsay.grid5000.fr
gdx0203.orsay.grid5000.fr
gdx0204.orsay.grid5000.fr
gdx0205.orsay.grid5000.fr
gdx0206.orsay.grid5000.fr
gdx0207.orsay.grid5000.fr
gdx0208.orsay.grid5000.fr
gdx0209.orsay.grid5000.fr
gdx0210.orsay.grid5000.fr
gdx0211.orsay.grid5000.fr
gdx0212.orsay.grid5000.fr
gdx0213.orsay.grid5000.fr
gdx0214.orsay.grid5000.fr
```

Grid’5000 Reconfiguration time

Time to reboot 1 cluster of Grid’5000 with Kadeploy



User Kernel reboot

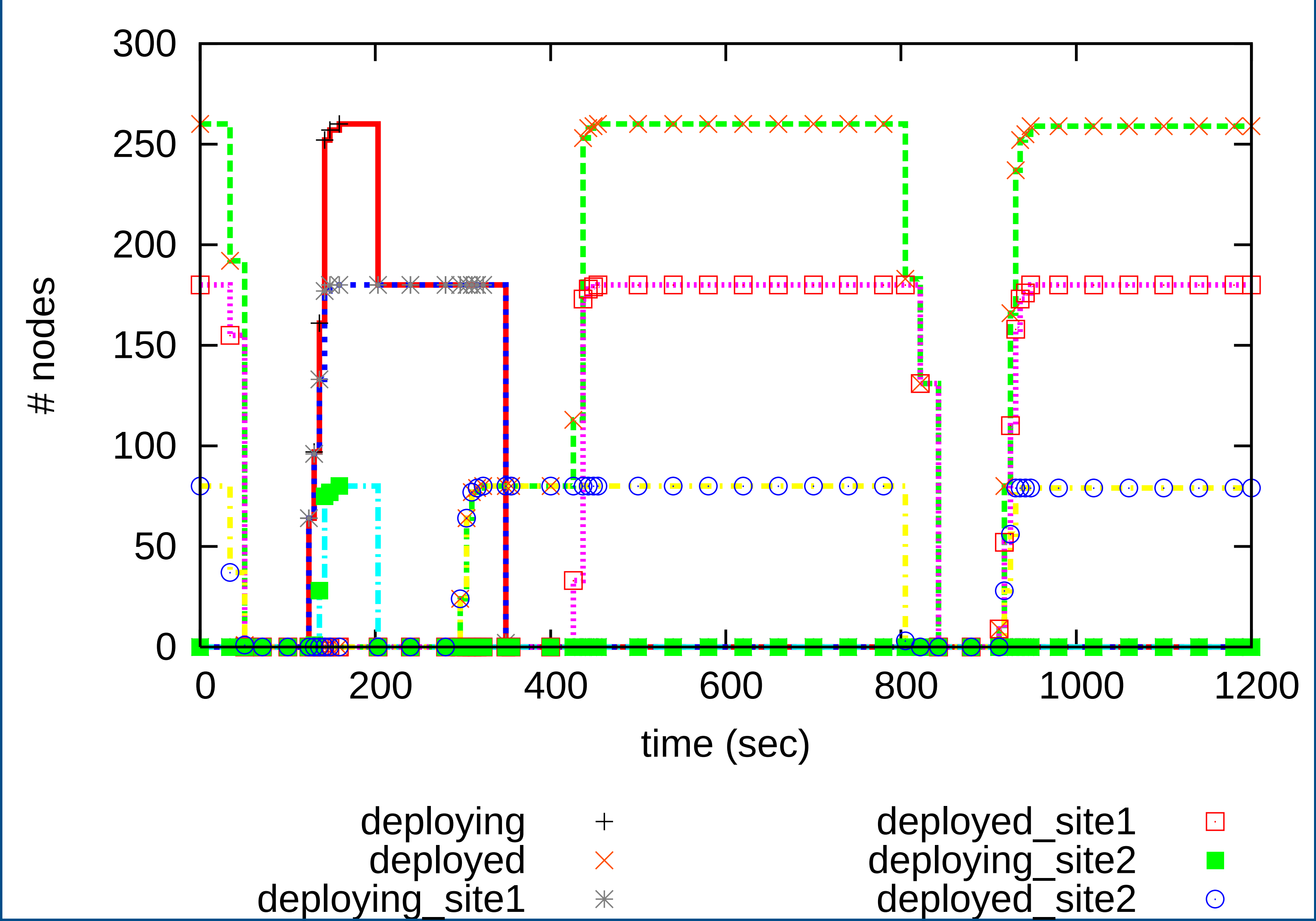
User Environment deployment

Partition prep.

Deploy. Kernel reboot

Grid'5000 Reconfiguration time

Time to reboot 2 clusters (Paris + Nice) of Grid'5000 (Kadeploy)



IFIP WG 10.4 on dependable Computing and Fault Tolerance

Grid’5000 Fault Generator: Fail

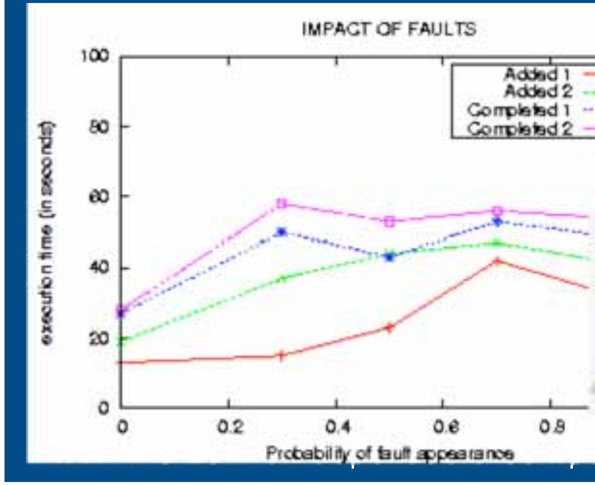
Objectives

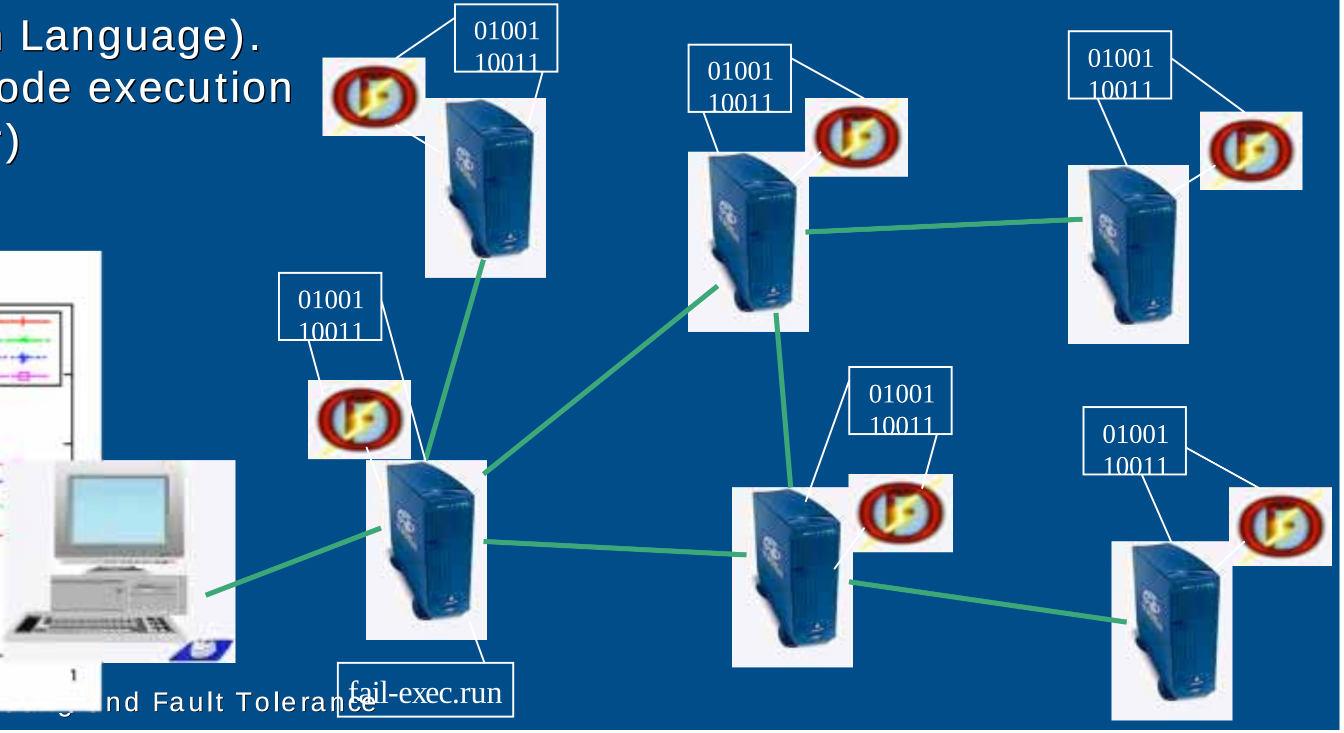
- •Probabilistic and deterministic (reproducible) fault injection.
- •Expressiveness of scenarios.
- •No code modification.
- •Scalable.

Concepts

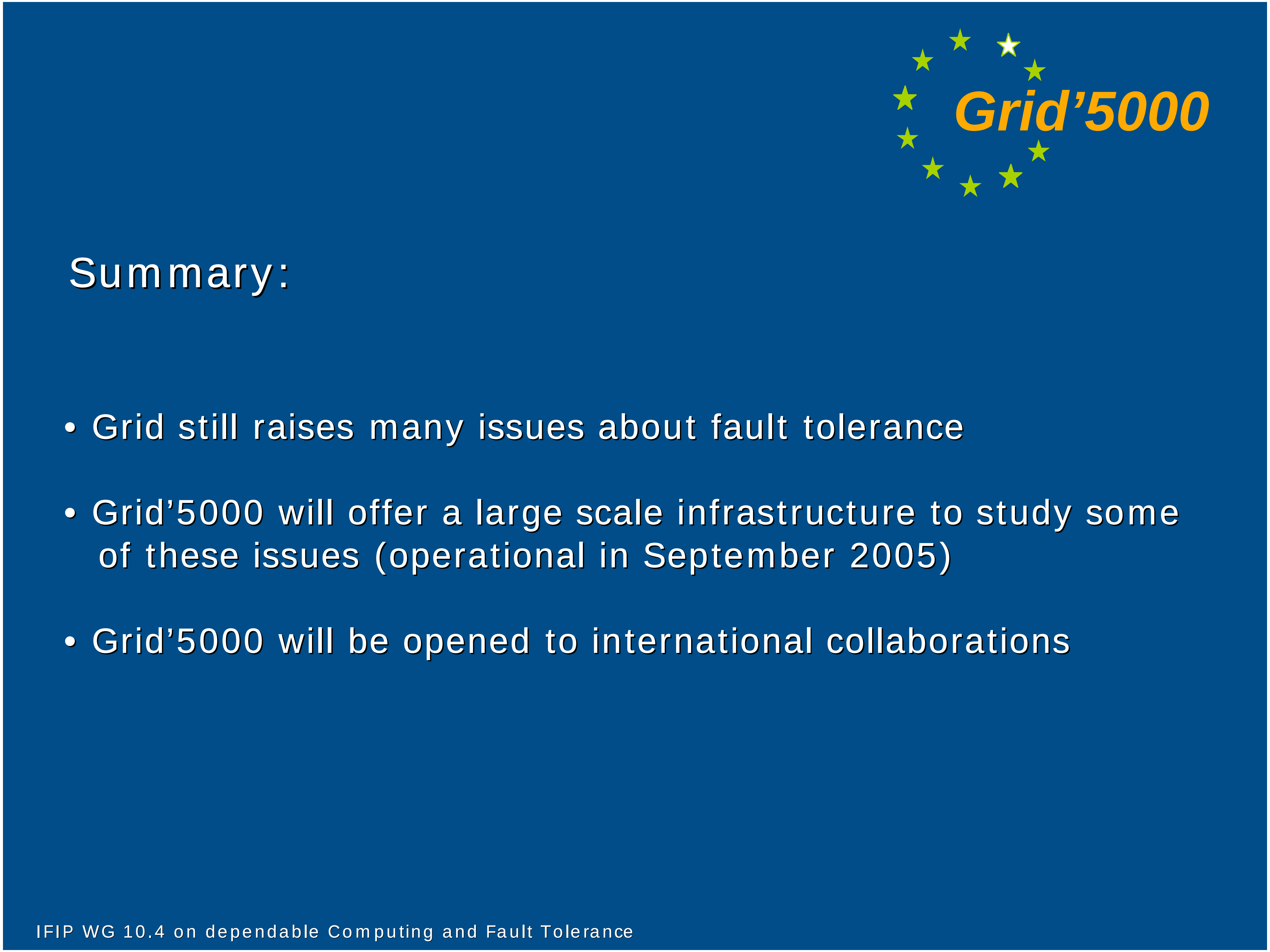
- •A dedicated language for fault scenario specification (FAIL: FAult Injection Language).
- •Fine control of the code execution (through a debugger)

```
Daemon ADV2
  {
  time_g timer = 5;
  node 1 :
    always int rand = FAIL_RANDOM (1,10);
    timer && rand < 2 -> halt, goto 2;
  node 2 :
    always int rand = FAIL_RANDOM (1,10);
    timer && rand > 7 -> restart, goto 3;
  node 3 :
  }
```





nd Fault Tolerance



Grid’5000

Summary:

- Grid still raises many issues about fault tolerance
- Grid’5000 will offer a large scale infrastructure to study some of these issues (operational in September 2005)
- Grid’5000 will be opened to international collaborations

Session 1.4

Security

Moderator and Rapporteur

Paulo J.E. Veríssimo, University of Lisbon, Portugal

Grid Security
: Authentication and Authorization

IFIP Workshop – 2/7/05

Jong Kim

Dept. of Computer Sci. and Eng.

Pohang Univ. of Sci. and Tech. (POSTECH)

Contents

- Grid Security
  - Grid Security Challenges
  - Grid Security Requirements
- Current Status of Grid Security
  - Authentication and Delegation
  - Authorization
  - Grid Security Infrastructure (GSI)
  - OGSA
  - Web Services Security
- Things need more study
  - Authentication Interoperability
  - Fine-grained Authorization
- Summary

Grid Security

- Grid Computing
  - Distributed computing infrastructure with a plenty of resources which are heterogeneous and scattered geographically
  - A controlled and coordinated resource sharing and resource use in dynamic, scalable, and distributed virtual organizations (VOs)

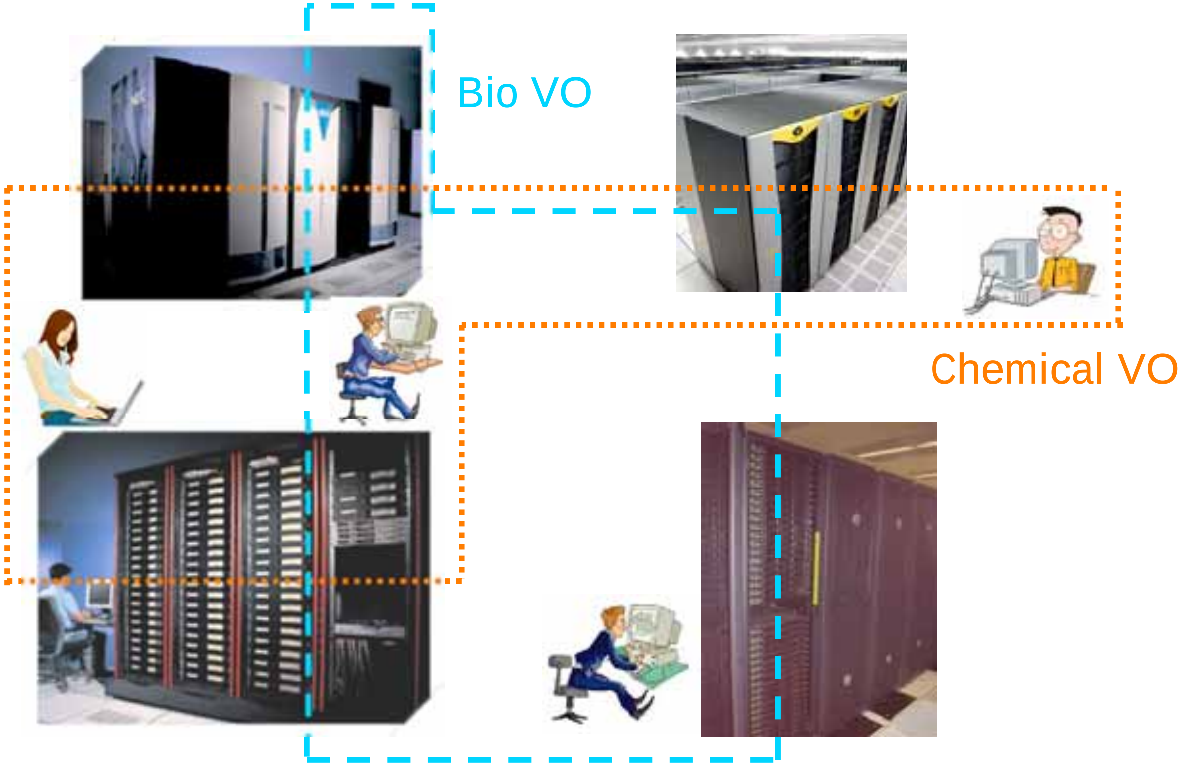
- Security for whom?
  - Resource Providers?
  - Virtual Organization?
  - End-user (participants)?

Grid Security

- What is Grid Security ?
  - Security architecture to enable dynamic, scalable, and distributed VOs protect resources for resource providers, computing entities for VOs, and end-processing for end-users
  - Thru
    - Authentication,
    - Delegation,
    - Authorization,
    - Confidentiality,
    - Privacy, ...

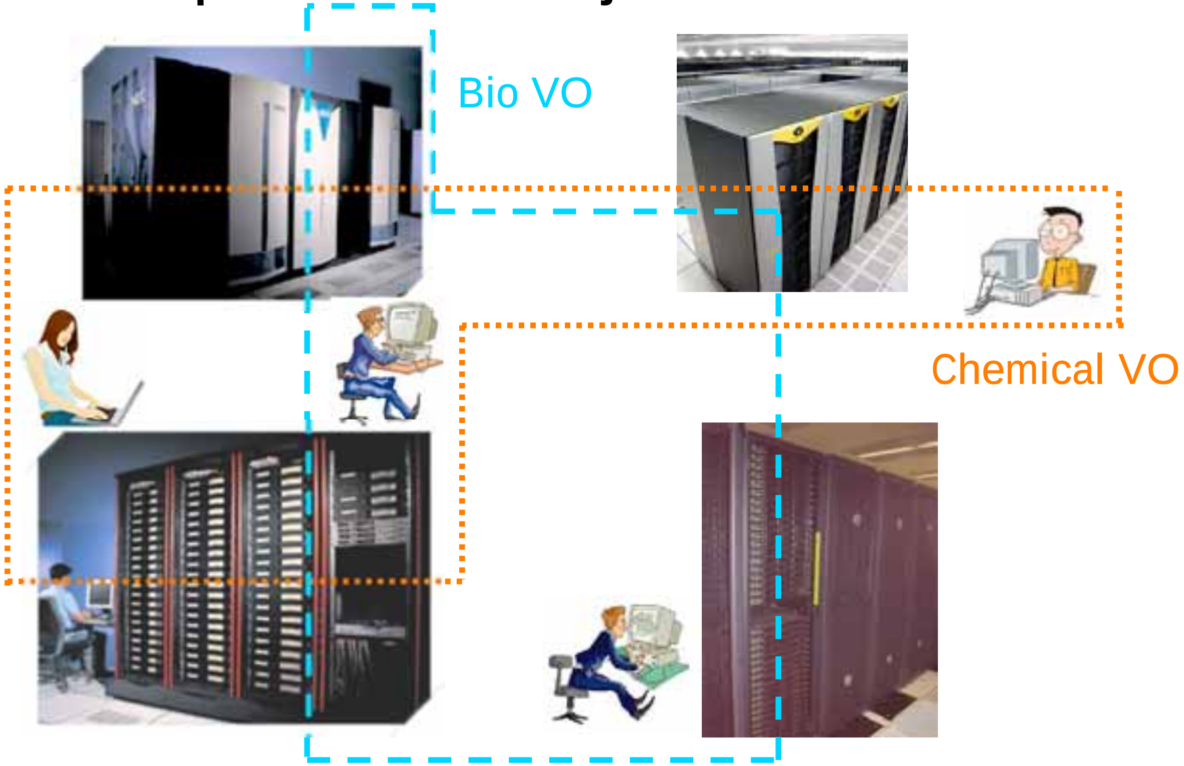
Dynamic VO in the Grid

- Virtual organizations (VOs) are collections of diverse and distributed individuals that seek to share and use diverse resources in a coordinated fashion.
- Users can join into several VOs, while resource providers also partition their resources to several VOs.



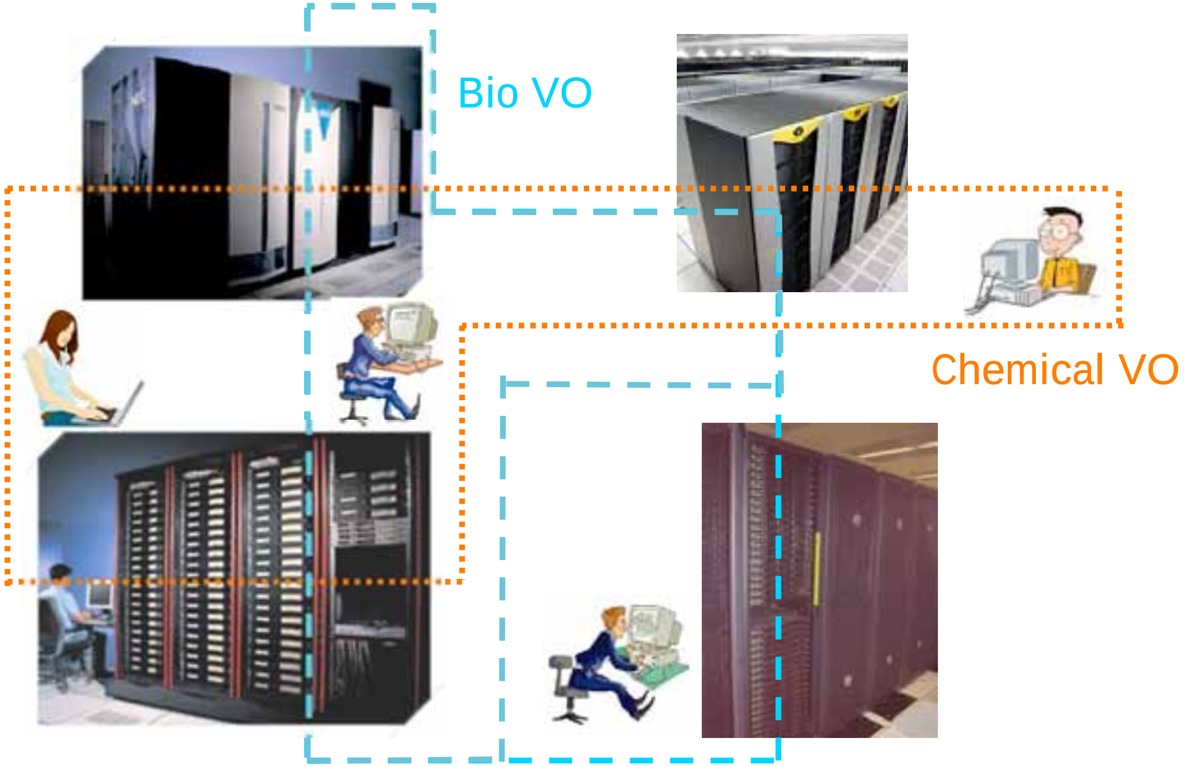
Grid Security Challenges

- Dynamic VO establishment
  - A VO is organized for some goal and disorganized after the goal is achieved.
  - Users can join into or leave VOs.
  - Resource providers can join into or leave VOs.



Grid Security Challenges

- Dynamic policy management
  - Resource providers dynamically change their resources policies.
  - VO managers manage VO users’ rights dynamically.



Grid Security Challenges

- Interoperability with different host environm ents
  - Security services for diverse domains and hosting environments should interact with each other.
  - At the *protocol* level, messages can be exchanged.
  - At the *policy* level, each entity can specify its policy and the policy can be mutually comprehensive.
  - At the *identity* level, a user can be identified from one domain in another domain.

Grid Security Challenges

- Integration with existing systems and technologies
  - It is unrealistic to use a single security technology to address Grid security issues.
  - Existing security infrastructures cannot be replaced.
  - Thus, a Grid security architecture must be
    - Implemental,
    - Extensible, and
    - Integrate

Grid Security Requirements

- Authentication
  - Entities are provided with plug points for multiple authentication mechanisms.
- Delegation
  - Users can delegate their access rights to services.
  - Delegation policies also can be specified.
- Single Logon
  - An entity is allowed to have continuous access rights for some reasonable period with single authentication.

Grid Security Requirements

- Credential Lifespan and Renewal
  - A job initiated by a user may take longer than the life time of the user’s initial credential.
  - In such case, the user needs to be notified prior to expiration of the credential, or be able to refresh it automatically.
- Authorization
  - Resources are used under a certain authorization policies.
  - A service provider can specify its own authorization policy, with which users can invoke those policies.

Grid Security Requirements

- Confidentiality
  - The confidentiality of the communication mechanism and messages or documents is supported.
- Message Integrity
  - It is ensured that unauthorized changes of messages or documents may be detected.
- Privacy
  - A service requester and a service provider enforce privacy policies.
- Other requirements
  - Policy exchange, secure logging, manageability, ...

Contents

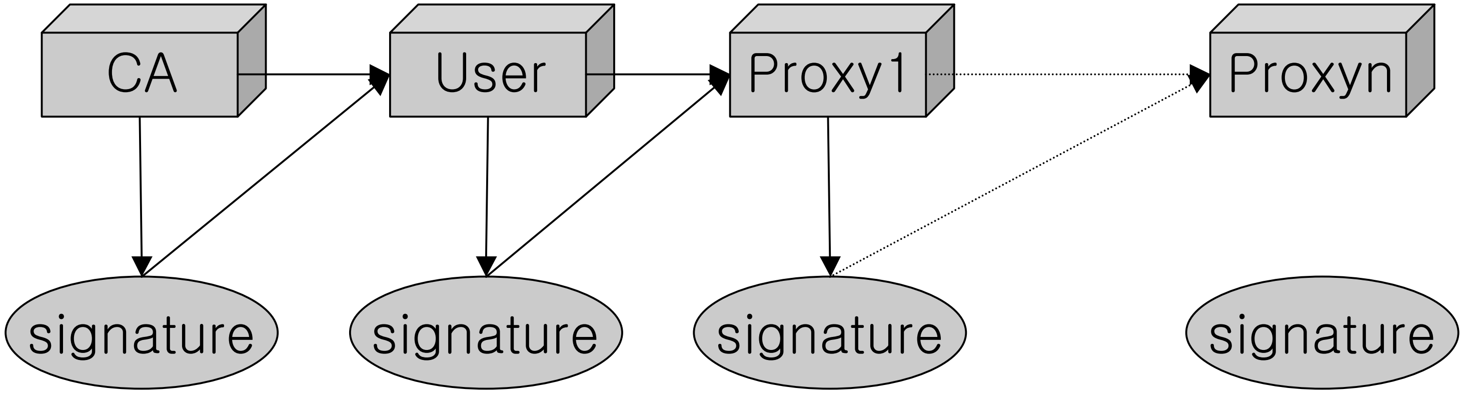
- Grid Security
  - Grid Security Challenges
  - Grid Security Requirements
- **Current Status of Grid Security**
  - Authentication and Delegation
  - Authorization
  - Grid Security Infrastructure (GSI)
  - OGSA
  - Web Services Security
- Things need more study
  - Authentication Interoperability
  - Fine-grained Authorization
- Summary

Authentication and Delegation (1/3)

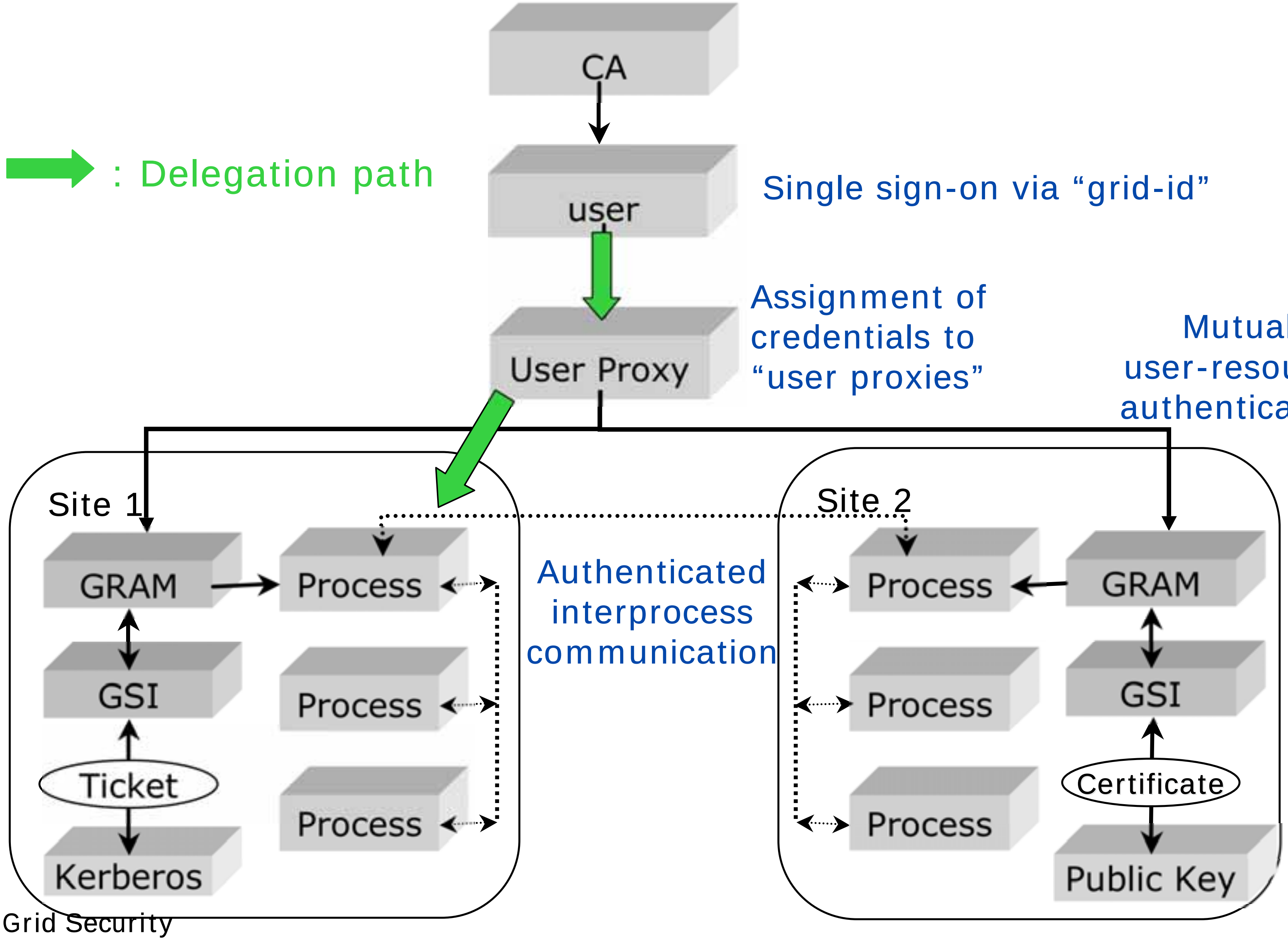
- The use of X.509 Certificates
  - Authentication by a distinguished name in a certificate under shared common CAs
  - Delegation and single sign-on through the use of X.509 proxy certificates
- Username and Password Authentication supported in GT4
  - Supporting WS-Security standard as opposed to X.509 credentials
  - Only providing authentication and not advanced features such as delegation, confidentiality, integrity, etc

Authentication and Delegation (2/3)

- Delegation of proxy certificates
  - Remote generation of user proxy
  - Generation of a new private key & certificate using the original key
  - Password or private key are not sent on network.



Authentication and Delegation (3/3)



Single sign-on via “grid-id”

Assignment of credentials to “user proxies”

Mutual user-resource authentication

Authenticated interprocess communication

Mapping to local ids

Authorization (1/4)

- Users want to delegate their rights to proxies in other systems.
- Resource providers need an authorization service for user proxies submitted to their systems.
- Delegation is the process of transferring rights of users to tasks or proxies.
    - When too much rights are delegated, the abuse of rights is possible.
    - When too less rights are delegated, proxies cannot be executed completely.
- Thus, we need an authorization service in which users delegate restricted rights to proxies and resource providers can check valid uses of delegated rights.

Authorization (2/4)

- Pull Model
  - Granting a user's rights only on the specific conditions
  - Delegating rights which a user specifies
  - Managing rights with a user and resource providers
  - Example : Akenti

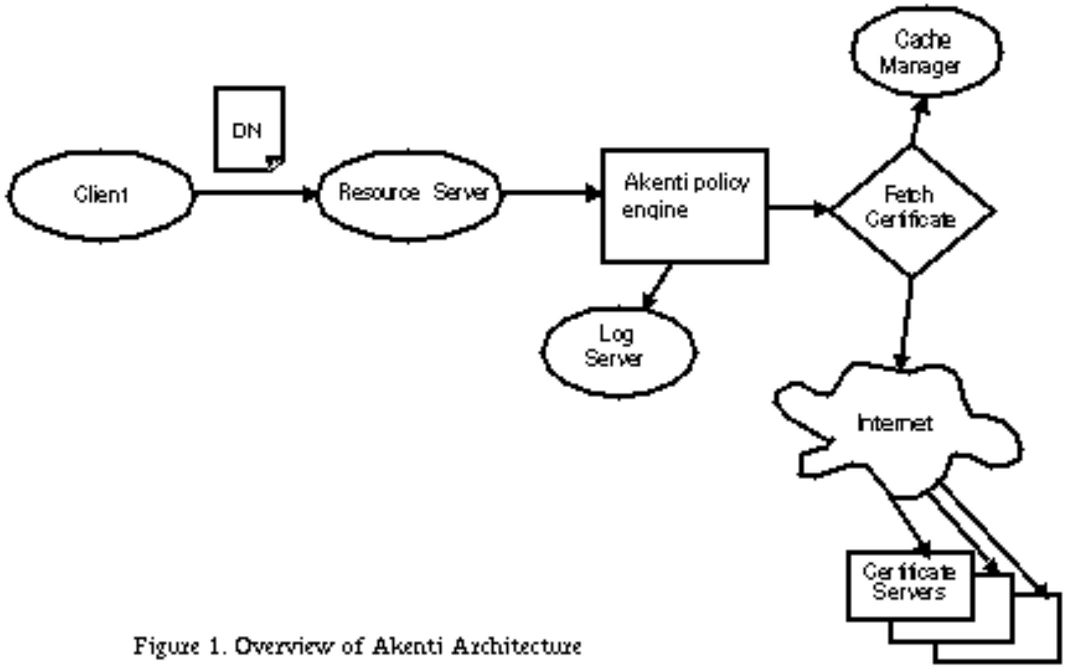
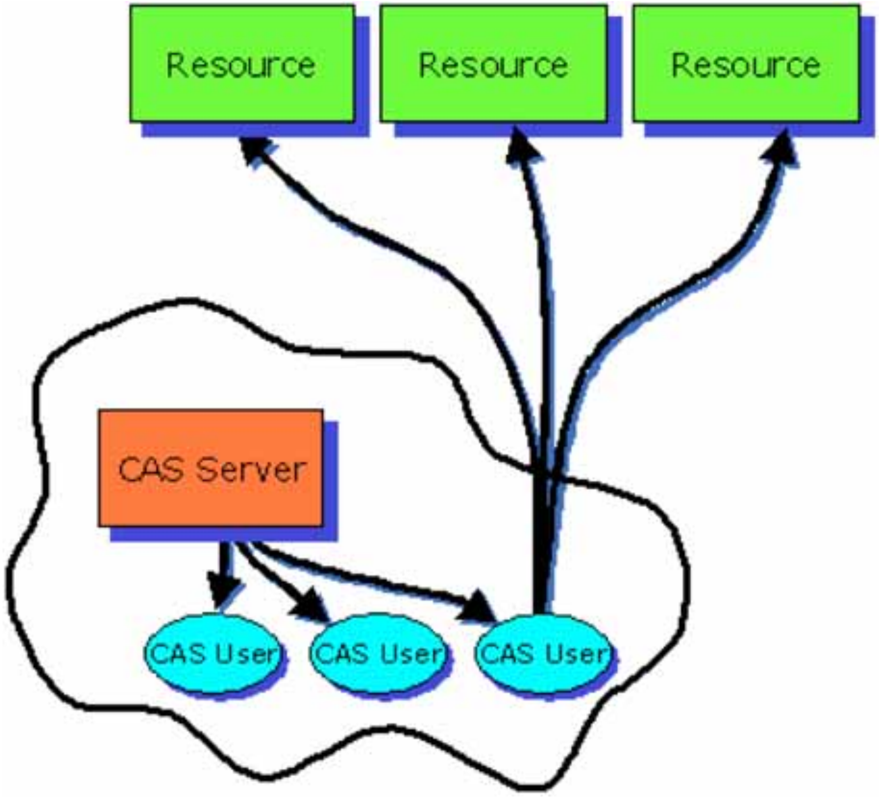


Figure 1. Overview of Akenti Architecture

Authorization (3/4)

- Push Model
  - Granting a user’s rights according to his or her role
  - Managing rights with a central administrator
  - Example : CAS, PERMIS, VOMS



Authorization (4/4)

- Problems in related works
  - Akenti
    - Writing specific conditions and rights manually
    - Managing rights by users and resource providers
  - CAS
    - Delegating all rights owned by user’s role
    - Not delegating restricted rights

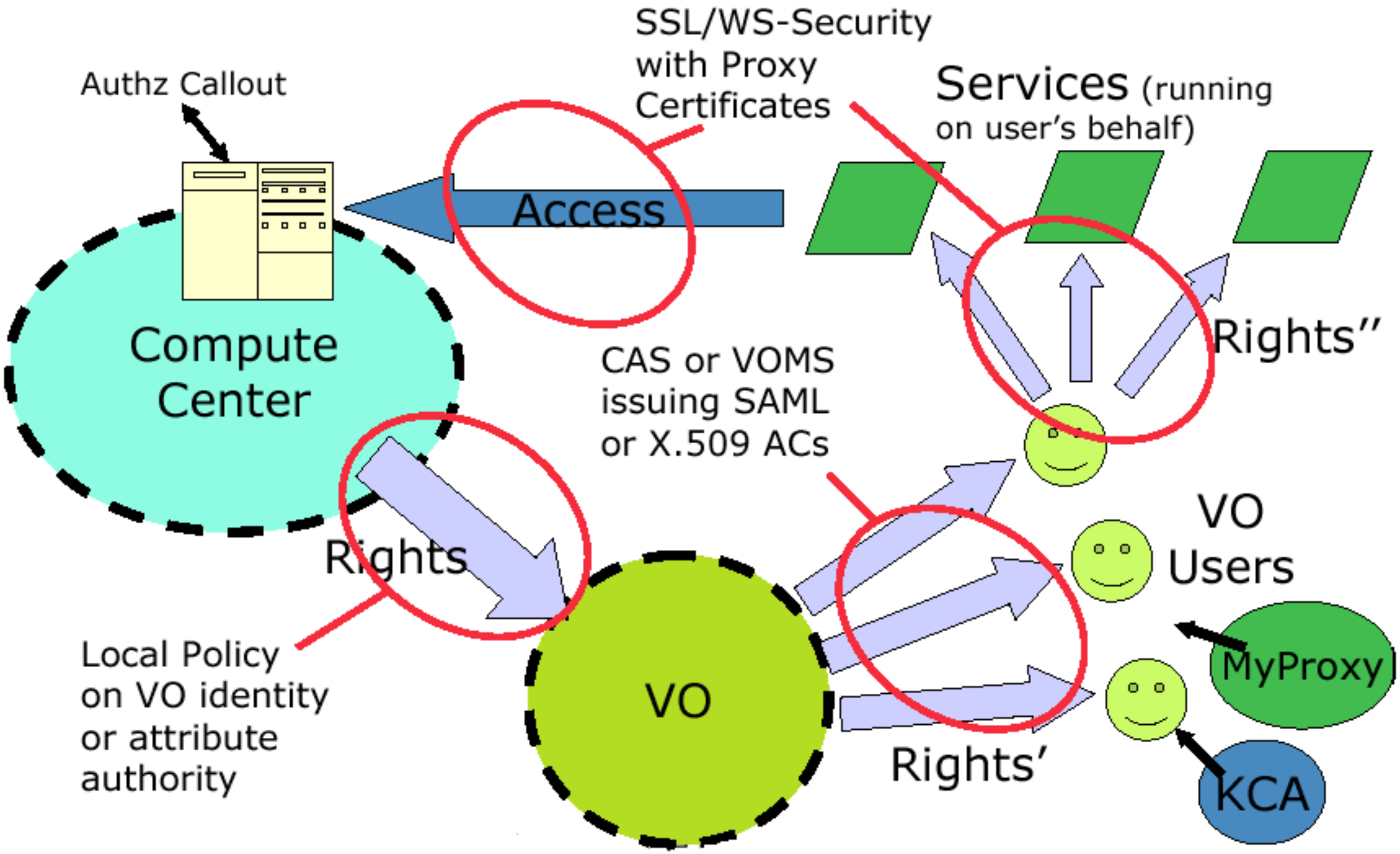
Grid Security Infrastructure (GSI)

- The fundamental security services in the Globus Toolkit
- Based on standard PKI technologies
  - SSL protocol for authentication, message protection
  - One-way, light-weight trust relationships by CAs
- X.509 Certificates for asserting identity
  - For users, services, hosts, etc
- Grid identity
  - A user is mapped to local identities using the distinguished name of the user’s certificate.

Grid Security Infrastructure (GSI)

- X.509 Proxy Certificates
  - Enables single sign-on.
  - Allows users to delegate their identities and rights to services.
- Community Authorization Service (CAS)
  - Enables fine-grained authorization policy.
  - Resource providers set course-grained policy rules for foreign domain on CAS-identity.
  - CAS sets policy rules for its local users.
  - Requestors obtain capabilities from their local CAS.

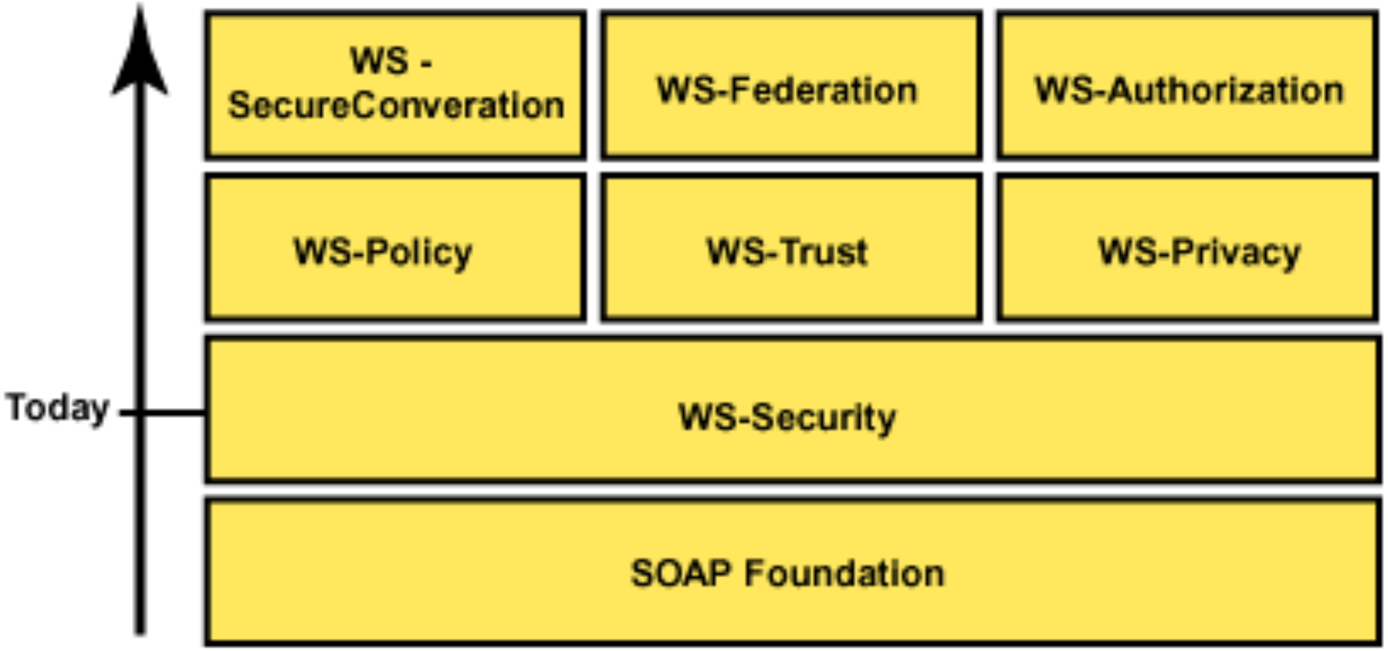
Grid Security Infrastructure (GSI)



Open Grid Services Architecture (OGSA)

- A Grid system architecture
  - Based on Web services and technologies
  - An open source collection of Grid services that follow OGSA principles are offered by the Globus project since GT3.0.
- WS-Resource Framework (WSRF)
  - A set of Web service specifications being developed by the OASIS organization
  - Describing how to implement OGSA capabilities using Web services
- Standardization
  - Underway in the Global Grid Forum (GGF) and OASIS
  - Many working groups on Grid security, such as OGSA Security, GSI, Authorization Frameworks and Mechanisms (AuthZ), Certificate Authority Operations (CAOPS), Grid Certificate Policy (GCP), and OGSA Authorization (OGSA-Authz)

Security in a Web Services World



- The Web services security roadmap provides a layered approach to address Web services.
- The OGSA security models needs to be consistent with Web services security model.

Contents

- Grid Security
  - Grid Security Challenges
  - Grid Security Requirements
- Current Status of Grid Security
  - Authentication and Delegation
  - Authorization
  - Grid Security Infrastructure (GSI)
  - OGSA
  - Web Services Security
- **Things need more study**
  - Authentication Interoperability
  - Fine-grained Authorization
- Summary

Authentication Interoperability

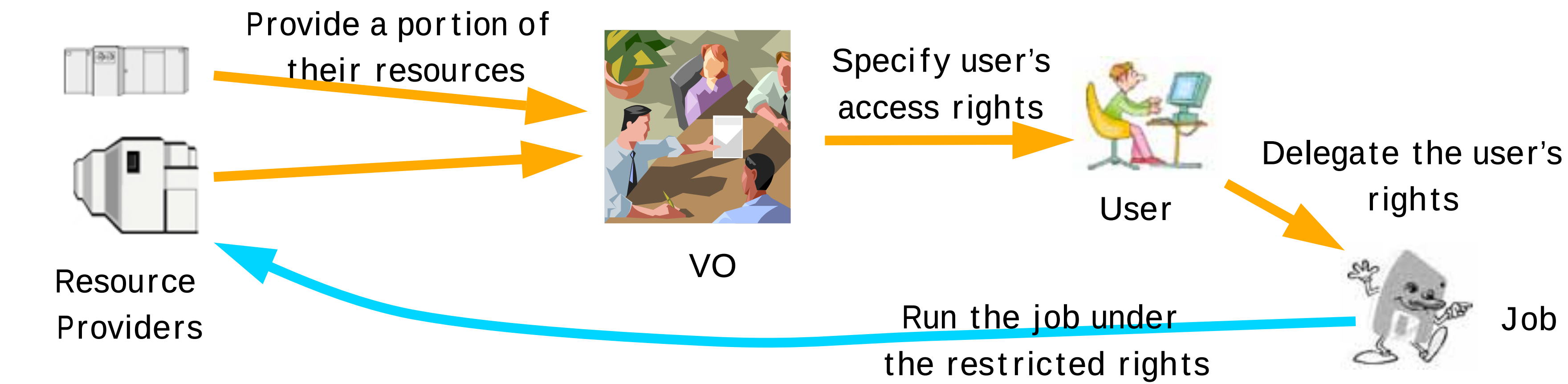
- Motivations
  - Use of different authentication schemes by different resource providers
  - Use of different policies for different resource providers and organizations
- Requirements
  - Need an interoperable authentication method
  - Need an automatic policy match and negotiation

Example Case

- Case
  - User A is given access rights to resources B and C when running a process D for some time.
  - How do we know he is accessing resources B and C for the process D?
  - How do we know he is not redoing the previously allowed job?
  - How do we know he has not exceeded his access time on using resources B and C in case that the resources given to the VO at which the user A belongs are larger than those given to the user A.
  - Etc...
- Need a fine control of resources
  - Also need for accounting

Fine-grained Authorization Service

- Motivations
  - Resource providers want their resources to be used by only VO members under their local polices.
  - VO managers specify user access rights.
  - A user delegates his or her rights to the job to run.
- Requirements
  - Combining polices from different sources
  - Fine-grained resource control
  - VO-based management of jobs and resources


Provide a portion of their resources
Specify user's access rights
Delegate the user's rights
Resource Providers
VO
User
Job
Run the job under the restricted rights

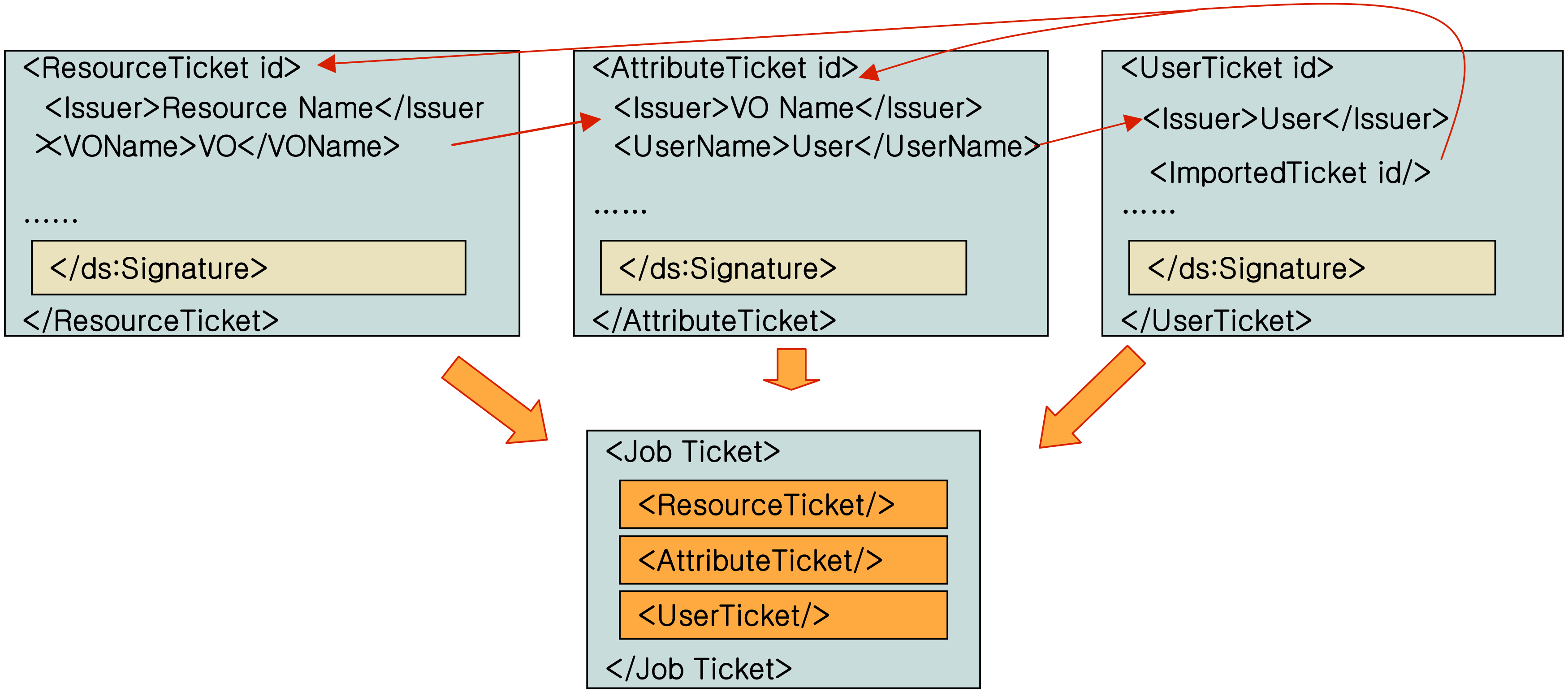
Grid Security

TAS : Tickets

- A ticket is an XML record asserting that the issuer specifies a policy.
    - A resource provider notifies the resource usage policy.
    - A VO manager issues VO users' attributes.
    - A user delegates his or her rights to the submitted job.
- Each ticket is signed by the private key of the issuer to protect the integrity of the ticket.
- Tickets are unforgeable and exchangeable among VO entities for resource control.
- Tickets are classified into
    - resource ticket,
    - attribute ticket,
    - user ticket, and
    - job ticket.

TAS : Job Ticket

- Generated by a user in order to request the rights
- Including necessary tickets for a job
  - Imported ticket field in the user ticket indicates other tickets.



<ResourceTicket id>
<Issuer>Resource Name</Issuer
><VOName>VO</VOName>
……
</ds:Signature>
</ResourceTicket>

<AttributeTicket id>
<Issuer>VO Name</Issuer>
<UserName>User</UserName>
……
</ds:Signature>
</AttributeTicket>

<UserTicket id>
<Issuer>User</Issuer>
<ImportedTicket id/>
……
</ds:Signature>
</UserTicket>

<Job Ticket>
<ResourceTicket/>
<AttributeTicket/>
<UserTicket/>
</Job Ticket>

TAS : Supported Grid Services

- Dynamic VO Management
  - A VO is easily managed by sharing resource and attribute tickets.
  - VO policies can be changed by re-issuing the corresponding tickets.
- Fine-grained Rights Delegation
  - Resource providers and VO managers delegate a set of permitted rights to users.
  - A user also delegates his or her rights to the job using the user ticket.

Summary

- Grid Security
  - Needs to solve many security issues to provide dynamic, scalable VOs in Grid computing environment.
  - Hard problem due to diversity, interoperability, integration, ...
- Fine-grained Authorization Services
  - As a Grid security service, it needs VO-wide fine-grained authorization of jobs and resources.

References

- F. Siebenlist, V. Welch, “Grid Security : The Globus Perspective,” GlobusWORLD 2004, http://www.globus.org/
- V. Welch, F. Siebenlist, I. Foster, J. Bresnahan, K. Czajkowski, J. Gawor, C. Kesselman, S. Meder, L. Pearlman, S. Tuecke, “Security for Grid Services,” HPDC-12, June 2003.
- N. Nagaratnam, P. Janson, J. Dayka, A. Nadalin, F. Siebenlist, V. Welch, I. Foster, S. Tuecke, “The Security Architecture for Open Grid Services,” OGSA-SEC-WG document, GGF.
- S. H. Kim, J. Kim, S. J. Hong, S. W. Kim, "Workflow based Authorization Service in Grid", 4th International workshop on Grid Computing (Grid 2003), pp 94-100, Novebmer 2003.
- S. H. Kim, K. H. Kim, J. Kim, S. J. Hong, S. W. Kim, “Workflow-based Authorization Service in the Grid”, Journal of Grid Computing, vol. 2, no. 1, pp. 43-55, 2004.
- B. J. Kim, S. J. Hong, J. Kim, "Ticket-Based Fine-Grained Authorization Service in the Dynamic VO Environment," ACM Workshop on Secure Web Services, October 2004.
- B. J. Kim, K. H. Kim, S. J. Hong, J. Kim, "Ticket-based Grid Services Architecture for Dynamic Virtual Organizations,“ LNCS 3470 (Advances in Grid Computing: EGC 2005), pp. 394-403, 2005.

Reliability and Security: From Measurements to Design

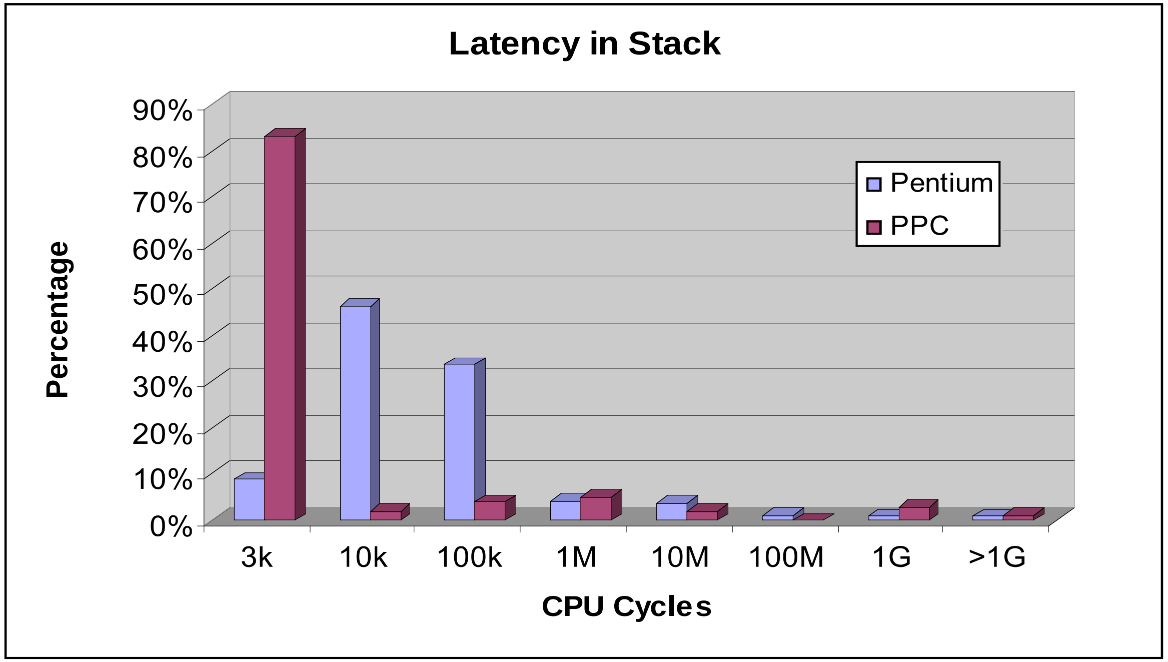
Ravi K. Iyer
Karthik Pattabiraman, Weining Gu,
Giampaolo Saggese, Zbigniew Kalbarczyk
Center for Reliable and High-Performance Computing
Coordinated Science Laboratory
University of Illinois at Urbana-Champaign

Supported by: NSF, SRC, DARPA, SUN, IBM, HP

http://www.crhc.uiuc.edu/DEPEND

Crash Latency Distributions for

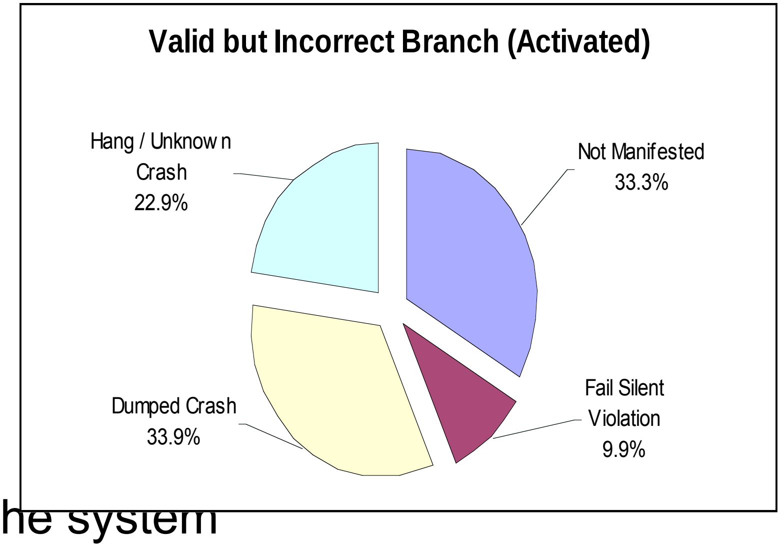
(Linux on Pentium P4 and PowerPC G4)



Early detection of kernel stack overflow on PPC major contributor to reduced crash latency

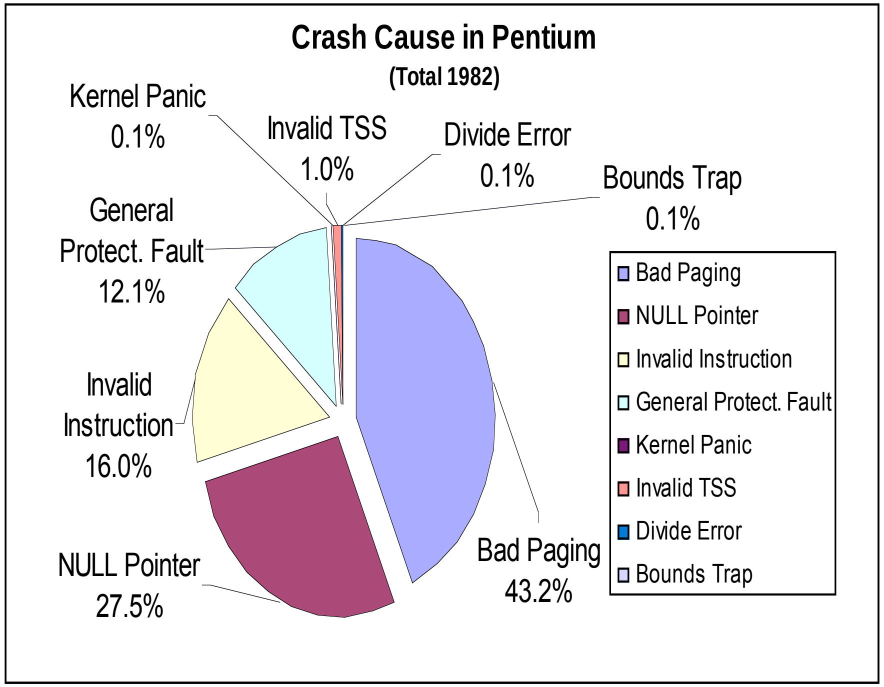
Crash Severity:
Linux Kernel on Pentium

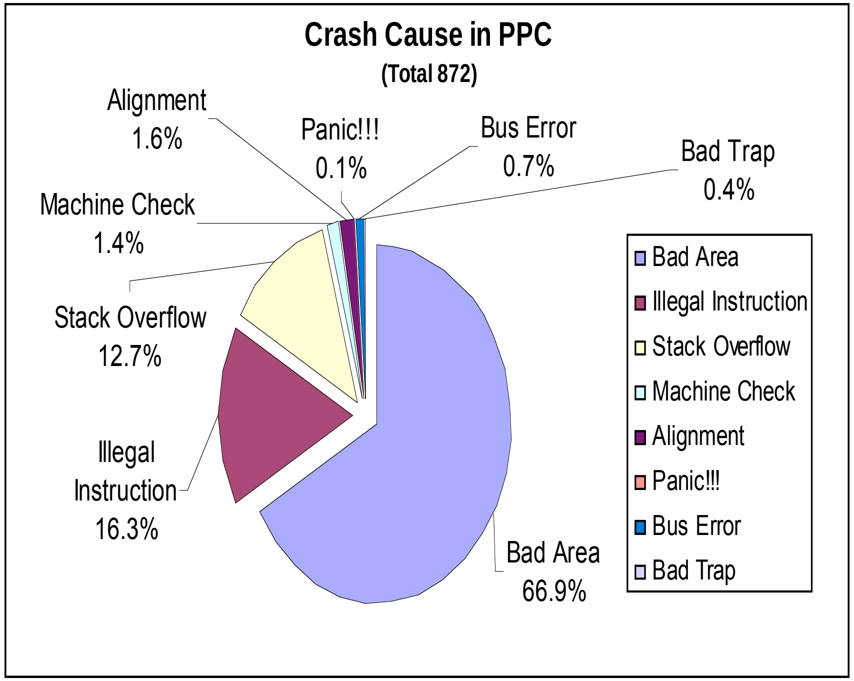
- Significant percentage (33%) of errors that alters the control path have no effect
  - Inherent redundancy in the code
- The most severe crashes are due to reversing the condition of a branch instruction
- The most severe crashes require a complete reformatting of the file system on the disk
  - Can take nearly an hour to recover the system
  - Profound impact on availability
  - To achieve 5NINES of availability (5 minutes/yr downtime) one can effort one such failure in 12 years



Crash Causes:

Linux on PowerPC G4 & Pentium 4

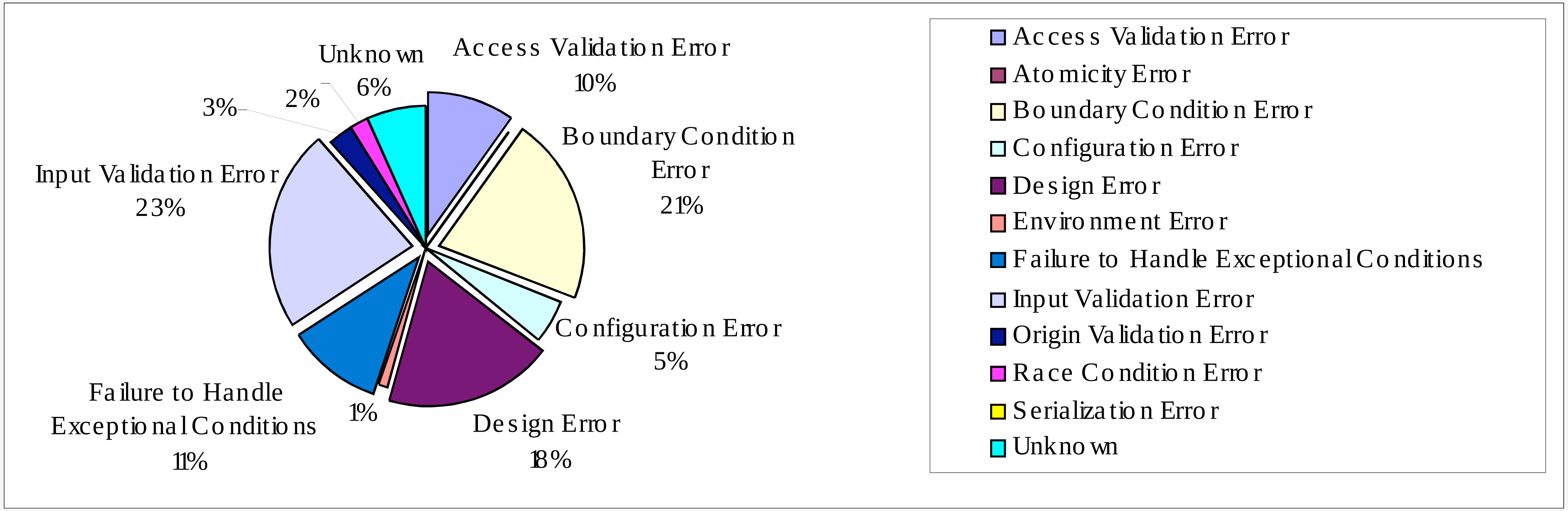




- NULL Pointer: NULL pointer de-reference;
- Bad Paging: Bad paging (except NULL pointer)
- General Protection Fault: Exceeding segment limit;
- Kernel Panic: Operating system detects an error;
- Invalid TSS: Selector, or code segment outside limit;
- Bounds Trap: Bounds checking error.

- Bad Area: Bad paging including NULL pointer;
- Stack Overflow: Stack pointer of a process out of range
- Machine Check: Errors on the processor-local bus;
- Alignment: Load/store operands not word-aligned;
- Bus Error: Protection faults;
- Bad trap: Unknown exceptions.

Breakdown of Vulnerabilities (*Bugtraq*)



- *Access Validation Error*: an operation on an object outside its access domain.
- *Atomicity Error*: code terminated with data only partially modified as part of a defined operation.
- *Boundary Condition Error*: an overflow of a static-sized data structure: a classic buffer overflow condition.
- *Configuration Error*: a system utility installed with incorrect setup parameters.
- *Environment Error*: an interaction in a specific environment between functionally correct modules.
- *Failure to Handle Exceptional Conditions*: system failure to handle an exceptional condition generated by a functional module, device, or user input.
- *Input Validation Error*: failure to recognize syntactically incorrect input.
- *Race Condition Error*: an error during a timing window between two operations.
- *Serialization Error*: inadequate or improper serialization of operations.
- *Design Error* and, *Origin Validation Error*: Not defined.

Bugtraq database included 5925 reports on software related vulnerabilities
(as of Nov.30 2002)

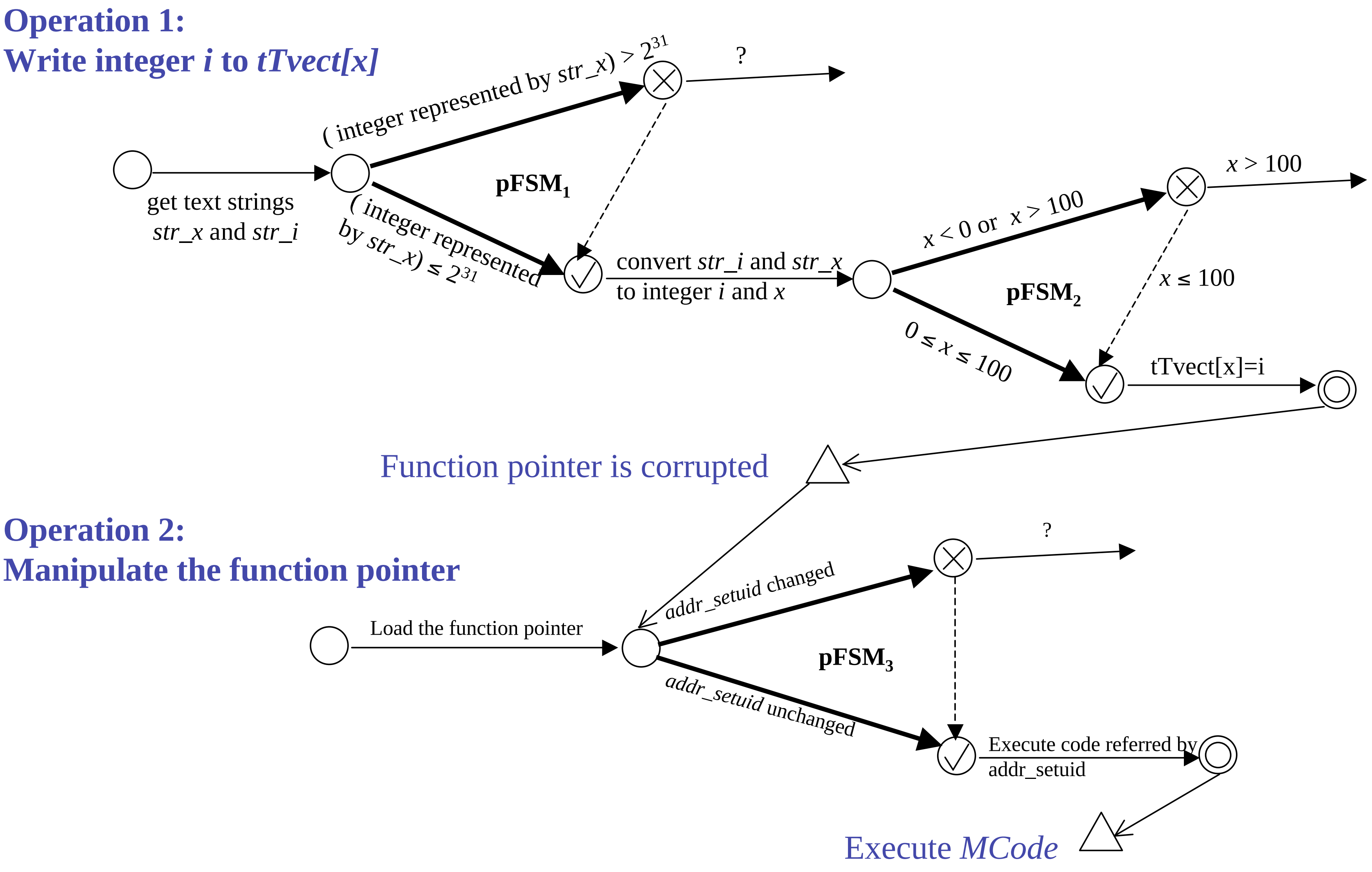
Observations from Vulnerability Analysis

- Exploiting a vulnerability involves multiple vulnerable **operations** on several objects.

- Exploits must pass through multiple **elementary activities**, each providing an opportunity for performing a security check.

- For each elementary activity, the vulnerability data and corresponding code inspections allow us to define a **predicate**, which if violated, naturally results in a security vulnerability.

Example: FSM Model for the *Sendmail* Vulnerability



Operation 1:
Write integer *i* to *tTvect[x]*

Function pointer is corrupted

Operation 2:
Manipulate the function pointer

Execute *MCode*

Some Lessons Learned

- Extracted common characteristics of a class of security vulnerabilities
- Developed an FSM methodology to model vulnerabilities.
- Only three pFSM types were required. Enforced reasoning indicate opportunities for security checking.
- Most vulnerabilities are in the interface between applications and library functions
- **Question: Can we develop *Vulnerability-Masking* schemes based on the observed characteristics**

Challenges: Understanding Failure Data

- Expectation is that transients will increase
  - Shrinking device size → Increased transient error rate
    - More error checking that is closer to processor needed
- System level impact of increase in transients
  - Increased error propagation → near-coincident (correlated) errors
  - More latent errors
  - Question: What are the corresponding high level fault models?
- Current recovery techniques oriented towards single isolated errors
- Recovery of correlated (or latent) errors is complex and adds significantly to unavailability

Challenges: Understanding Attack Data

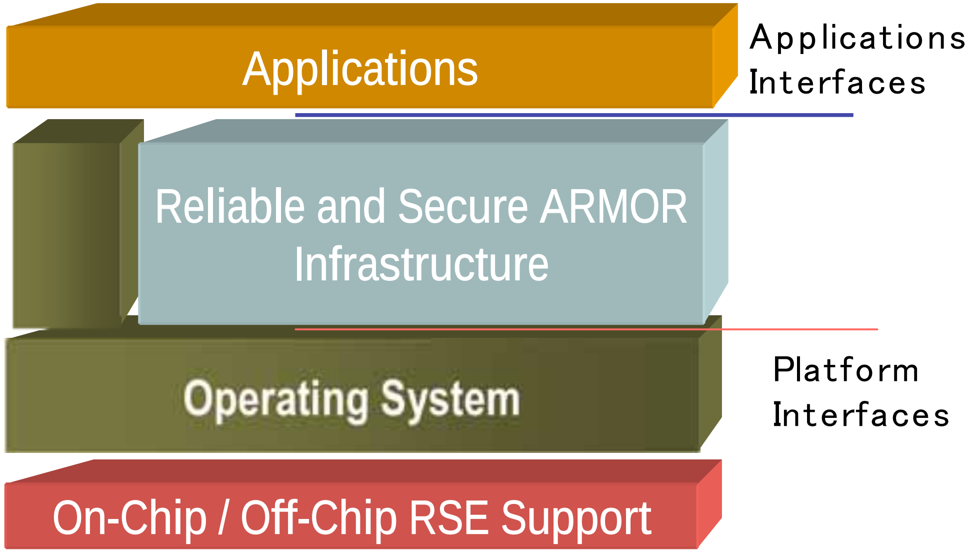
- Analysis of data (from *Bugtraq*) on security attacks to:
  - identify vulnerabilities and to classify the attacks according to attacks causes
  - understand potential inconsistencies in application/system specifications resulting in security vulnerabilities of an actual application/system implementation
- Measurement-based models depicting the attack process
- Software (e.g., compiler-based) and hardware (e.g., processor embedded) vulnerability masking/prevention techniques

What is Needed?

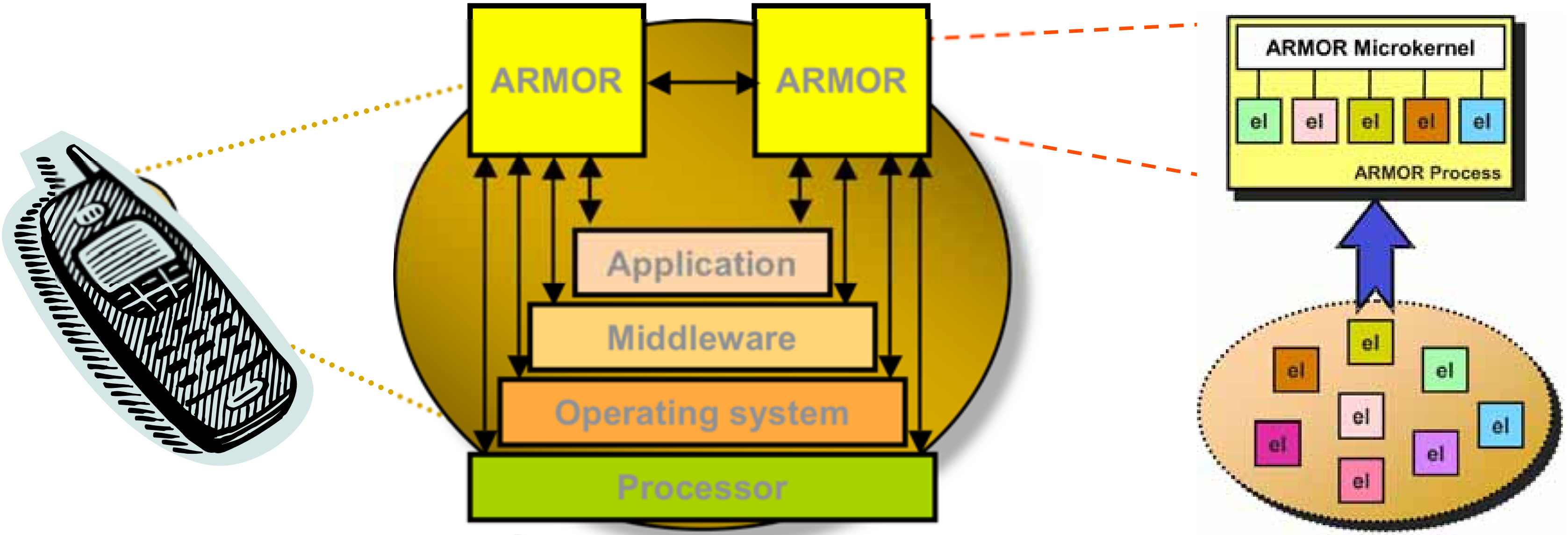
- Application aware detection mechanisms
  - generic fault-tolerance and security techniques, targeting a particular fault/attack-model provide limited coverage
  - application cannot selectively take advantage of mechanisms, which best meet the needs
- Extract application properties that can be used as an indicator of correct behavior
- Exploit the knowledge of such properties to derive efficient error detection
  - application-specific checks can complement the coverage provided by generic techniques
- Assess the benefits (tradeoffs) of software or hardware implementaion

Application Aware Checking in Software:
ARMOR Self-checking Middleware

- Adaptive Reconfigurable Mobile Objects of Reliability
  - Processes composed of replaceable software modules.
  - Provide error detection, recovery and security services to user applications.
- ARMORs Hierarchy form runtime environment:
  - System management, detection, and recovery services distributed across ARMOR processes.
  - ARMORs resilient to their own failures.


Applications
Applications Interfaces
Reliable and Secure ARMOR Infrastructure
Operating System
Platform Interfaces
On-Chip / Off-Chip RSE Support

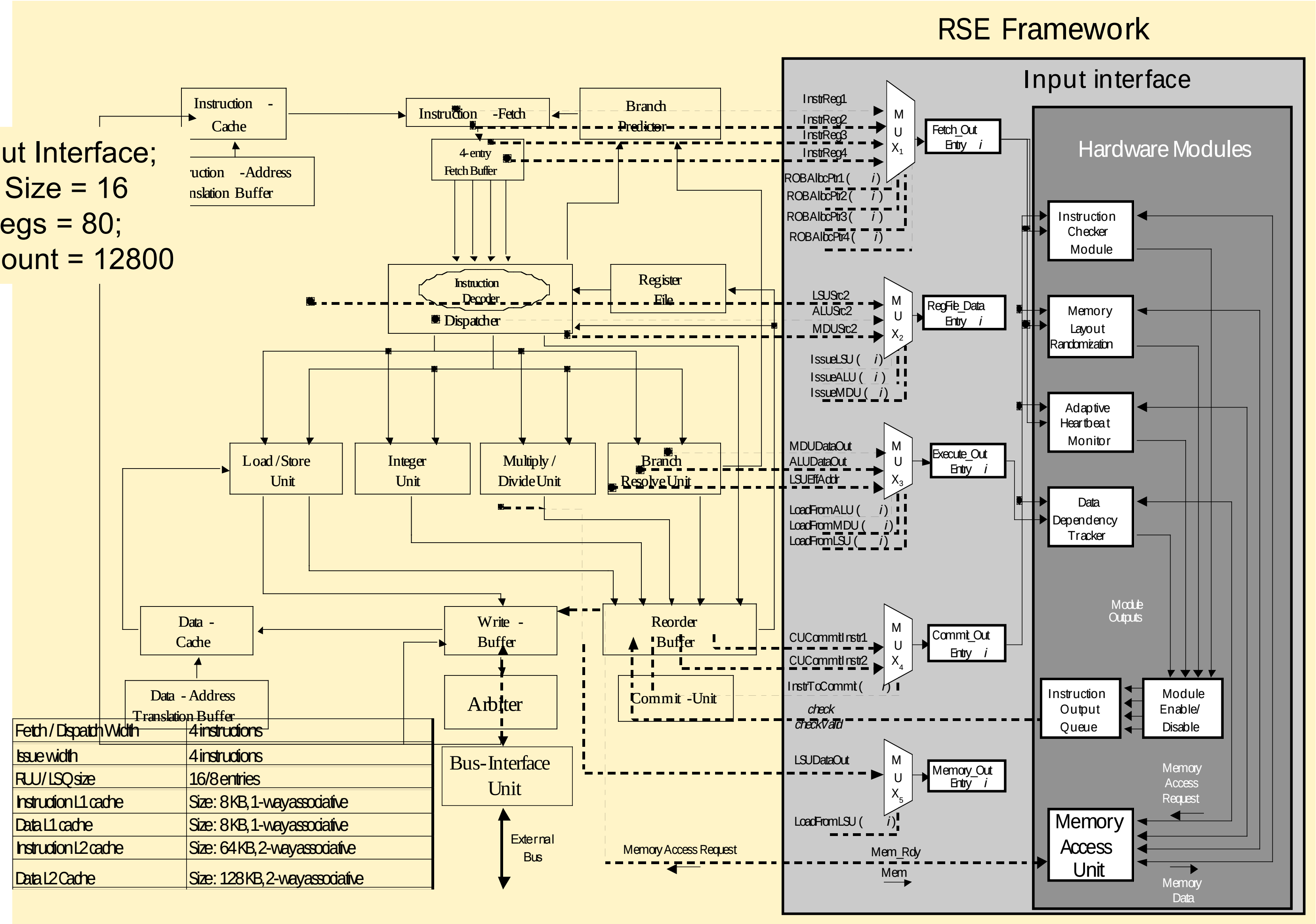
ARMOR Self-checking Middleware: "Embedded Solution"



- Modular design of ARMOR processes around elements lends itself well to small footprint solutions.
- Special versions of elements optimized for memory and performance requirements.
- Specialized microkernel:
  - Remove support for inter-ARMOR communication through regular messaging infrastructure
  - Static configuration of elements; no need to dynamically add/remove elements

Application Aware Checking in Hardware: Reliability and Security Engine

For Input Interface;
Queue Size = 16
32-bit regs = 80;
Gate Count = 12800



Fetch / Dispatch Width	4 instructions
Issue width	4 instructions
RUU / LSQ size	16/8 entries
Instruction L1 cache	Size: 8 KB, 1-way associative
Data L1 cache	Size: 8 KB, 1-way associative
Instruction L2 cache	Size: 64 KB, 2-way associative
Data L2 Cache	Size: 128 KB, 2-way associative

N. Nakka, J. Xu, Z. Kalbarczyk, R. K. Iyer, “An Architectural Framework for Providing Reliability and Security Support”, DSN2004.

Reliability and Security Engine

- A common framework to provide a variety of application-aware techniques for error-detection, masking of security vulnerabilities and recovery under one umbrella, in a uniform, low overhead manner.
- FPGA implementation as an integral part of a superscalar microprocessor
- Hardware-implemented error-detection and security mechanisms embedded as FPGA modules in the framework
- The framework serves two purposes
  - Hosts hardware modules that provide reliability and security services, and
  - Implements interface of the modules with the main pipeline and the executing software (OS and application)

TRUSTED *ILLIAC*

COMBINING HIGH PERFORMANCE WITH APPLICATION-AWARE RELAIBILTY AND SECURITY

HTTP://WWW.CSL.UIUC.EDU

Goal: Application-Aware Trusted Computing

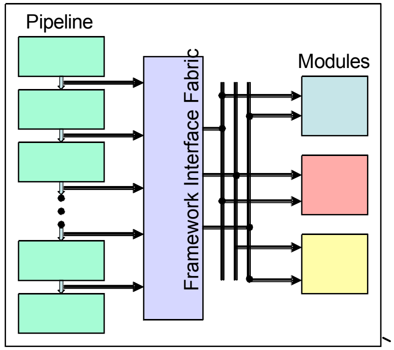
- Create a large, demonstrably-trustworthy, enterprise computing platform
  - Application aware reliability and security
  - Reconfigurable
  - High performance
  - Easy programming
- Support for
  - Enterprise computing with seamless extension across wireline-wireless domains
  - Significant number of applications that co-exist and share the HW/SW resources
- State of the Art: Provide HW and SW with a *one-size-fits-all* approach
  - Creating a trustworthy environment is complex, expensive to implement and difficult to validate

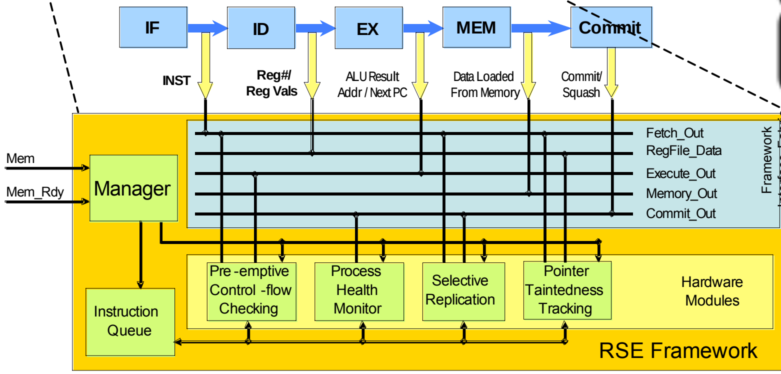
Application Aware Trusted Computing

- Applications-specific level of reliability and security provided in a transparent manner, while delivering optimal performance
- Customized levels of trust (specified by the application)
    - enforced via an integrated approach involving
        - re-programmable hardware,
        - compiler methods to extract security and reliability properties
        - configurable OS and middleware
- Scale from few nodes to large networked systems
- Enable inclusion of ad-hoc wireless nodes

Application-Aware Checking: An Example

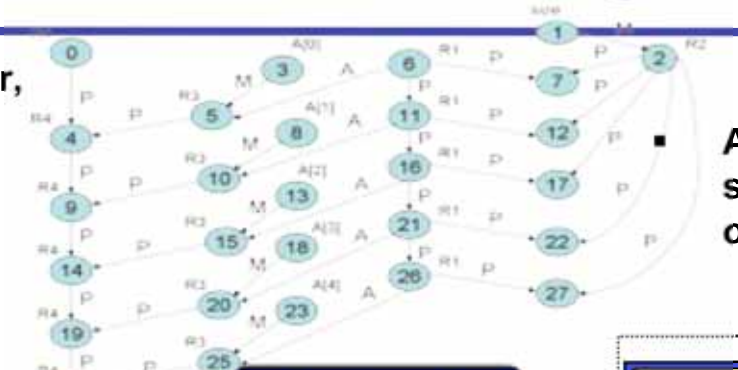
On-core approach – processor, framework, and modules part of the same core.

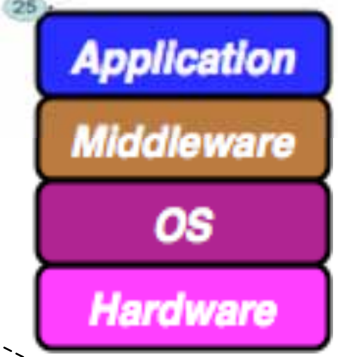




A Reliability and Security Engine (RSE)

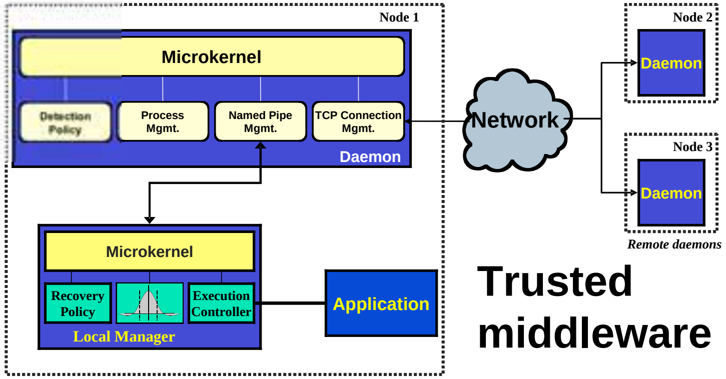
- **Reconfigurable processor-level hardware framework**
- **Provides HW modules for reliability and security**
- **Modules and framework interface configured to meet application demands**



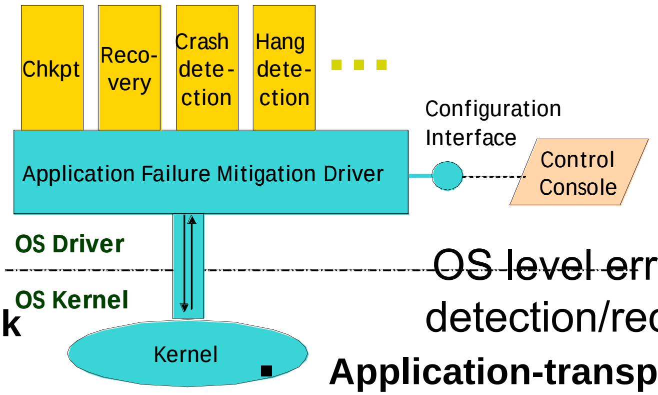


Assertion-Based Checking

- **Automatic generation and software/hardware implementation of error detectors**



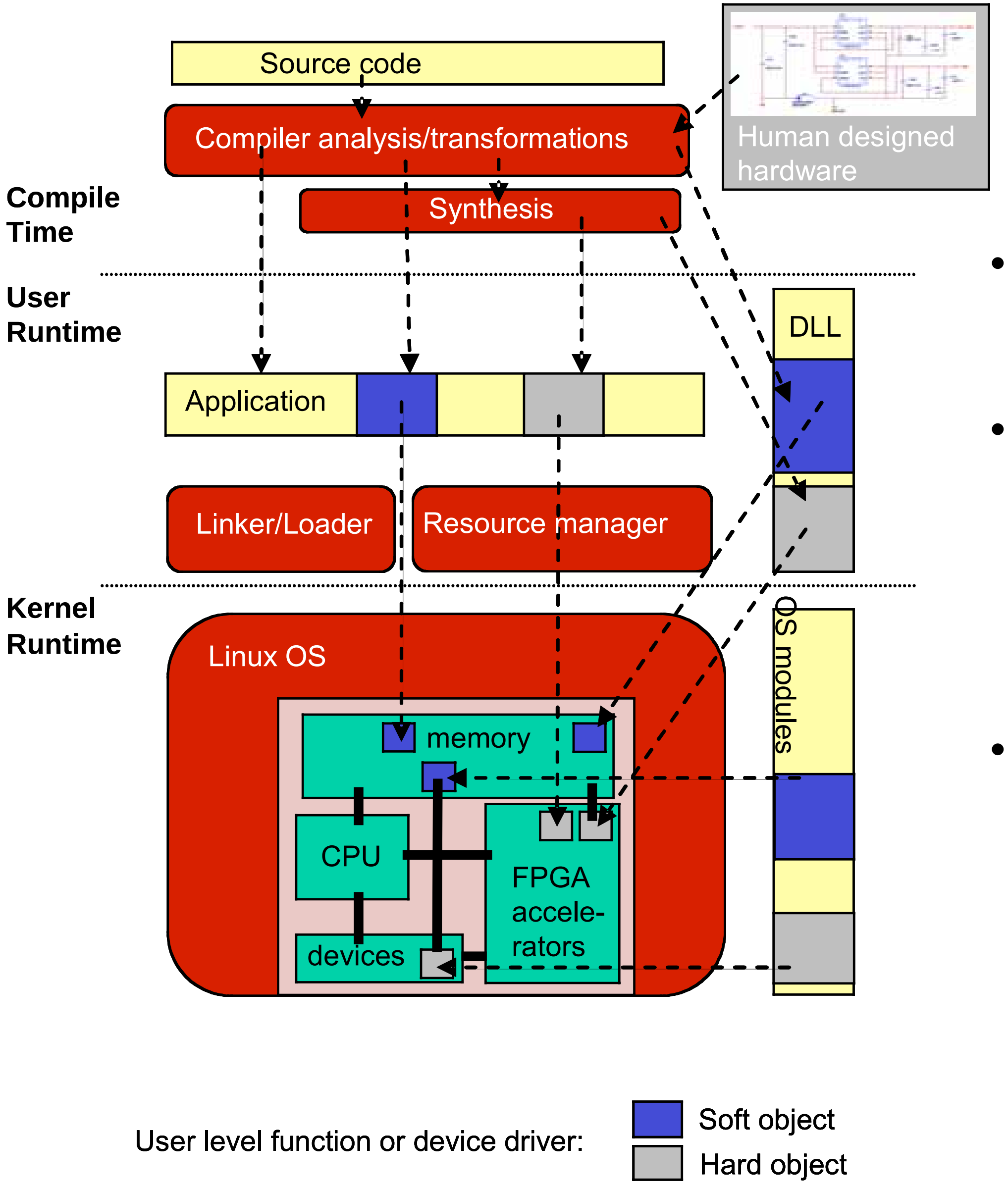
Trusted middleware



OS level error detection/recovery

- **Application-transparent OS-level checkpointing**
- **OS health monitoring**

Hardware/Software Execution Model



- Seamless integration of hardware accelerators into the Linux software stack
- Compiler supported deep program analysis and transformations to generate CPU code, hardware library stubs and synthesized components
- OS resource management

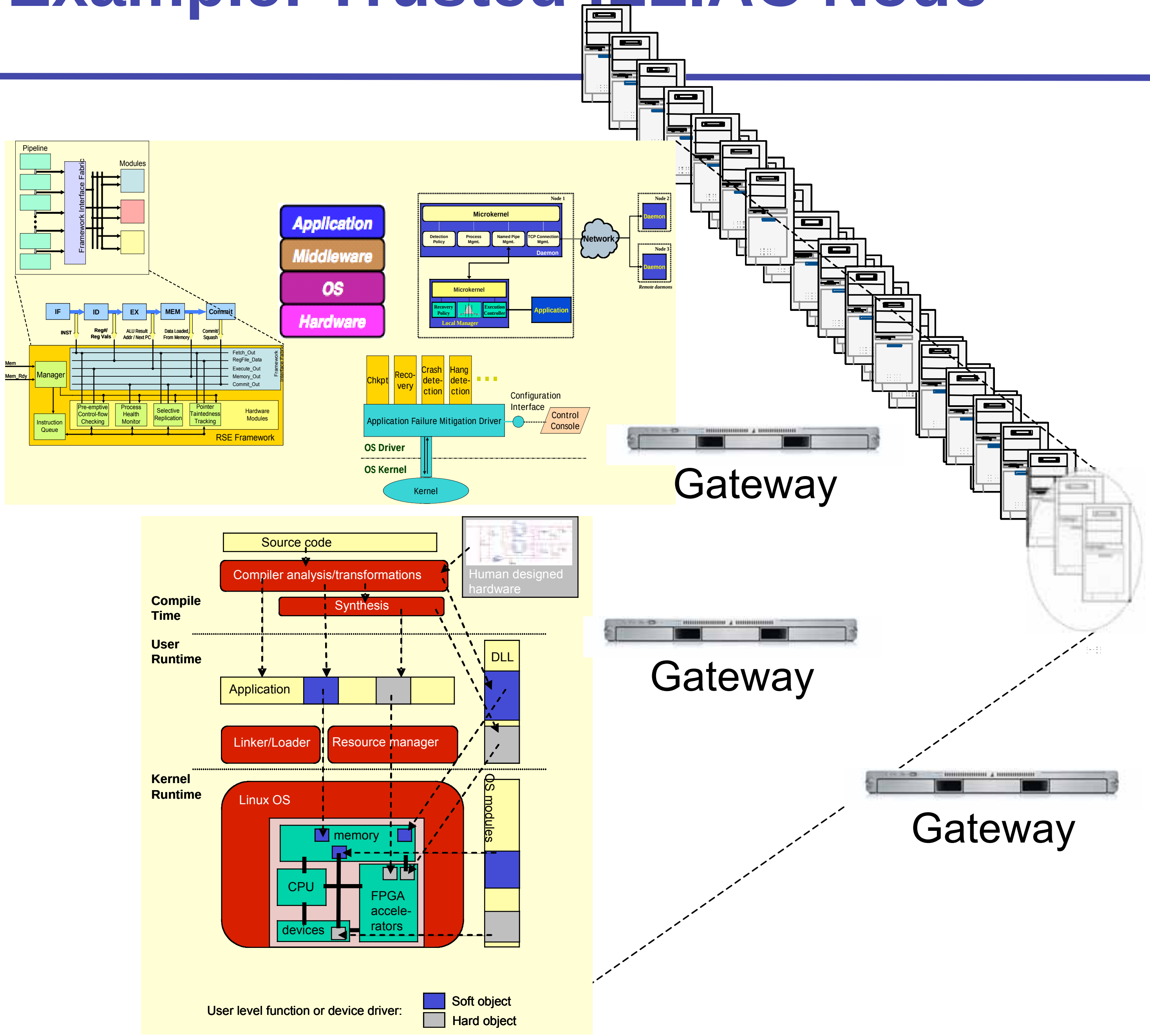
Validation Framework

- An integral part of the Trusted ILLIAC
- Quantitative assessment of alternative designs and system solutions
- Provides tools for
  - Analytical models (e.g., MOBIUS)
  - Simulation (e.g., RINSE)
  - Experimental validation (e.g., NFTAPE)
    - Fault/error injection
    - Attack generation
  - Monitoring
  - Measurement
- Crucial in making design decisions, which require understanding tradeoffs such as cost (in terms of complexity and overhead) versus efficiency of proposed mechanisms.

Trusted *ILLIAC*: The Broader Context

- New experience in system building: reliable and secure processing architectures, smart compilers combined with configurable OS and hardware
- Pushing the boundaries in customizable trusted computing technologies
- Enable university, industry, and government collaboration
- Train the next generation of students and professionals
- (See next slide)

Example: Trusted ILLIAC Node



Secure Grid Computing: an Empirical View

IFIP WG 10.4 Workshop on Grid Computing and Dependability

Carl Landwehr (clandweh@nsf.gov**)**
Cyber Trust Coordinator
National Science Foundation

… with thanks to Matti Hiltunin, Bill Cheswick & Brian LaMacchia

July 2, 2005

Grid Computing Application area:

Matti’s talk definitions:

Grid computing: collaborative use of computers, networks, databases, scientific instruments, and data; potentially owned and managed by multiple organizations. –

- Scale: thousands of machines common
- Geographic: worldwide distribution common; transfer large volumes of data across the world
- Administrative: span multiple domains
- Trust: execute tasks on untrusted computers

- Most successful grid application in practice?
- Perhaps it’s controllers and zombies conducting DDOS attacks and sending spam!
- Multiple domains, encrypted signals, coordinated computation, shifting sets of processors, …

NETWORKWORLD

Shakedown on the ‘Net

By Ellen Messmer, NetworkWorld.com, 05/16/05

… extortionist launches a distributed-denial-of service (DDoS) attack, flooding the access to your Web site with unwanted traffic or knocking it offline.

…Does the business pay up?

… it appears that all too often victimized businesses are giving in to the shakedown.

… it’s hard to bring these ‘Net shakedown artists to justice.

…The cost for a few months of anti-DDoS service can add up to a payment to an extortionist, so some see it as an equal burden monetarily.

That’s a sad state for the industry to be in.

http://www.networkworld.com/weblogs/security/008861.html

The Marketplace: Botnet Rental Rates

- From IFIP WG 10.4 47th mtg, Brian La Macchia
- Data fall 2004
- 6 cents per bot-week on offer:
  - Price: $350 weekly, $1,000 monthly
  - Type of service:
    - exclusive (one slot only)
    - Always online (5,000-6,000)
    - Update every (10 minutes)
- Other examples:
  - 3.6 cents per bot-week
  - 2.5 cents per bot-week

How the Market for Zombies Can Lead to Secure Grid Computing

- Today, unmonitored, unpatched home PCs are a big source of zombies used in DDoS attacks
- How to improve patch rate on these PCs?
- Possible service: vendor provides free remote patching and updating service for PCs

Late 1990's: Venture Capital Approach

- Startup company offers the service
- Make it "free":
  - Customer just downloads a bit of software
  - Software occasionally shows the user an ad to pay for itself
  - Periodically visits server with latest patches, etc., and downloads them
- Company lives on the advertising revenue
- User endures slight annoyance or perhaps pays a small annual subscription fee to avoid ads
- Company goes public and investors get rich!

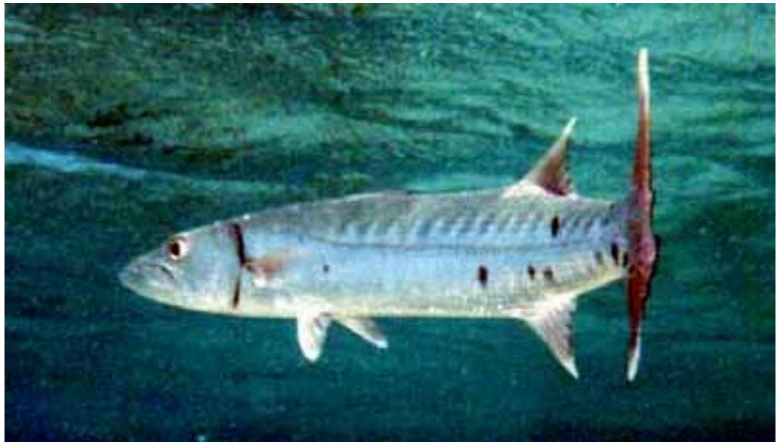
2005: Internet Cyber Security Ecology

- Blackmailer locates potential victim business
- Blackmailer seeks source of zombies
- Home user connects new PC to the Internet
  - Maybe it has a flaw, a weak password, misconfiguration
  - Maybe user browses to web site that installs flawed spyware
- Hacker scanning for victims exploits flaw to compromise machine and turn it into a zombie

Cyber Security Ecology (concluded)

- Hacker strategy: make money by selling access to the machine for spamming, DDoS attacks, etc.
- Hacker tactics:
  - 1. **Close other holes on the machine** so that other competing hackers can't seize his asset
  - 2. **Use just enough of the machine to make money** without bothering home user (or user will discover his exploits and kick him out)
- Sell to the blackmailer
- Victim business pays blackmailer,
- Blackmailer pays zombie-provider ("herder")
- Home user's computer stays patched, produces revenue by computing functions for others

Symbiosis!



- Hacker sells cycles on machine that user didn’t need anyway
- In return, hacker protects user from everyone else, like a barracuda shepherding a school of scissortails

The future we want?

- **Maybe not, but it may be the future we get!**

- **We need to get out of this box!**

Session 1.5

Synthesis and Wrap Up

Moderator
Yoshihiro Tohma

Session 1
Evolution of Grid Computing and Dependability

summery by Moderator Hiro Ihara

1 Dependability Issues in Emerging Web Services-Based Grid Computing by Matti A Hiltunen

Overview、On- going R/D groups and their Activities

Security , Standard, Pricing etc.

from Web Service domain

2 Grid on Future Blade Data Center Infrastructure

by T Basil Smith

R/D issues from commercial business and IBM approach

Accounting and Dependability

from Business domain

3 Grid Computing Evolution and Challenges for

Resilience, Performance and Scalability

by Luca Simoncini

On -going research report and Challenging issues to be done

From Academic domain

What Session 1 could expose

- GC seems reasonable next step
- Many investigating groups have already formed
- Technical evolution in networking is pushing force
- Demand of high performance computing is pulling force
- Strong vulnerability exists
- Strong vulnerability exists
- Can you find any similar system in human activity？
- Strong vulnerability exists
- Are standardization and dependability vital keys?
- When does GC appear in real world?
- More challenge is necessary for Paradigm shift

Session 2 Summary: Practice & Experiments

Jay Lala

Grid Computing

Customer Interest, Expectation, and Requirement for Grid in Dependability Context

48th Meeting of IFIP Working Group 10.4

Takanori Seki, Distinguished Engineer
Technical Sales Support, IBM Japan

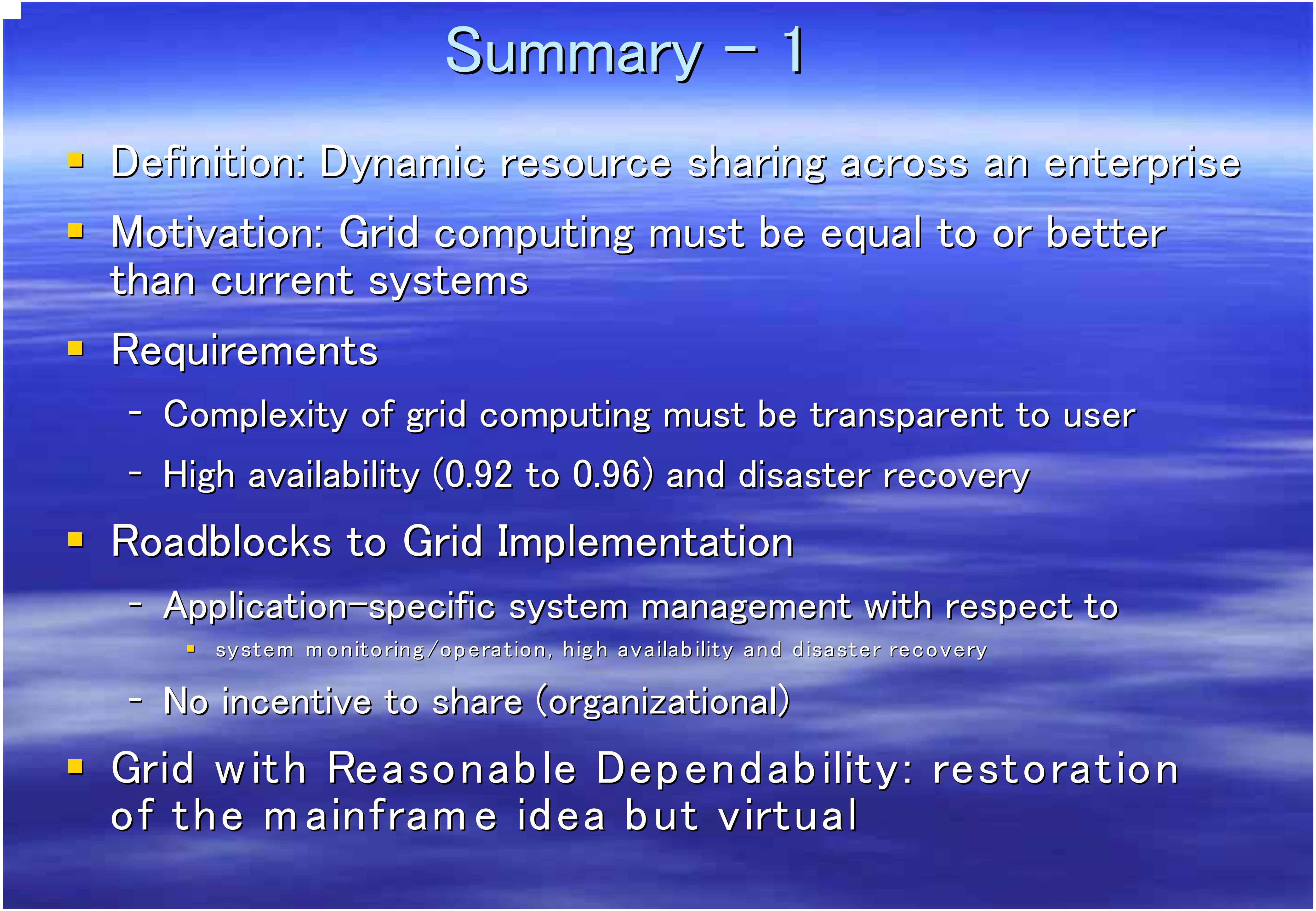
30 (4*7.5)

Japanese Business Grid Project Objectives & Key Technical Issues

IFIP Conference, July 2005

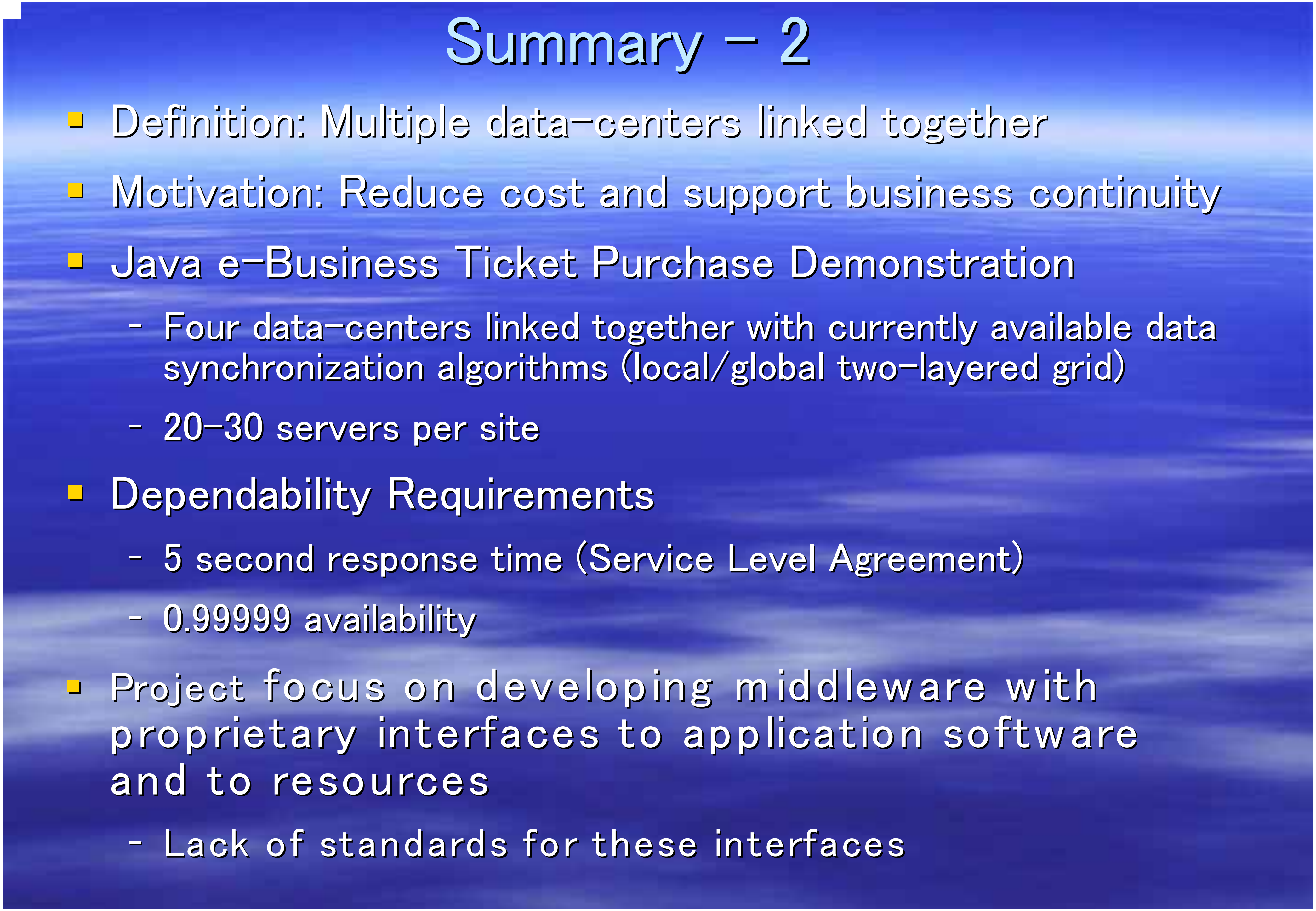
Nobutoshi Sagawa (Hitachi Ltd)
Toshiyuki Nakata (NEC Corporation)
Hiro Kishimoto (Fujitsu Ltd)

Thanks to all the teams in the BUSINESS GRID COMPUTING PROJECT



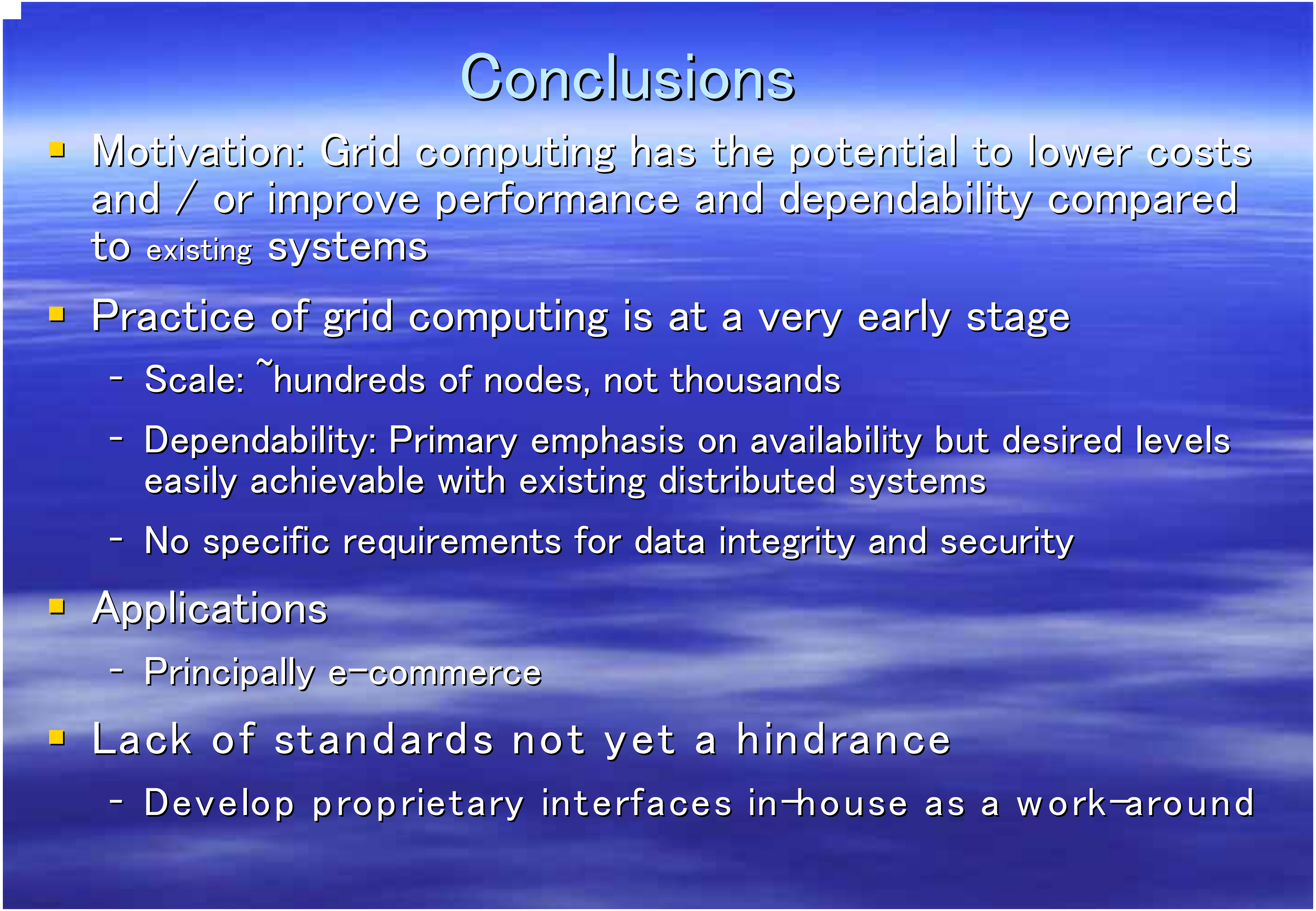
Summary – 1

- Definition: Dynamic resource sharing across an enterprise
- Motivation: Grid computing must be equal to or better than current systems
- Requirements
  - Complexity of grid computing must be transparent to user
  - High availability (0.92 to 0.96) and disaster recovery
- Roadblocks to Grid Implementation
  - Application-specific system management with respect to
    - system monitoring/operation, high availability and disaster recovery
  - No incentive to share (organizational)
- Grid with Reasonable Dependability: restoration of the mainframe idea but virtual



Summary – 2

- Definition: Multiple data-centers linked together
- Motivation: Reduce cost and support business continuity
- Java e-Business Ticket Purchase Demonstration
  - Four data-centers linked together with currently available data synchronization algorithms (local/global two-layered grid)
  - 20-30 servers per site
- Dependability Requirements
  - 5 second response time (Service Level Agreement)
  - 0.99999 availability
- Project focus on developing middleware with proprietary interfaces to application software and to resources
  - Lack of standards for these interfaces



Conclusions

- Motivation: Grid computing has the potential to lower costs and / or improve performance and dependability compared to existing systems
- Practice of grid computing is at a very early stage
  - Scale: ~hundreds of nodes, not thousands
  - Dependability: Primary emphasis on availability but desired levels easily achievable with existing distributed systems
  - No specific requirements for data integrity and security
- Applications
  - Principally e-commerce
- Lack of standards not yet a hindrance
  - Develop proprietary interfaces in-house as a work-around

Challenge

- Not sure how all this is different from distributed computing and justifies a new name
- If Grid Computing is truly something new and unique, the community needs to define crisply
  - What is it
  - What benefits it might offer
  - What are the unique problems posed by grid computing, especially from the dependability viewpoint

Session 4
Security in GRID Computing

Summary by
Paulo Veríssimo

Security Issues in Grid: Authentication and Authorisation

Jon Kim

- Security in grids very much concerned with Virtual Organisations
  - use grid resources in coordinated fashion
- Key issues:
  - Provide authentication and authorisation
  - Promote integration with existing systems and technologies

- Grid Security Requirements:
  - Authentication, Delegation, Single logon
  - Credential Lifespan/Renewal
  - Authorisation, Confidentiality, Integrity, Privacy
- State of Play in Grid Security
  - Authent and delegation; authorisation
  - Grid Security Infrastructure
  - Open Grid Services Architecture
- Research topics:
  - Authorisation interoperability, Fine-grained authorisation

Reliability and Security: An application aware approach

Ravi Iyer et al.

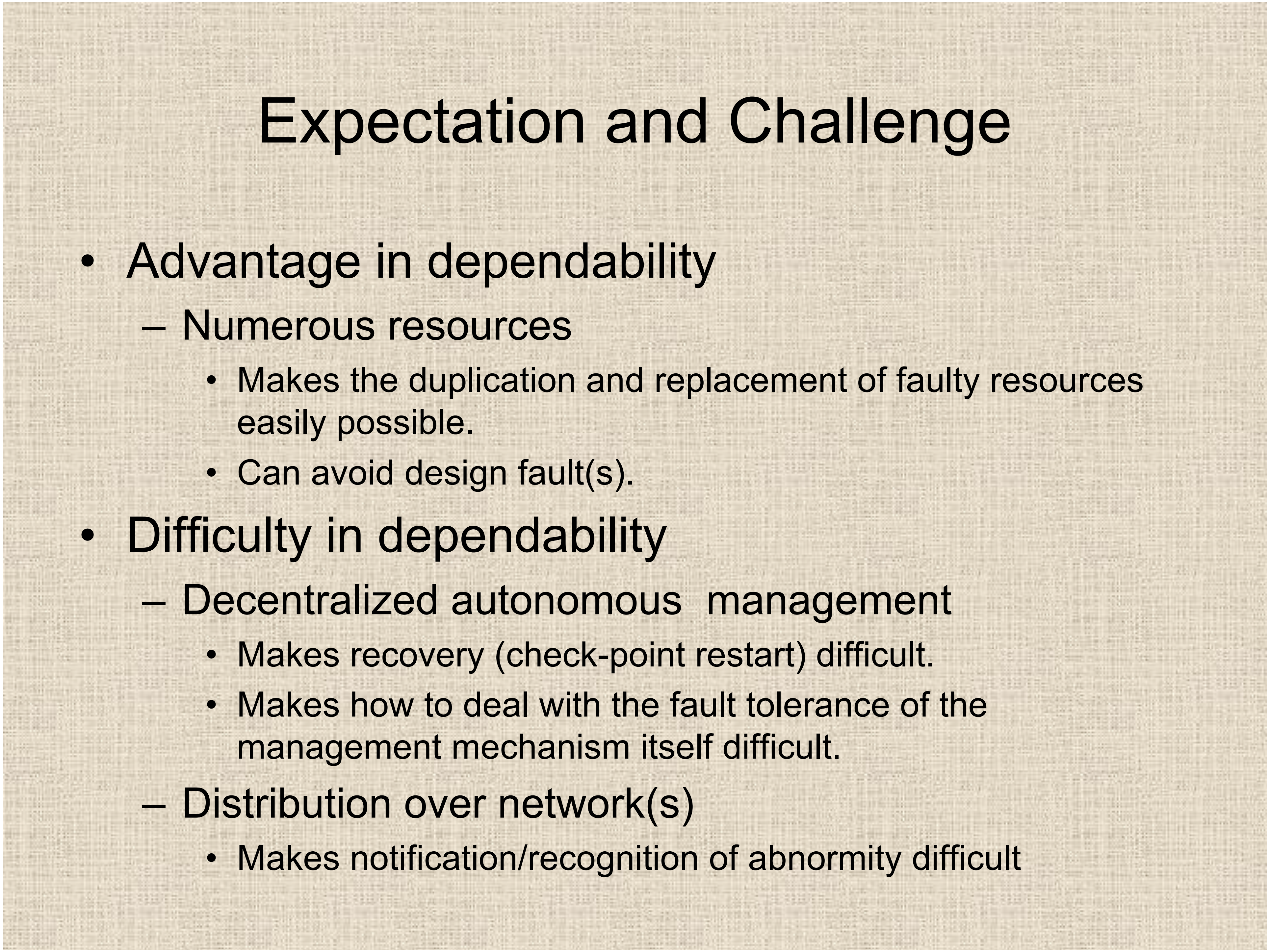
- Crash latency and severity distributions show:
  - Failures are not clean crashes: latency, control flow errors
  - Sometimes the after-failure damage impacts availability (time to restore)

- Solutions:
  - Fine-grained detectors
  - Detector placement strategies
  - Detector semantics: value and time
  - Metrics: e.g., fanout, lifetime, etc.
- Word of caution:
  - Crash in this presentation does not really mean ‘crash’

Security in the grid world

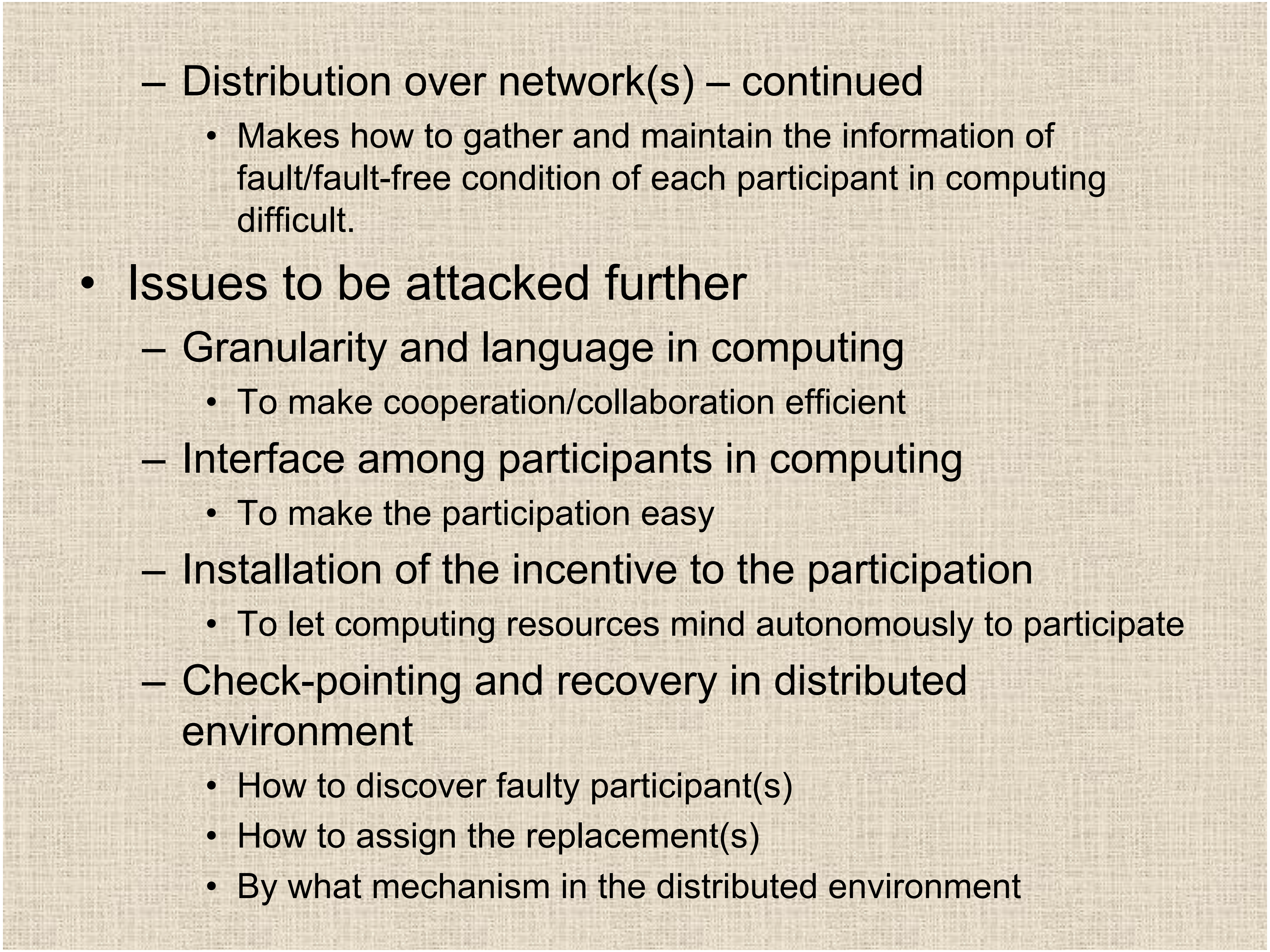
Carl Landwehr

- Perspectives on a model for Grid Security
- or
- How Grid can put zombies out of business...
- Or
- Vice-versa



Expectation and Challenge

- Advantage in dependability
  - Numerous resources
    - Makes the duplication and replacement of faulty resources easily possible.
    - Can avoid design fault(s).
- Difficulty in dependability
  - Decentralized autonomous management
    - Makes recovery (check-point restart) difficult.
    - Makes how to deal with the fault tolerance of the management mechanism itself difficult.
  - Distribution over network(s)
    - Makes notification/recognition of abnormity difficult



- Distribution over network(s) – continued
  - Makes how to gather and maintain the information of fault/fault-free condition of each participant in computing difficult.
- Issues to be attacked further
  - Granularity and language in computing
    - To make cooperation/collaboration efficient
  - Interface among participants in computing
    - To make the participation easy
  - Installation of the incentive to the participation
    - To let computing resources mind autonomously to participate
  - Check-pointing and recovery in distributed environment
    - How to discover faulty participant(s)
    - How to assign the replacement(s)
    - By what mechanism in the distributed environment

- How to deal with intrusion
  - Vulnerability of the distribution over network(s)
- Addition?

Diverse definition of grid

Need of standardization

Need of unifying many grid projects

Heterogeneity and dynamism of resources

Overhead

Licensing

Charging

Need experimental tools for Grid Computing

Proper placement of error detectors can help even in Grid environment

Workshop 2

Nomadic Computing and Dependability

Coordinator

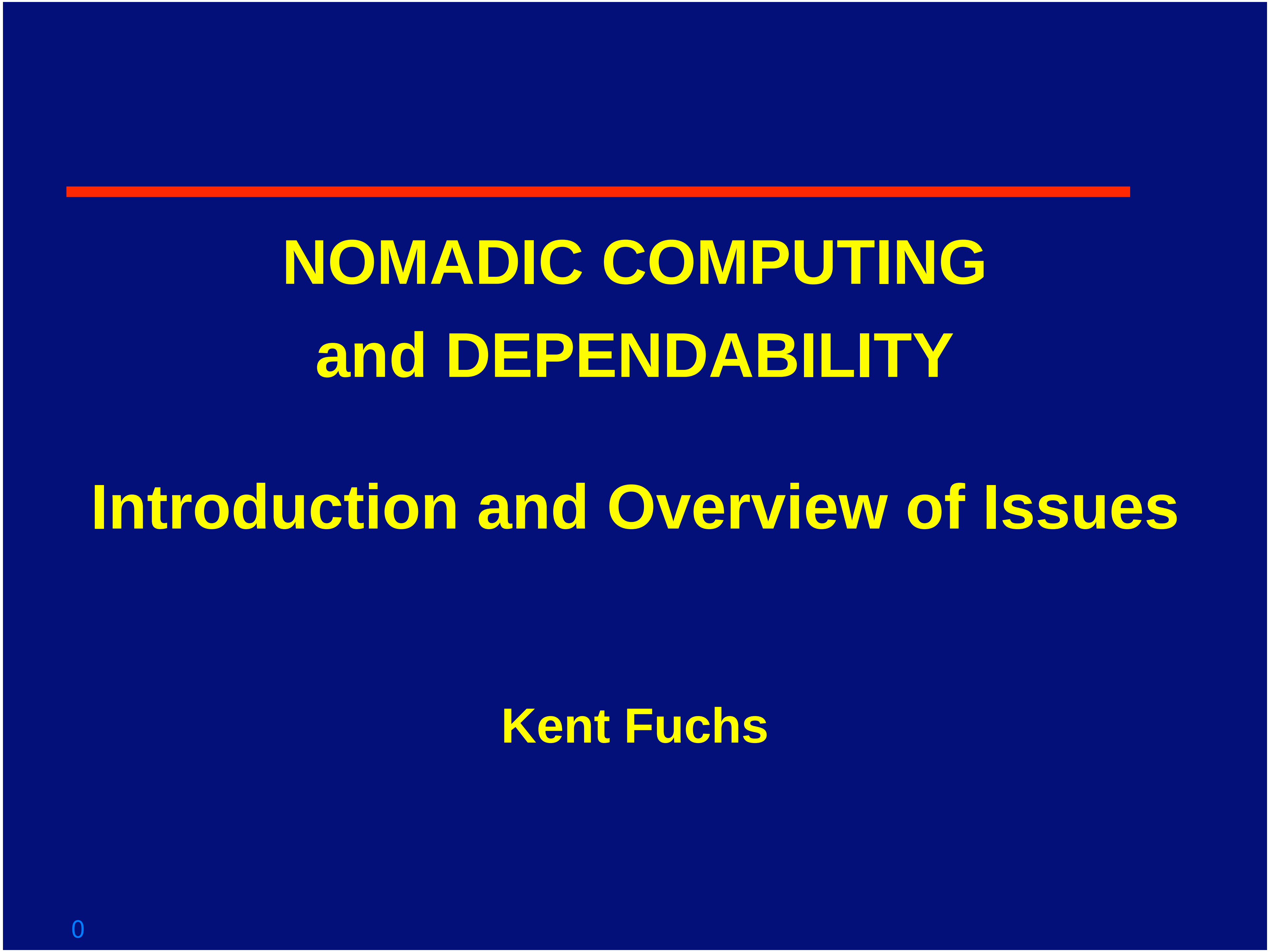
W. Kent Fuchs, Cornell University, Ithaca, NY, USA

Session 2.1

Nomadic Devices and Dependability

Moderator and Rapporteur

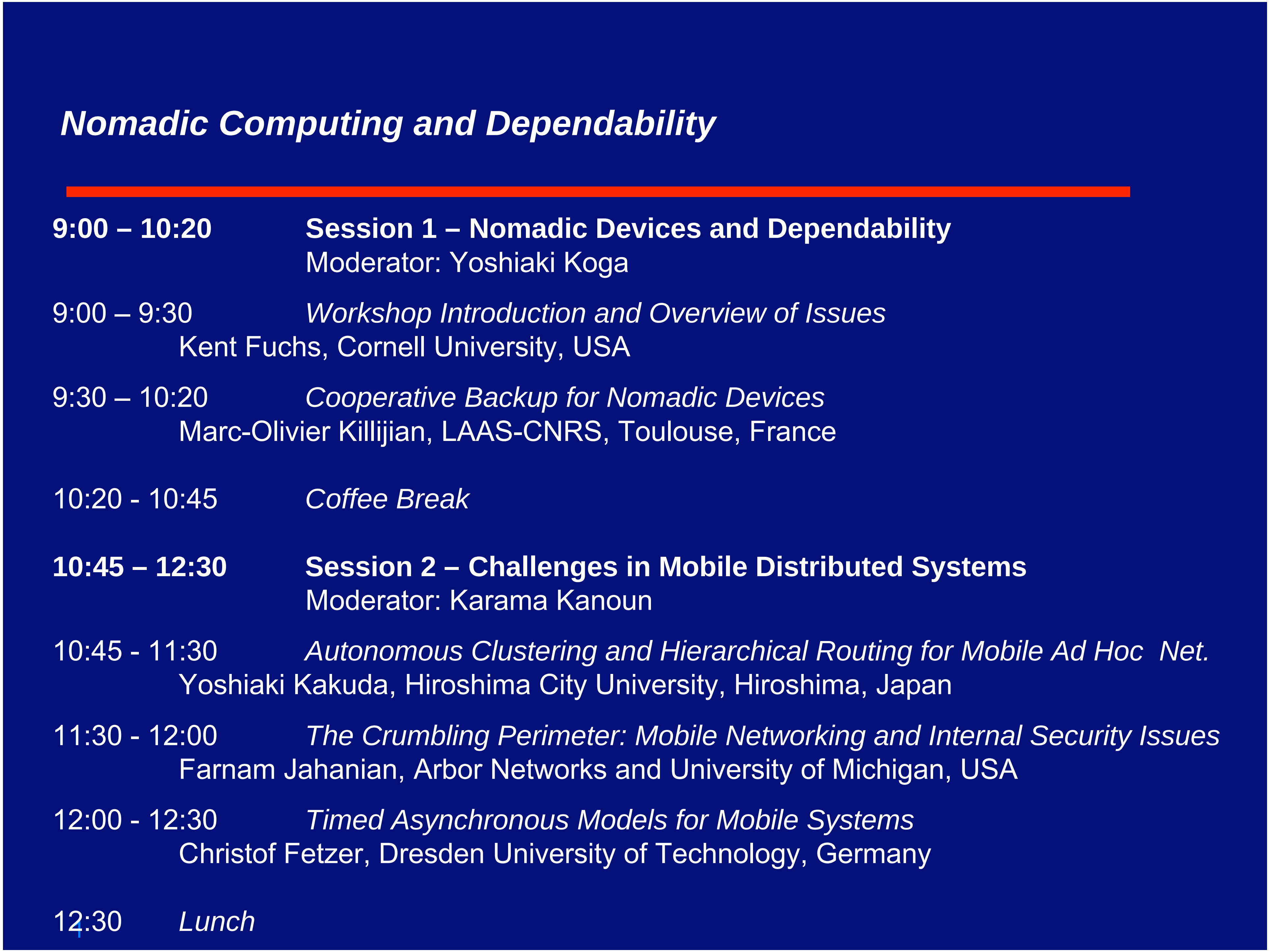
Yoshiaki Koga, Acad. & Educ. Foundation for NDA, Yokohama, Japan



NOMADIC COMPUTING and DEPENDABILITY

Introduction and Overview of Issues

Kent Fuchs



Nomadic Computing and Dependability

9:00 – 10:20 **Session 1 – Nomadic Devices and Dependability**
Moderator: Yoshiaki Koga

9:00 – 9:30 *Workshop Introduction and Overview of Issues*
Kent Fuchs, Cornell University, USA

9:30 – 10:20 *Cooperative Backup for Nomadic Devices*
Marc-Olivier Killijian, LAAS-CNRS, Toulouse, France

10:20 - 10:45 *Coffee Break*

10:45 – 12:30 **Session 2 – Challenges in Mobile Distributed Systems**
Moderator: Karama Kanoun

10:45 - 11:30 *Autonomous Clustering and Hierarchical Routing for Mobile Ad Hoc Net.*
Yoshiaki Kakuda, Hiroshima City University, Hiroshima, Japan

11:30 - 12:00 *The Crumbling Perimeter: Mobile Networking and Internal Security Issues*
Farnam Jahanian, Arbor Networks and University of Michigan, USA

12:00 - 12:30 *Timed Asynchronous Models for Mobile Systems*
Christof Fetzer, Dresden University of Technology, Germany

12:30 *Lunch*

15:30 – 16:45 **Session 3 – Mobility and Ubiquitous Computing**

Moderator: Henrique Madeira

15:30 – 16:15 *A Comprehensive Localization Framework for Self-Organizing Nomadic*
Sys. Emin Gün Sirer, Cornell University, Ithaca, NY, USA

16:15 – 16:45 *A Network Service Provider's View of Ubiquitous Computing*
Rick Schlichting, AT&T Research, Florham Park, NJ, USA

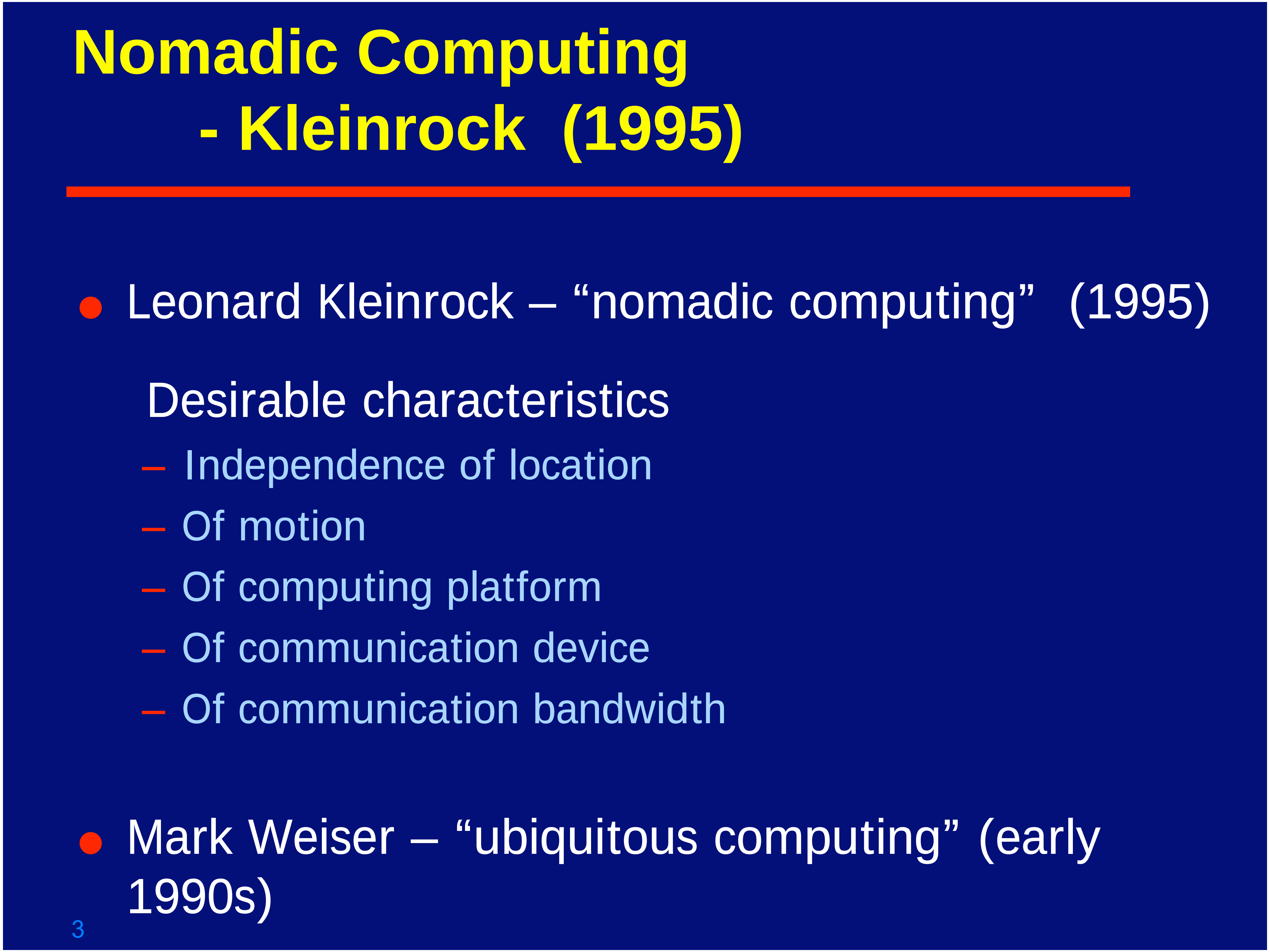
16:45 – 17:10 *Coffee*

17:10 - 17:40 **Session 4 – Synthesis and Wrap Up**

Moderator: Kent Fuchs

- *Reports by Session Moderators*
- *Discussion on Challenges and Expectations*

2

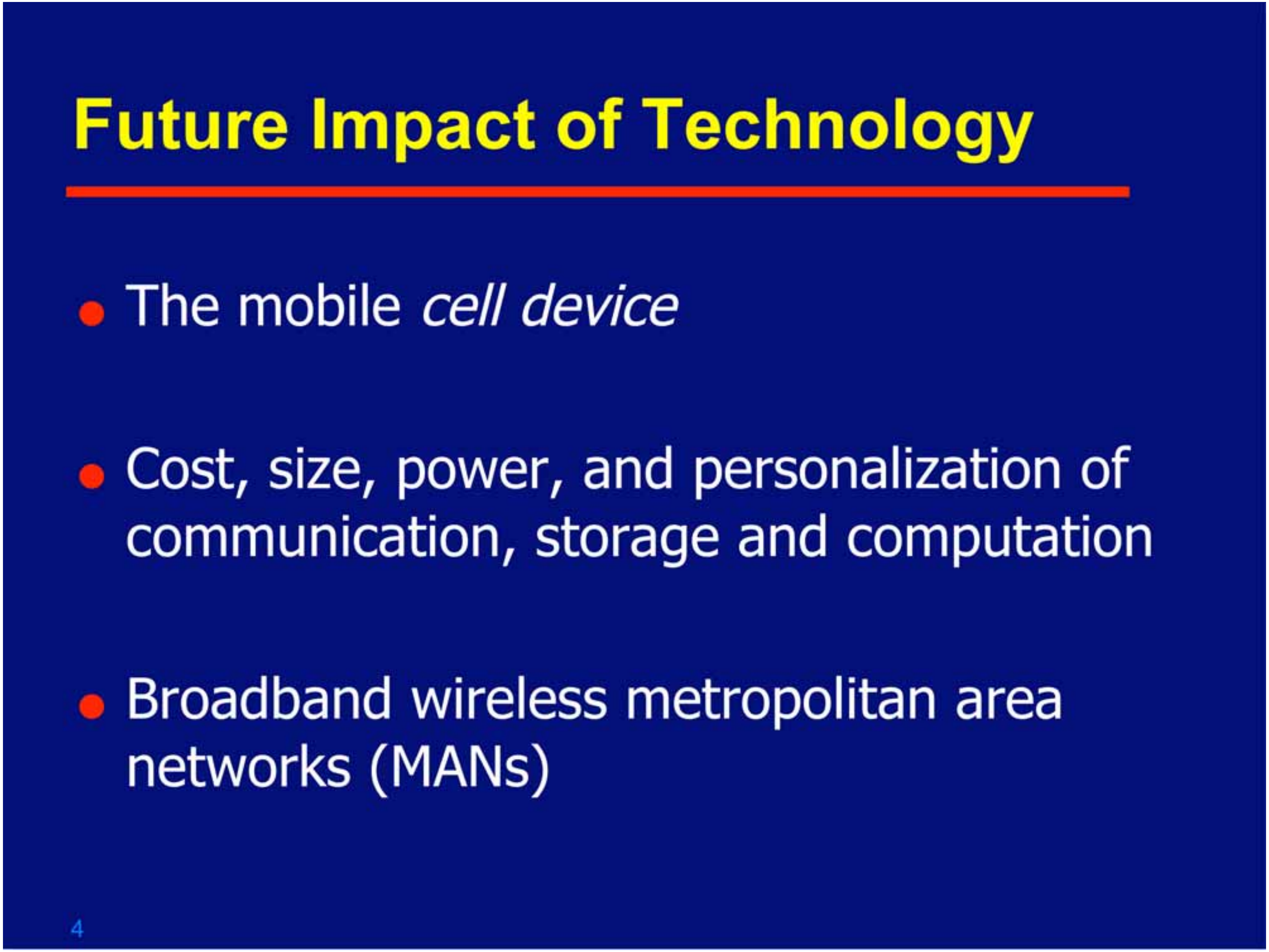


Nomadic Computing - Kleinrock (1995)

- Leonard Kleinrock – “nomadic computing” (1995)

  Desirable characteristics
  - Independence of location
  - Of motion
  - Of computing platform
  - Of communication device
  - Of communication bandwidth

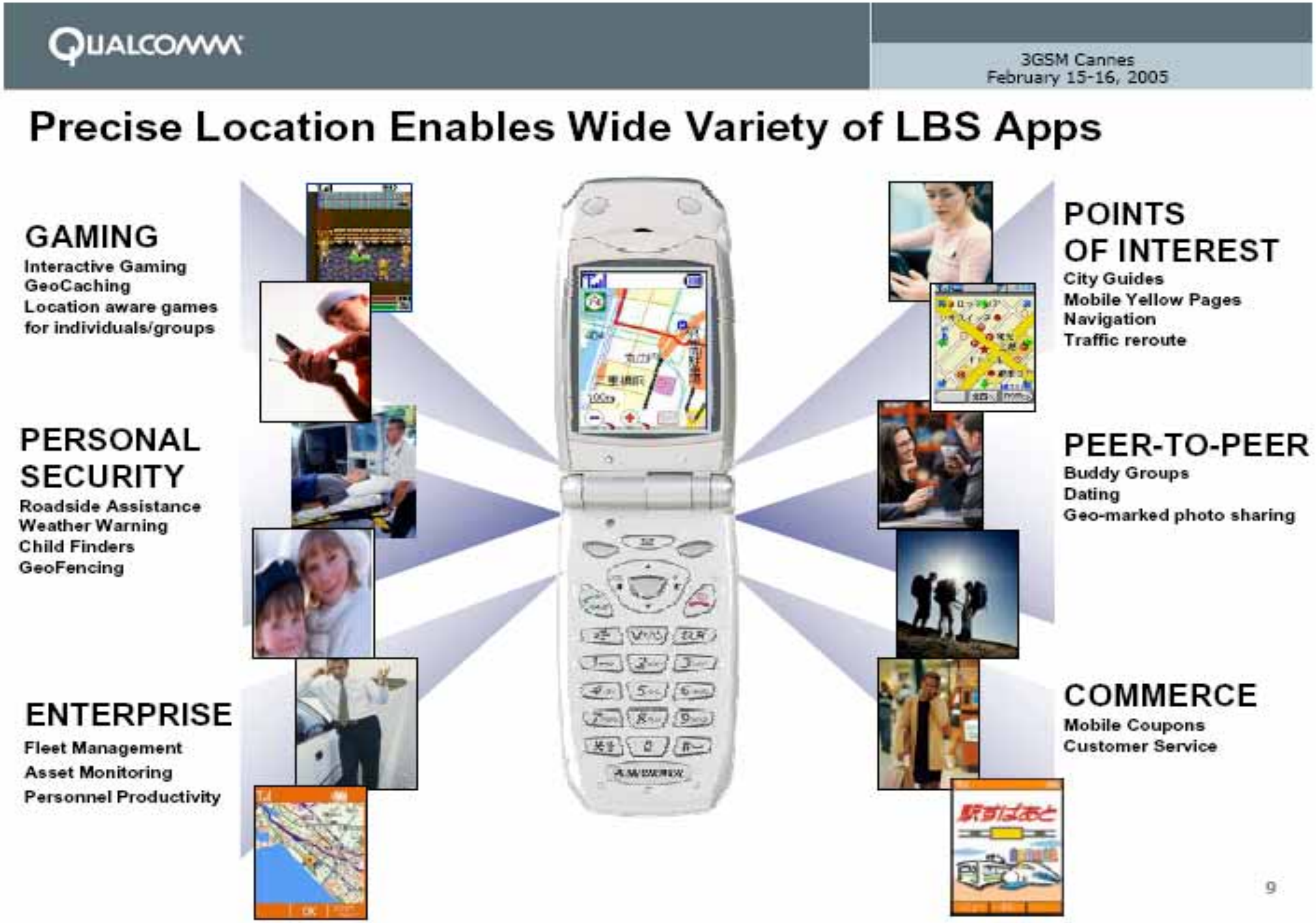
- Mark Weiser – “ubiquitous computing” (early 1990s)



Future Impact of Technology

- The mobile *cell device*
- Cost, size, power, and personalization of communication, storage and computation
- Broadband wireless metropolitan area networks (MANs)

4



QUALCOMM

3GSM Cannes
February 15-16, 2005

Precise Location Enables Wide Variety of LBS Apps

GAMING
Interactive Gaming
GeoCaching
Location aware games for individuals/groups

PERSONAL SECURITY
Roadside Assistance
Weather Warning
Child Finders
GeoFencing

ENTERPRISE
Fleet Management
Asset Monitoring
Personnel Productivity

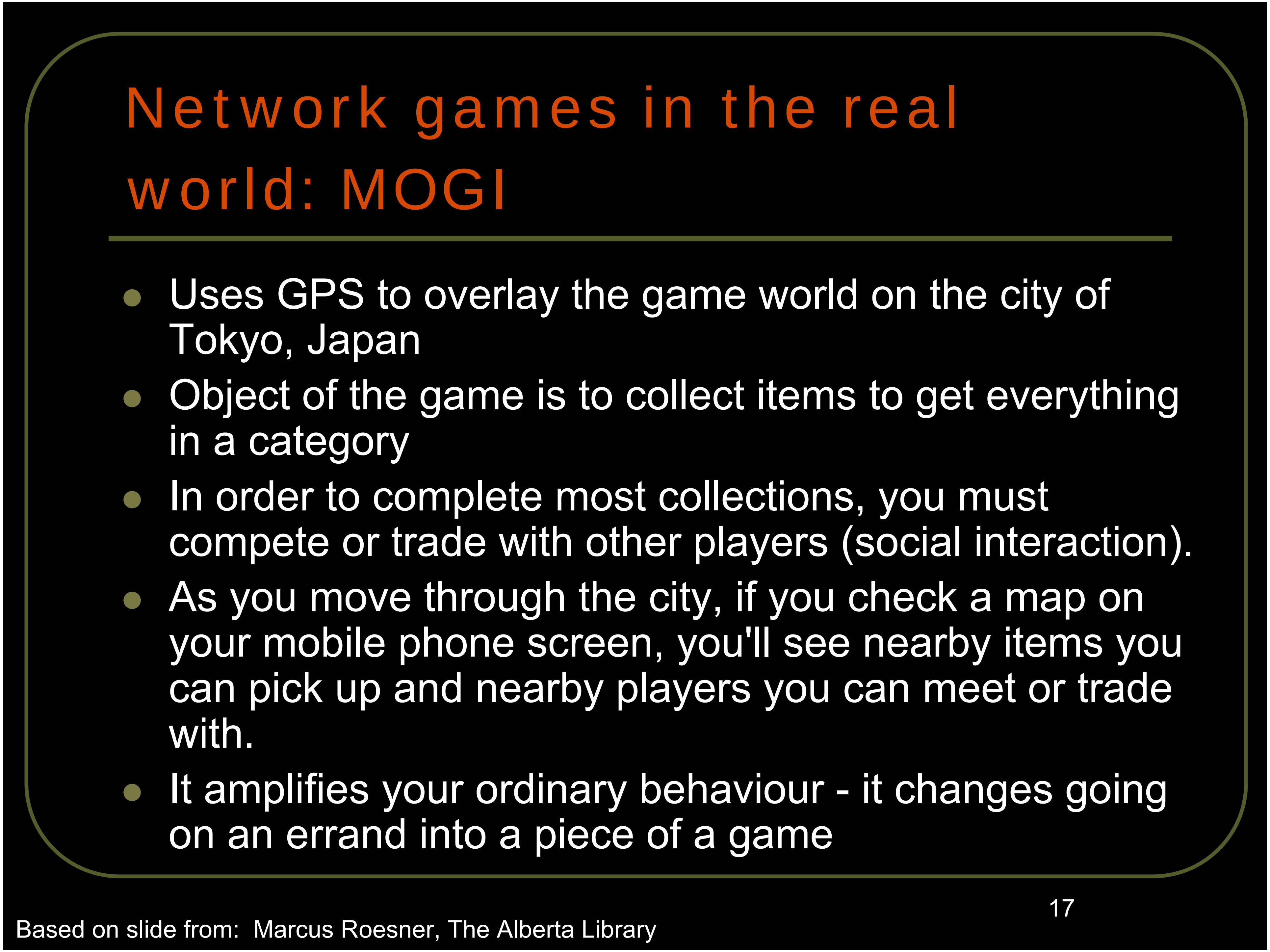
POINTS OF INTEREST
City Guides
Mobile Yellow Pages
Navigation
Traffic reroute

PEER-TO-PEER
Buddy Groups
Dating
Geo-marked photo sharing

COMMERCE
Mobile Coupons
Customer Service

9

Slide from: Qualcomm, 3GSM Cannes, February, 2005



Network games in the real world: MOGI

- Uses GPS to overlay the game world on the city of Tokyo, Japan
- Object of the game is to collect items to get everything in a category
- In order to complete most collections, you must compete or trade with other players (social interaction).
- As you move through the city, if you check a map on your mobile phone screen, you'll see nearby items you can pick up and nearby players you can meet or trade with.
- It amplifies your ordinary behaviour - it changes going on an errand into a piece of a game

17

Based on slide from: Marcus Roesner, The Alberta Library

www.mogimogi.com

18

Based on slide from: Marcus Roesner, The Alberta Library

Annual Meeting of Stockholders, March 8, 2005

Convergence With Consumer Electronics

Voice
Walkie-Talkie
PDA
Television
Camcorder
Photo Album
Glucometer
Camera
Wallet
FM Radio
Bar Scanner
Game Console
PC
MP3 Player
Newspaper
GPS Device
Rolodex
Pager

Slide from: Qualcomm, Annual Meeting of Stockholders, March, 2005

More *nomadic and smart storage*.

21

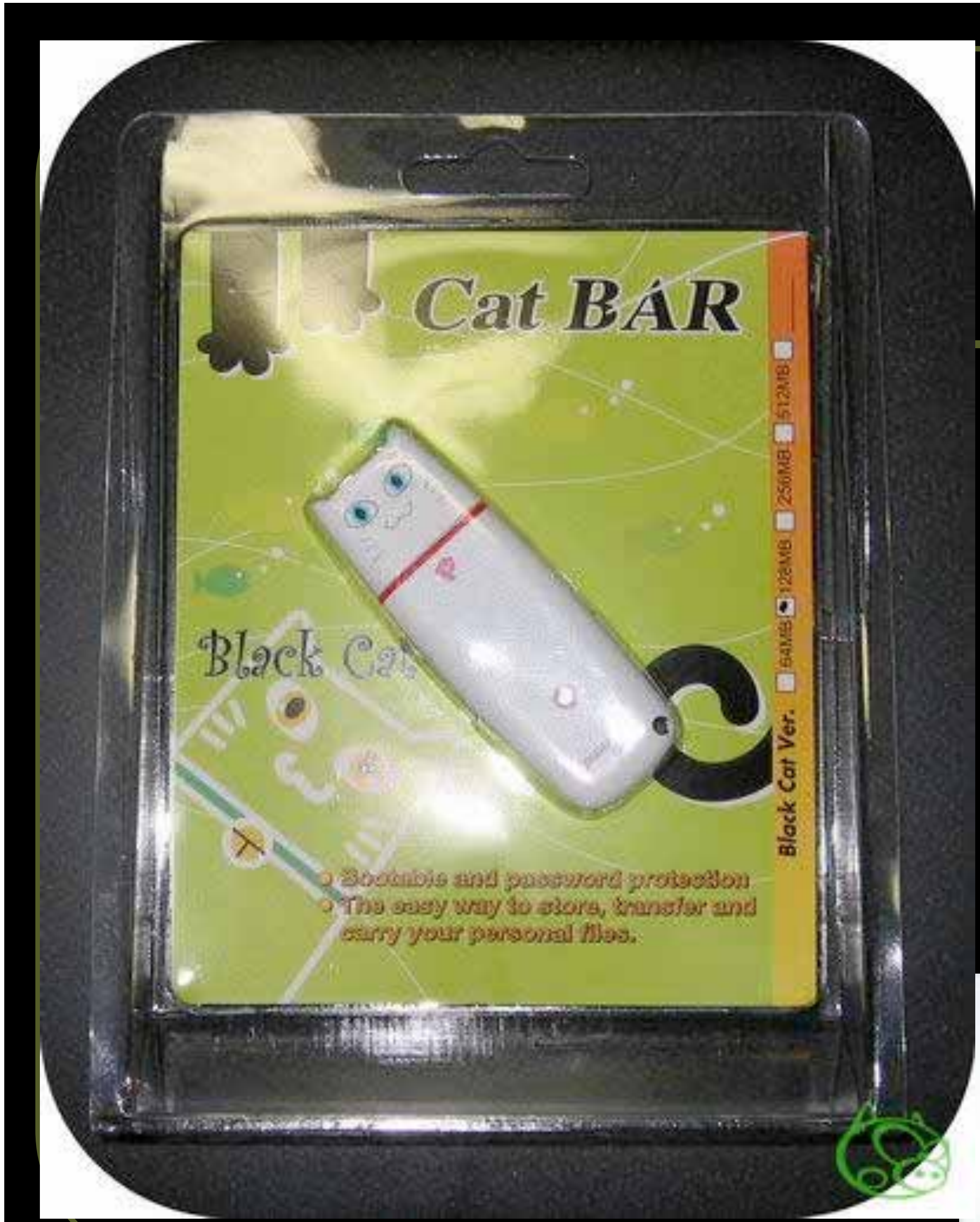
Based on slide from: Marcus Roesner, The Alberta Library

This functional pen not only has 128 MB of *storage* but also has a USB *connection* and a *connector* for SD memory cards.

Or get the one that adds in an MP3 Player.

22

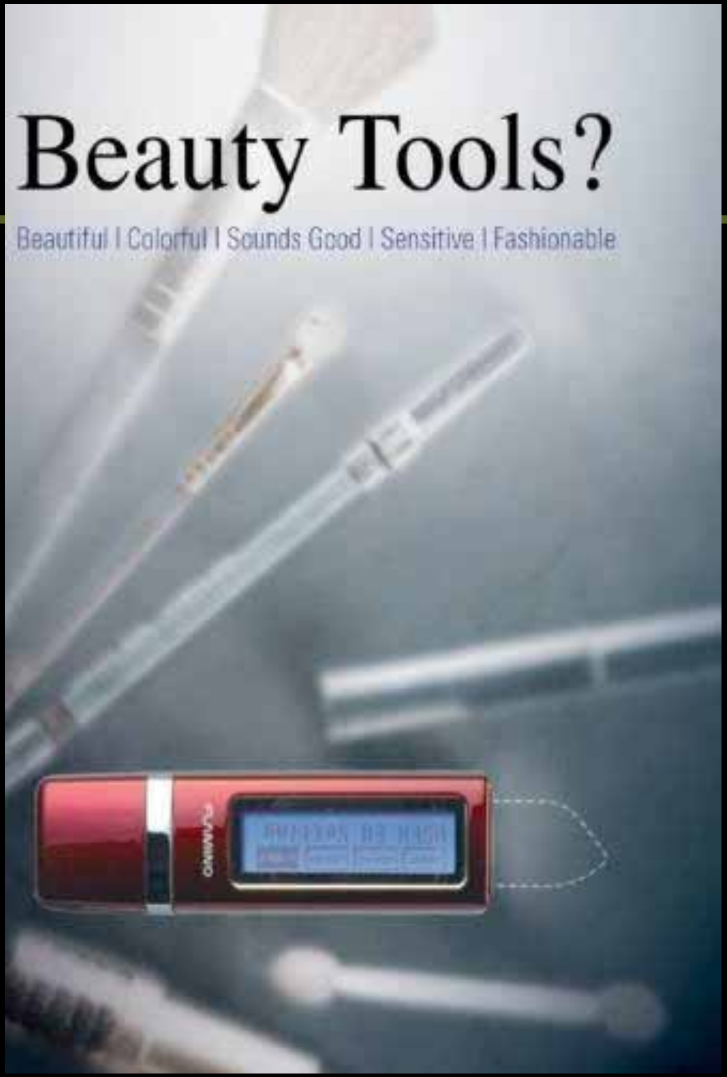
Slide from: Marcus Roesner, The Alberta Library


Cat BAR
Black Cat
Bootable and password protection
The easy way to store, transfer and carry your personal files.
Black Cat Ver. □ 64MB ☑ 128MB □ 256MB □ 512MB

A Fun USB Memory Stick (Portable *Storage*)

Based on slide from: Marcus Roesner, The Alberta Library

23

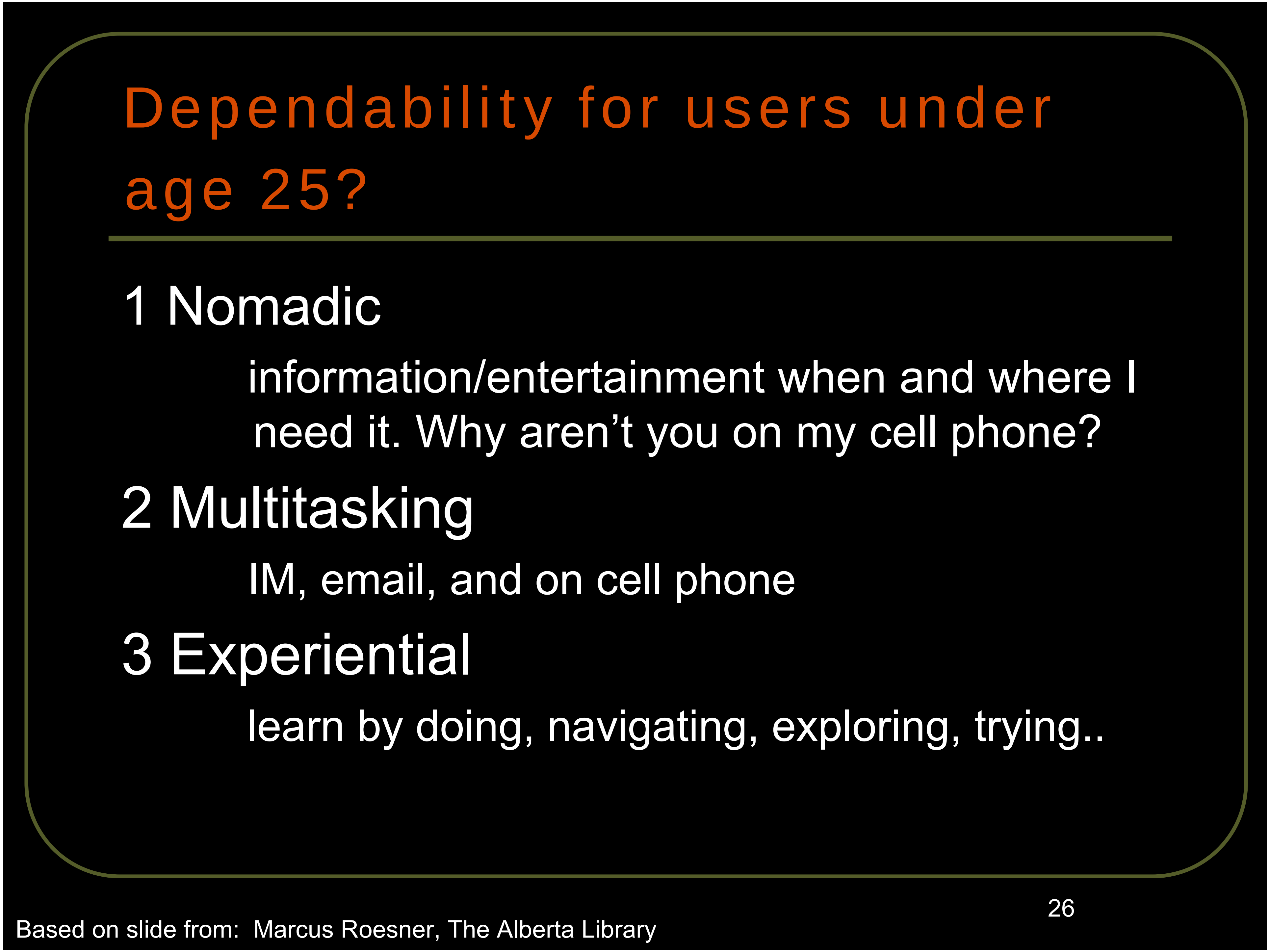

Beauty Tools?
Beautiful | Colorful | Sounds Good | Sensitive | Fashionable
FLAMINGO

24

Based on slide from: Marcus Roesner, The Alberta Library

Personalization

Based on slide from: Marcus Roesner, The Alberta Library



Dependability for users under age 25?

1 Nomadic

information/entertainment when and where I need it. Why aren't you on my cell phone?

2 Multitasking

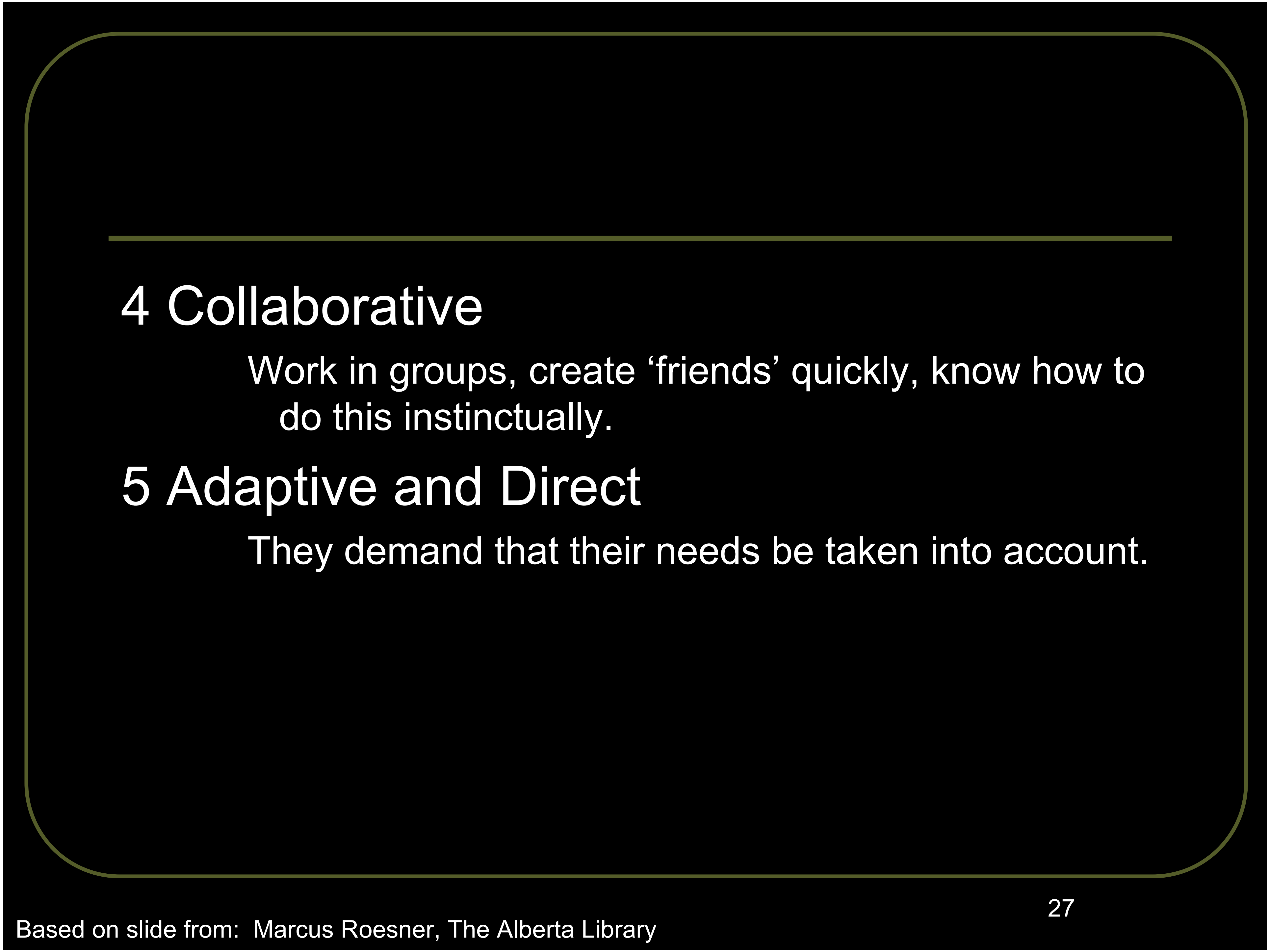
IM, email, and on cell phone

3 Experiential

learn by doing, navigating, exploring, trying..

26

Based on slide from: Marcus Roesner, The Alberta Library



4 Collaborative

Work in groups, create ‘friends’ quickly, know how to do this instinctually.

5 Adaptive and Direct

They demand that their needs be taken into account.

Based on slide from: Marcus Roesner, The Alberta Library

CMP
Chew On This Wrigley looks to integrate business processes with SAP PAGE 18
Data Warehouse Redux Real-time access means more strategic role PAGE 20
Cutting Edge America Online begins Linux blade-server test PAGE 22
InformationWeek
Business Innovation Powered By Technology
Utility Computing
THOUGH MOMENTUM BUILDS, NO BIG-BANG ACCEPTANCE FOR AMBITIOUS TECH MODEL
40
The 5¢ Solution
WAL-MART, PROCTER & GAMBLE, OTHERS LOOK TO CHEAP RADIO-I.D. CHIPS (ABOVE, IN VIAL) TO REVOLUTIONIZE THEIR SUPPLY C
16
$10B Welcome Massive integration is first step in border ID program
Offshore Arguments The politics of outsourcing heat up
Time To Deliver FedEx aims for growth, but not in its IT budget
InformationWeek
Business Innovation Powered By Technology
Jan. 12, 2004
RFID MEETS MAD COW
The livestock industry is counting on a nearly $600 million identification project to help track animals in hours or even mi
18
Will It Fly? Boeing project involves new ways to collaborate
Soft Drinks & Software Coke bottler teams with SAP to refresh supply chain
Trading Up Exchanges foresee advantages from SEC mandate
InformationWeek
Business Innovation Powered By Technology
Feb. 16, 2004
RFID: Now's The Time
PLAN FOR A BOOM IN DATA VOLUME
30
FACE UP TO GROWING PRIVACY CONCERNS
39

www.rfidinsights.com

RFID and Wal-Mart

- Wal-Mart now has 100+ suppliers shipping cases and pallets with RFID tags.
- Wal-Mart is scheduled to expand its RFID initiative to 12 distribution centers and 600 stores by end of 2005.
- In January 2005, Wal-Mart has installed RFID equipment in 104 stores.
- By the beginning of 2006, Wal-Mart's top 300 suppliers will be required to tag cases and pallets of selected products with RFID tags. By the end of 2006, the retailer expects its entire supplier base (up to 20,000 suppliers) to be "engaged in RFID in some form or fashion."
- Deploying RFID equipment across 35 distribution centers and approximately 1,300 retail outlets by Fall 2005.

Slide from Anita Campbell, University of Akron

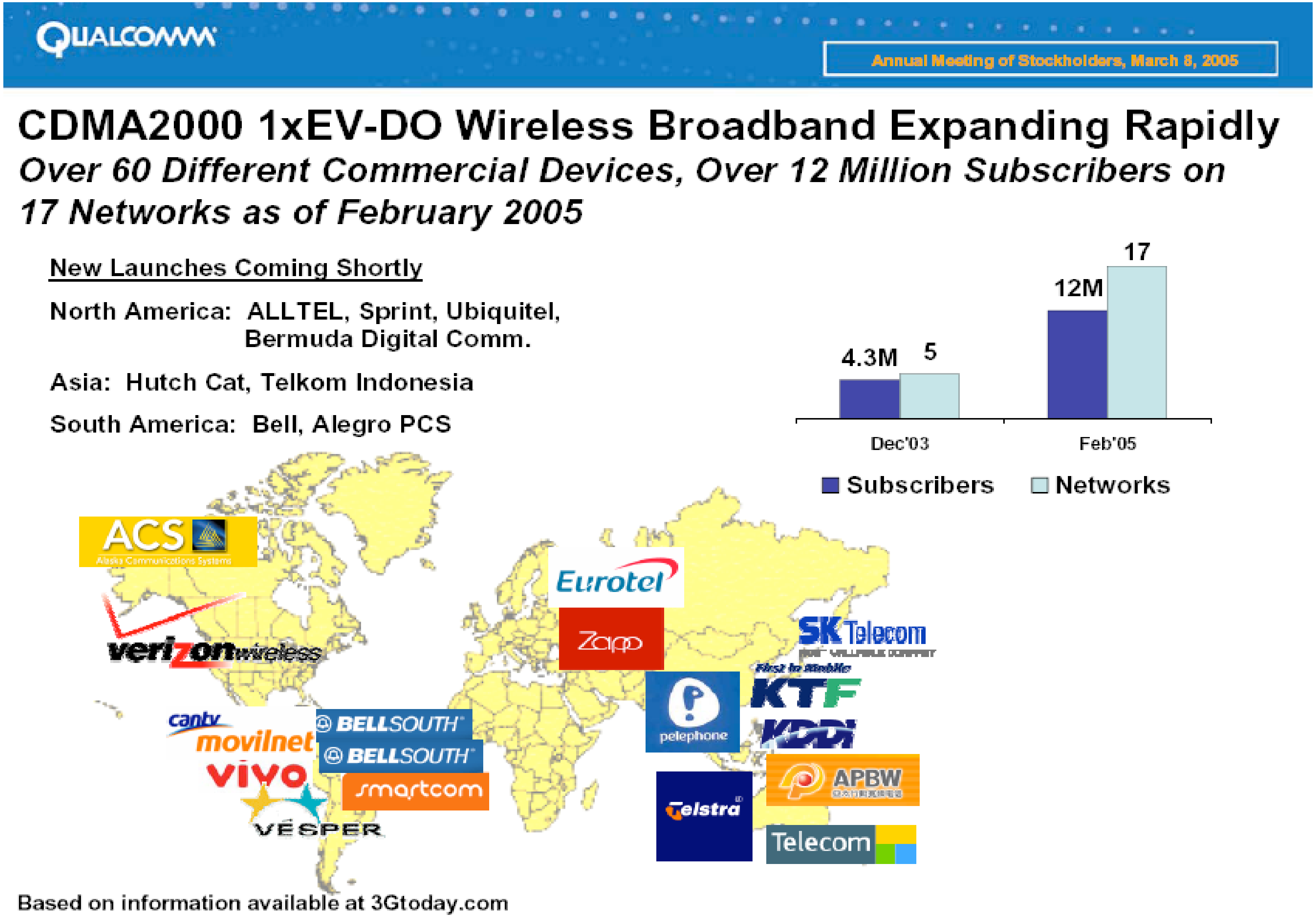
Issue: Privacy concerns

- □ Item level tagging
- □ Tagging people

"Mark of the Beast"

privacy.org
the site for news, information, and action

Based on slide from Anita Campbell, University of Akron



QUALCOMM

Annual Meeting of Stockholders, March 8, 2005

CDMA2000 1xEV-DO Wireless Broadband Expanding Rapidly

Over 60 Different Commercial Devices, Over 12 Million Subscribers on 17 Networks as of February 2005

New Launches Coming Shortly

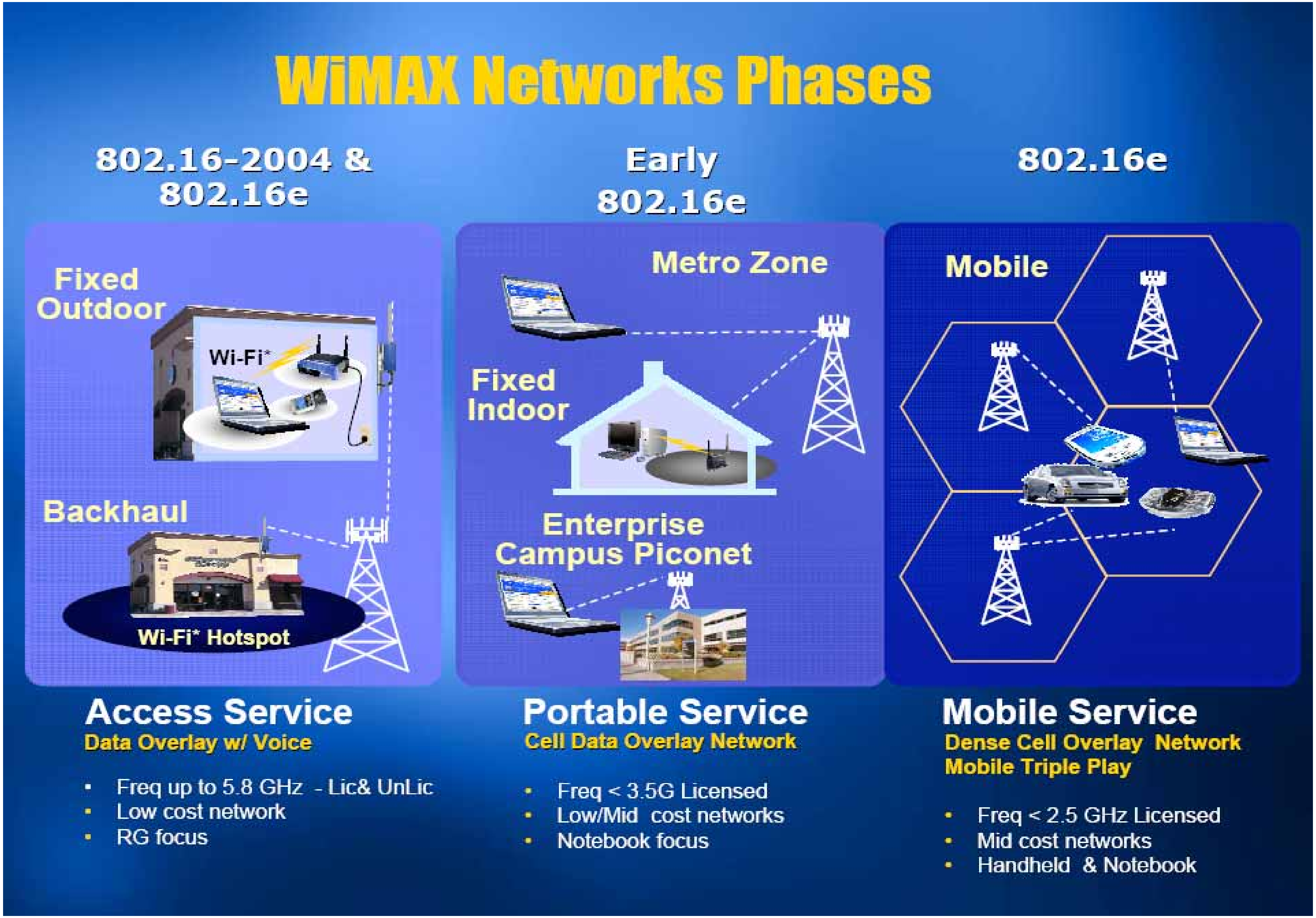
North America: ALLTEL, Sprint, Ubiquitel,
Bermuda Digital Comm.

Asia: Hutch Cat, Telkom Indonesia

South America: Bell, Alegro PCS

Based on information available at 3Gtoday.com

Slide from: Qualcomm, Annual Meeting of Stockholders, March, 2005



WiMAX Networks Phases

802.16-2004 & 802.16e

Early 802.16e

802.16e

Fixed Outdoor

Wi-Fi*

Backhaul

Wi-Fi* Hotspot

Metro Zone

Fixed Indoor

Enterprise Campus Piconet

Mobile

Access Service

Data Overlay w/ Voice

- Freq up to 5.8 GHz - Lic& UnLic
- Low cost network
- RG focus

Portable Service

Cell Data Overlay Network

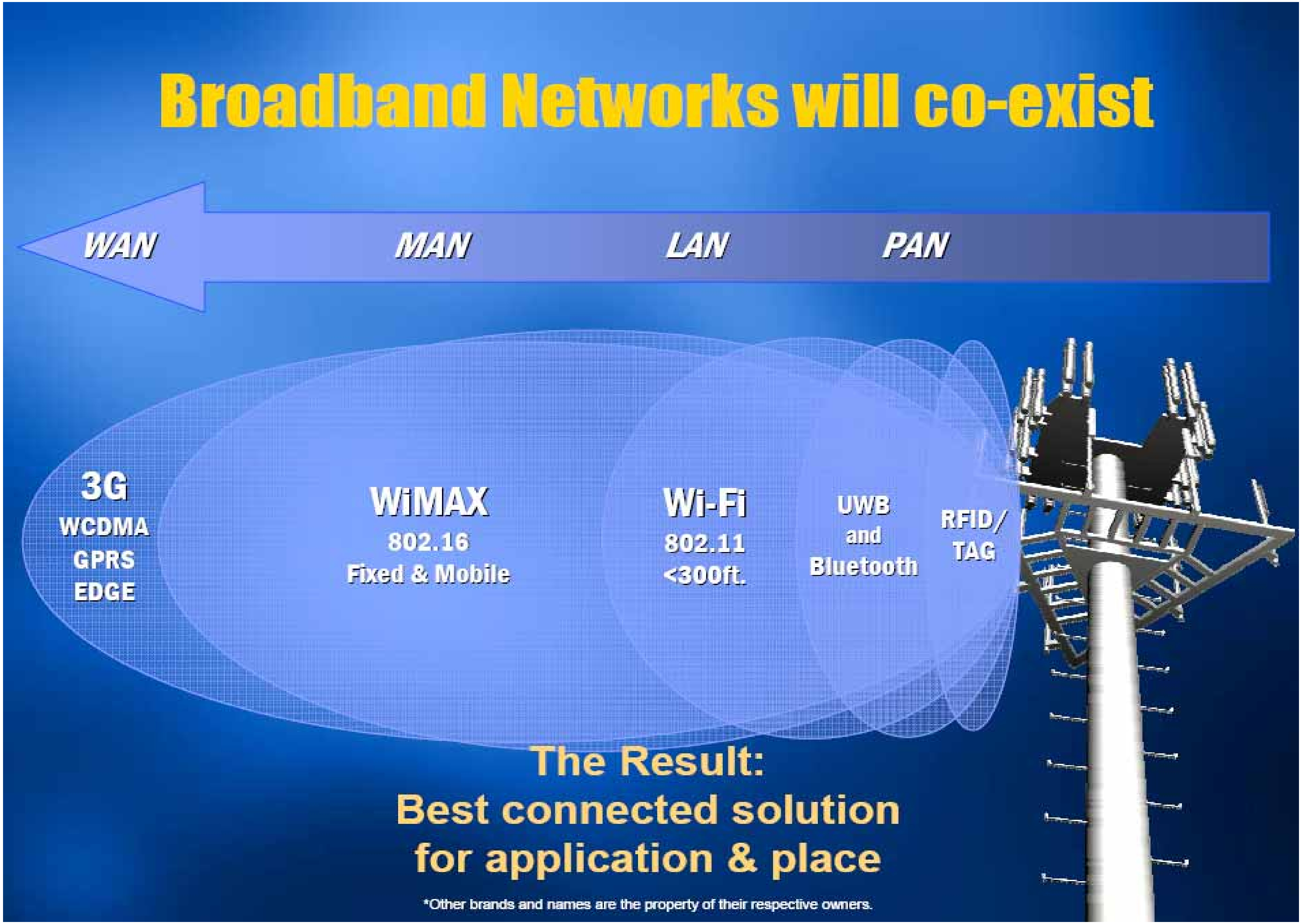
- Freq < 3.5G Licensed
- Low/Mid cost networks
- Notebook focus

Mobile Service

Dense Cell Overlay Network
Mobile Triple Play

- Freq < 2.5 GHz Licensed
- Mid cost networks
- Handheld & Notebook

Slide from Julie Coppernoll, Intel



Broadband Networks will co-exist

WAN MAN LAN PAN

3G
WCDMA
GPRS
EDGE

WiMAX
802.16
Fixed & Mobile

Wi-Fi
802.11
<300ft.

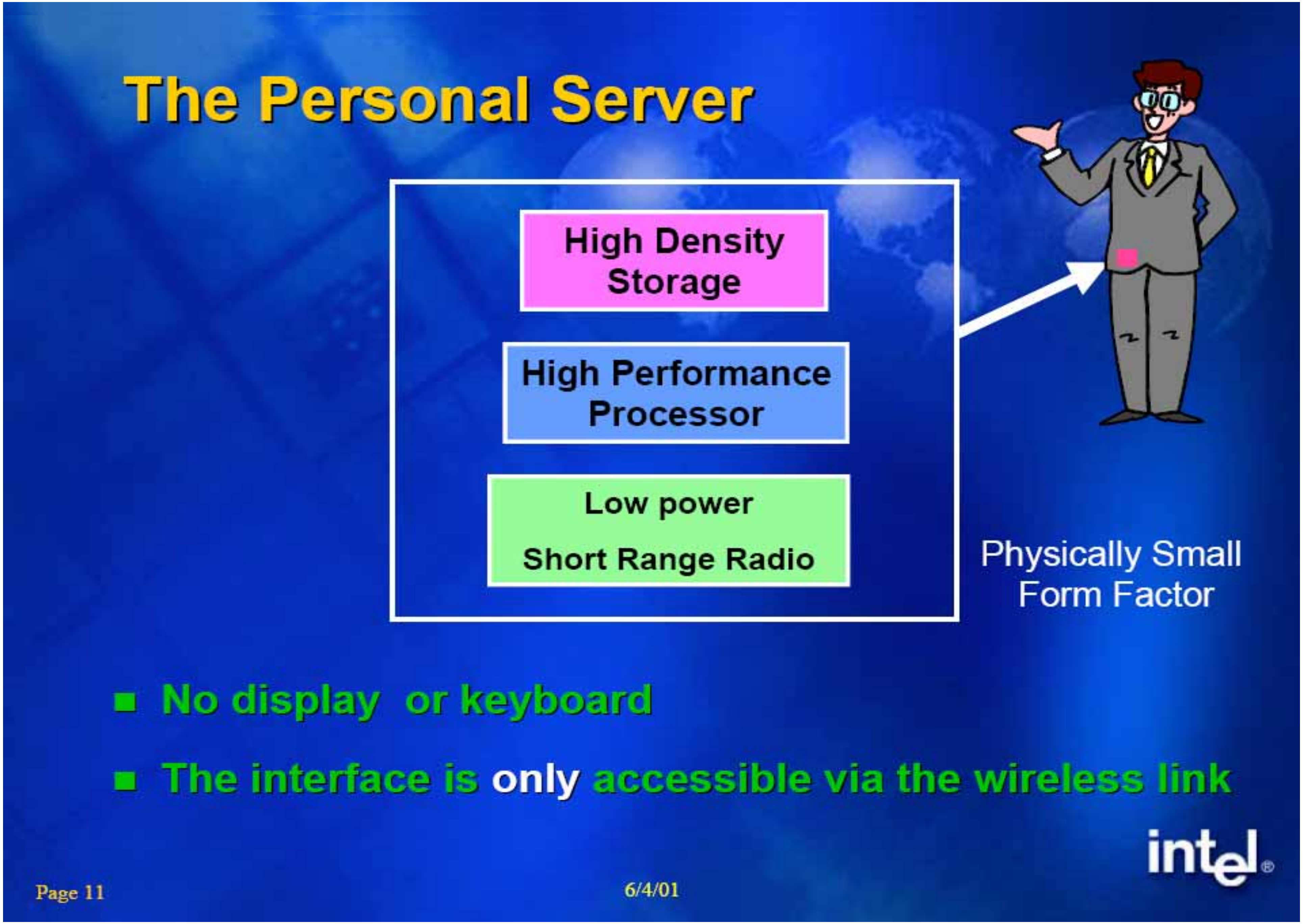
UWB
and
Bluetooth

RFID/
TAG

The Result:
Best connected solution
for application & place

*Other brands and names are the property of their respective owners.

Slide from Julie Coppernoll, Intel



The Personal Server

High Density Storage

High Performance Processor

Low power

Short Range Radio

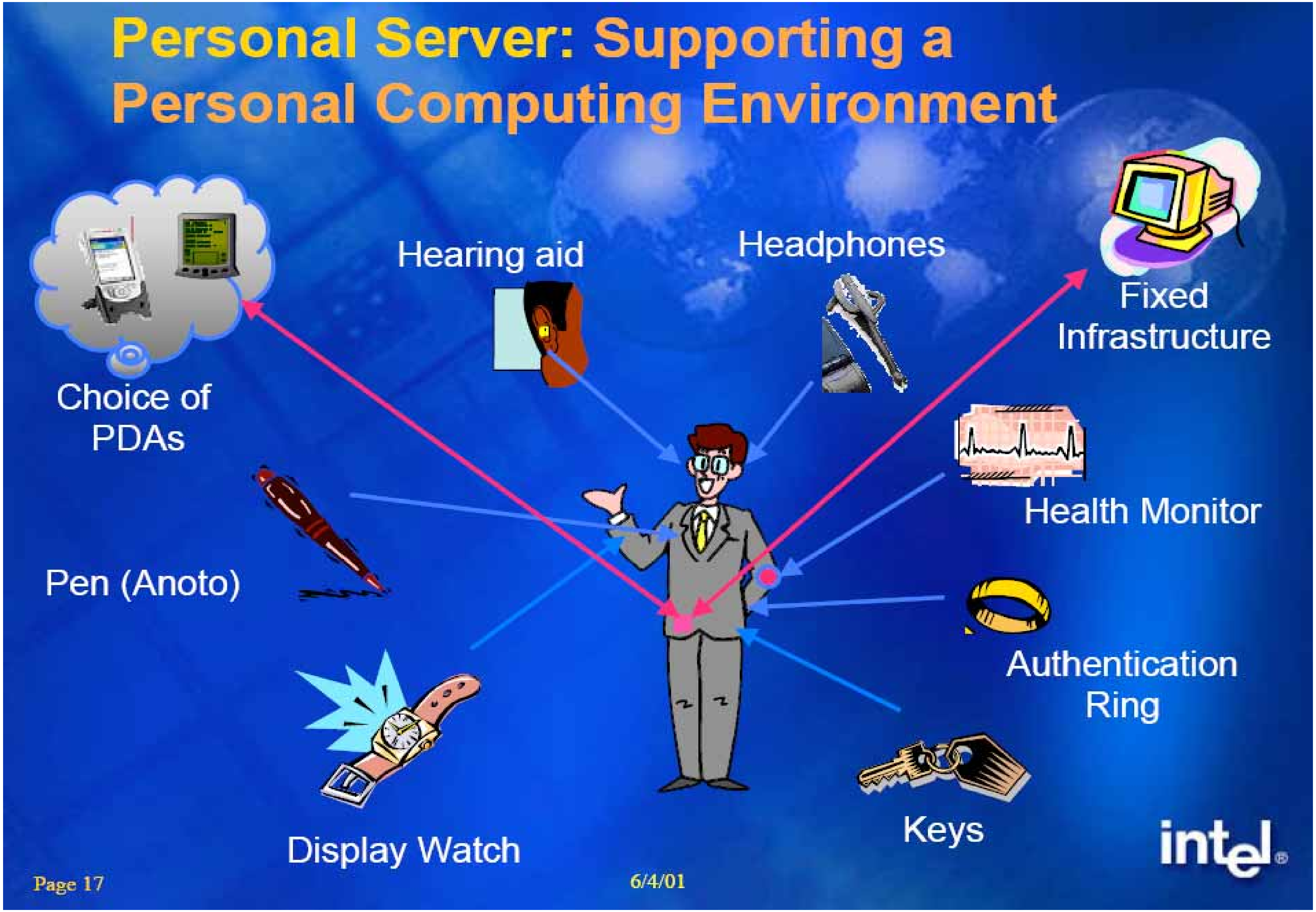
Physically Small Form Factor

- No display or keyboard
- The interface is only accessible via the wireless link

Page 11

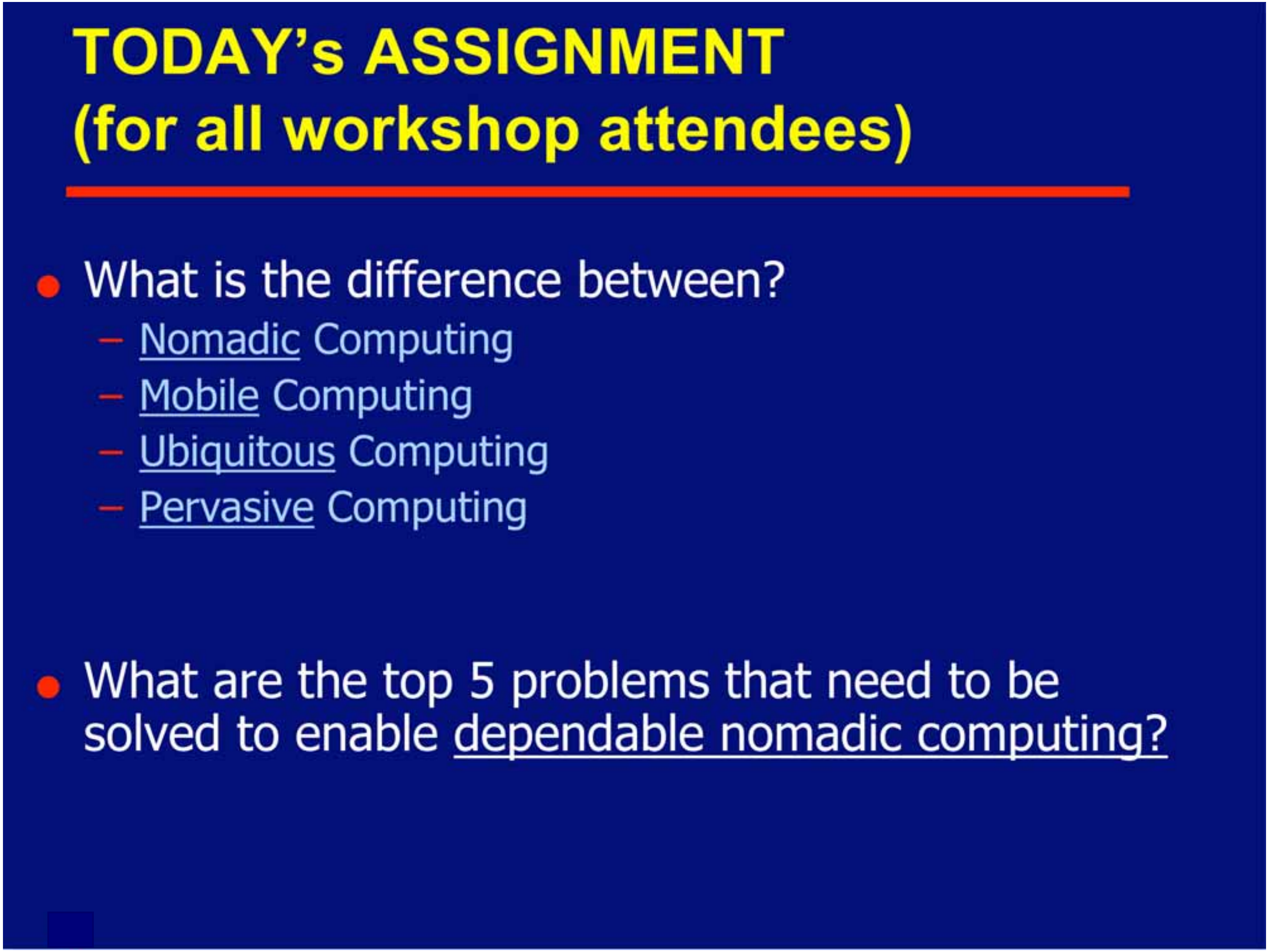
6/4/01

Slide from Roy Want, Intel



Personal Server: Supporting a Personal Computing Environment

Slide from Roy Want, Intel



TODAY's ASSIGNMENT (for all workshop attendees)

- What is the difference between?
  - Nomadic Computing
  - Mobile Computing
  - Ubiquitous Computing
  - Pervasive Computing

- What are the top 5 problems that need to be solved to enable dependable nomadic computing?

Collaborative Backup for Nomadic Devices

M.O.Killijian, D.Powell

Context

- The MoSAIC project
  - Mobile Systems Availability Integrity and Confidentiality
- 3 years, 3 partners: LAAS, Eurécom, IRISA
  - Officially started September 2004
  - Funded by French Ministry of Research
- Nomadic device scenario
  - Mostly disconnected operations
  - Opportunistic wireless communication with similar devices
  - Peer-to-peer model of interactions

- Secure Collaborative Backup for Nomadic Devices

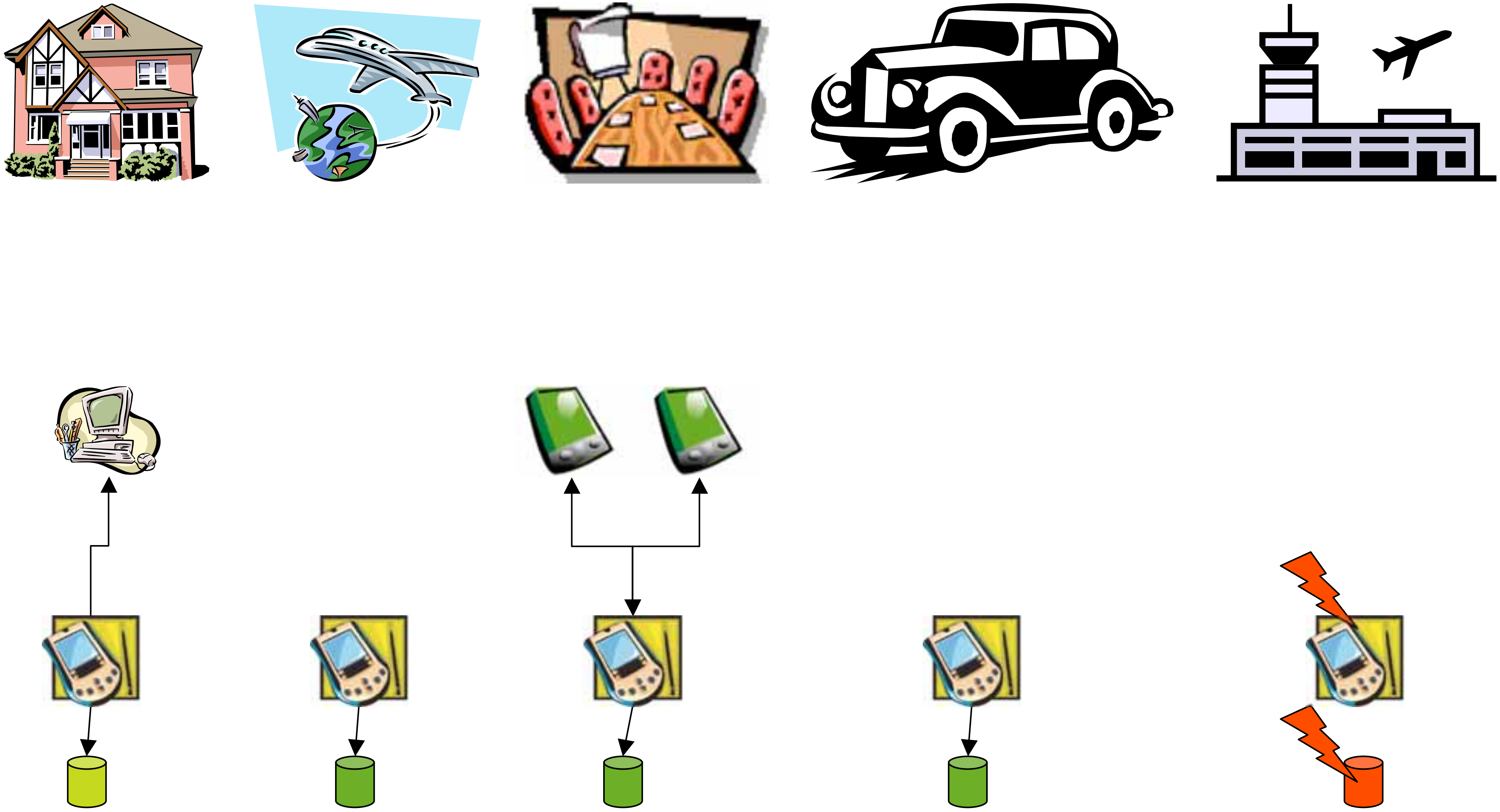
MoSAIC Goals

- In this context
  - new distributed algorithms and mechanisms for the tolerance of
    - accidental faults
    - malicious faults
  - without usual strong assumptions
    - synchronous communication
    - global clocks
    - Infrastructure
- New middleware for dependable mobile systems

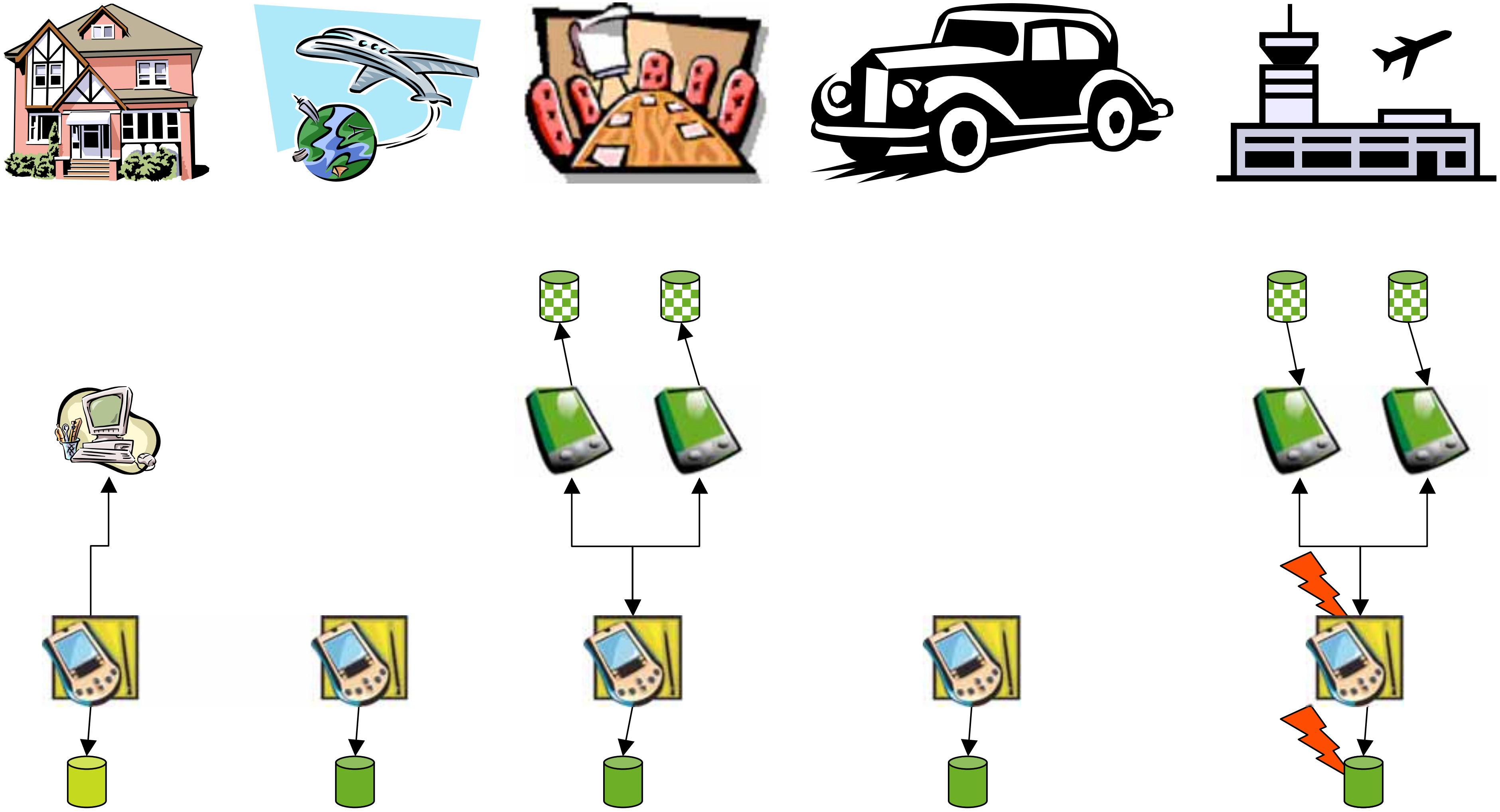
Overview

- **Overview of MoSAIC project**
- Collaborative Backup Systems
- Trust Management
- Current Status

Scenario without MoSAIC



Scenario with MoSAIC

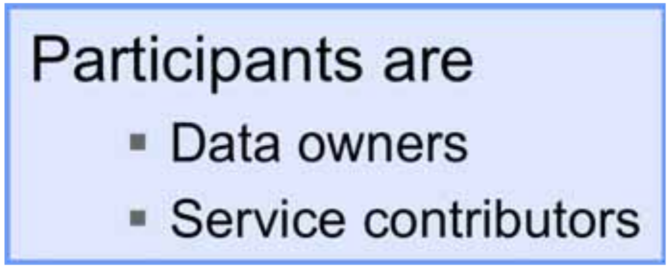


Challenges for Dependability

- Limited energy, computation and storage
- Only intermittent access to a fixed infrastructure
- No prior organization
- Ephemeral interactions
- Critical private data

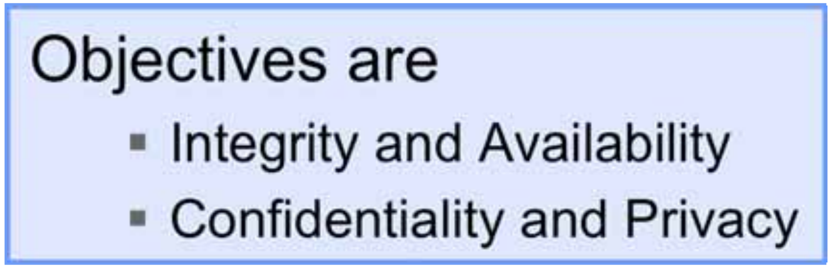
+ Usual criteria for classic functionalities
  - User transparency
  - Usability
  - etc.

Collaborative Backup



Participants are

- Data owners
- Service contributors



Objectives are

- Integrity and Availability
- Confidentiality and Privacy

Potential faults are

- Permanent and transient faults affecting a data owner
- Theft or loss of a data owner

- Accidental or malicious faults affecting availability of data backups
- Accidental or malicious modification of data backups
- Malicious read access to data backups

- Malicious denial of service (sabotage)
- Selfish denial of service (refusal to cooperate)

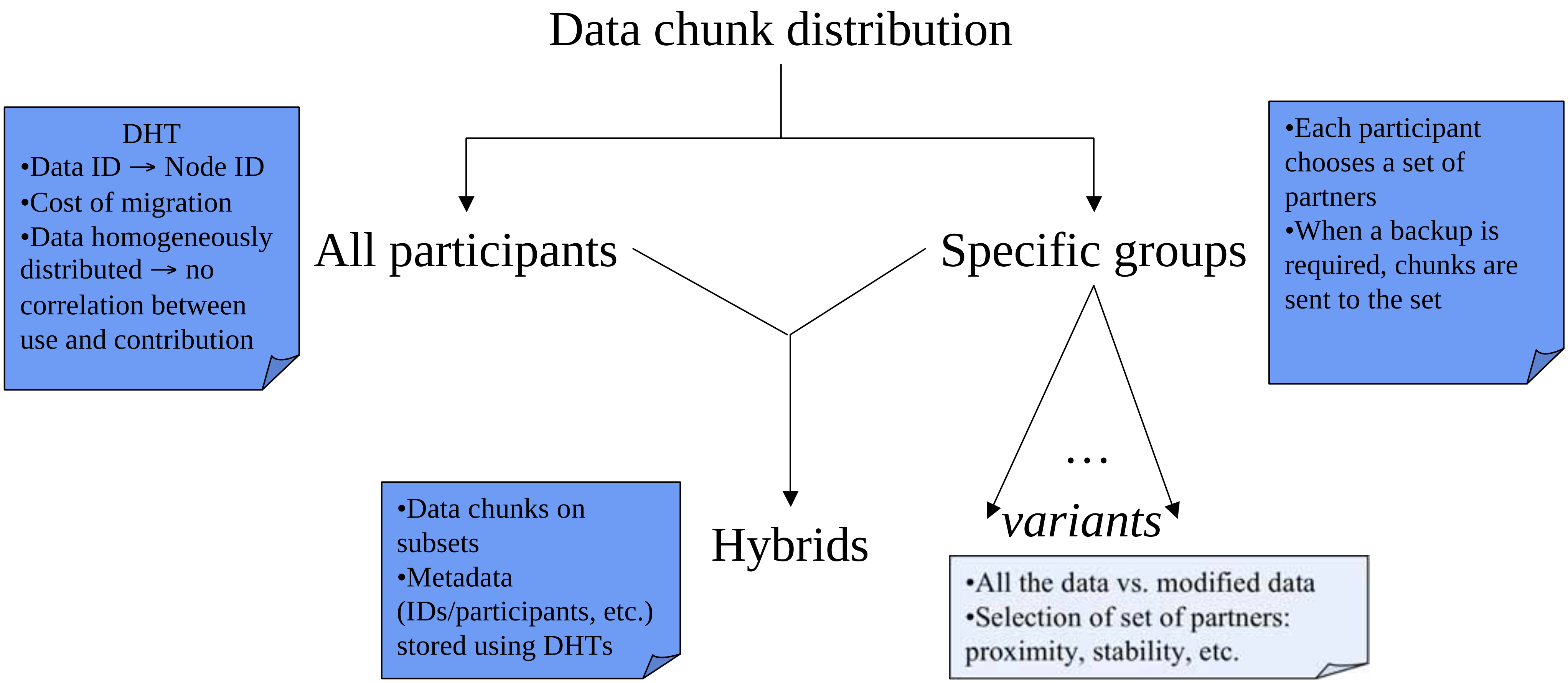
Overview

- Overview of MoSAIC project
- **Collaborative Backup Systems**
- Trust Management
- Current Status

P2P Storage Systems

- Peer-to-peer file sharing systems
  - ➢ *Overlay networks, DHT, unstructured*
    - GNUnet
    - FreeNet
    - OceanStore
- Peer-to-peer backup systems
  - ➢ *Cooperation incentives, trust*
    - Elnikety, Pastiche, PeerStore, pStore for WANs
    - Flashback for PANs

Storage space discovery and allocation



Data chunk distribution

All participants

Specific groups

Hybrids

…

variants

DHT
- Data ID → Node ID
- Cost of migration
- Data homogeneously distributed → no correlation between use and contribution

- Each participant chooses a set of partners
- When a backup is required, chunks are sent to the set

- Data chunks on subsets
- Metadata (IDs/participants, etc.) stored using DHTs

- All the data vs. modified data
- Selection of set of partners: proximity, stability, etc.

Elnikety *et al.*

- Peer-to-peer backup system on the Internet
  - No unique ID, no certified public keys, no routing
  - Set of partners, point-to-point reciprocal relationships
- Enforces
  - Confidentiality: secret key cryptography (IDEA)
  - Robustness: block redundancy using erasure codes (Reed-Solomon)
  - Integrity: self-checking sub-blocks, crypto hash-keys (HMAC-MD5)
  - Authentication: pairwise shared secret keys (Diffie-Hellman)
- Attacks
  - Selfish DoS: periodic challenges, grace and commitment periods
  - Malicious DoS: protocol against man-in-the-middle attacks

Flashback

- Devices are part of a Personal Area Network (PAN)
  - Same owner: a priori mutual trust
- Permanent fault (or theft) of the data owner
  - Same ID assigned to a new device
  - Reinitialized from backed-up data
- Optimization of the restorable data
  - Limitation of # of copies (function of block priority)
  - Replication rate function of current number of copies
  - Taking into account heterogeneity (energy, storage)
- Backup contracts: notion of lease
  - Duration of lease > expected duration of disconnection
  - Lease renewal at 50% expiry time

P2P vs. MoSAIC

- Fixed and unique IDs: not available
- Bandwidth, duration of connections: not known a priori
- Mobility: partnerships have to change and adapt
- Resource and node discovery: knowing one participant/repository is not enough
- Intermittent connection to fixed infrastructure: mostly disconnected
- Trust mechanisms for disconnected operation: reputation (e.g., using trusted HW)

Overview

- Overview of MoSAIC project
- Collaborative Backup Systems
- **Trust Management**
- Current Status

Tragedy of the Commons

- Why do we need cooperation incentives?
- “Tragedy of the Commons” [Hardin68]
  - Resource sharing
    - Naturally there are disincentives
    - Cooperation implies consumption of ones own resources
  - Selfish users behave as free-riders
    - Consumption without contribution
  - Very common behavior especially in large networks
    - 70% of Gnutella network users do not contribute

Routing in ad hoc networks 1

- Forwarding/routing packets costs
  - Energy, bandwidth, CPU cycles
- Different misbehaving nodes
  - Selfish DoS (passive) - priority is energy
    - Don’t forward packets
  - Malicious DoS (active) - priority is damage
    - Drop packets
    - Send wrong routes
- No a priori trust/confidence
- Enforce cooperation
  - Detection of misbehaving nodes
  - Isolation of misbehaving nodes
  - Stimulate and encourage cooperation

Without excessive resource consumption

Routing in ad hoc networks 2

- Use redundant routes for every packet
  - Increased energy consumption
- Consider false route information as old routes
  - Need a majority of honest nodes
- Use localization information for routing (GPS)
  - Privacy attacks
- **Money** as an incentive
  - Exchange virtual money for routing (e.g., Buttyan’s nuglets)
  - Requires secure kernels/trusted hardware
- Detect misbehavers, give them bad **reputation**
  - Global reputation requires access to servers
  - Local reputation (e.g., Marti’s watchdogs)

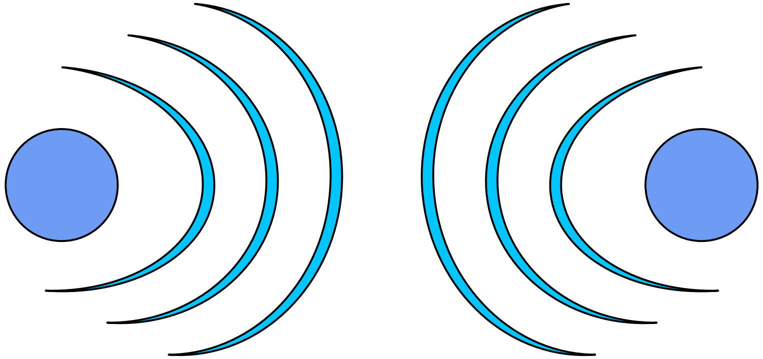
Trust Mechanisms

- Traditional key management
  - Public Key Infrastructure (PKI)
  - Trust authority to establish trust between mutually distrusting entities
  - **Centralized trust servers**
- Trust established using long-term accountability
  - Micro-payment against free-riding [Golle]
  - Contributor ratings [eBay, bizrate, etc.]
  - **Centralized rating/bank servers**
- Web of trust
  - Distributed trust model, PGP-like
  - Used primarily for key management
  - Content-centric for reputation-guided searching [Poblano]
  - Peer-centric [Law-Governed Interaction] needs trusted kernels/HW

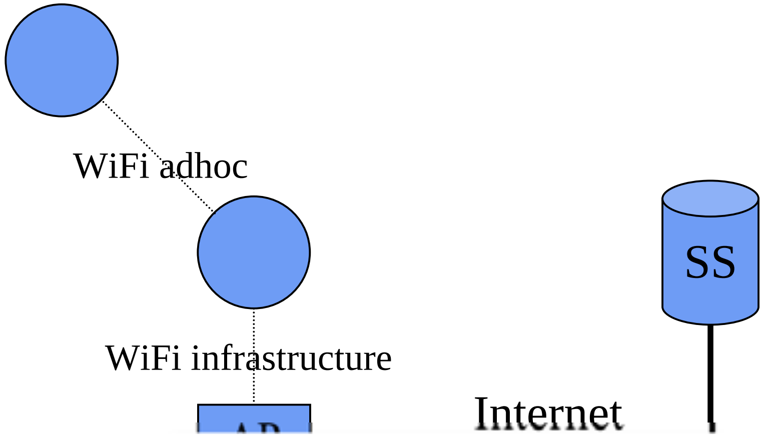
Overview

- Overview of MoSAIC project
- Collaborative Backup Systems
- Trust Management
- **Current Status**

Node discovery

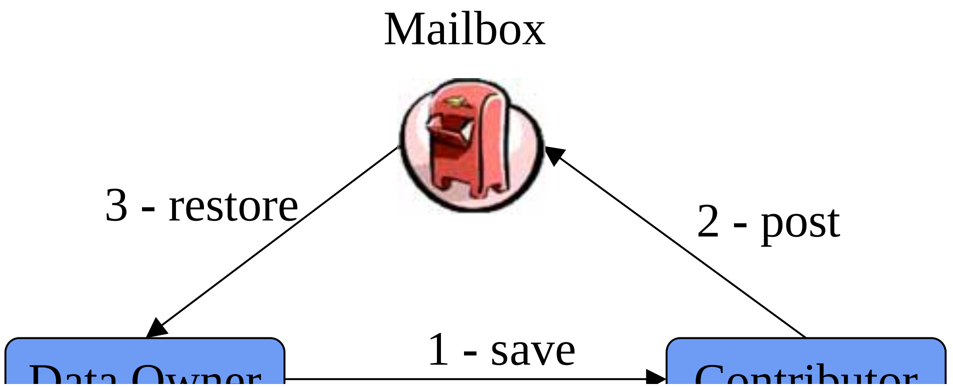


- Discovery of MoSAIC nodes
  - Online
  - Creation of ad hoc network
  - Active beaconing:
    low latency vs energy economy
- Discovery of Internet access
  - Be able to backup on reliable storage service
- Ad hoc and infrastructure mode at the same time
  - Cooperation + storage service access



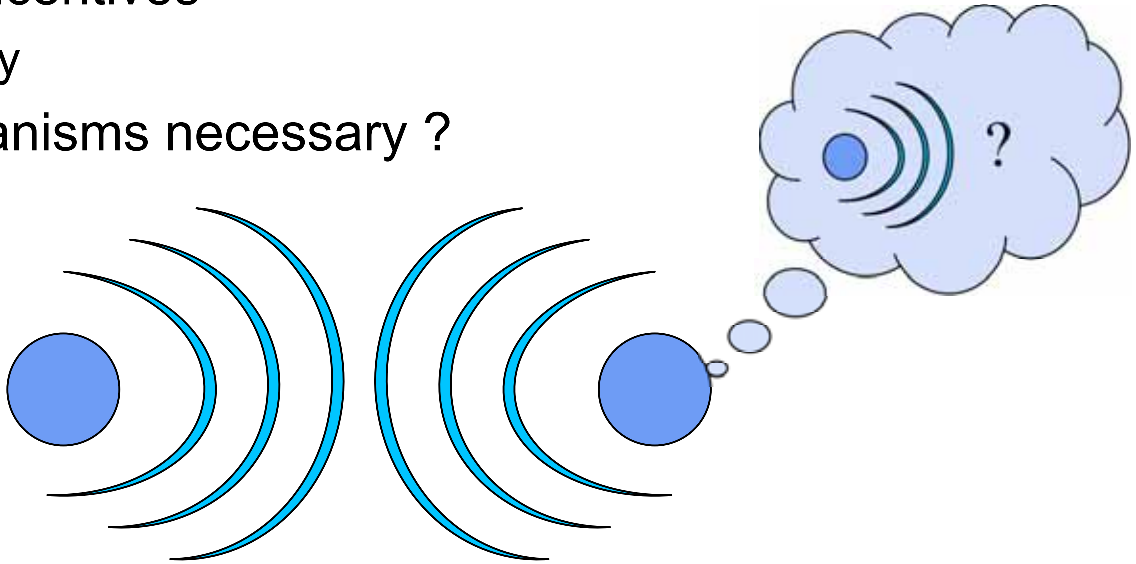
Being Opportunistic

- Opportunistically use connection to Internet
  - “Mailbox” for storing the backup chunks
  - Accommodate several restoration models
    - Push: the contributor sends the chunks back home
      - Internet access, mailbox at the owner’s home
    - Pull: the data owner searches for the data when necessary
      - Ad hoc network, mailbox hosted by the contributor
    - Push-pull: storage service as an intermediary
      - Internet access, mailbox hosted by the storage service

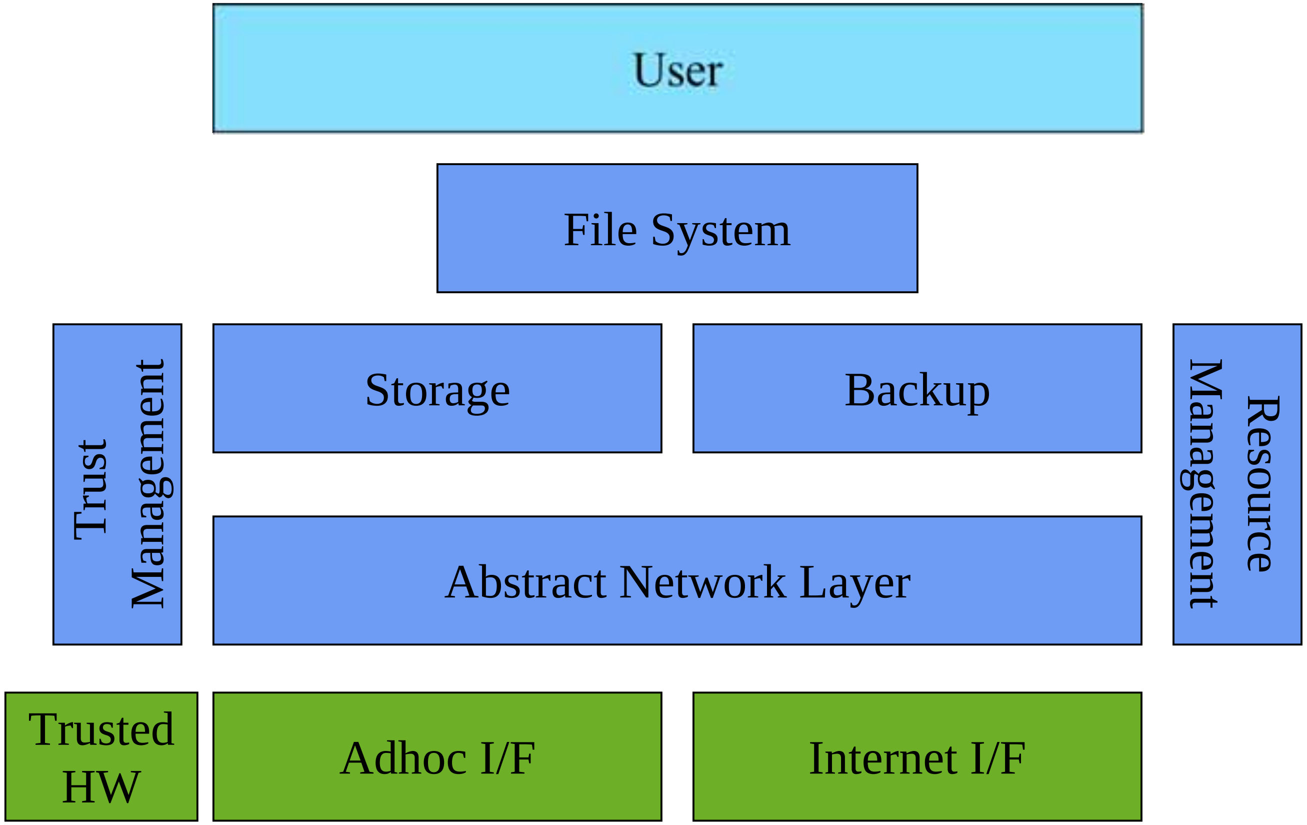


Trust Management

- Classic solutions
  - Participants are almost always connected
- Strong mobility, ephemerous connections, etc.
  - Self-carried reputation (using trusted HW)
    - Checked by other participants
    - Link with the mailbox implementation
  - Collaboration incentives
    - Virtual money
  - Are both mechanisms necessary ?



Architecture



Conclusion

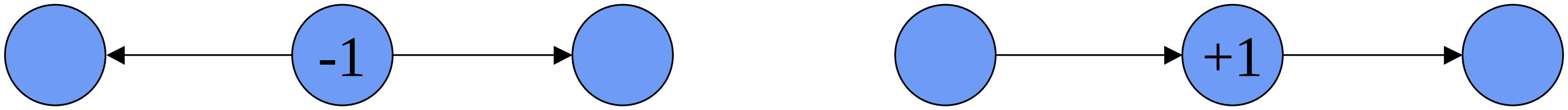
- Scenario for
  - Designing new algorithms
  - Developing new middleware
- Implies fault-tolerance
  - Classic faults
    - Devices: crash of devices (owners and contributors), etc.
    - Data: integrity, confidentiality
  - Interaction faults (selfishness, maliciousness)
- New FT-enabling mechanisms
  - Self-carried reputation, virtual money, etc.
  - Opportunistic Internet backup, P2P interactions
- Project is 10 months old, still a lot to do ….

Collaborative Backup for Nomadic Devices

M.O.Killijian, D.Powell

Buttyan's nuglets

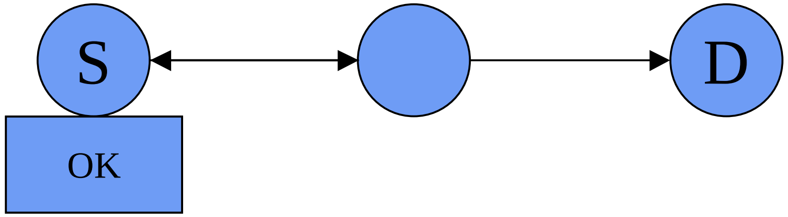
- Each node maintains a counter (nuglet)
  - Decreased when sending its own packet
  - Increased when forwarding a packet
  - The counter must remain positive



- The policy must be enforced
  - Use of tamperproof hardware
    - SIMcards, JavaCards, etc.
    - TPM

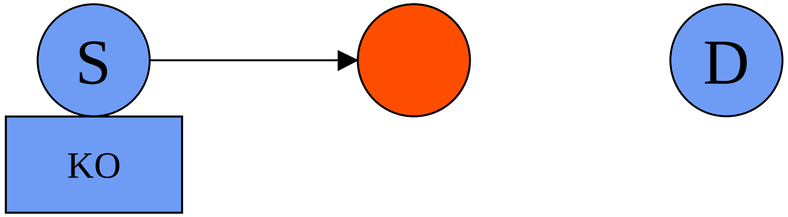
Marti’s Watchdogs

- Each node possesses a watchdog
  - When a node sends a packet, the watchdog verifies that the neighbors forward it



Marti's Watchdogs

- Each node possesses a watchdog
  - When a node sends a packet, the watchdog verifies that the neighbors forward it



- Misbehaving nodes are detected: bad reputation
- Limits
  - Collisions
  - Low transmission power attacks
  - False positives
  - Collusion
  - Partial propagation

Session 2.2

Challenges in Mobile Distributed Systems

Moderator and Rapporteur

Karama Kanoun, LAAS-CNRS, Toulouse, France

48th Meeting of IFIP Working Group 10.4
Workshop “Nomadic Computing and Dependability”
Hakone, Japan – Monday July 4, 2005

Autonomous Clustering and Hierarchical Routing for Mobile Ad Hoc Networks

Yoshiaki Kakuda

Hiroshima City University

Experiences of My Research Activities on Dependability (1)

- FTCS-10, 1980, Kyoto, Japan
- FTCS-12, 1982, Santa Monica, USA
- Workshop on Responsive Computer Systems, 1992, Kamifukuoka, Japan

General Chairs: Miroslaw Malek, Tohru Kikuno

Program Chairs: Hermann Kopetz, Yoshiaki Kakuda

Experiences of My Research Activities on Dependability (2)

- Workshop on Dependability in Advanced Computing Paradigms, 1996, Hitachi, Japan

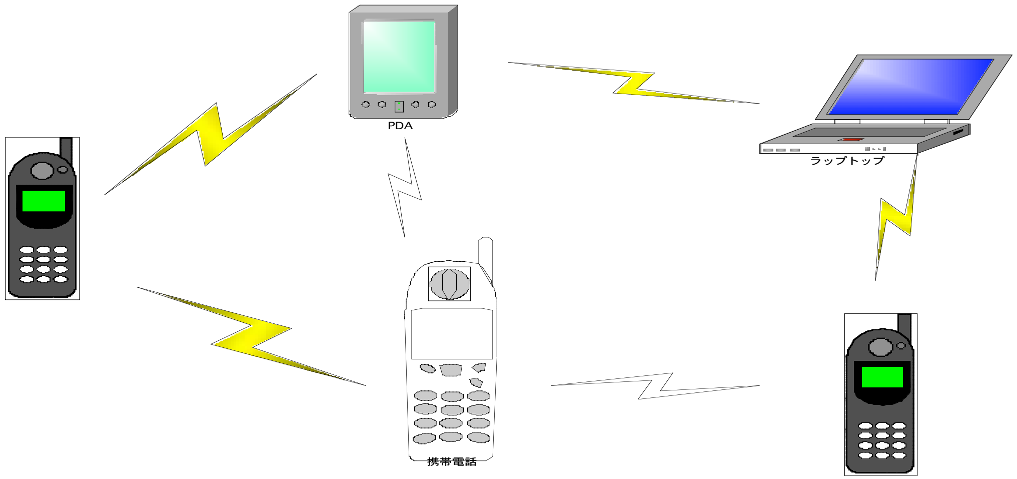
General Chairs: Jack Goldberg, Yoshihiro Tohma

Program Chairs: Hermann Kopetz,
Richard Schlichting, Yoshiaki Kakuda

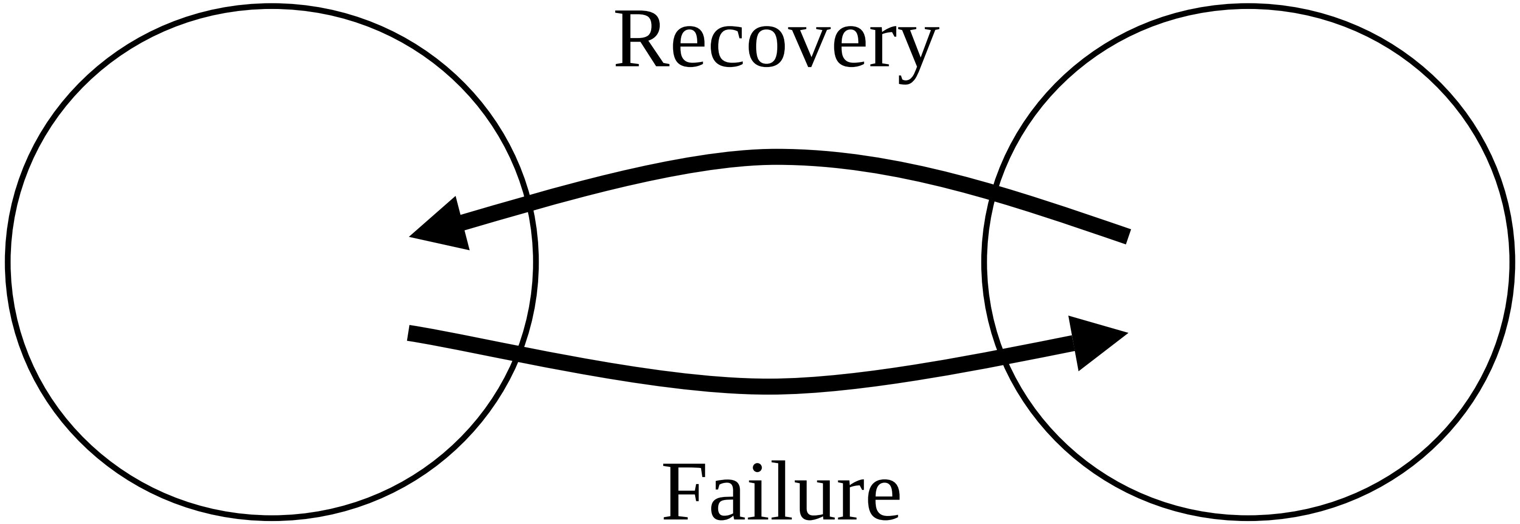
- IFIP Conference on DCCA (Dependable Computing for Critical Applications)-7, Program Committee, 1999, San Jose, USA
- DSN-2005, DCCS Program Committee, 2005, Yokohama, Japan

Mobile Ad Hoc Networks

- Wireless mobile network without the aid of any base stations
- Each mobile node has the function of router
- Each mobile node can move around the network



Characterization of Mobile Ad Hoc Networks by Dependability



Recovery

Failure

Normal system states

Abnormal system states

Challenging Issues in Routing for Mobile Ad Hoc Networks

- Routing for large-scale networks
- Routing for asymmetric networks
- Location-based routing
- Energy efficient routing
- Secure routing
- QoS routing

Scalability Issue in Routing for Mobile Ad Hoc Networks

- Why does the scalability issue occur?
  - Increase of the numbers of mobile nodes and pairs of a source and a destination
  - Frequent node movement

⇨ **Stable routes are required.**

Research Funds from MIAC

- Ministry of Internal Affairs and Communications in Japan
- Stragetic Information and Communications R&D Promotion Programme (SCOPE)
- Research and Development Promoting Info-Communications Technology for Community Development (SCOPE-C)

Joint Project of University and Industries

- Project Title: R&D on Scalable Technology for Confirming Group Members in Mobile Ad Hoc Networks
- Project Members: Hiroshima City University, KDDI Corporation, National Institute of Advanced Industrial Science and Technology (Information Technology Research Institute), The Chugoku Electric Power Co., Inc. (Technical Research Center), Chuden Engineering Consultants

Table-Driven Routing

- Each node always has the routing table for the destination node because it periodically exchanges route information with each other.
- Distance-vector and link-state types
- OLSR, TBRPF, DSDV

On-demand Routing

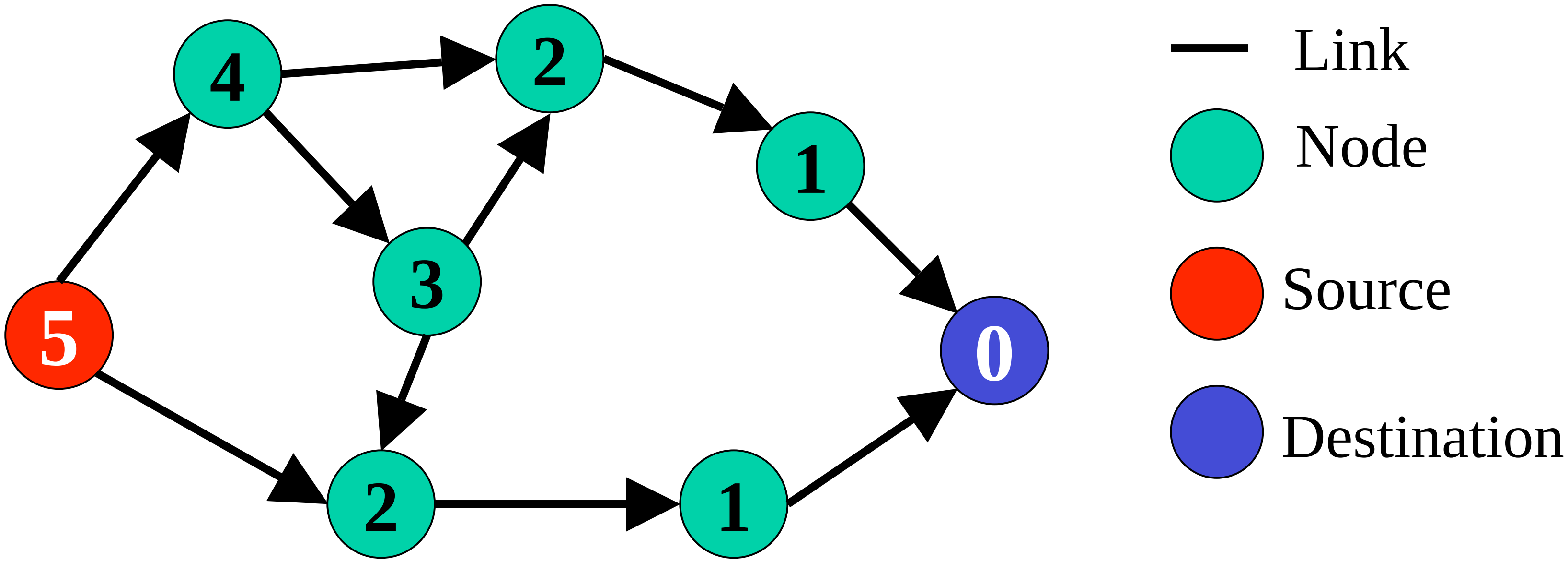
- A route to a destination is required only when a source node wants to send data packets
- Utilizing the route cache
- Overhead to create the route is lower
- It takes longer time to start to send data packets
- TORA, DSR, AODV

Motivation

- Ad hoc network routing protocols
  - TORA, DSR, AODV（**Flat routing**）
  - The performance becomes worse along with the increase of the network size
- Hierarchical routing protocols based on the autonomous clustering
  - Hi-TORA, Hi-DSR, Hi-AODV（Hierarchical routing）

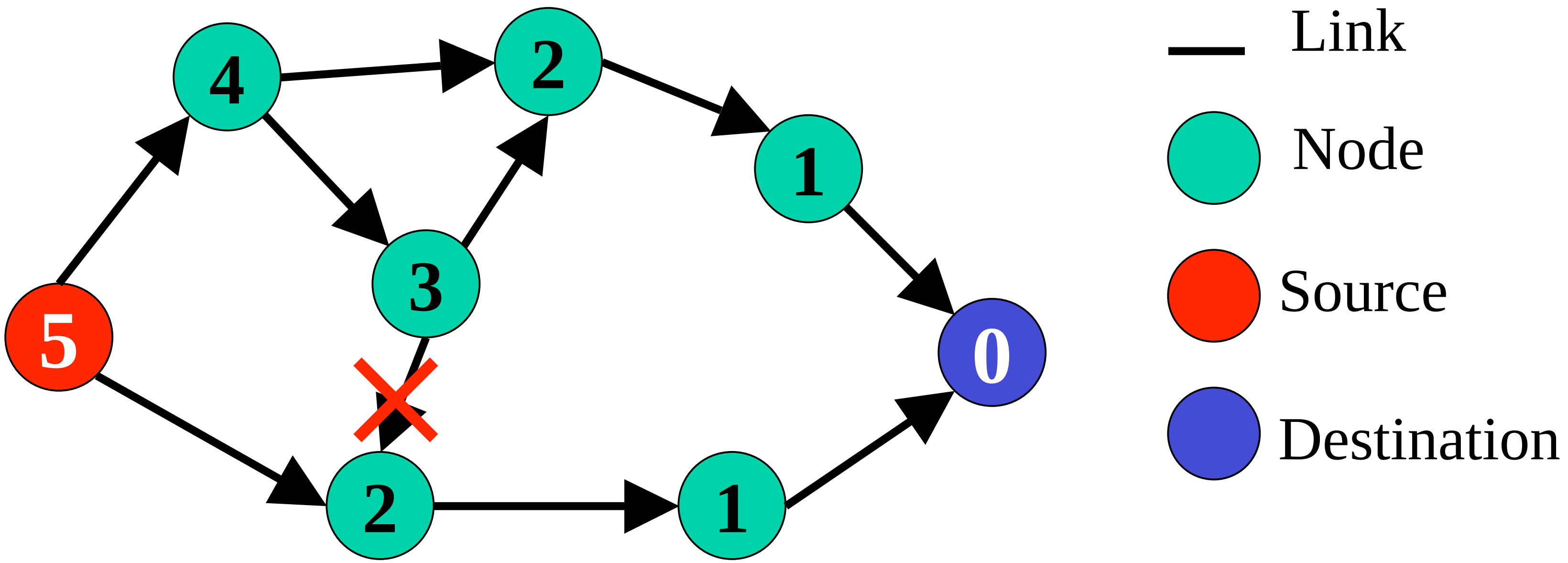
Proposal and evaluation of hierarchical routing protocols based on the autonomous clustering

Route Discovery in TORA (Temporally-Ordered Routing Algorithm)

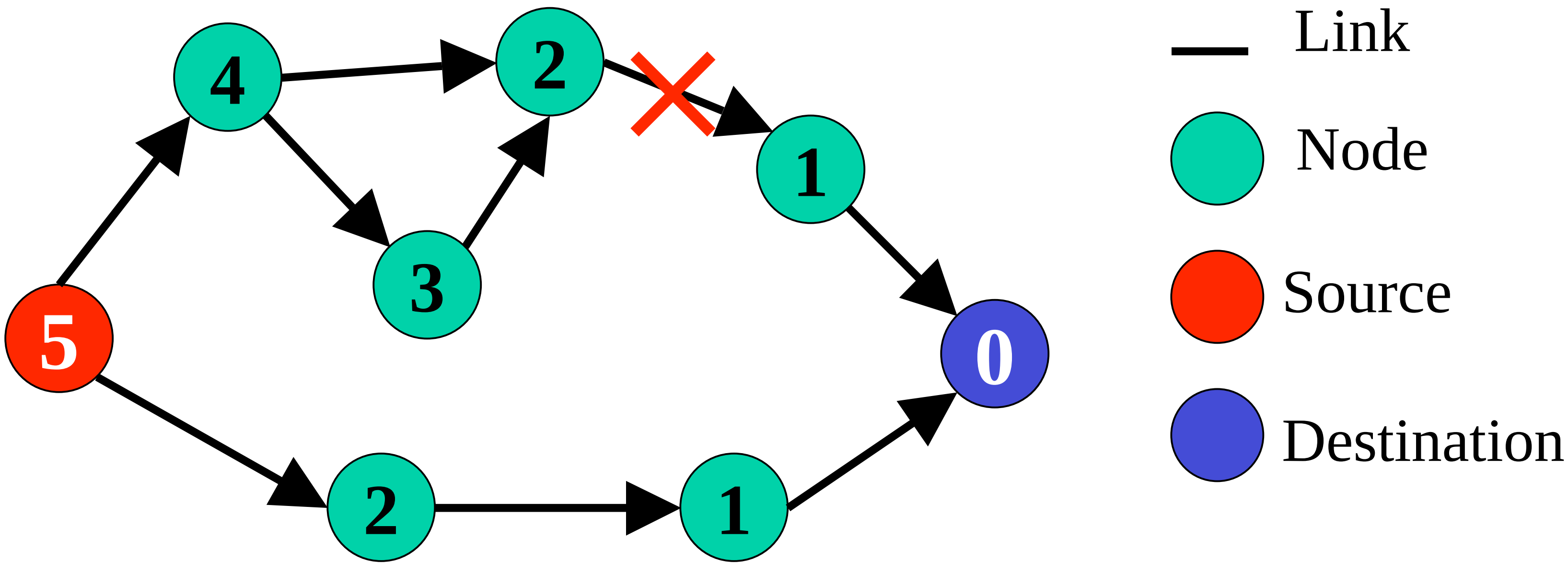


- A source node broadcasts REQUEST packets to all nodes and the notation of *height* is assigned to them to create the route.

Route Maintenance in TORA

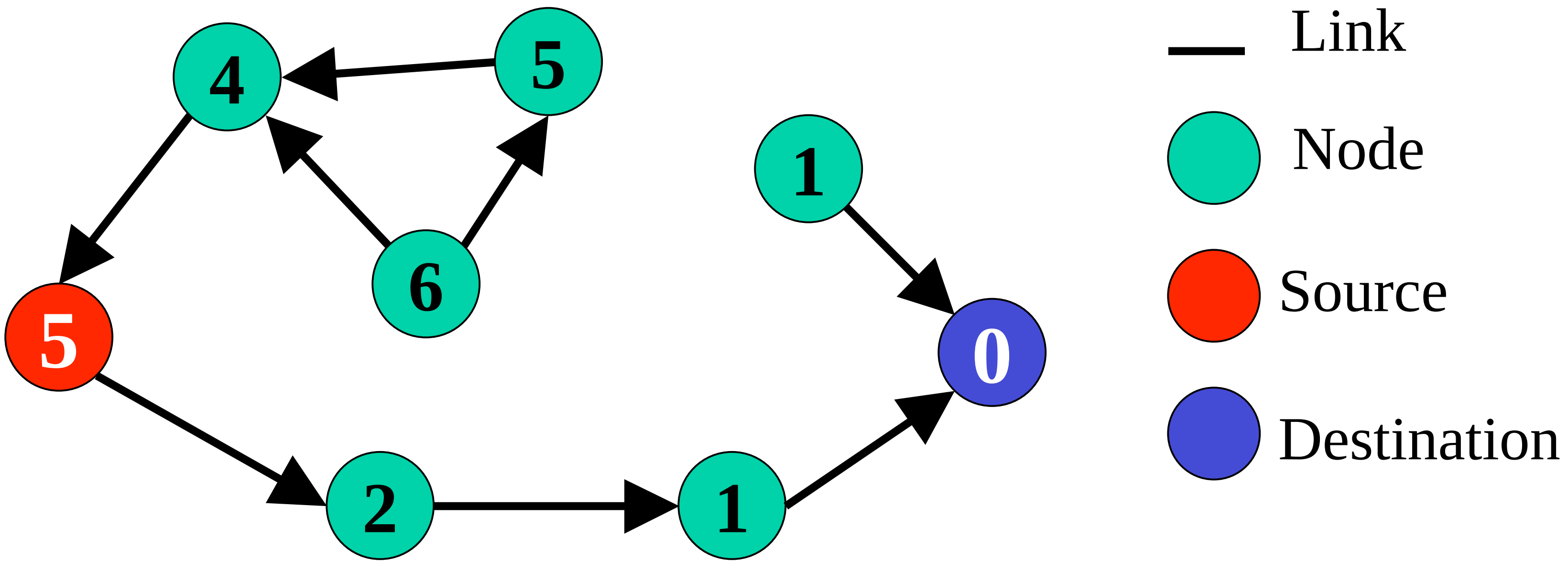


Route Maintenance in TORA



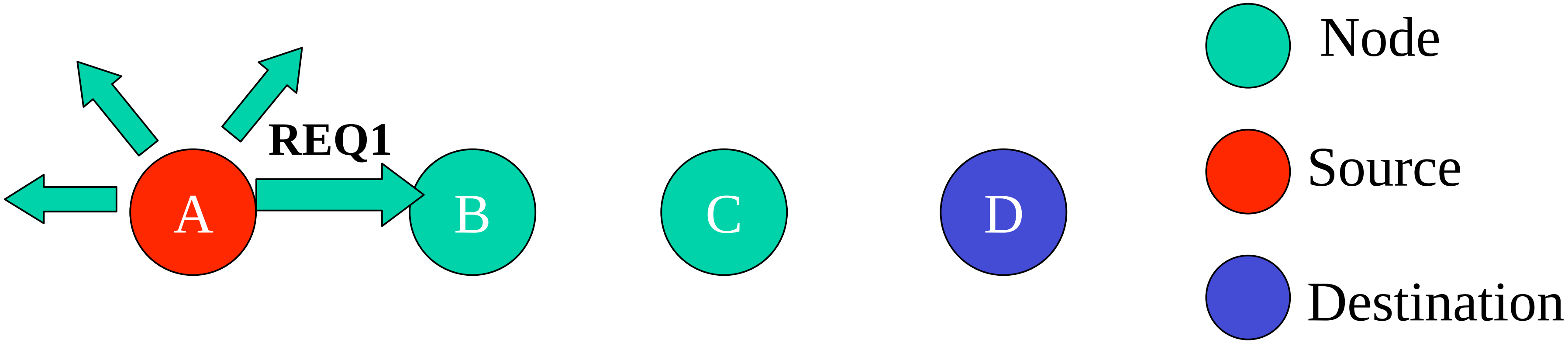
- Nothing to do because there is another route.

Route Maintenance in TORA



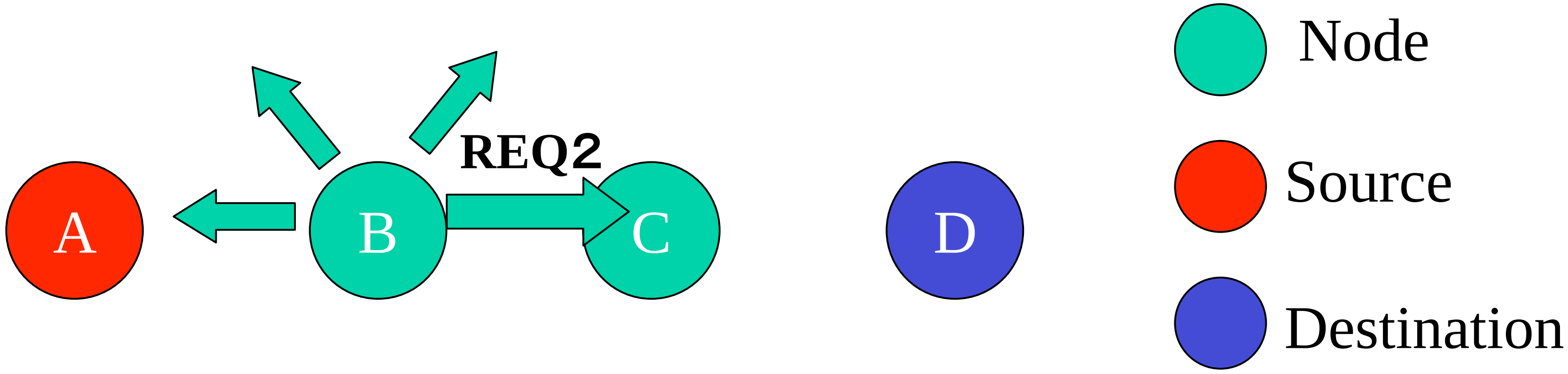
- There is possibility that the number of hops between a source node and a destination node becomes long because each node repairs the route locally.

Dynamic Source Routing(DSR)



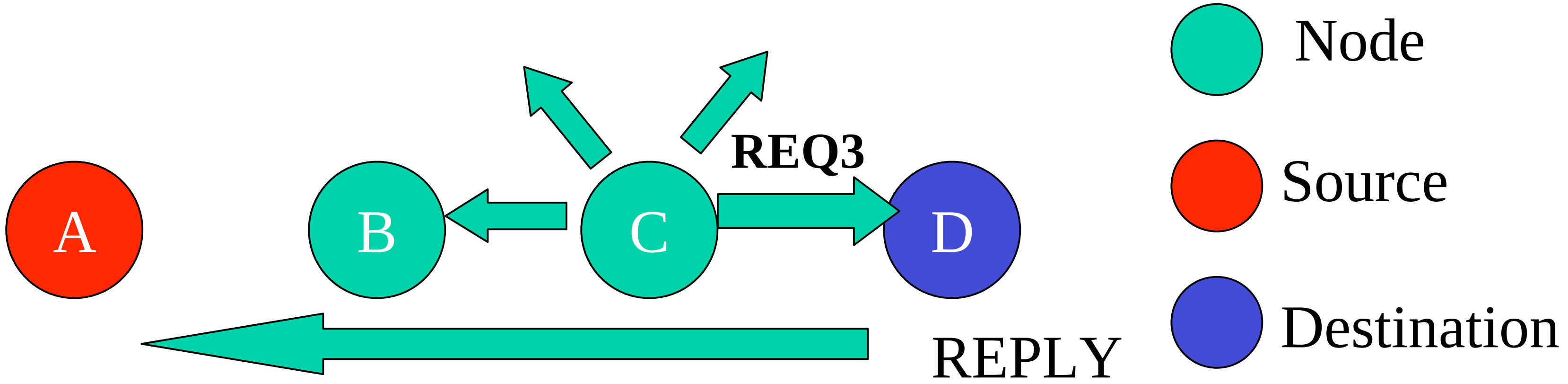
	Route Information
REQ1	A
REQ2	A→B
REQ3	A→B→C
REPLY	A→B→C→D

Dynamic Source Routing(DSR)



	Route Information
REQ1	A
REQ2	A→B
REQ3	A→B→C
REPLY	A→B→C→D

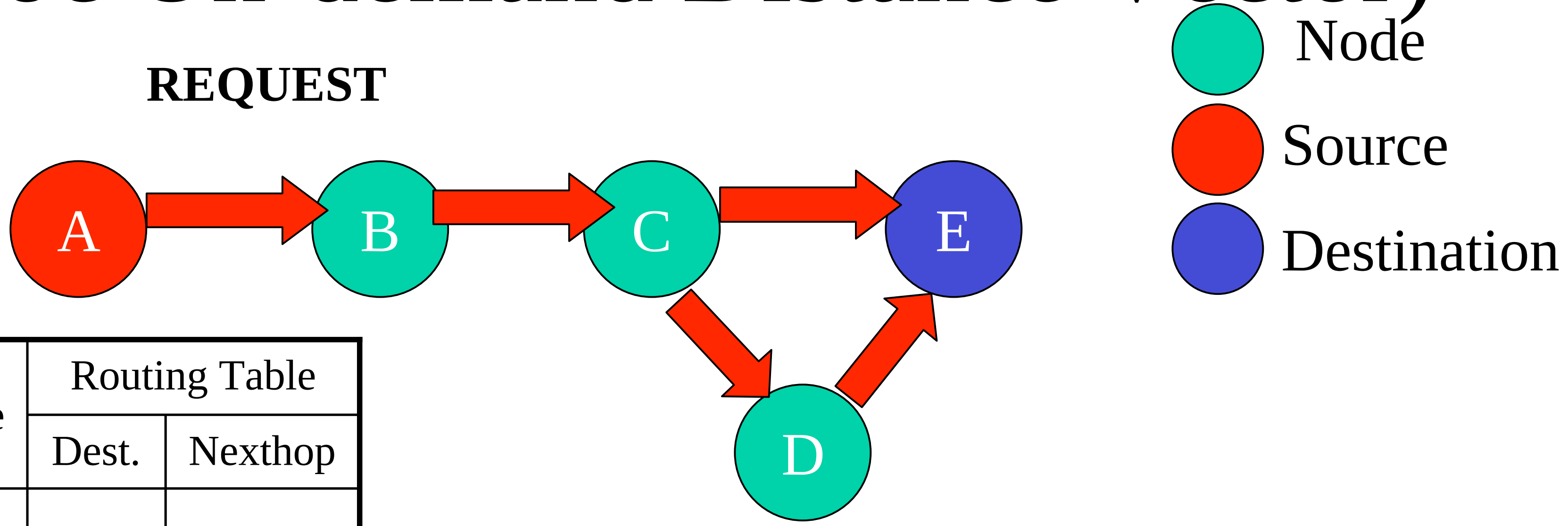
Dynamic Source Routing(DSR)



	Route Information
REQ1	A
REQ2	A→B
REQ3	A→B→C
REPLY	A→B→C→D

In DSR, if the route between the source and the destination disappeared, a source node invokes route discovery again.

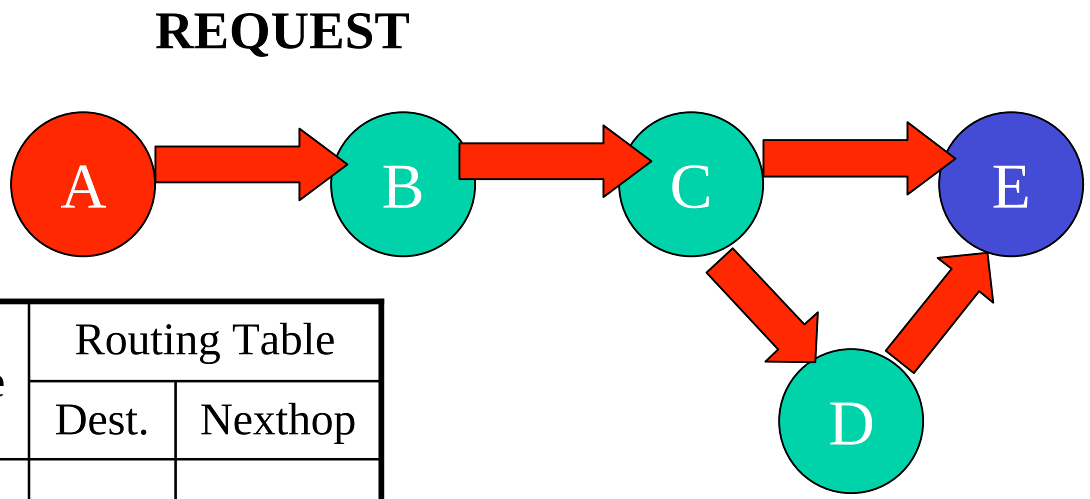
Route Discovery in AODV (Ad hoc On-demand Distance Vector)



Node	Routing Table	
	Dest.	Nexthop
A		
B		
C		
D		
E		

- Nodes which received REQUEST or REPLY packets update the routing table and forward it to the neighbors.
- Data packets are delivered along with the routing table in each node.

Route Discovery in AODV



● Node

● Source

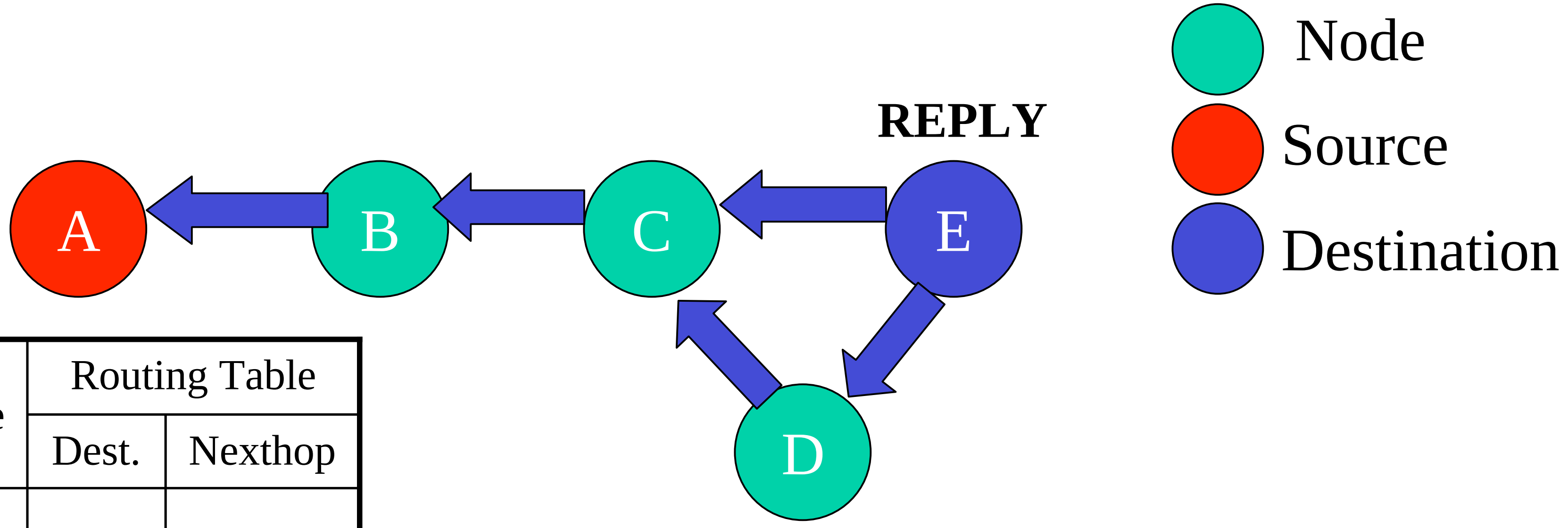
● Destination

Node	Routing Table	
	Dest.	Nexthop
A		
B	A	A
C	A	B
D	A	C
E	A	C

- Nodes which received REQUEST or REPLY packets update the routing table and forward it to the neighbors.
- Data packets are delivered along with the routing table in each node.

21

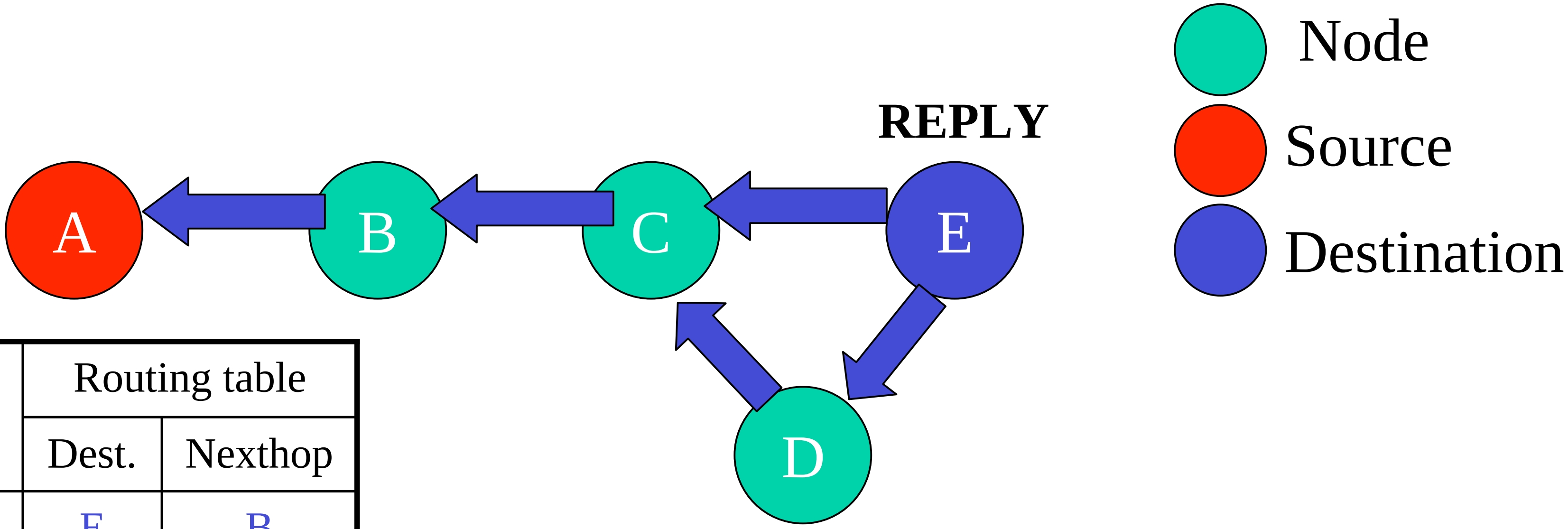
Route Discovery in AODV



Node	Routing Table	
	Dest.	Nexthop
A		
B	A	A
C	A	B
D	A	C
E	A	C

- Nodes which received REQUEST or REPLY packets update the routing table and forward it to the neighbors.
- Data packets are delivered along with the routing table in each node.

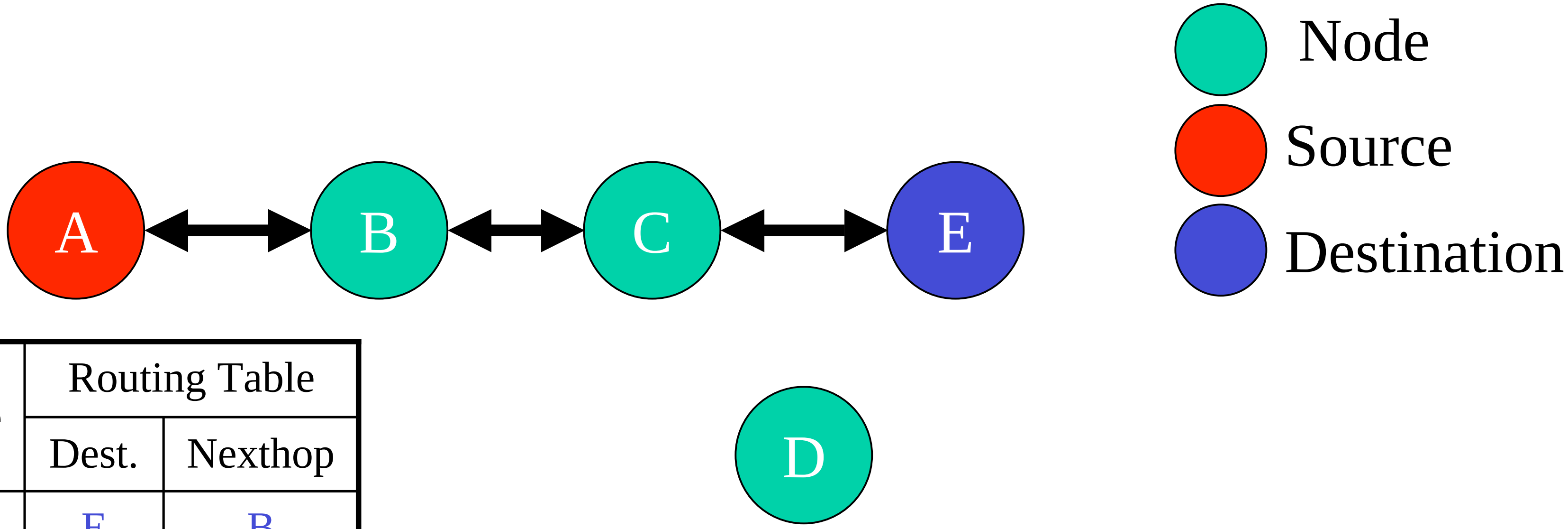
Route Discovery in AODV



Node	Routing table	
	Dest.	Nexthop
A	E	B
B	A	A
	E	C
C	A	B
	E	E
D	A	C
	E	E
E	A	C

- Nodes which received REQUEST or REPLY packets update the routing table and forward it to the neighbors.
- Data packets are delivered along with the routing table in each node.

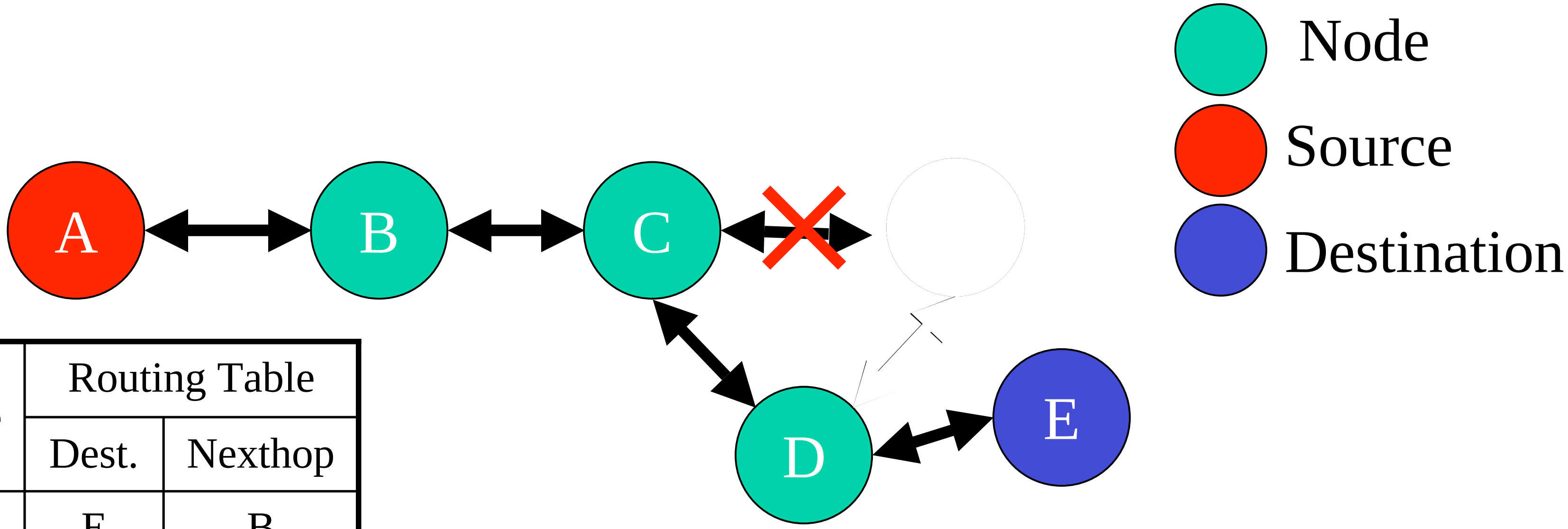
Route Discovery in AODV



Node	Routing Table	
	Dest.	Nexthop
A	E	B
B	A	A
	E	C
C	A	B
	E	E
D	A	C
	E	E
E	A	C

- Nodes which received REQUEST or REPLY packets update the routing table and forward it to the neighbors.
- Data packets are delivered along with the routing table in each node.

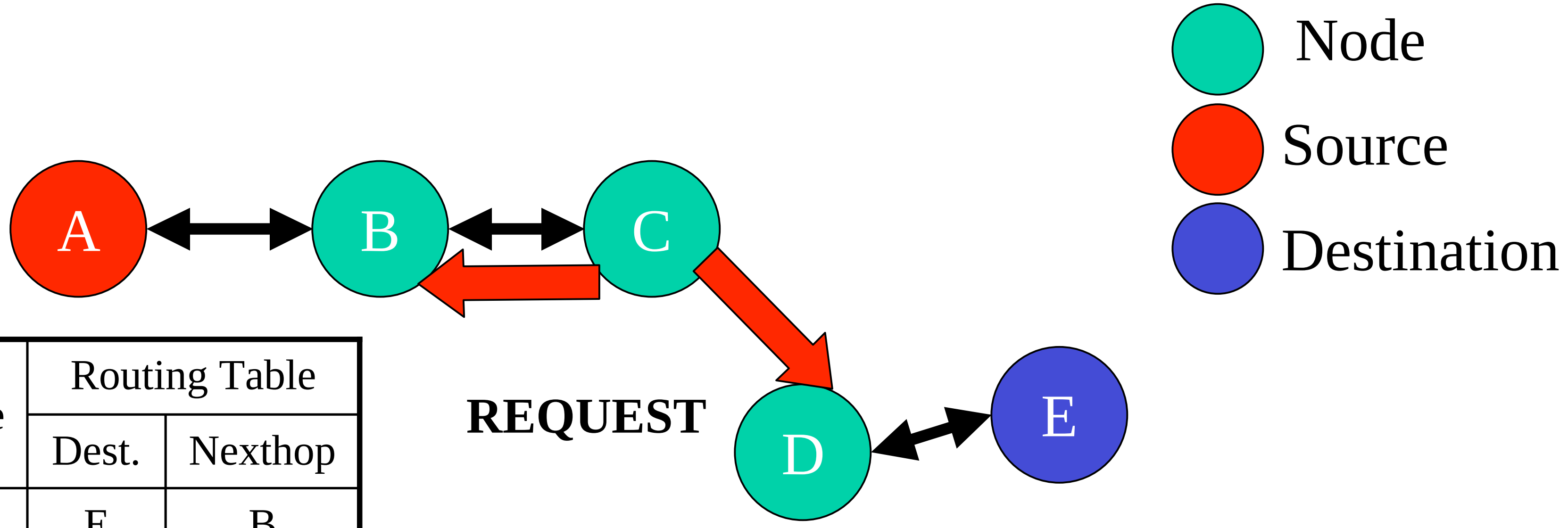
Route Maintenance in AODV



Node	Routing Table	
	Dest.	Nexthop
A	E	B
B	A	A
	E	C
C	A	B
	E	E
D	A	C
	E	E
E	A	C

•If a node at which the route disappeared is close to the destination node, it repairs the route locally.

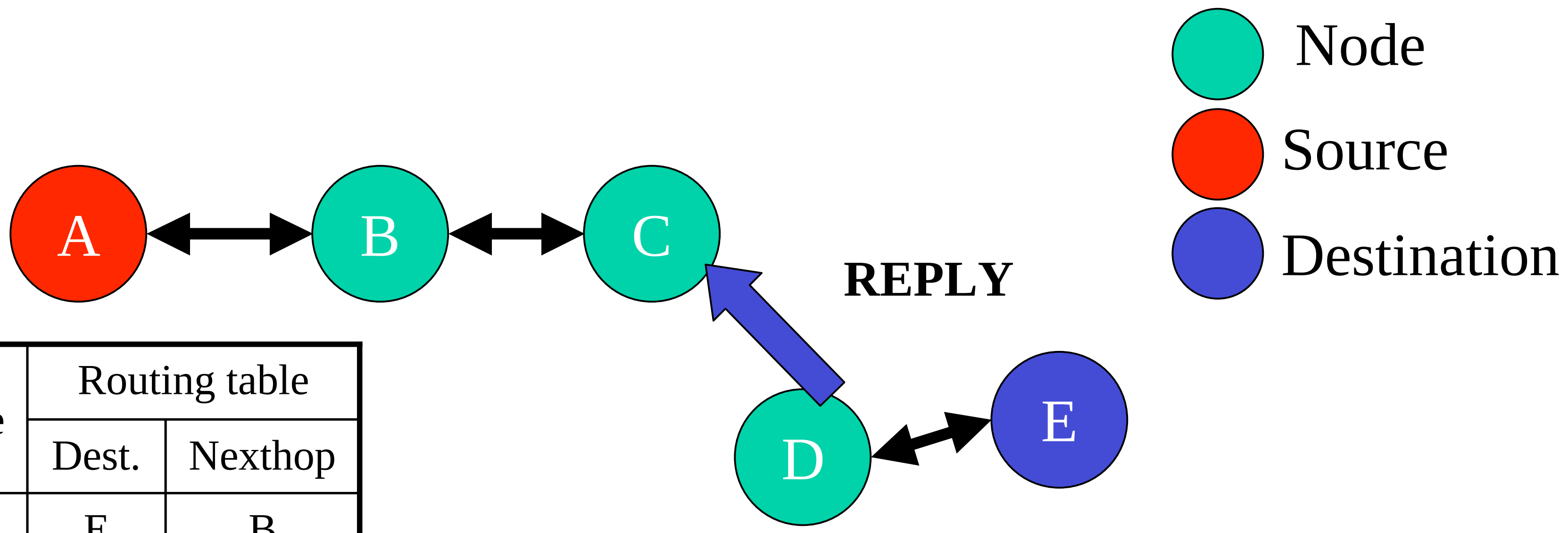
Route Maintenance in AODV



Node	Routing Table	
	Dest.	Nexthop
A	E	B
B	A	A
	E	C
C	A	B
	E	--
D	A	C
	E	E
E	A	C

- •Node C which detected the route disappearance tries to repair the route locally.
- •Node C broadcasts REQUEST packets within TTL.

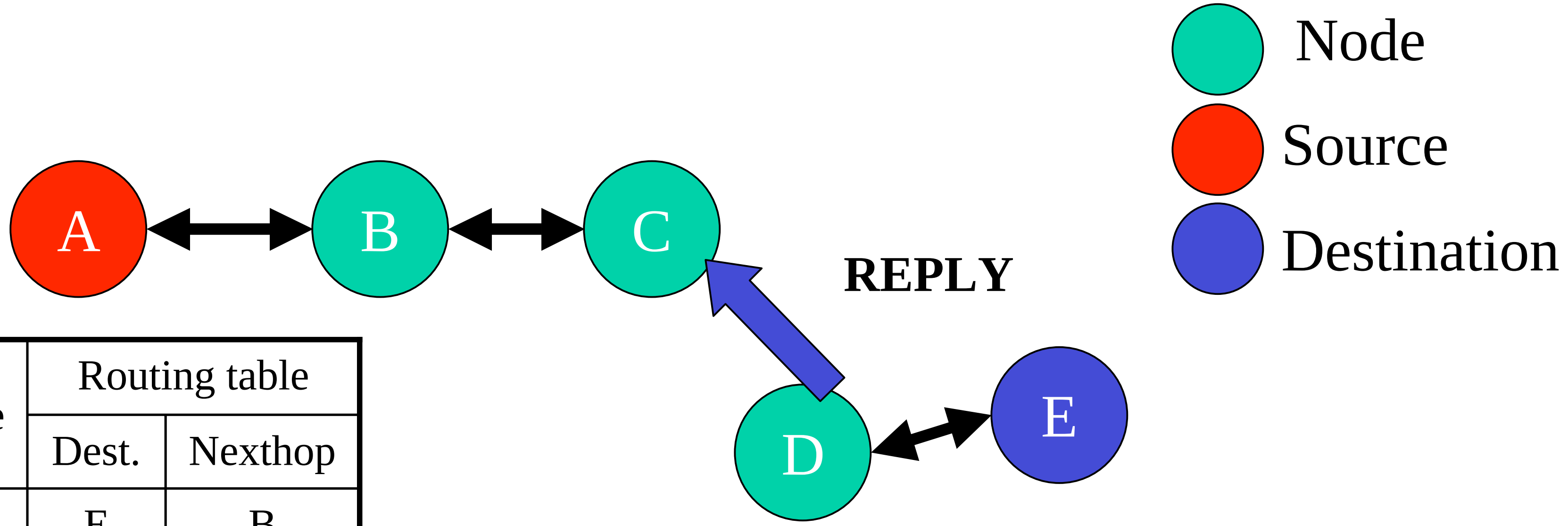
Route Maintenance in AODV



Node	Routing table	
	Dest.	Nexthop
A	E	B
B	A	A
	E	C
C	A	B
	E	--
D	A	C
	E	E
E	A	C

- Node C which detected the route disappearance tries to repair the route locally.
- Node C broadcasts REQUEST packets within TTL.

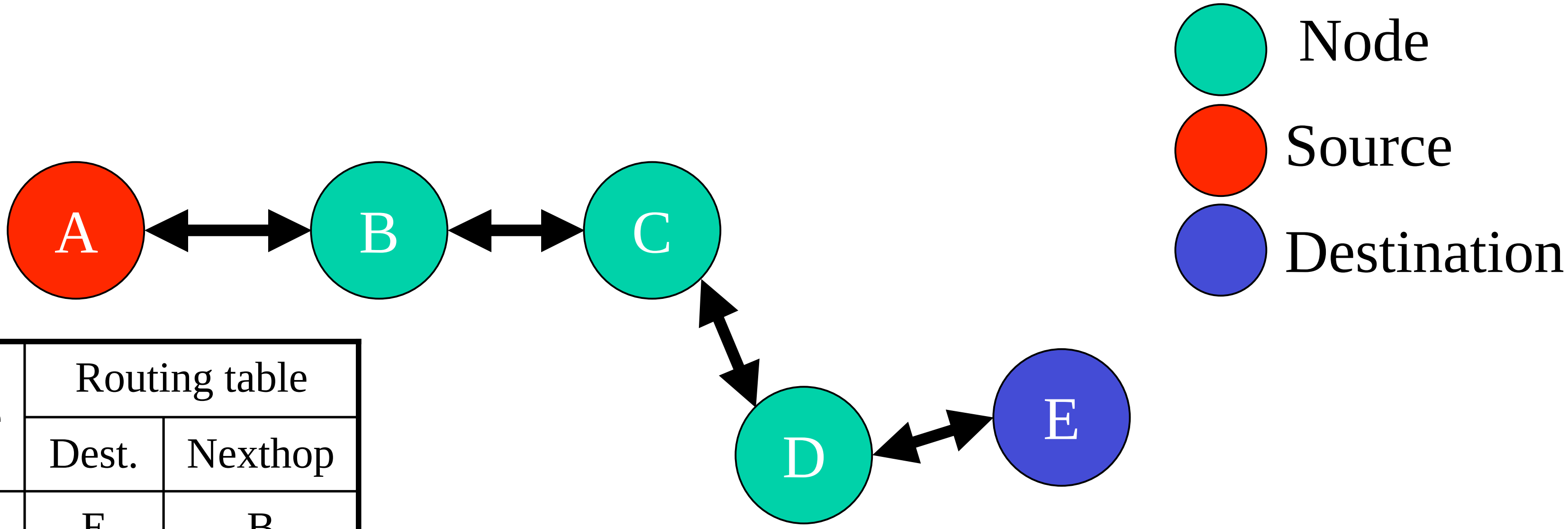
Route Maintenance in AODV



Node	Routing table	
	Dest.	Nexthop
A	E	B
B	A	A
	E	C
C	A	B
	E	D
D	A	C
	E	E
E	A	C

- Node C which detected the route disappearance tries to repair the route locally.
- Node C broadcasts REQUEST packets within TTL.

Route Maintenance in AODV

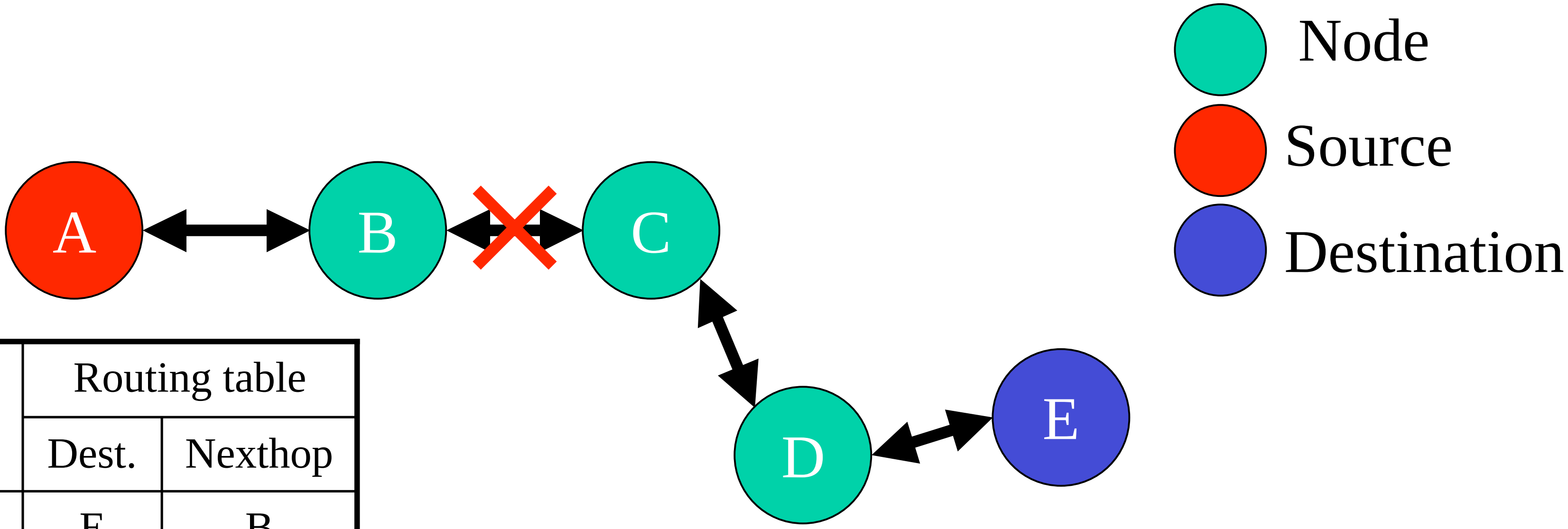


Node	Routing table	
	Dest.	Nexthop
A	E	B
B	A	A
	E	C
C	A	B
	E	D
D	A	C
	E	E
E	A	C

•Node C which detected the route disappearance tries to repair the route locally.

•Node C broadcasts REQUEST packets within TTL.

Route Maintenance in AODV



<table>
<tr><th rowspan="2">Node</th><th colspan="2">Routing table</th></tr>
<tr><th>Dest.</th><th>Nexthop</th></tr>
<tr><td>A</td><td>E</td><td>B</td></tr>
<tr><td rowspan="2">B</td><td>A</td><td>A</td></tr>
<tr><td>E</td><td>C</td></tr>
<tr><td rowspan="2">C</td><td>A</td><td>B</td></tr>
<tr><td>E</td><td>D</td></tr>
<tr><td rowspan="2">D</td><td>A</td><td>C</td></tr>
<tr><td>E</td><td>E</td></tr>
<tr><td>E</td><td>A</td><td>C</td></tr>
</table>

- If a node at which the route disappeared is close to the source node, it sends ERR packets back to the source node and the source node invokes route discovery again.

Problems of Flat Routing Protocols - Route Discovery -

- A source node broadcasts REQUEST packets over the entire network to create the route.
  - Due to the heavily control packets, a stable route is not provided
- TORA
  - It takes considerable control packets to create the route on all nodes and maintain it.

Problems of Flat Routing Protocols - Route Maintenance -

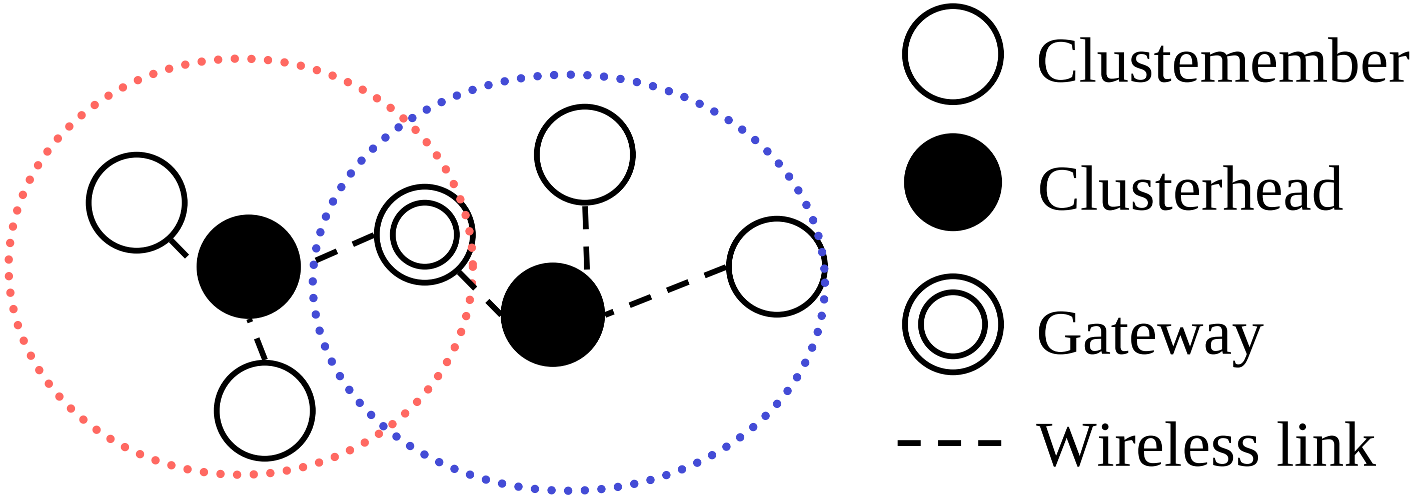
- TORA
  - The route is locally maintained while there is a possibility that the route distance becomes long.
- DSR
  - Due to node movement, the route disappearance occurs at an intermediate node and the source node invokes the route discovery again. If the route disappearance occurs frequently, the number of control packets becomes large because the source node invokes the route discovery frequently.
- AODV
  - When the route disappearance occurs near the source node, the source node invokes route discovery again.

Clustering and Hierarchical Routing

- Scalability issue
  - Hierarchical routing based on clustering (e.g. ZRP)
- Conventional clustering scheme
  - Each cluster is overlapped with each other.
- Autonomous clustering
  - True hierarchy because each cluster is not overlapped with each other.

Conventional Clustering Scheme

- A cluster consists of a clusterhead and its all neighboring nodes connected by one hop number.
- A node which has neighboring different clusterheads becomes a gateway which connects them.



○ Clustemember

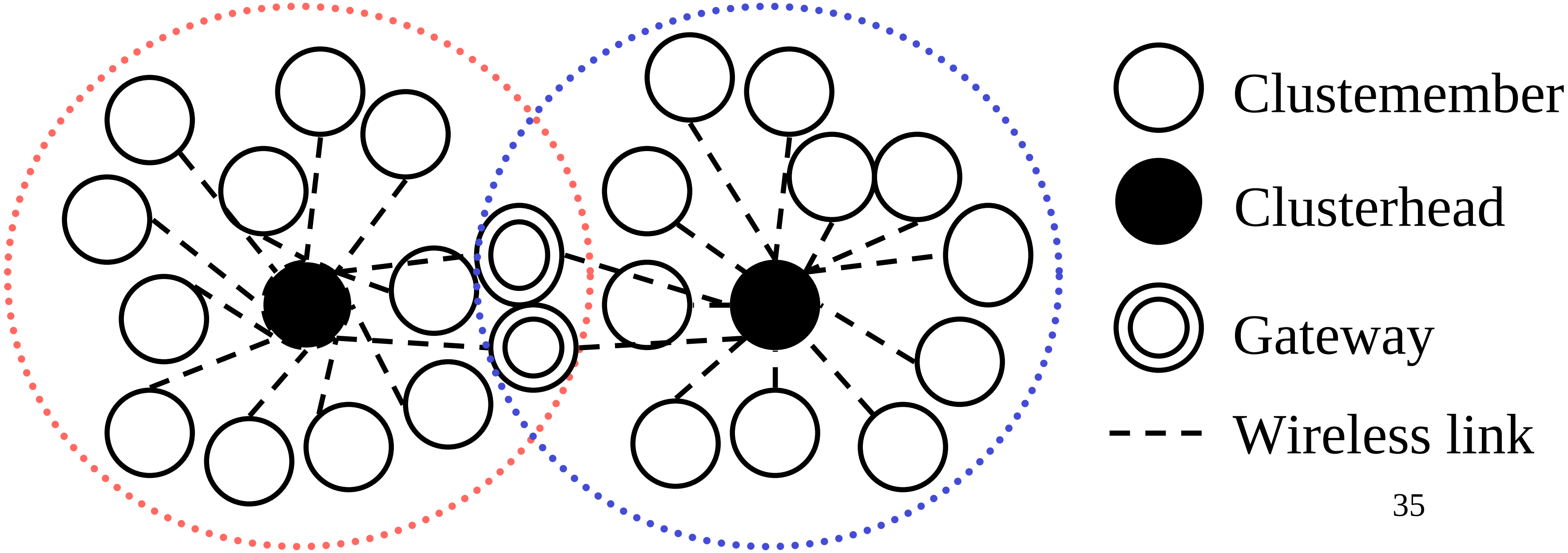
● Clusterhead

◎ Gateway

- - - Wireless link

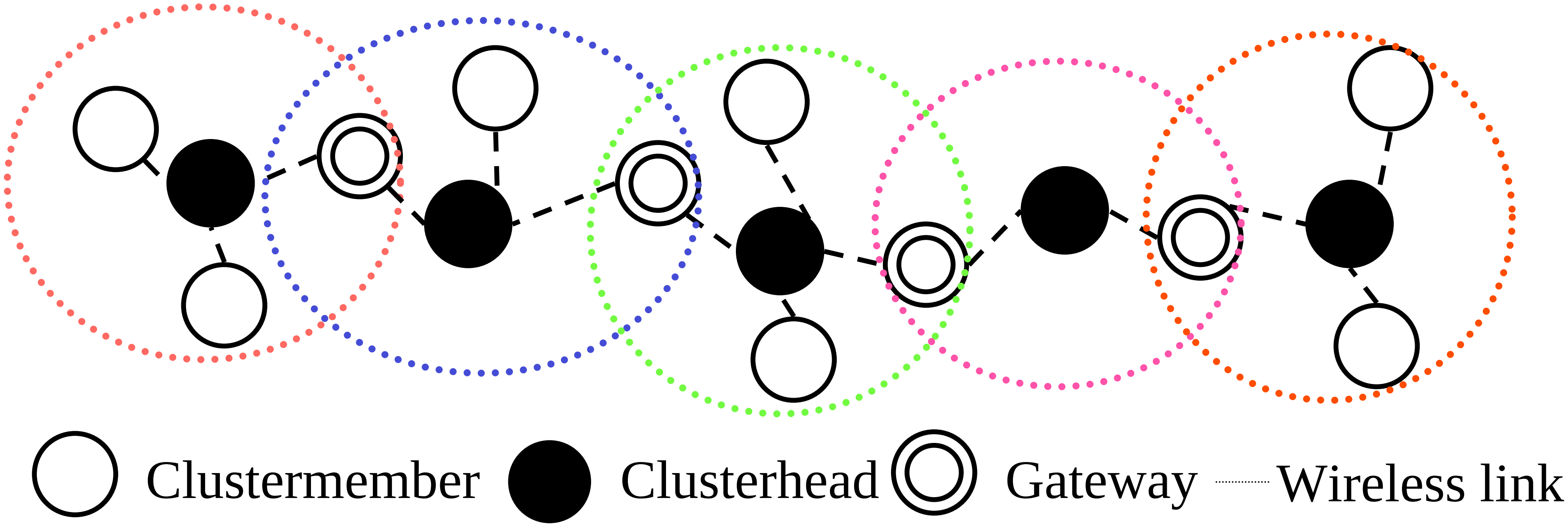
Deficiency of Conventional Scheme (1)

- Unevenly distributed node density has a big impact on network performance.
  - Too High-Density The control node has a large overhead from managing its routing table.

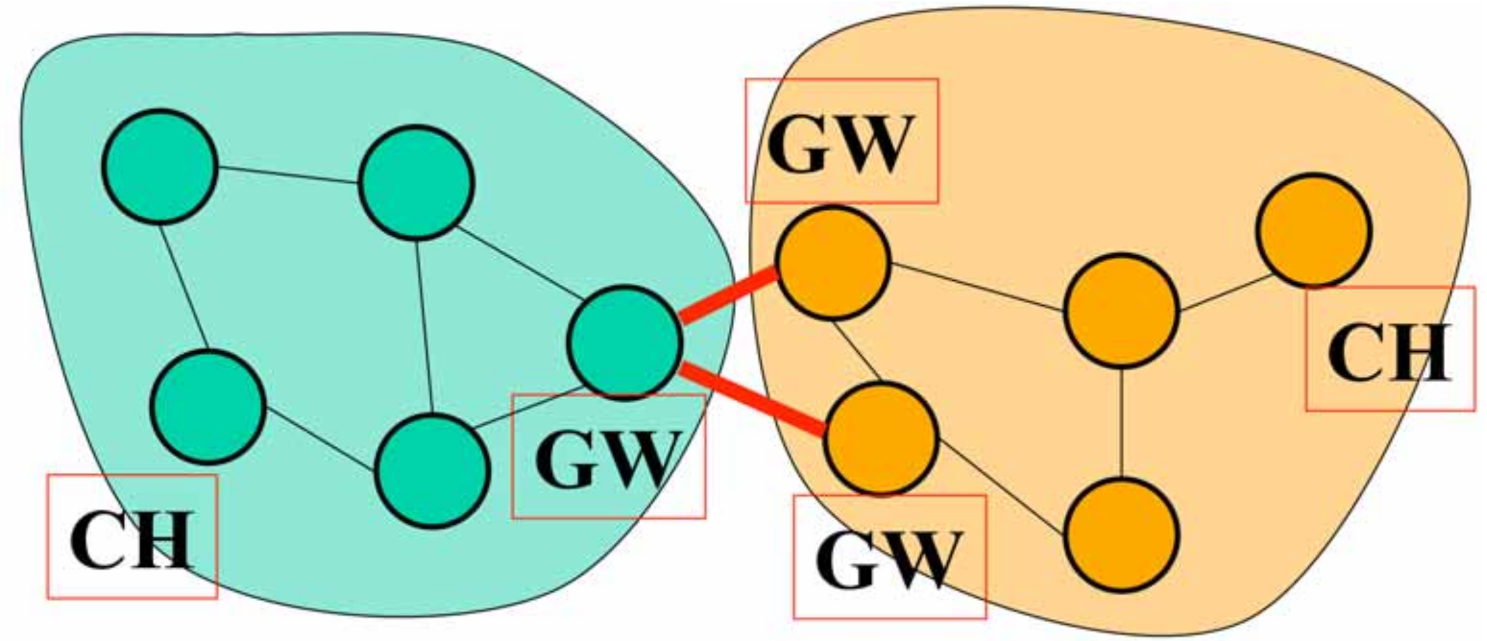


Deficiency of Conventional Scheme (2)

- Too Low-Density The benefits of a hierarchical structure are not apparent, because there are many small clusters in the network.

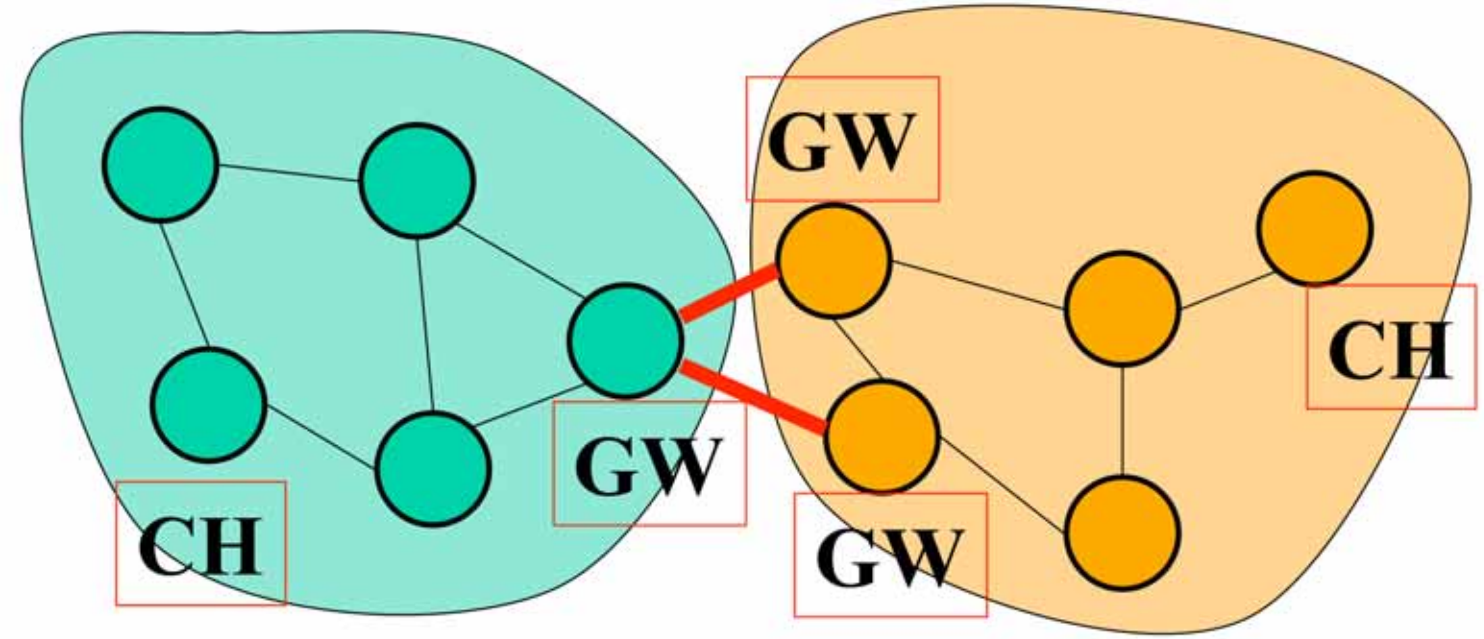


Proposed Clustering Scheme (1)



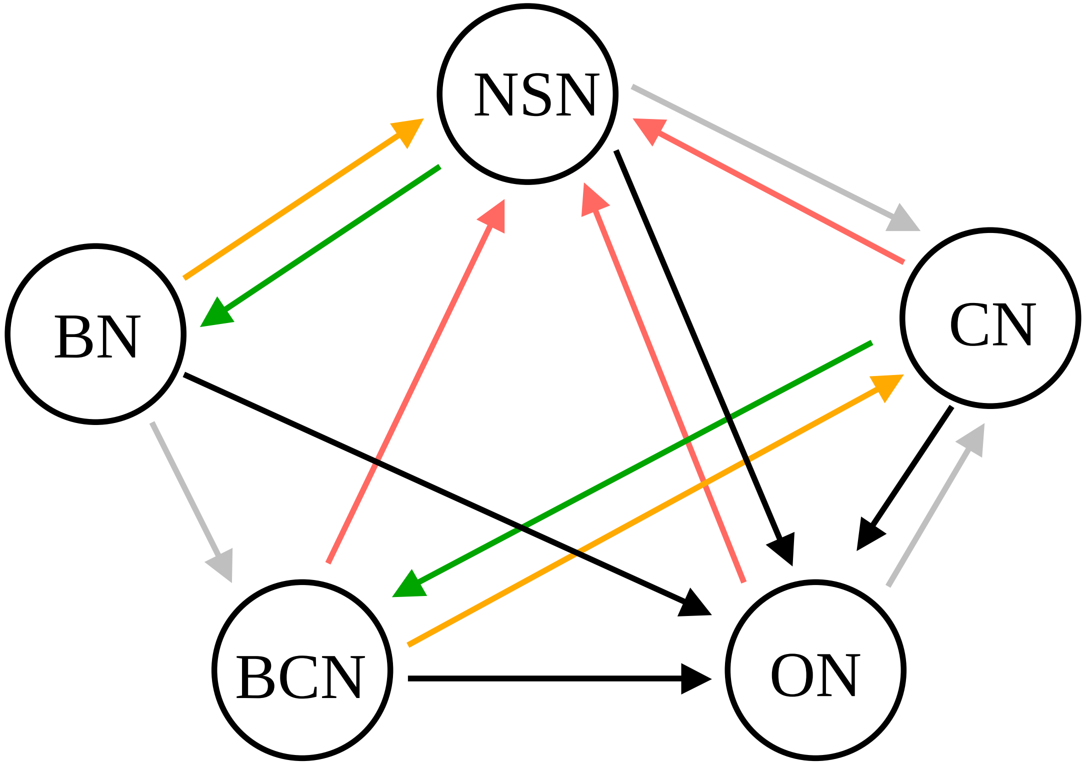
- The cluster consists of one clusterhead (CH), one or more gateways (GW), and clustermembers.
- When a node in a cluster communicates with a node in its neighboring cluster, packets are forwarded through only the GWs.

Proposed Clustering Scheme (2)



- Clusterhead works to manage the cluster.
- Gateway works to get the information of a neighboring cluster.

State Transition Diagram in Each Node



Transition A : Add the role of a gateway to a node.

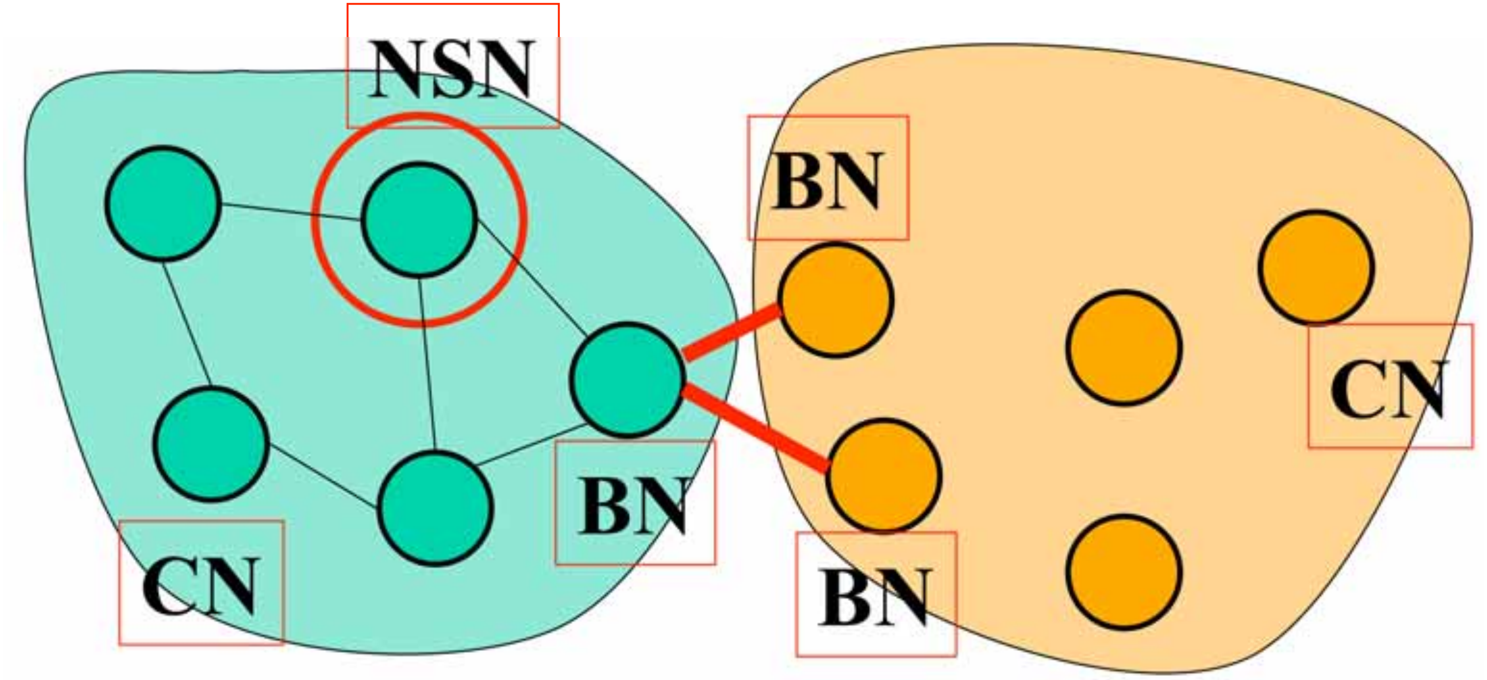
Transition B : Change to a clustermember.

Transition C : Add the role of a gateway to a node.

Transition D : Delete the role of a gateway from a node.

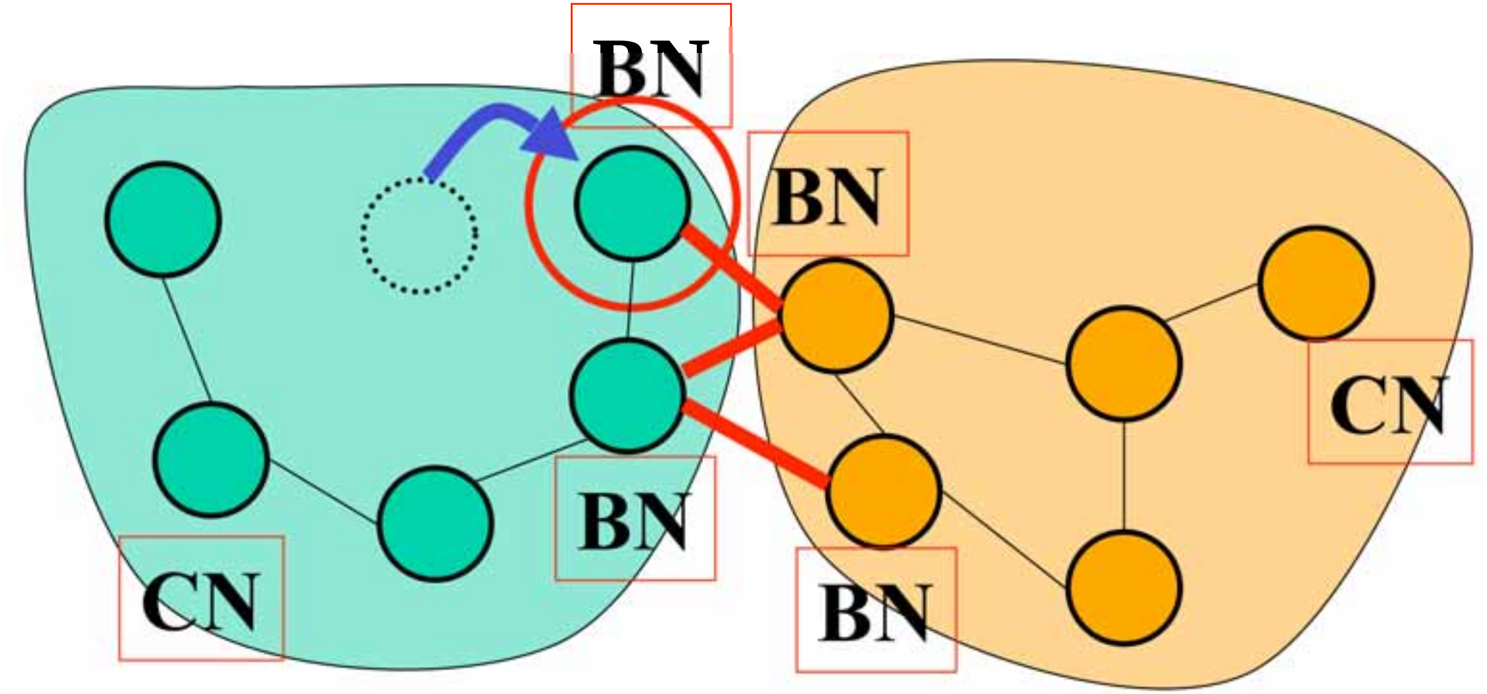
Transition E : Change to Orphan Node.

Example of Maintenance (1)



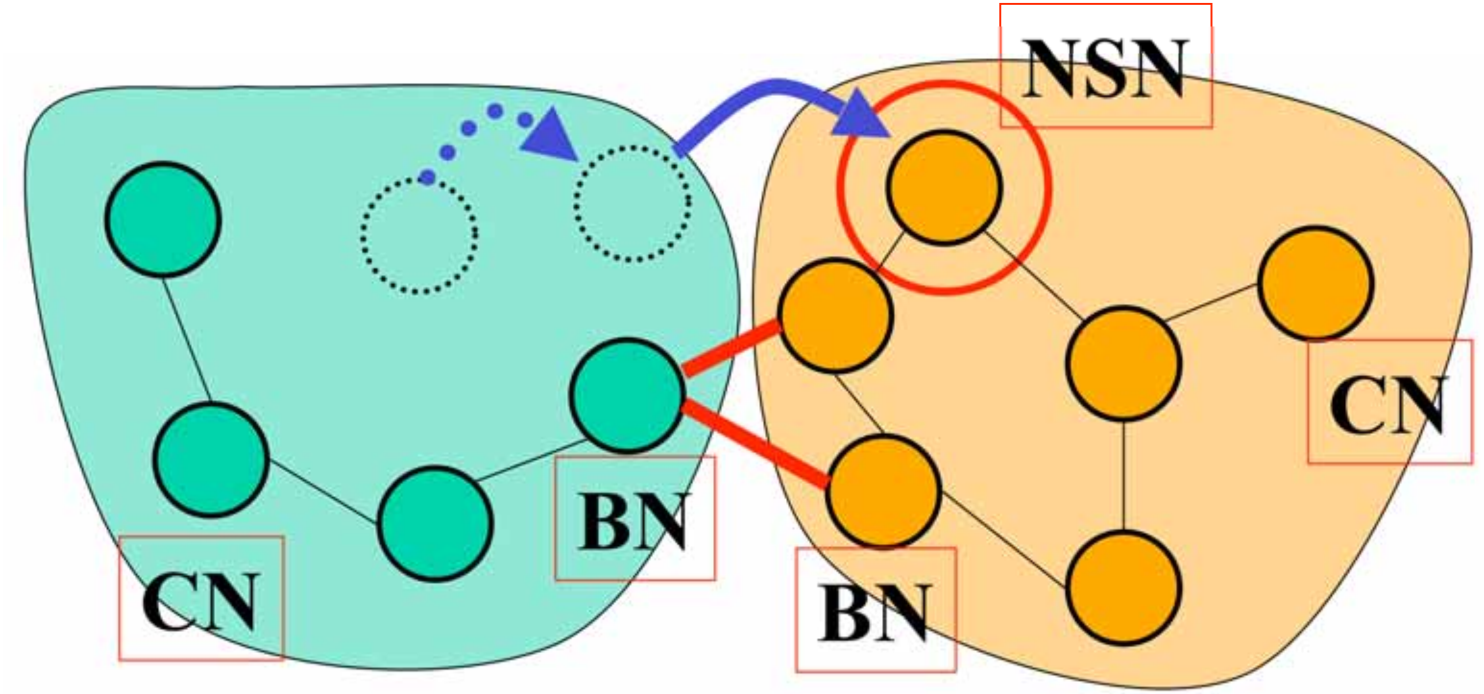
- The current state of ◉ is NSN because the node has neighboring nodes all of which belong to green cluster.

Example of Maintenance (2)



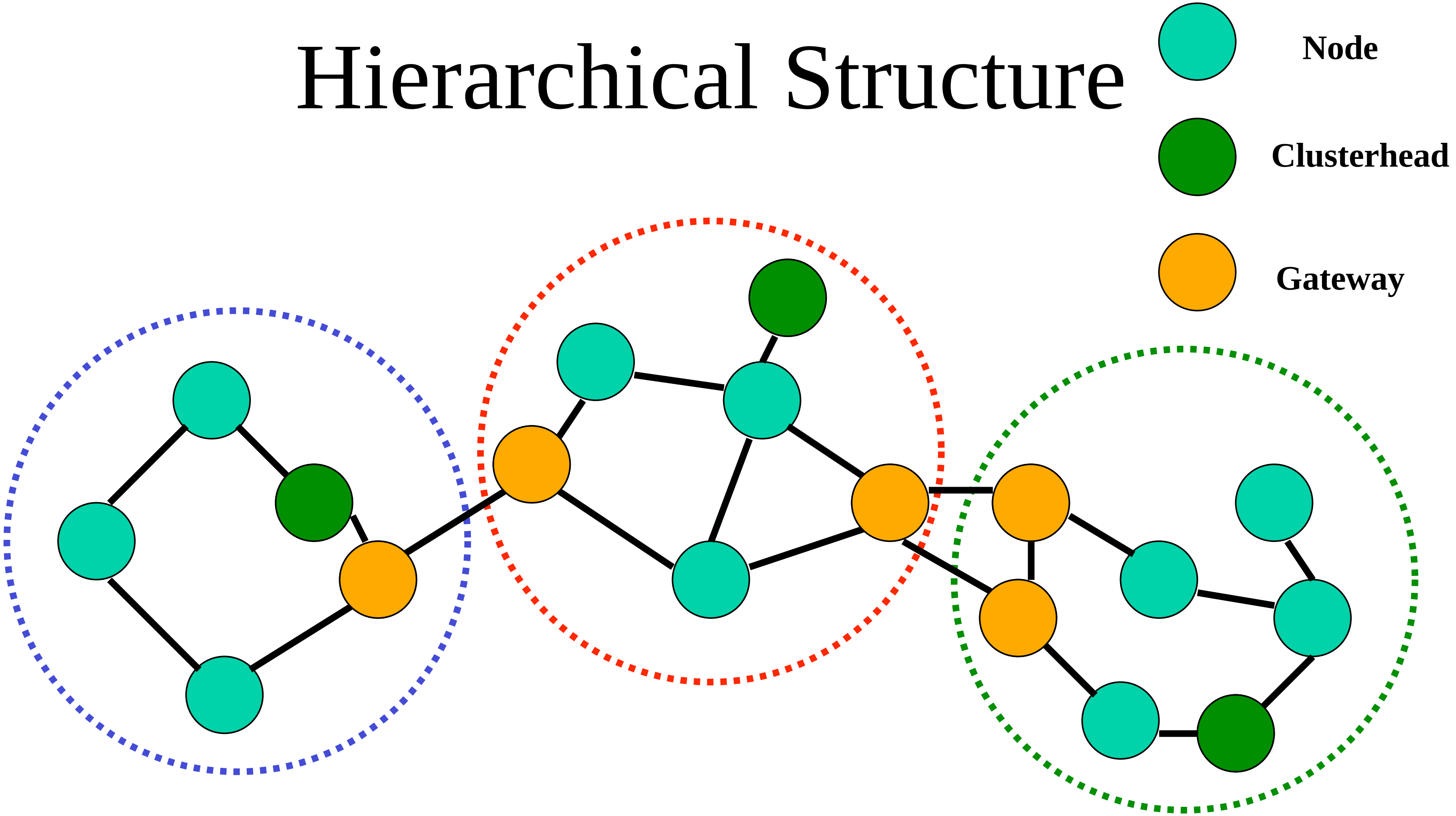
- The node changes the state to the border node because the node has some neighboring nodes which belong to orange cluster.
- It works to get the information of the neighboring cluster represented by orange.

Example of Maintenance (3)



- The node changes the cluster ID to orange cluster because the node has neighboring nodes all of which belong to orange cluster.

Hierarchical Structure



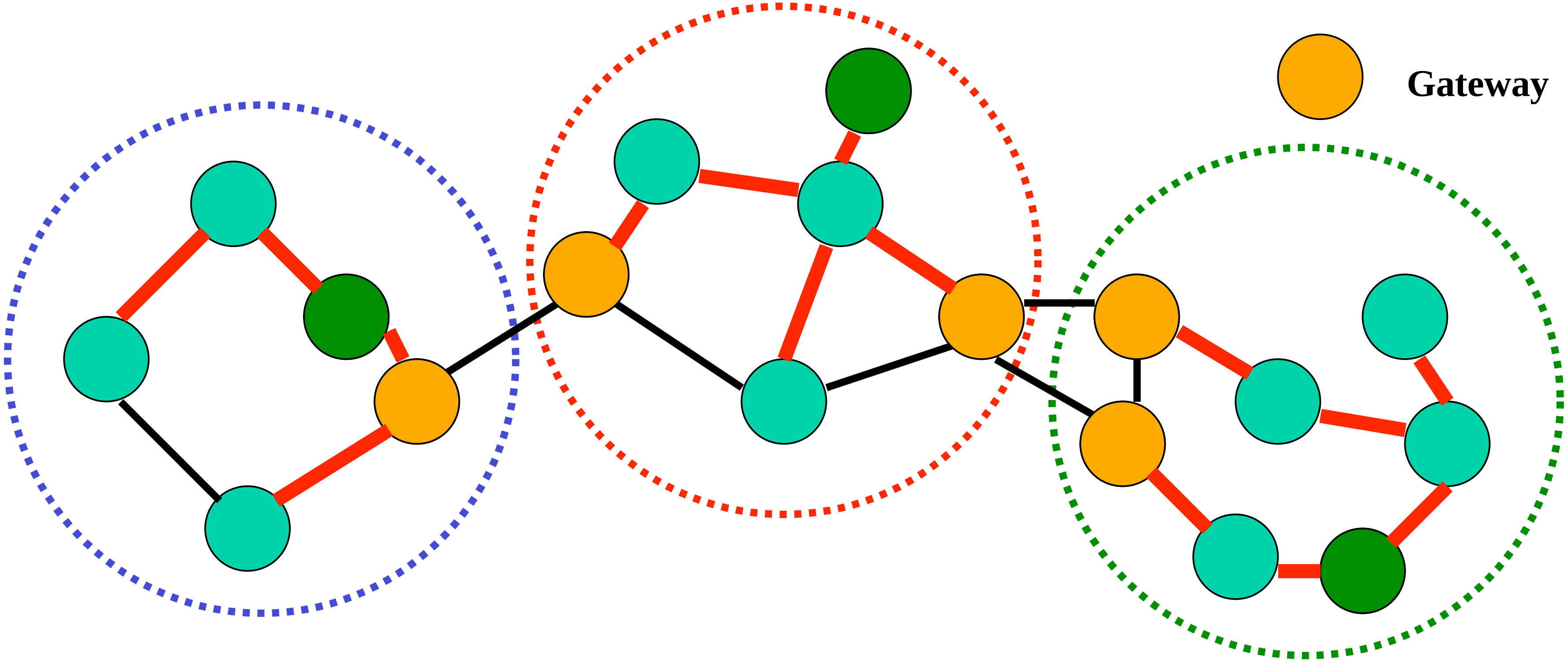
- The entire network is divided into multiple clusters.
- The cluster size is managed by the number of nodes in the cluster (Upper bound and Lower bound).

Hierarchical Structure

● **Node**

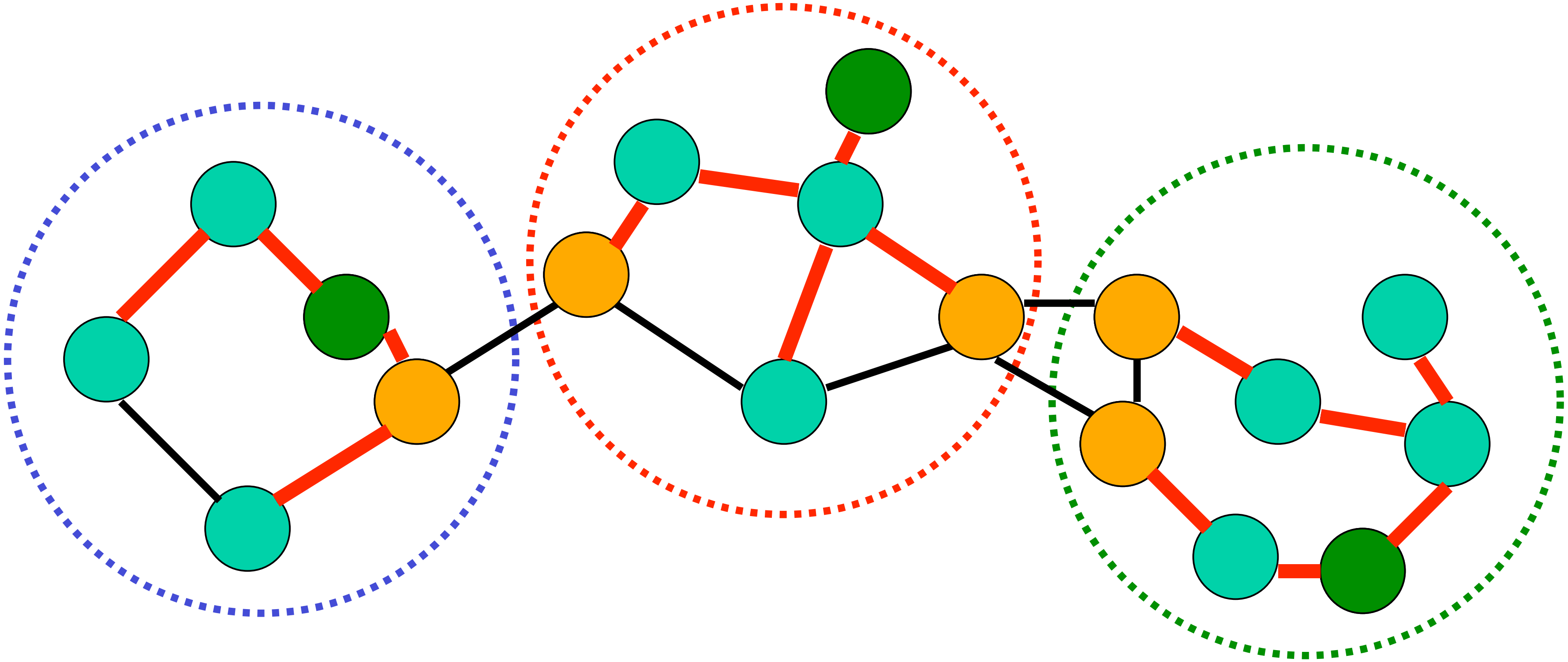
● **Clusterhead**

● **Gateway**



- A spanning tree at which the clusterhead is rooted is constructed.

Hierarchical Routing Protocol Based on Autonomous Clustering



- By regarding each cluster as one node, the route are constructed.

Within cluster ･･･ Spanning tree is used.

Among clusters ･･･ TORA, DSR, or AODV is used.

Effect of Autonomous Clustering Scheme

- Among clusters
  - By regarding each cluster as a virtual node, the routing protocol works just like in the small network.
- Within cluster
  - The route within cluster is stable because the clustering is provided by the autonomous clustering scheme and the proper cluster size.

Evaluation Purpose

- In large mobile ad hoc network environment, we compare proposed hierarchical routing protocols with conventional flat routing protocols.
  - Overhead
    - Measuring the number of control packets to maintain the route between a source node and a destination node.
  - Stability of route
    - Measuring the number of data packets which destination nodes could receive.

Network Simulator 2(NS2)

Node Mobility Model

- Random waypoint model
  1. A node moves at a specified speed to a position which is selected randomly.
  2. At the position, the node stays for a specified period (which is called “pause time”).
  3. Return 1.

Pause time is 0 in our simulation.

Simulation Models

- Conventional simulation models
  - – Field size ･･･ 1200mx300m
  - – Number of nodes ･･･ 50
- Our simulation model
  - – Field size ･･･ 2000mx1500m（8.3 times）
  - – Number of nodes ･･･ 150（3 times）

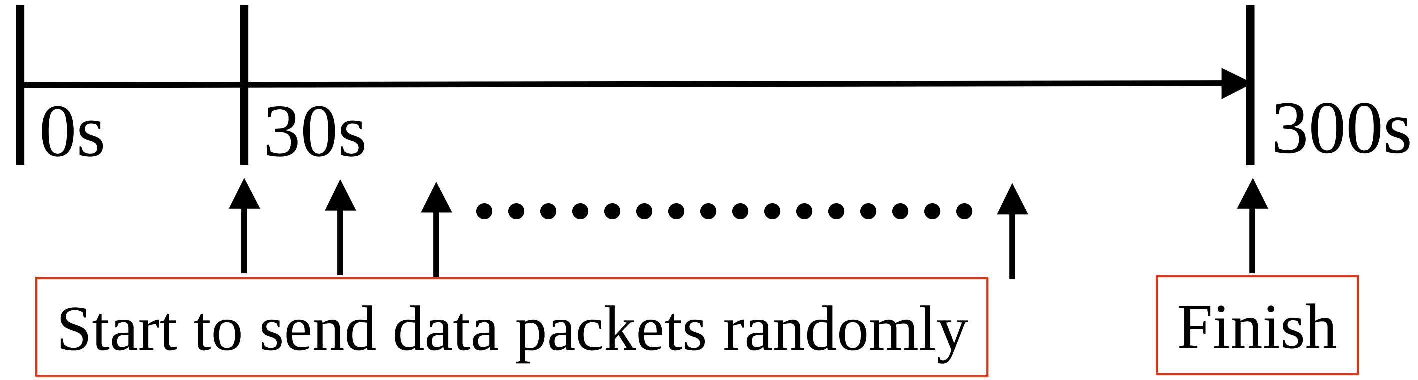
Simulation Method

- Number of nodes ･･･150
- Movement model･･･Random waypoint model
- Field size･･･2000m x 1500m
- Range of wireless link ･･･ 250m
- Cluster size･･･Upper 50, Lower 20

In comparison with conventional simulation models, the field size is 8.3 times and the number of nodes is 3 times.

Simulation Method（cont.）

- Simulation time：300 sec.
- # of SD pairs：10，20，30



0s

30s

300s

Start to send data packets randomly

Finish

- Maximum Node Moving Speed
  - **1m/s (3.6km/h), 2m/s (7.2km/h), 3m/s (10.8km/h), 4m/s (14.4km/h), 5m/s (18.0km/h), 10m/s (36.0km/h), 15m/s (54.0km/h), 20m/s (72.0km/h)**

About Data Packets

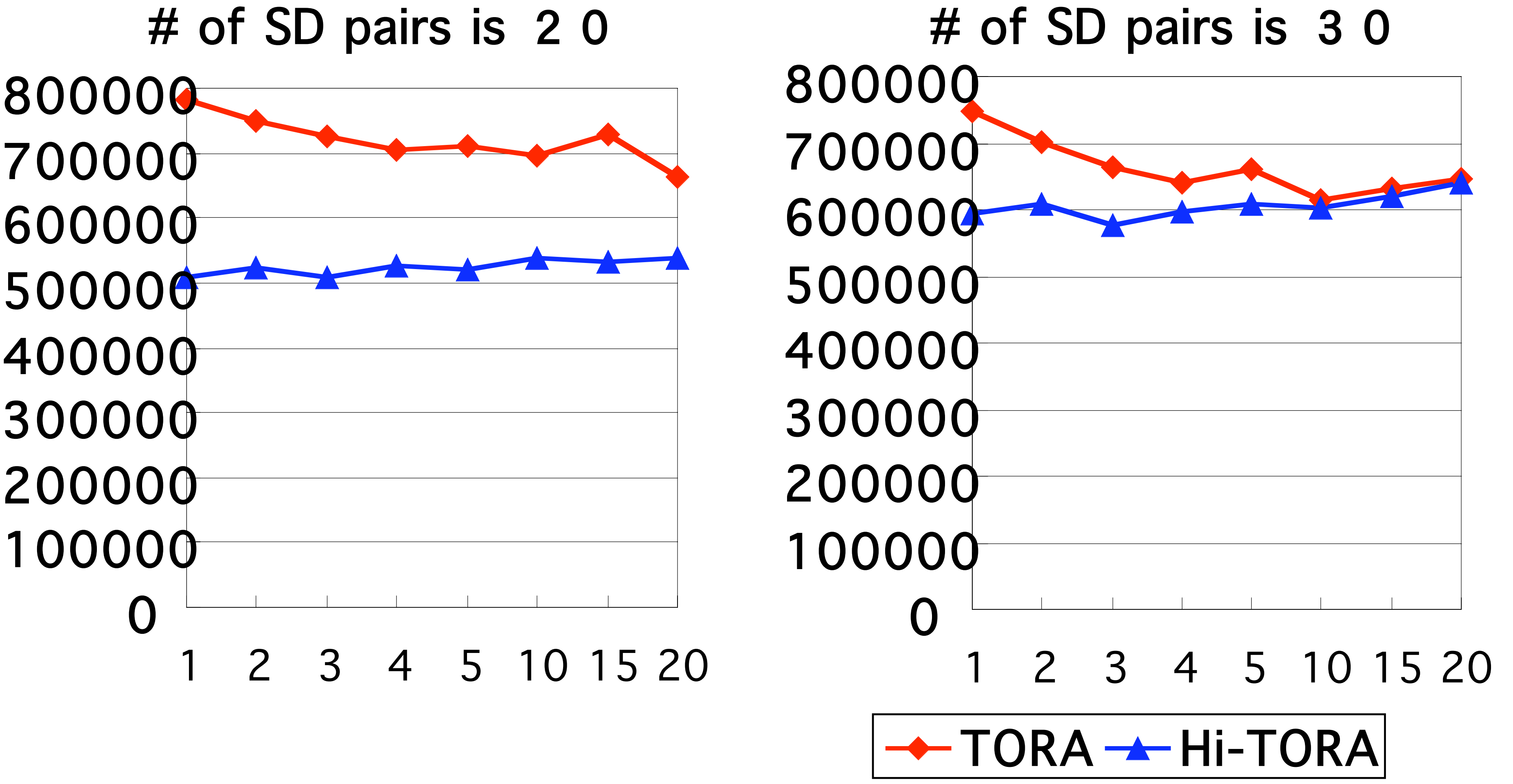
- Total number of data packets which source nodes send
  - Packet size･･･512byte
  - # of SD pair is 10 ･･･about 9000
  - # of SD pair is 20 ･･･about 18000
  - # of SD pair is 30 ･･･about 27000

Interval of sending : 250msec

Simulation Experiment 1

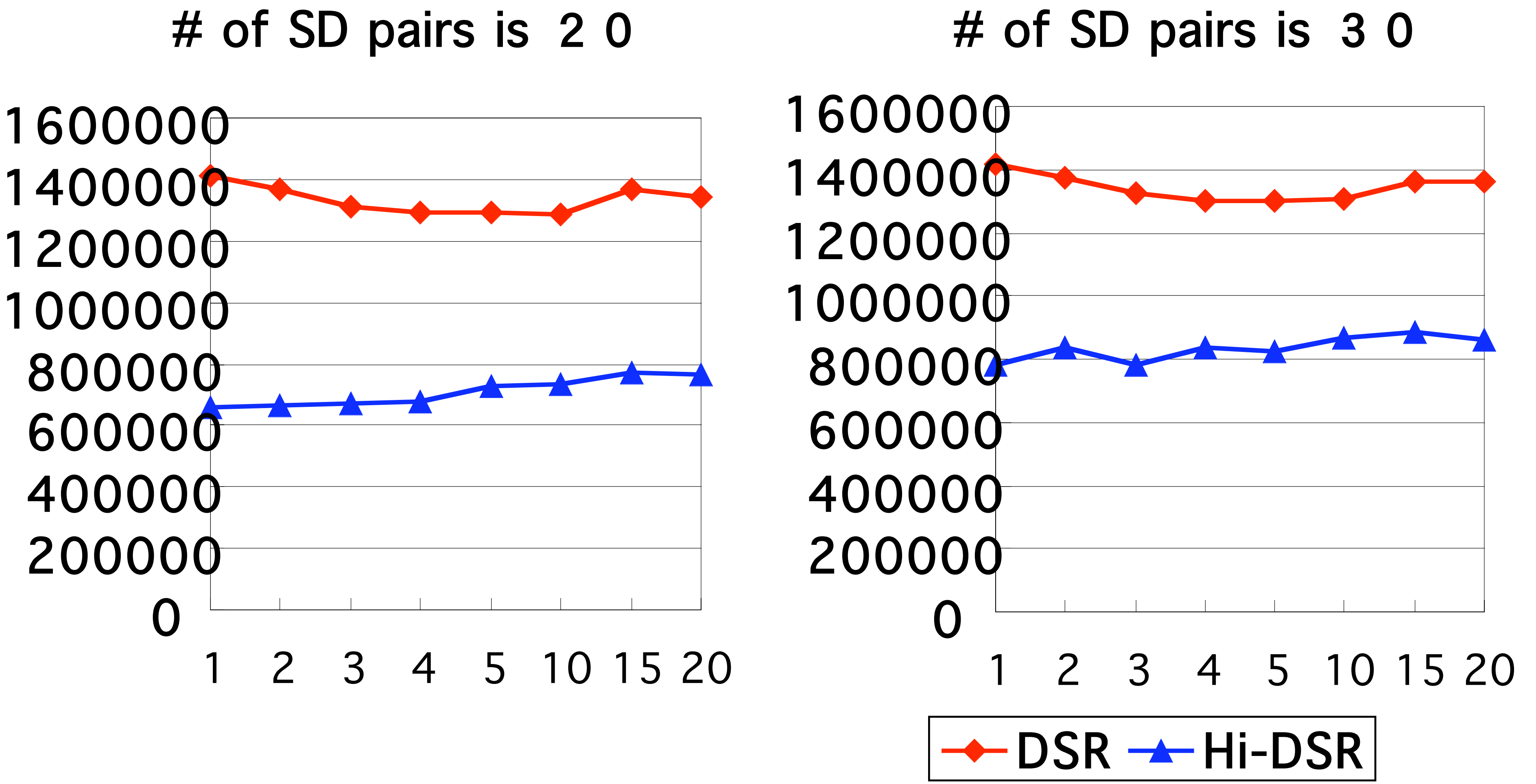
- We evaluated the total number of control packets.
- Types of control packet
  - Control packets for autonomous clustering
  - Control packets for routing

Number of Control Packets TORA vs. Hi-TORA



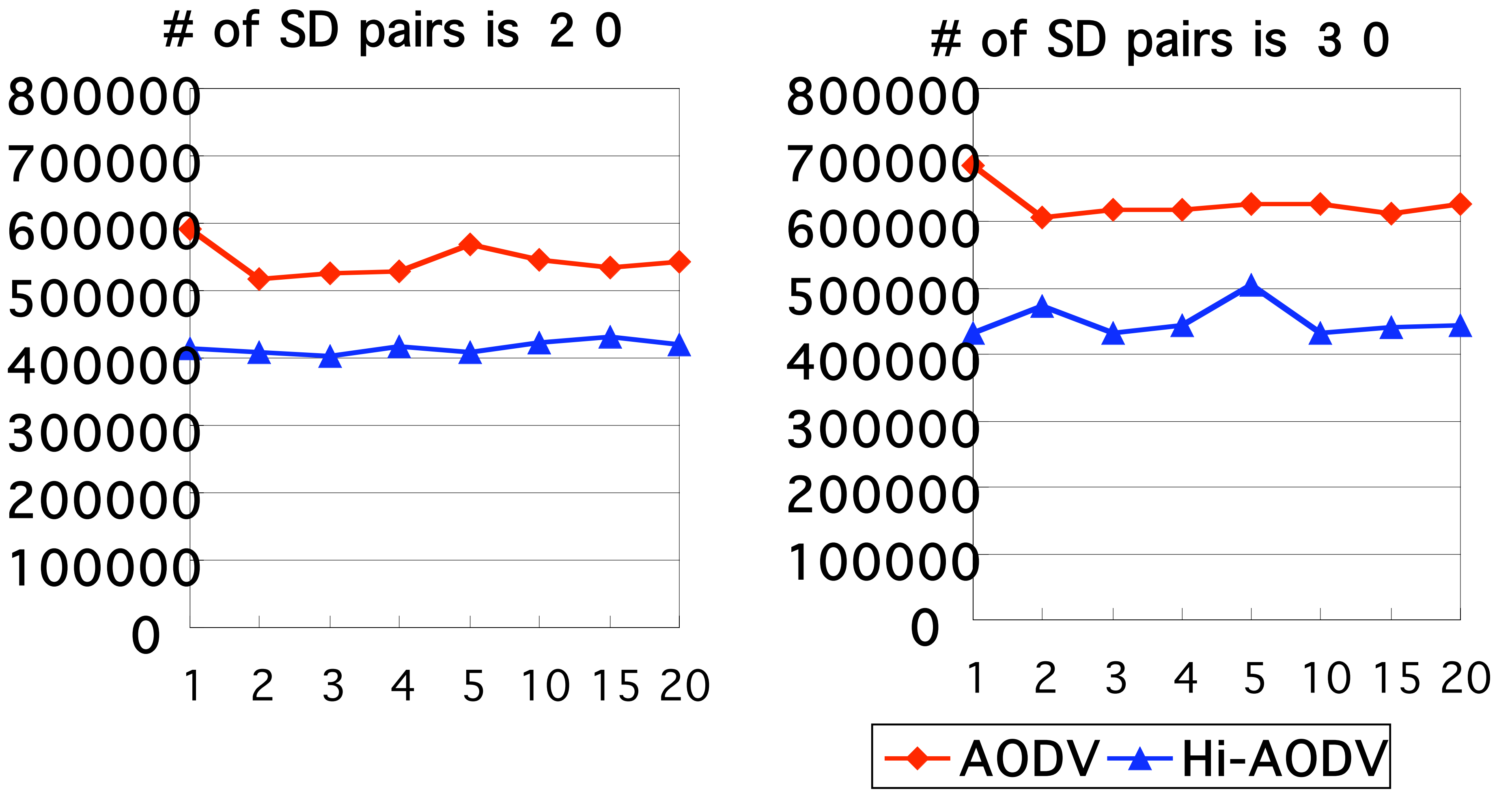
X-axis:Node moving speed (m/s) Y-axis:# of control packets

Number of Control Packets DSR vs. Hi-DSR



X-axis:Node moving speed (m/s) Y-axis:# of control packets

Number of Control Packets Hi-AODV vs. AODV

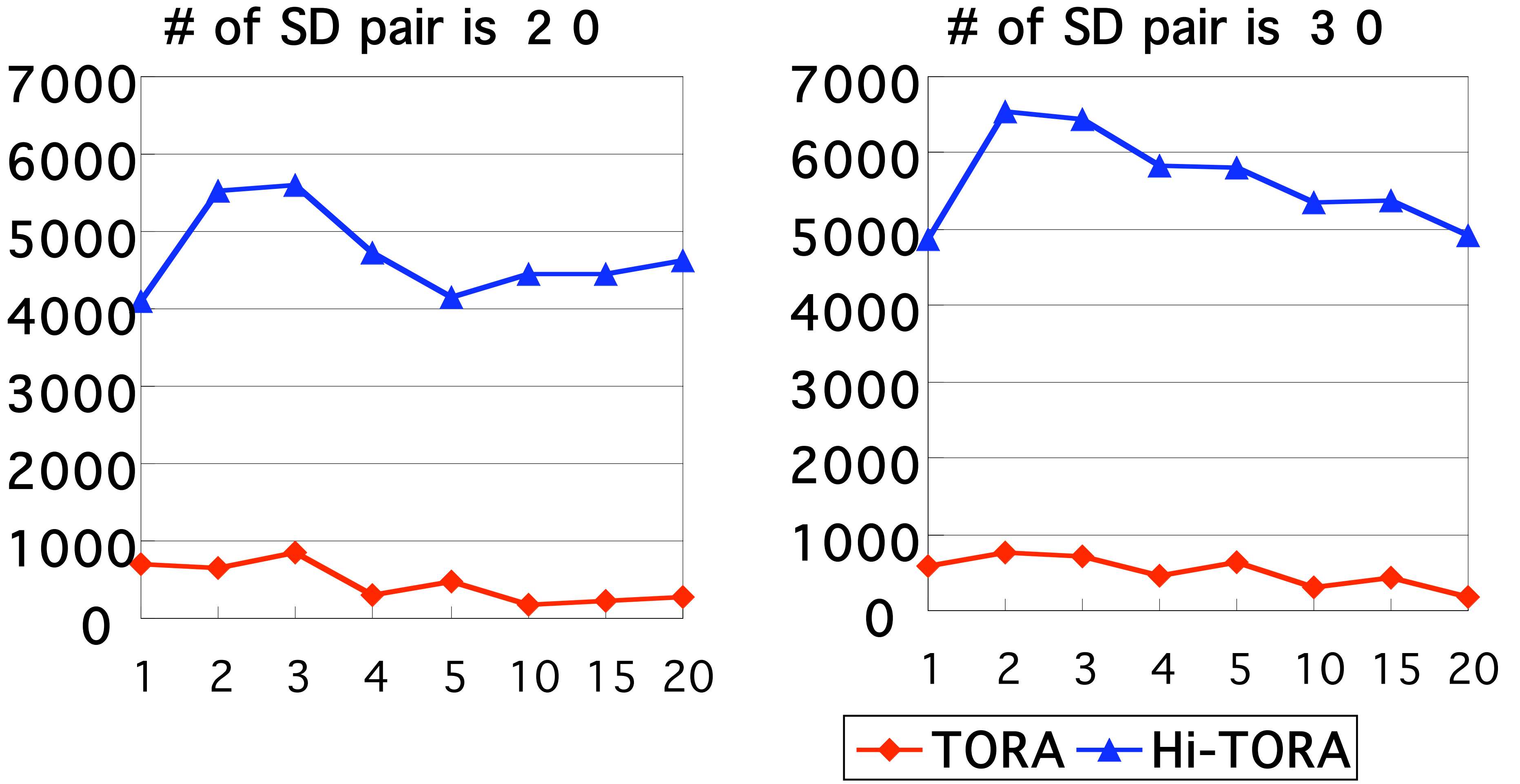


X-axis:Node moving speed (m/s) Y-axis:# of control packets

Simulation Experiment 2

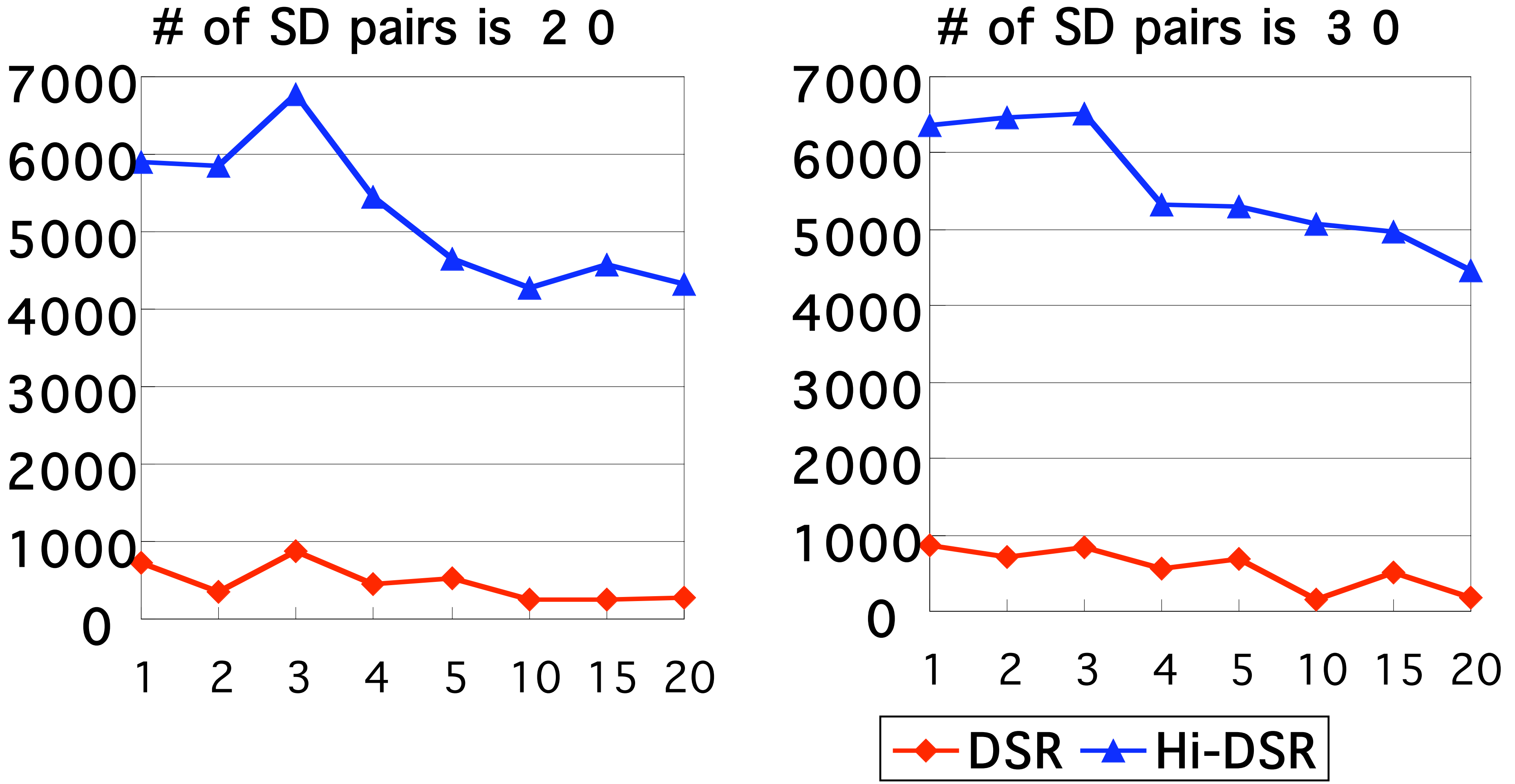
- We evaluated the number of delivered data packets.

Number of Delivered Data Packets TORA vs. Hi-TORA



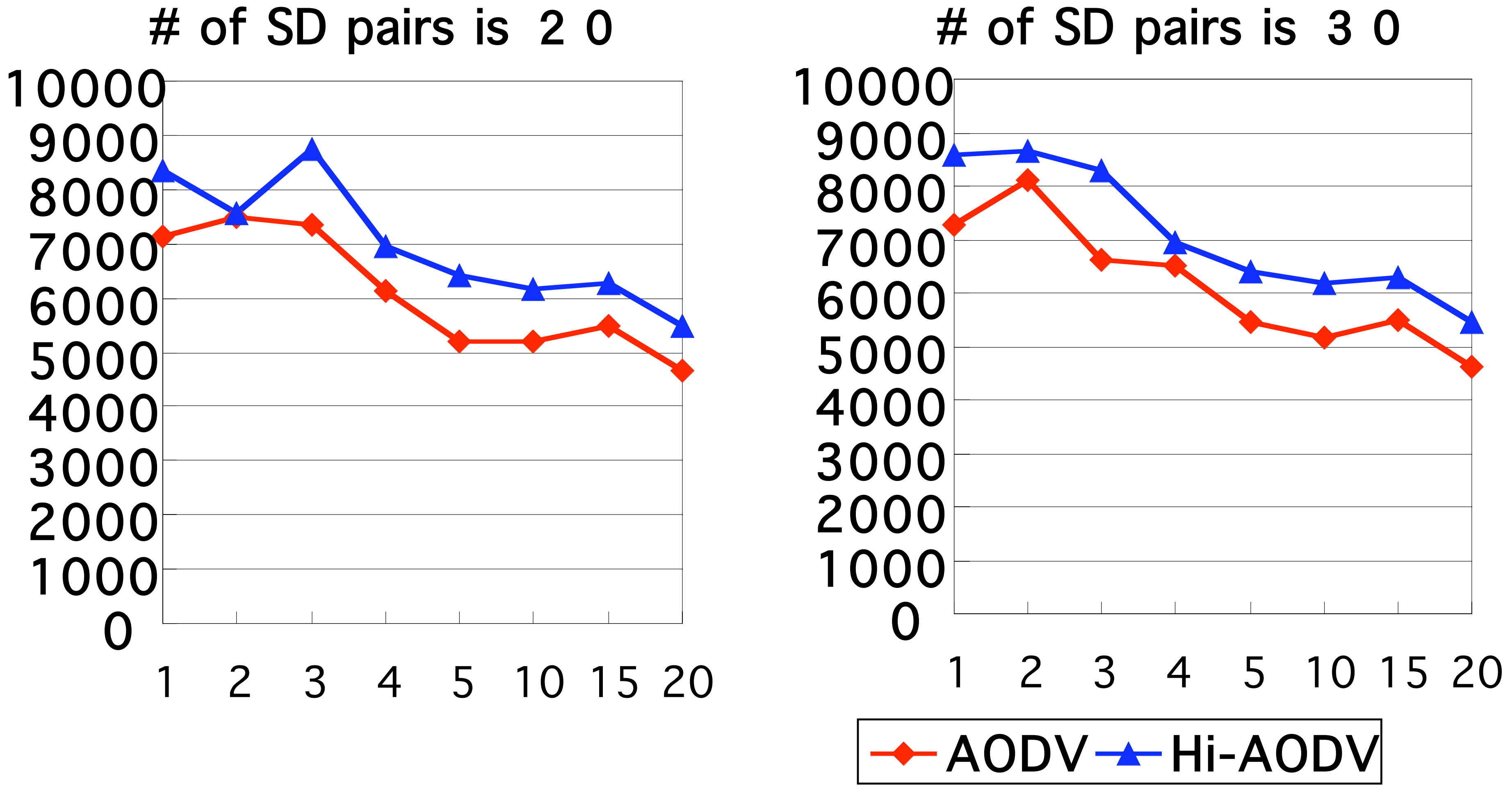
X-axis:Node moving speed (m/s) Y-axis:# of delivered data packets

Number of Delivered Data Packets DSR vs. Hi-DSR



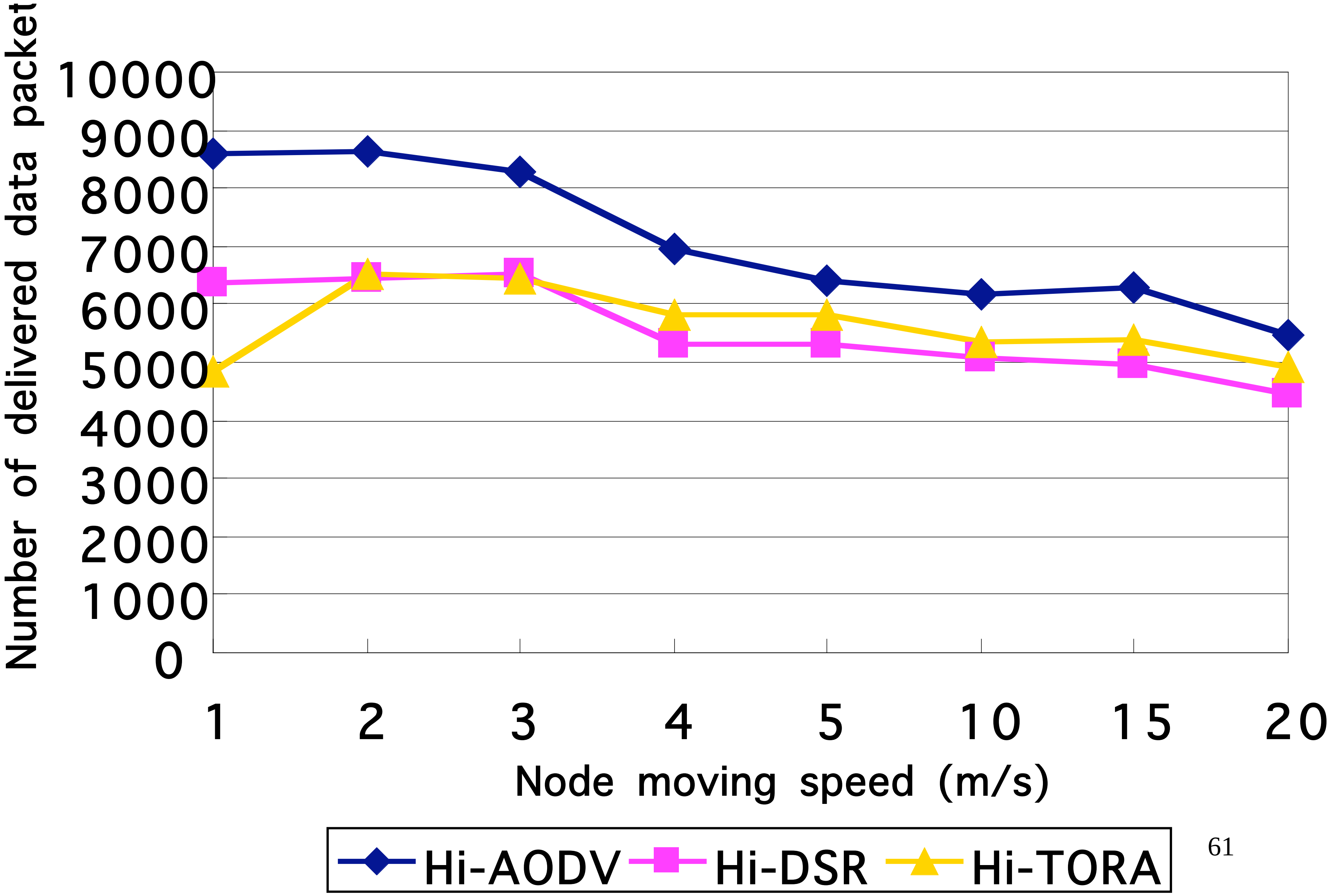
X-axis:Node moving speed (m/s) Y-axis:# of delivered data packets

Number of Delivered Data Packets AODV vs. Hi-AODV



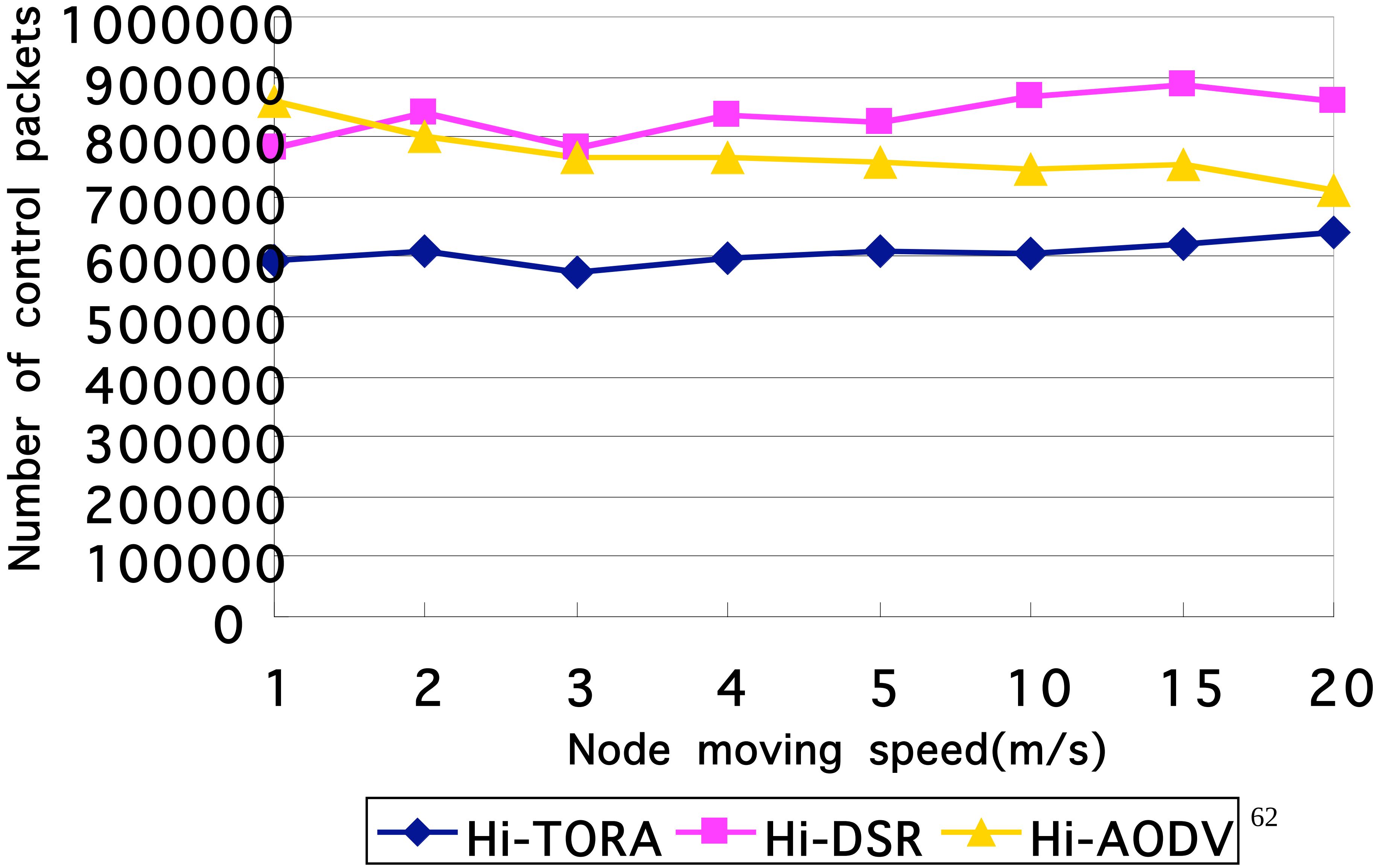
X-axis:Node moving speed (m/s) Y-axis:# of delivered data packets

Number of Delivered Data Packets



61

Number of Control Packets



62

Observation - Hierarchical Routing -

- Effect of autonomous clustering
  - By regarding each cluster as one node, the routing protocol works just like in the small network.
  - The route within cluster is stable because the clustering is provided by the autonomous clustering scheme and the proper cluster size.

Observations - Hi-AODV -

- Hi-AODV is the best hierarchical routing protocol as shown in the result of delivered data packets
  - Effect of autonomous clustering
  - Different from Hi-TORA and Hi-DSR, when the route disappeared in an intermediate cluster, the overhead becomes low because the intermediate cluster repairs the route locally. As a result, Hi-AODV provides the most stable routes.

Evaluation of Hierarchical Routing Protocols (Control packets)

	Hi-TORA	TORA	Hi-DSR	DSR	Hi-AODV	AODV
Node Moving Speed	○	×	○	×	○	○
# of SD	○	×	○	×	○	△
Decreasing Rate	20%	—	50%	—	20%	—

Evaluation of Hierarchical Routing Protocols (Data packets)

	Hi-TORA	TORA	Hi-DSR	DSR	Hi-AODV	AODV
Node Moving Speed	○	×	○	×	○	△
# of SD	○	×	○	×	○	△
Increasing Rate	50%	—	50%	—	10%	—

Conclusion and Future Work

- Conclusion
  - Hierarchical routing protocols based on the autonomous clustering scheme provide the stable route in comparison with flat routing protocols.
  - We have applied for a patent on the autonomous clustering.
- Future Work
  - Developing a framework of hierarchical routing protocol based on the autonomous clustering scheme.

Challenging Issues in Routing for Mobile Ad Hoc Networks

- Routing for large-scale networks
- Routing for asymmetric networks
- Location-based routing
- Energy efficient routing
- Secure routing
- QoS routing

The Crumbling Perimeter: Mobile Computing and Internal Security Issues

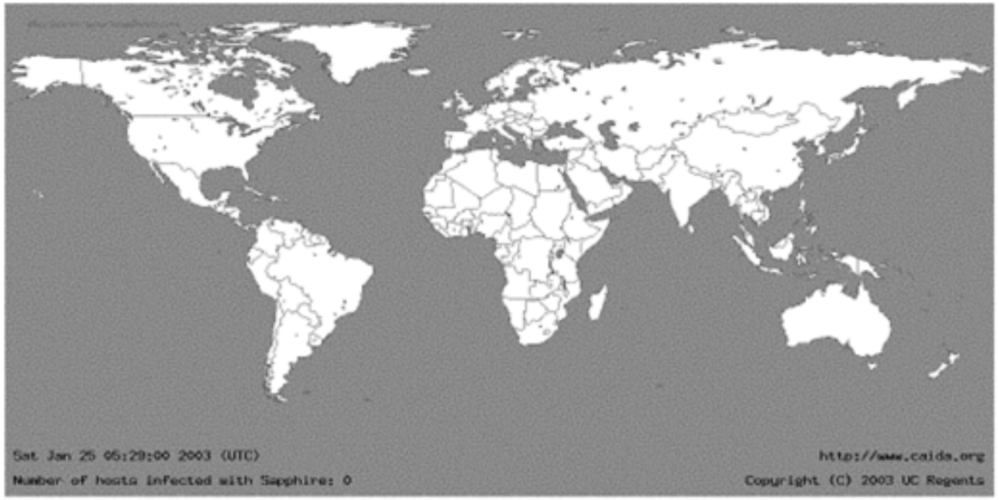
Farnam Jahanian
Arbor Networks and University of Michigan

IFIP Working Group 10.4
July 1-5, 2005

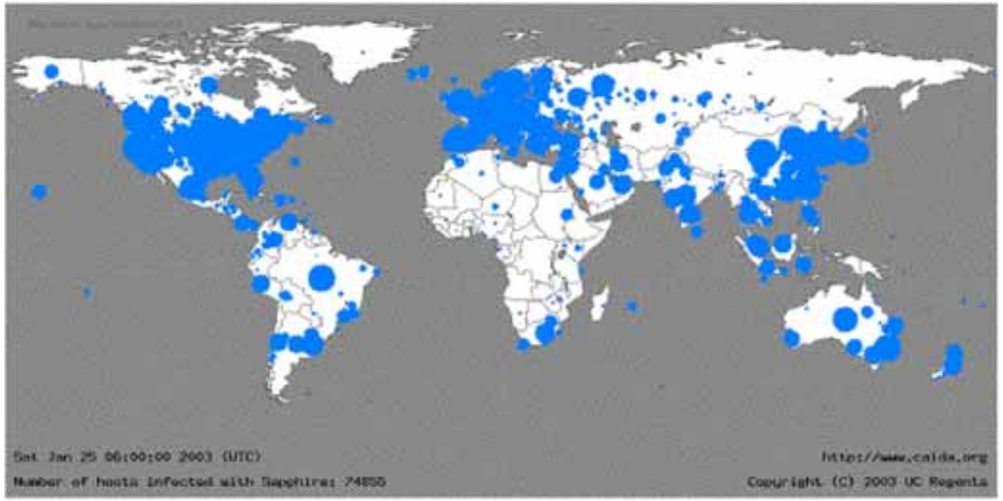
Trends in Internet Security Threats

- ***Globally scoped***, respecting no geographic or topological boundaries.
  - At peak, 5 Billion infection attempts per day during Nimda including significant numbers of sources from Korea, China, Germany, and the US. [Arbor Networks, Sep. 2001]
- Exceptionally ***virulent***, propagating to the entire vulnerable population in the Internet in a matter of minutes.
  - During Slammer, 75K hosts infected in 30 min. [Moore et al, NANOG February, 2003]
- ***Zero-day*** threats, exploiting vulnerabilities for which no signature or patch has been developed.
  - In Witty, "victims were compromised via their firewall software the day after a vulnerability in that software was publicized”

SQL Slammer Attack Propagation


Sat Jan 25 05:29:00 2003 (UTC)
Number of hosts infected with Sapphire: 0
http://www.caida.org
Copyright (C) 2003 UC Regents

0 hosts infected at the start


Sat Jan 25 06:00:00 2003 (UTC)
Number of hosts infected with Sapphire: 74855
http://www.caida.org
Copyright (C) 2003 UC Regents

75,000 hosts infected in 30 min.

Infections doubled every 8.5 sec.

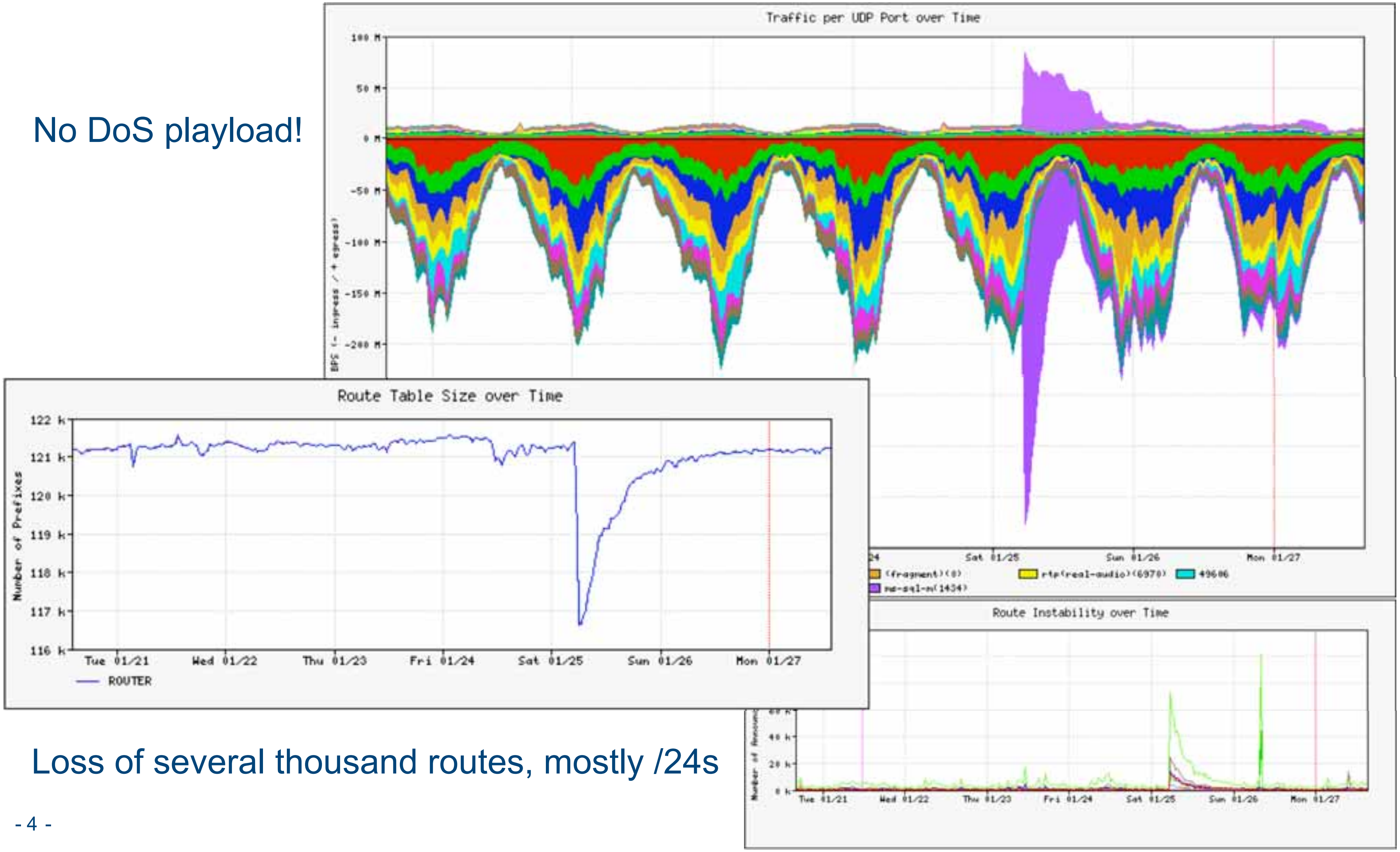
Spread 100X faster than Code Red

At peak, scanned 55M hosts per sec.

[Moore, Paxson, et al; NANOG February, 2003]

Impact of Slammer on the Internet

No DoS playload!

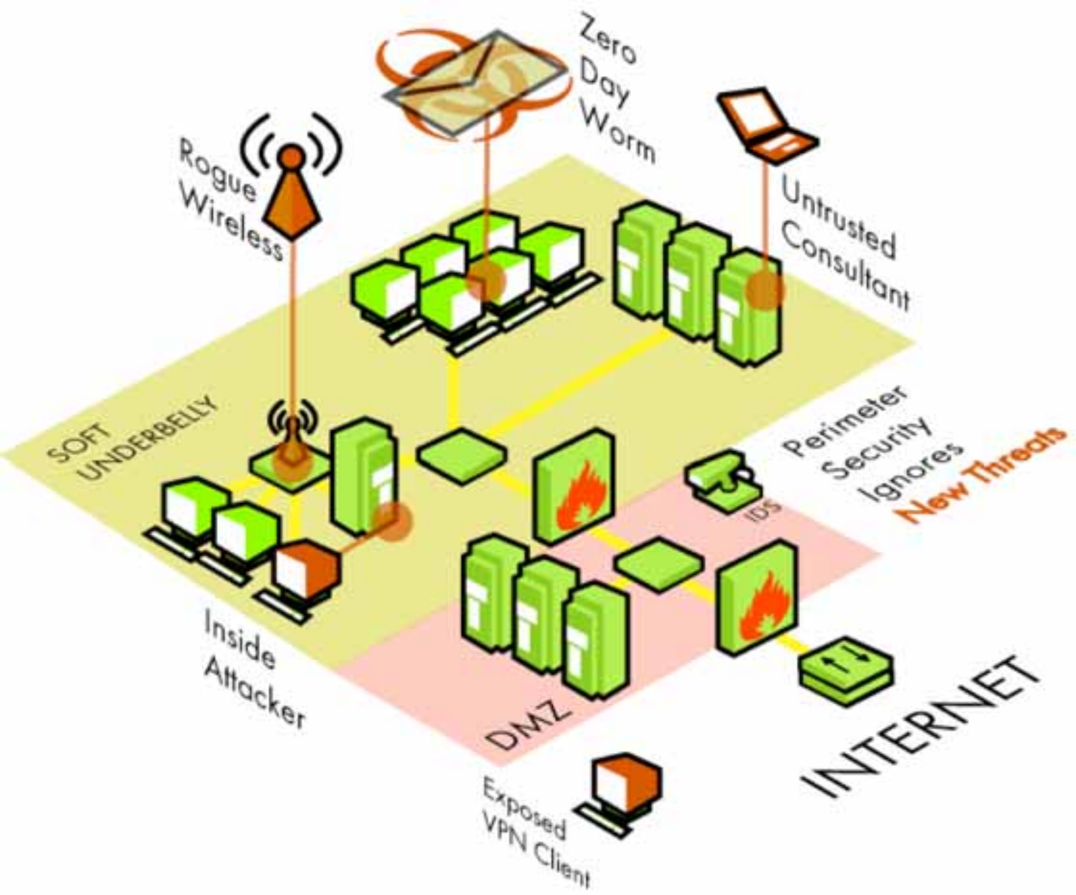


Loss of several thousand routes, mostly /24s

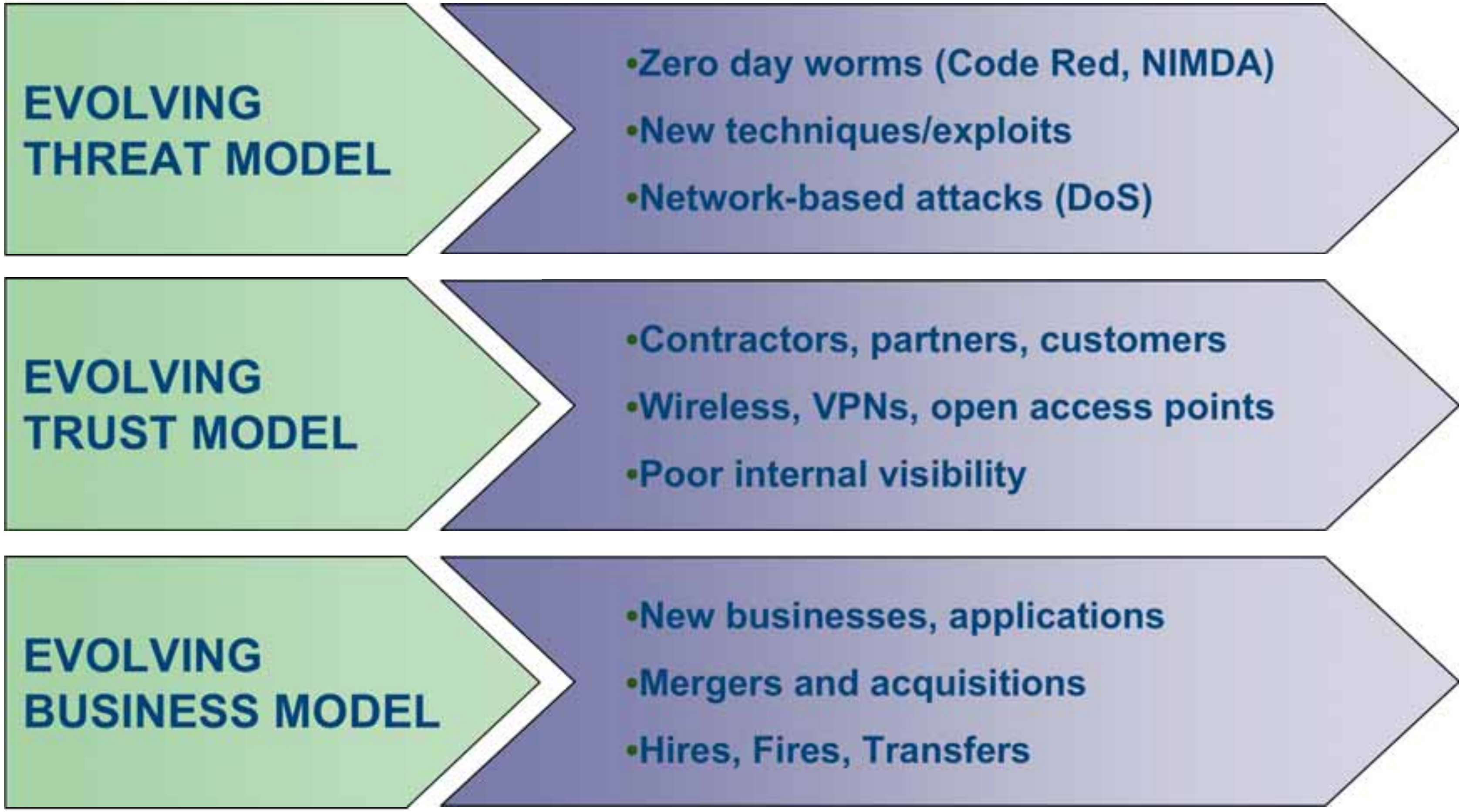
The Crumbling Perimeter

Much of perimeter security problem addressed by making perimeter vulnerability-aware (IDS, smart firewall, VA)

With crumbling perimeter (wireless, tunnels, etc) and near-zero visibility, internal network security has emerged as the most pressing IT security issue



Internal Security Challenge: The Soft Underbelly



EVOLVING THREAT MODEL

- Zero day worms (Code Red, NIMDA)
- New techniques/exploits
- Network-based attacks (DoS)

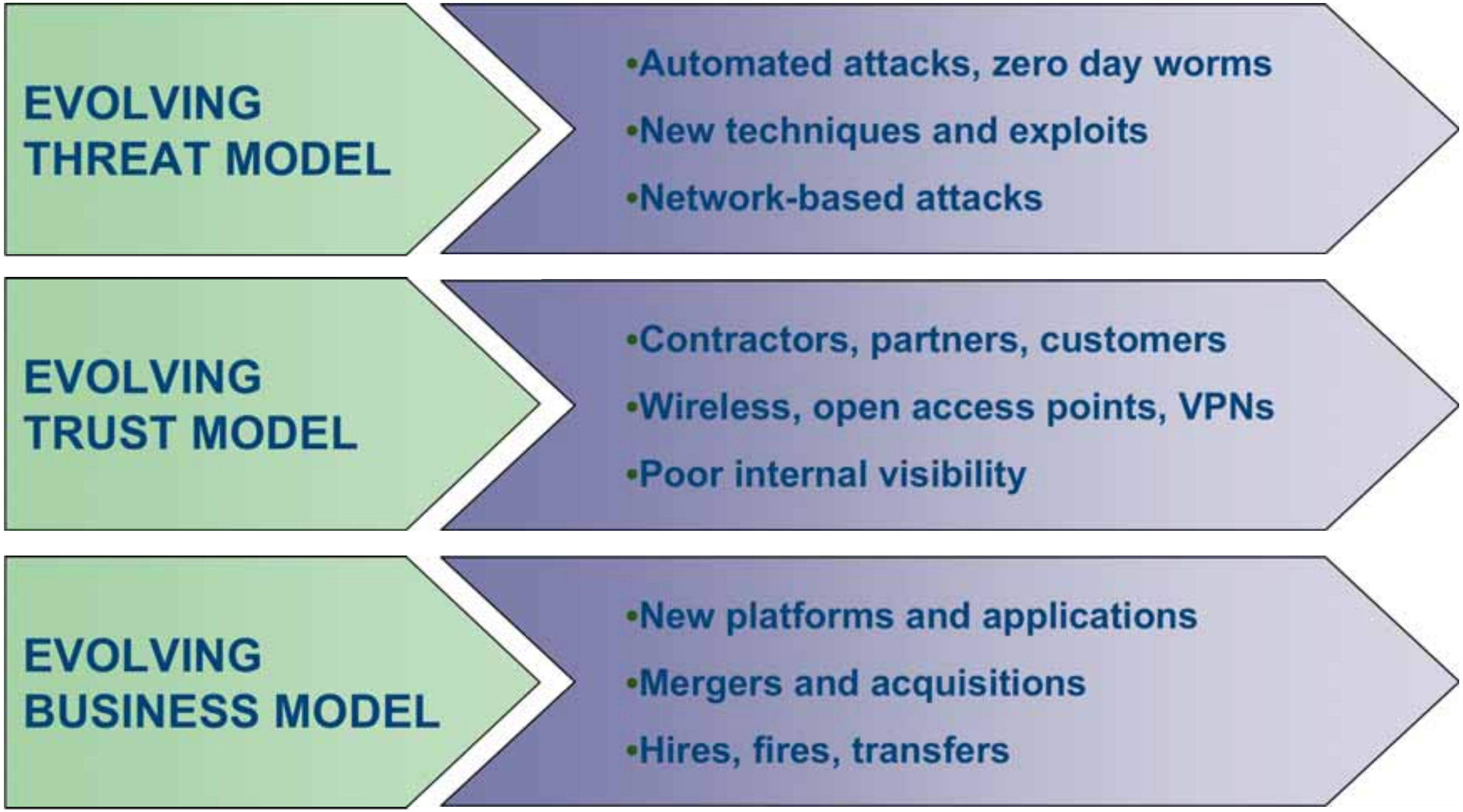
EVOLVING TRUST MODEL

- Contractors, partners, customers
- Wireless, VPNs, open access points
- Poor internal visibility

EVOLVING BUSINESS MODEL

- New businesses, applications
- Mergers and acquisitions
- Hires, Fires, Transfers

Internal Security Challenge: The Soft Underbelly



EVOLVING THREAT MODEL

- Automated attacks, zero day worms
- New techniques and exploits
- Network-based attacks

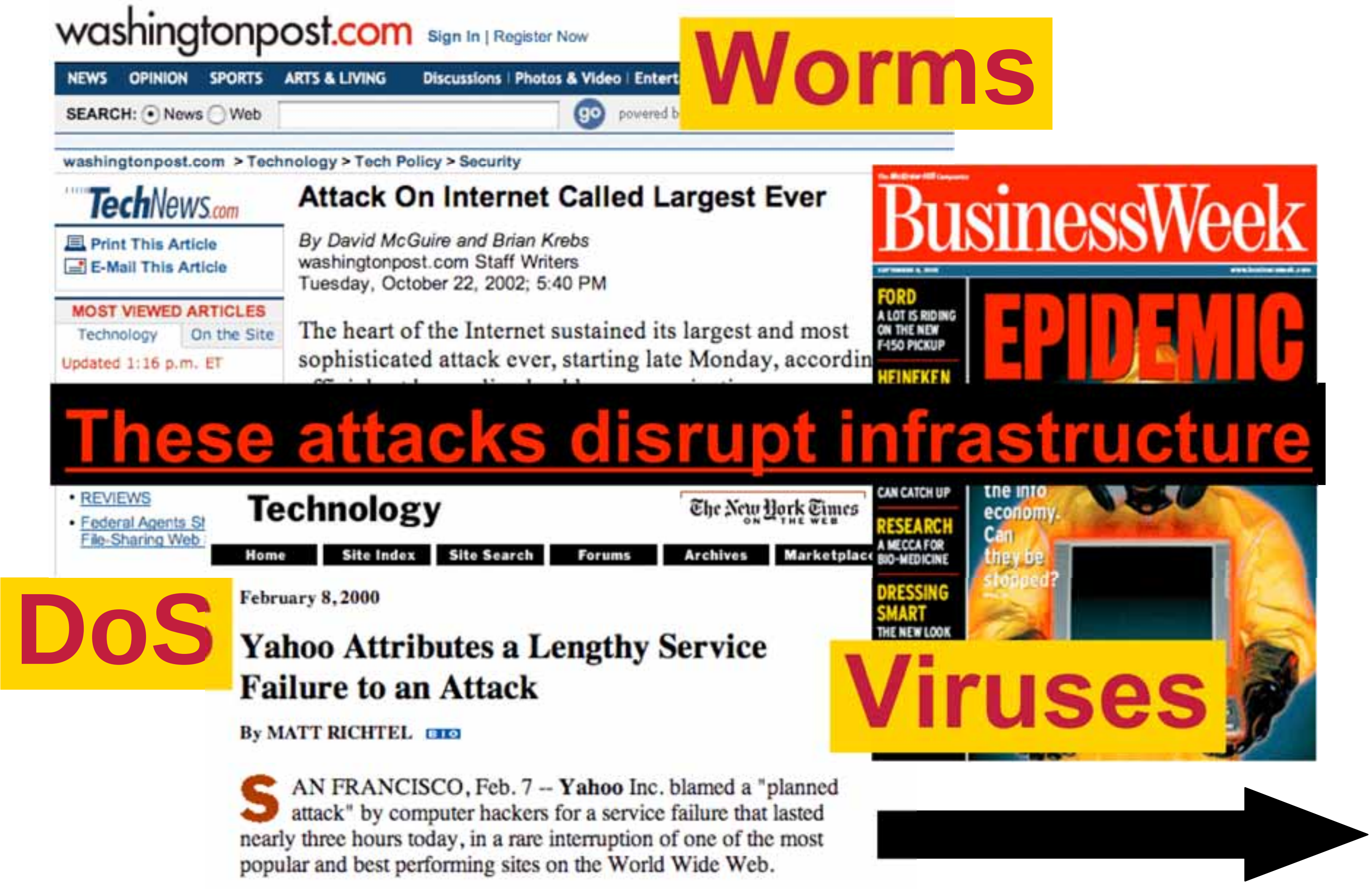
EVOLVING TRUST MODEL

- Contractors, partners, customers
- Wireless, open access points, VPNs
- Poor internal visibility

EVOLVING BUSINESS MODEL

- New platforms and applications
- Mergers and acquisitions
- Hires, fires, transfers

Yesterday … Availability Attacks



Worms

washingtonpost.com Sign In | Register Now

NEWS OPINION SPORTS ARTS & LIVING Discussions | Photos & Video | Entert

SEARCH: ● News ○ Web go powered b

washingtonpost.com > Technology > Tech Policy > Security

TechNews.com

Print This Article
E-Mail This Article

MOST VIEWED ARTICLES
Technology On the Site
Updated 1:16 p.m. ET

Attack On Internet Called Largest Ever

By David McGuire and Brian Krebs
washingtonpost.com Staff Writers
Tuesday, October 22, 2002; 5:40 PM

The heart of the Internet sustained its largest and most sophisticated attack ever, starting late Monday, accordin

BusinessWeek

FORD A LOT IS RIDING ON THE NEW F-150 PICKUP
HEINEKEN

EPIDEMIC

CAN CATCH UP
RESEARCH A MECCA FOR BIO-MEDICINE
DRESSING SMART THE NEW LOOK

the info economy. Can they be stopped?

These attacks disrupt infrastructure

- REVIEWS
- Federal Agents St File-Sharing Web

Technology

The New York Times ON THE WEB

Home | Site Index | Site Search | Forums | Archives | Marketplace

DoS

February 8, 2000

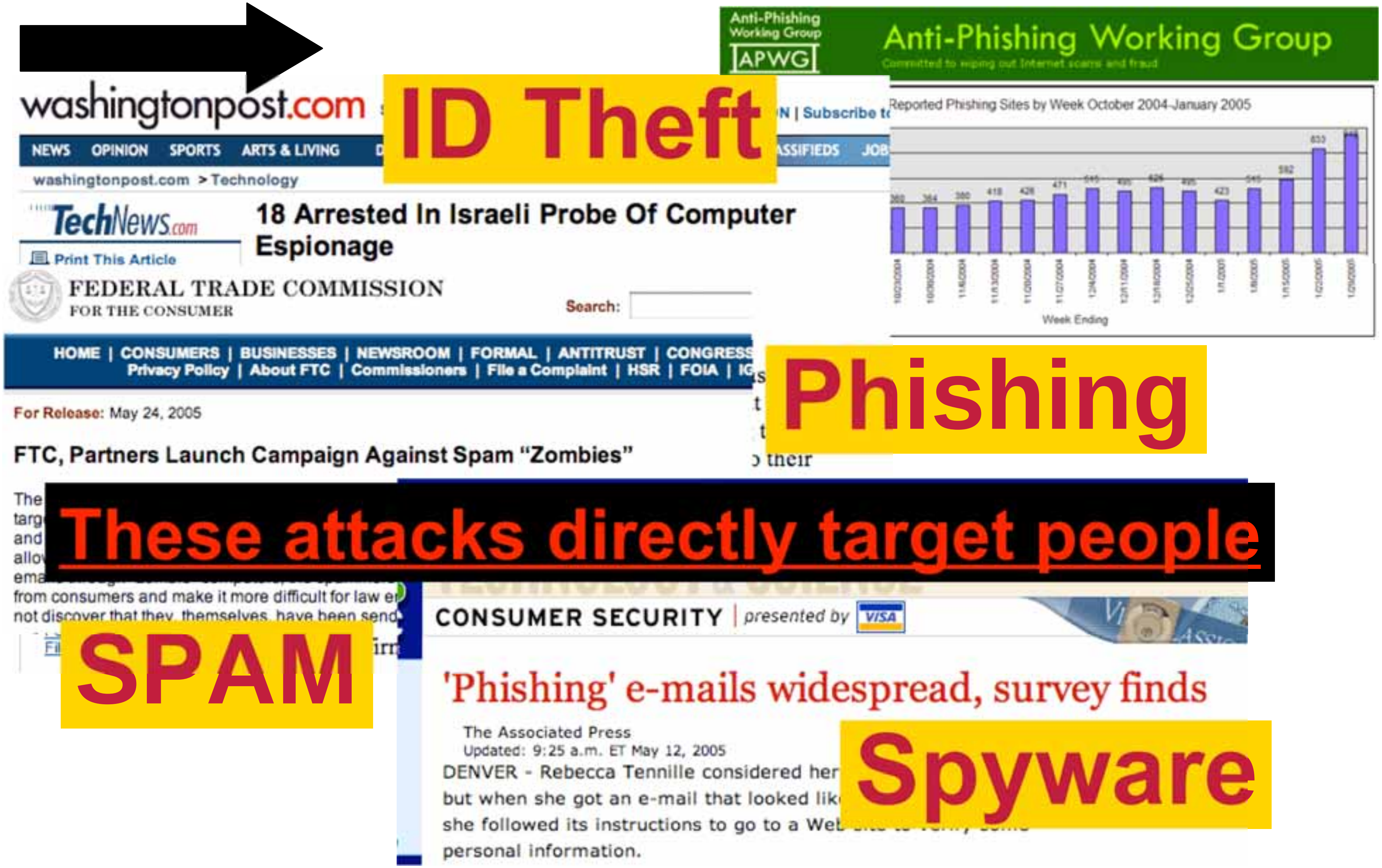
Yahoo Attributes a Lengthy Service Failure to an Attack

By MATT RICHTEL BIO

SAN FRANCISCO, Feb. 7 -- **Yahoo** Inc. blamed a "planned attack" by computer hackers for a service failure that lasted nearly three hours today, in a rare interruption of one of the most popular and best performing sites on the World Wide Web.

Viruses

A Dramatic Transformation and Escalation



washingtonpost.com

NEWS OPINION SPORTS ARTS & LIVING

washingtonpost.com > Technology

TechNews.com

Print This Article

18 Arrested In Israeli Probe Of Computer Espionage

FEDERAL TRADE COMMISSION
FOR THE CONSUMER

Search:

HOME | CONSUMERS | BUSINESSES | NEWSROOM | FORMAL | ANTITRUST | CONGRESS
Privacy Policy | About FTC | Commissioners | File a Complaint | HSR | FOIA | IG

For Release: May 24, 2005

FTC, Partners Launch Campaign Against Spam "Zombies"

The
targ
and
allo
ema
from consumers and make it more difficult for law e
not discover that they, themselves, have been send

Anti-Phishing Working Group
APWG

Anti-Phishing Working Group
Committed to wiping out Internet scams and fraud

N | Subscribe to

SSIFIEDS JOB

CONSUMER SECURITY | presented by VISA

'Phishing' e-mails widespread, survey finds

The Associated Press
Updated: 9:25 a.m. ET May 12, 2005

DENVER - Rebecca Tennille considered her
but when she got an e-mail that looked lik
she followed its instructions to go to a Web
personal information.

ID Theft

Phishing

These attacks directly target people

SPAM

Spyware

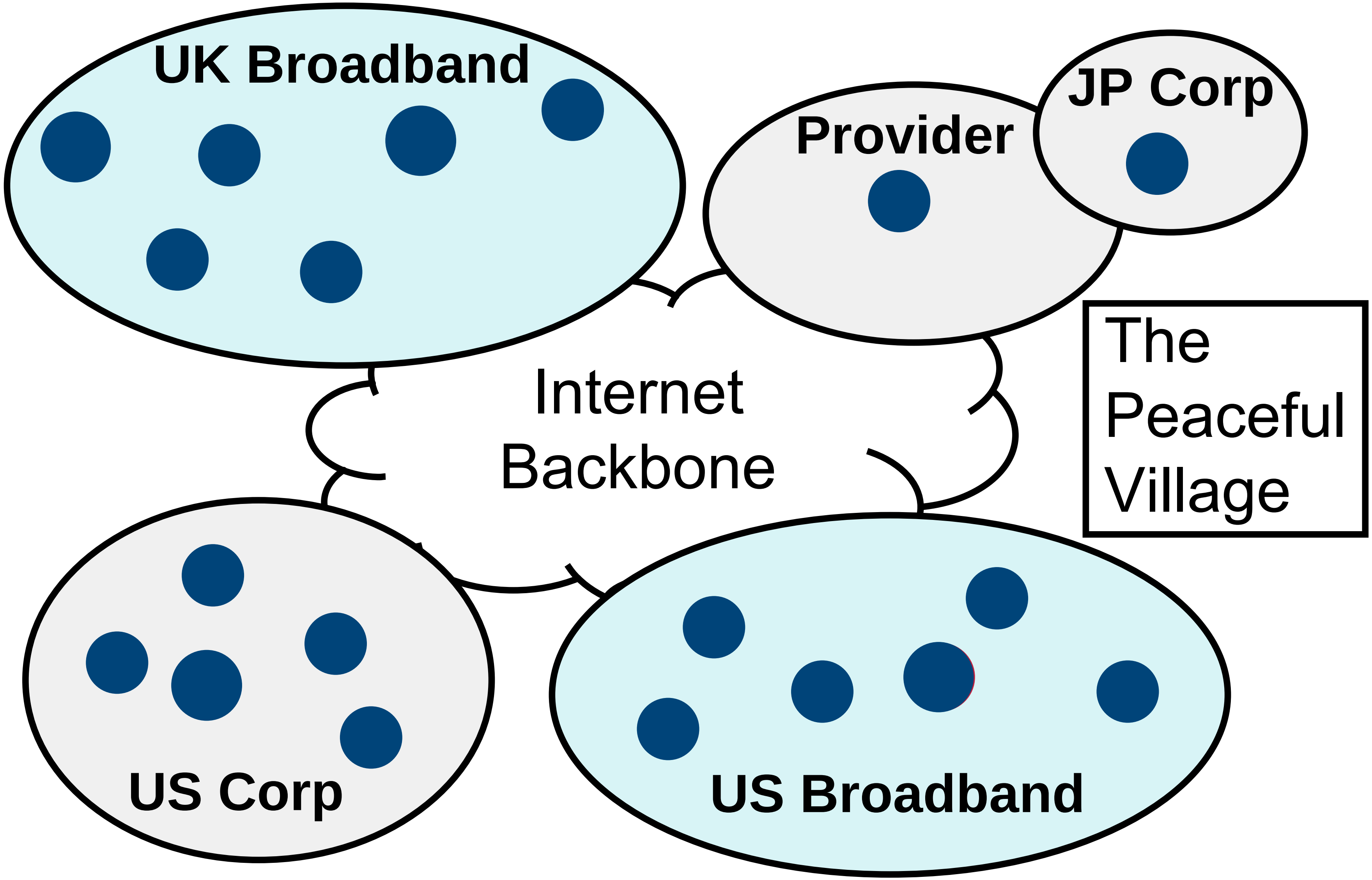
Rise of the Botnets (Zombie Armies)

- 1000’s of new bots each day [Symantec 2005]
- Over 900,000 infected bots as phishing attacks are growing at 28% per month [Anti-Phishing Working Group 2005]
- A single botnet comprised of more than 140,000 hosts was observed “in the wild” [CERT Advisory CA-2003-08, March 2003]

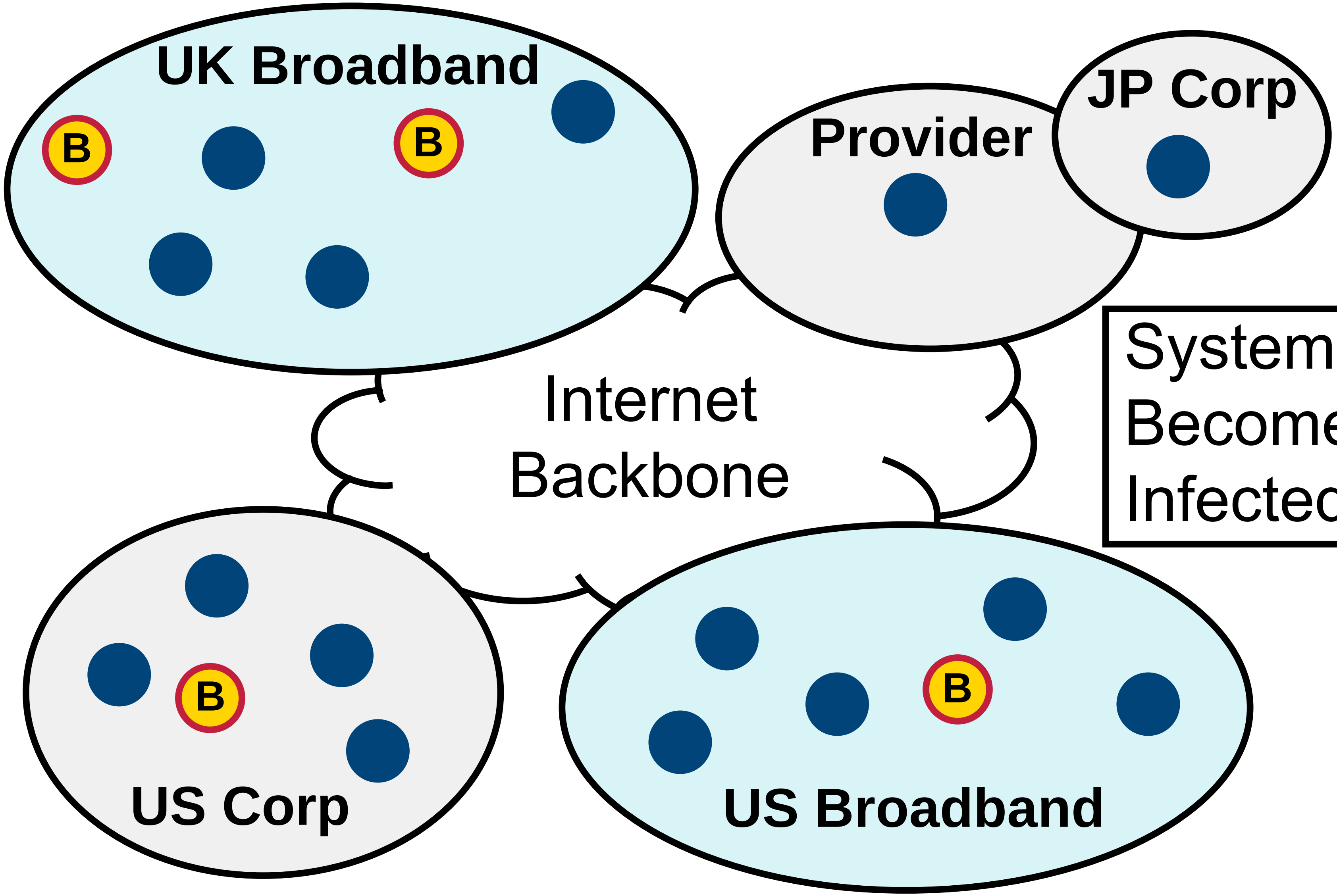
Attackers have learned a compromised system is more useful alive than dead!

  (evasion/economics?)
  - Significant more firepower: *Broadband (1Mbps Up) x 100s == OC3!!!*
- An entire economy is evolving around bot ownership
  - Sell and trade of bots ($0.10 for “generic bot”, $40 or more for an “interesting bot; e.g., a .mil bot)
  - Bots are a commodity - no significant resource constraints

The Botnet

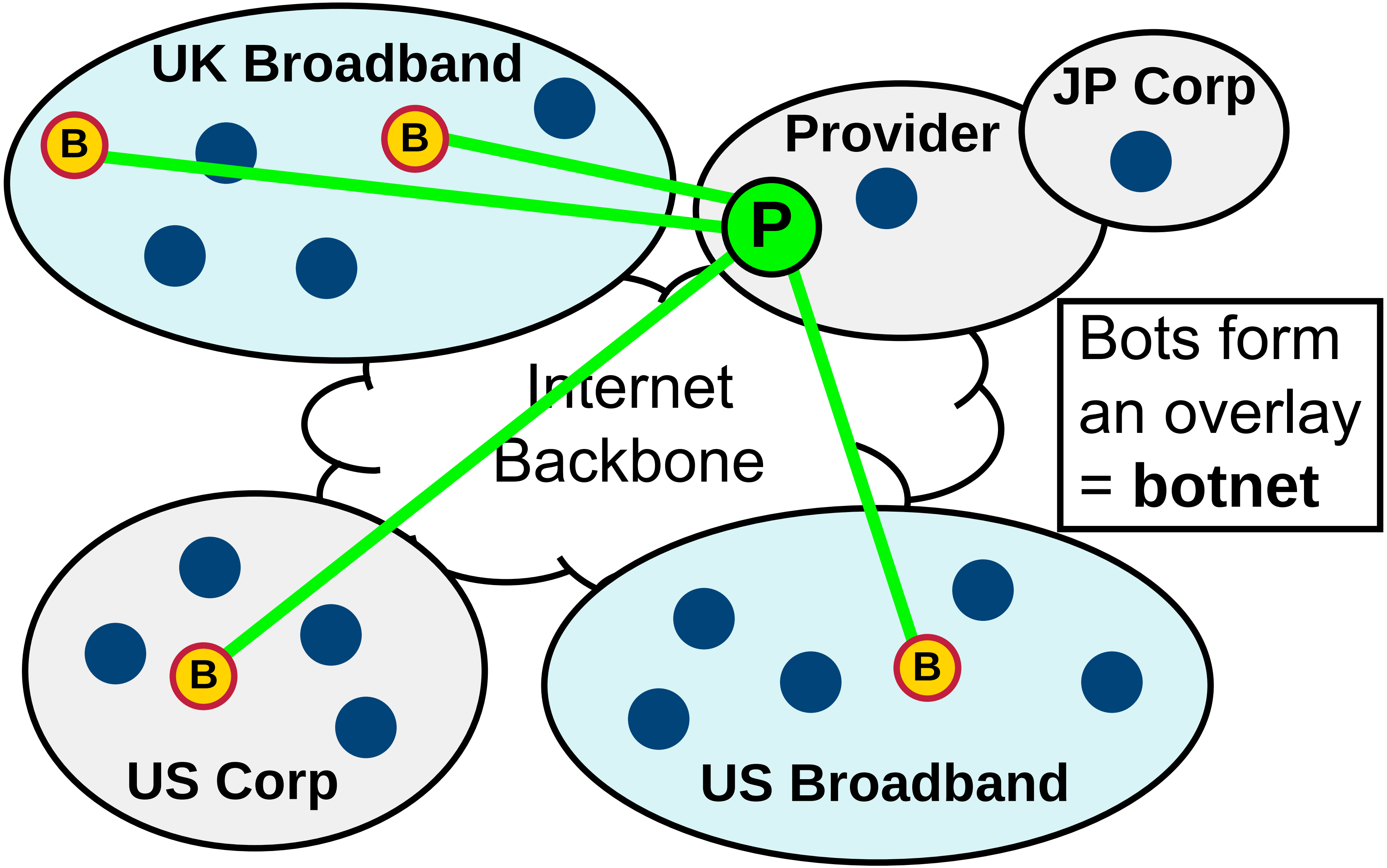

UK Broadband
Provider
JP Corp
Internet Backbone
The Peaceful Village
US Corp
US Broadband

The Botnet



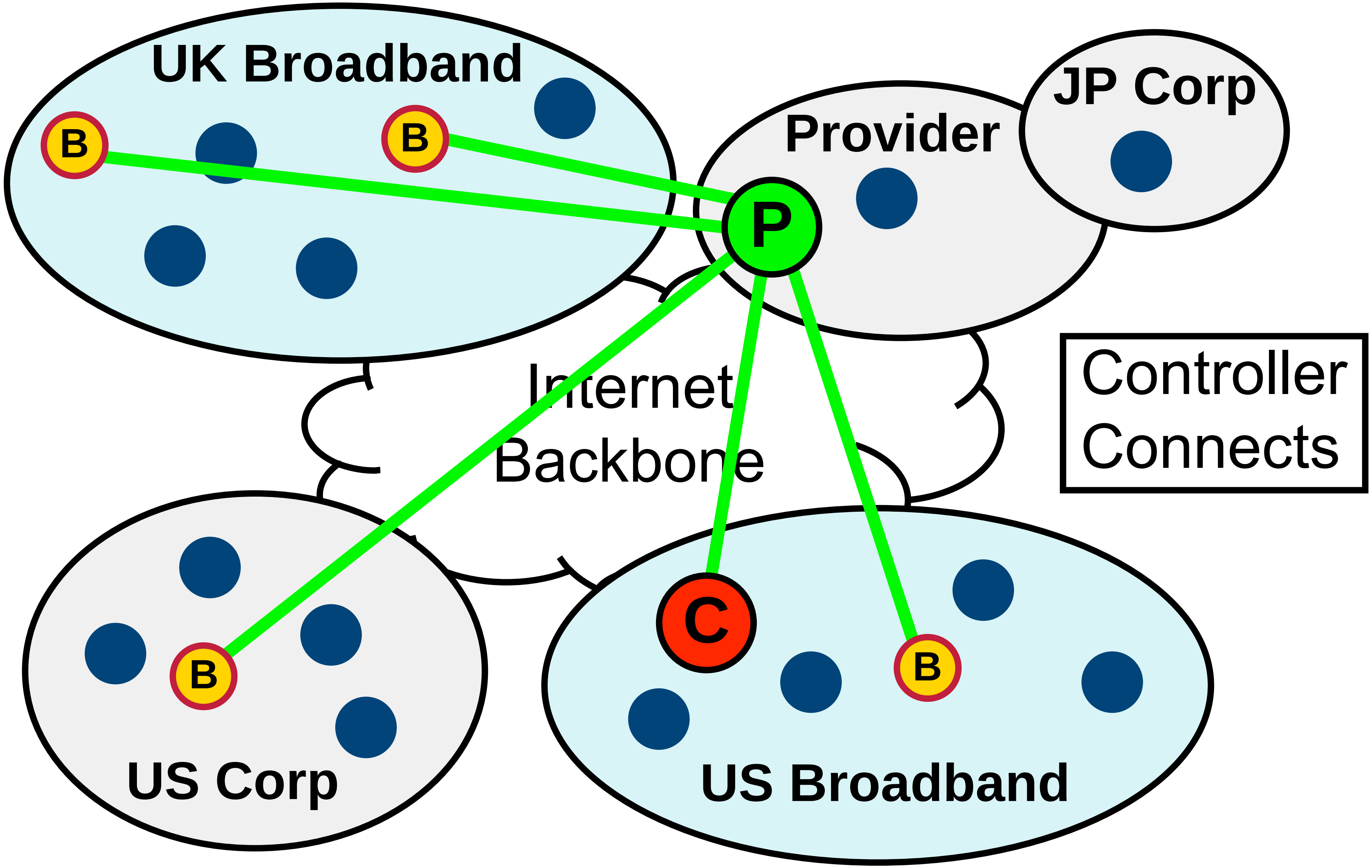
Systems Become Infected

The Botnet

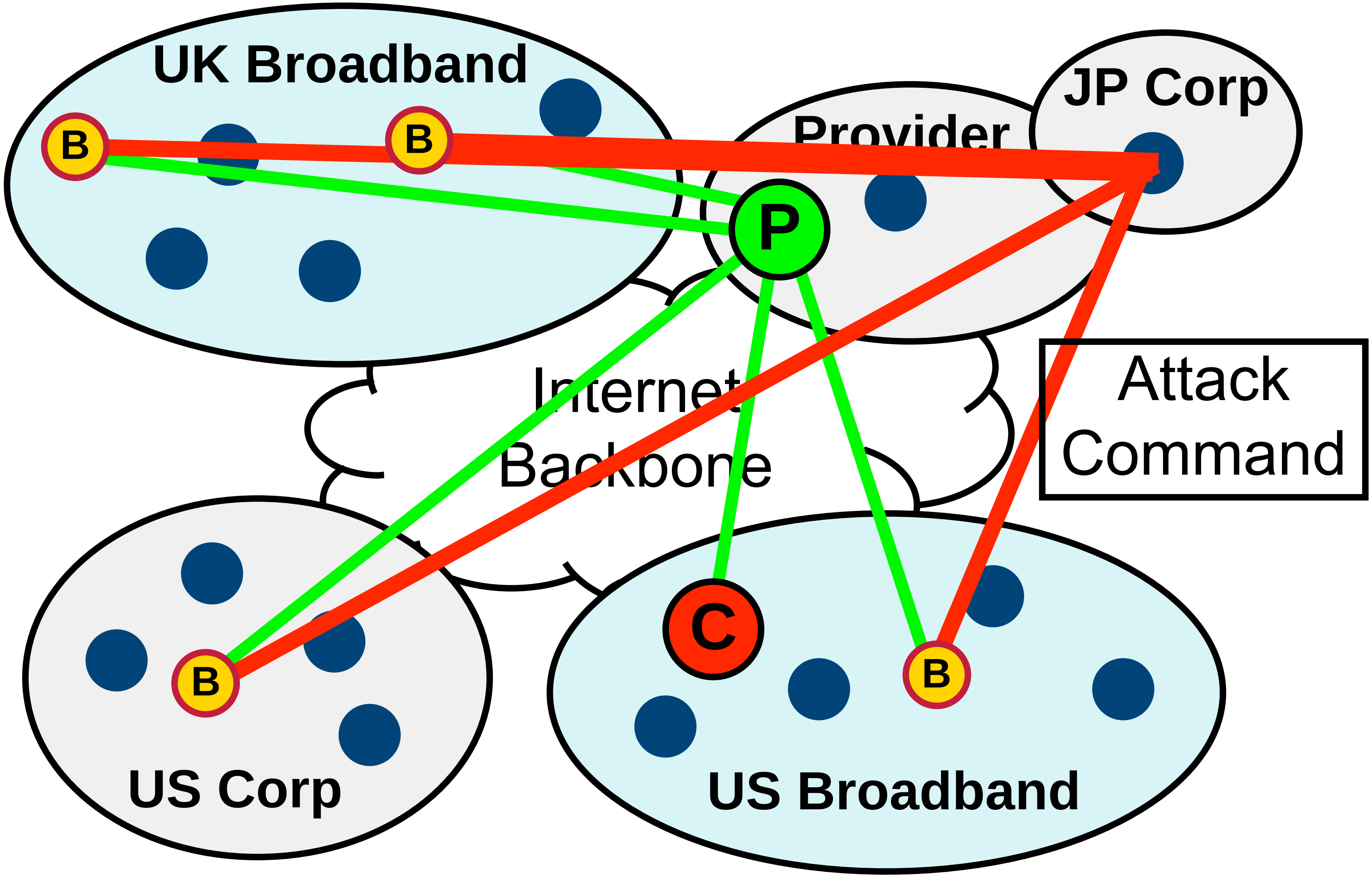


Bots form an overlay = **botnet**

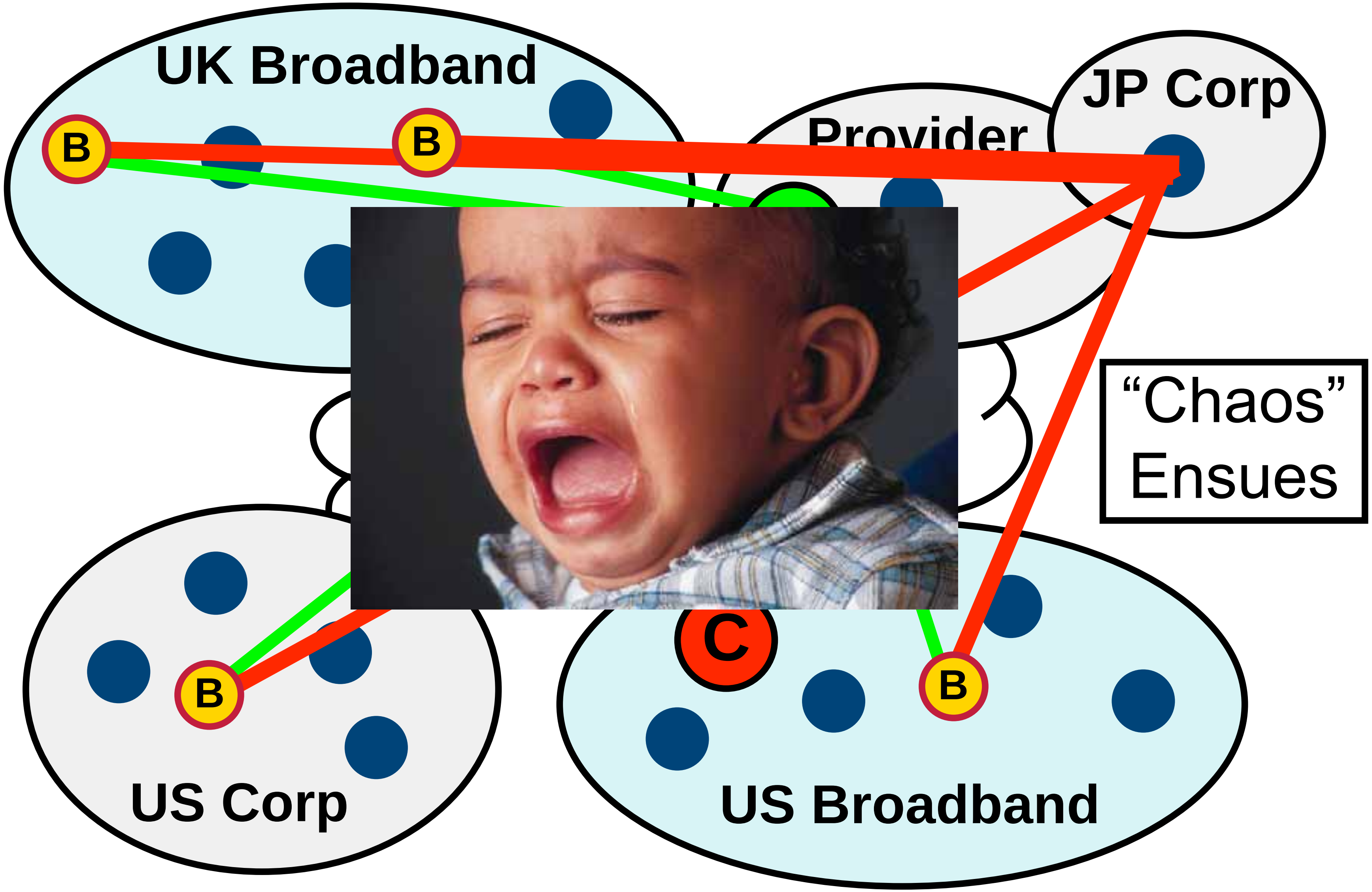
The Botnet


UK Broadband
B
B
Provider
JP Corp
P
Internet
Backbone
Controller
Connects
C
B
B
US Corp
US Broadband

The Botnet


UK Broadband
B
B
Provider
JP Corp
P
Internet
Backbone
Attack
Command
B
C
B
US Corp
US Broadband

The Botnet



“Chaos” Ensues

Mobile Computing

Distinguishing Characteristics:

- Relatively resource-poor mobile elements
- Potential variability in network connectivity
- Constraints on power consumption and energy source
- Inherent vulnerability of mobile devices
- Increased tension between autonomy & interdependence: application-aware vs. application-transparent [Satyanarayanan et. al.]
- Not so subtle: wireless medium and node mobility

Impact on Security Design

- Stringent resource constraints (cpu, power) may lead to weaker protection
- Low-end devices can hardly perform computation-intensive tasks such as asymmetric cryptographic alg.
- Shared medium (wireless channel) is accessible to both legitimate users and malicious attackers
- Preservation of location discovery and privacy for mobile users

Note on perimeter defense and best practices!

Rethinking the Classic Client-Server Model

- Small set of trusted servers augmented by end-to end authentication and encrypted transmission

- Mobility may temporarily blur the distinction between client and server to achieve performance or availability
  - Sensitive data cached on client
  - Client emulating server functions when limited/no connectivity
  - Shipping client functions to resource-rich server

Mobile Ad Hoc Networks

Distinguishing Characteristics:

- Self-configuration and self-maintenance
- Open peer-to-peer architecture
- Lack of dedicated network (routing) infrastructure
- Routing and packet forwarding done by mobile nodes
- Lack of a centralized monitoring or management point
- Absence of a certification authority

Impact on Security Design

- No clear line of defense: boundary between inside and outside blurred
- No well-defined place for deploying security monitoring (IDS) or access control mechanisms (firewall)
- Internal security issue if a mobile node is compromised
- Potential disruption of routing substrate
- Highly dynamic topology with frequent joins and departures

Classification of Attacks

- Attacks on the wireless infrastructure
- Using wireless network to gain foothold into the wired network
  - Internal security attacks
  - Jumping-off point for launch attacks
- Attacks on mobile devices

Infrastructure Attacks

- Packet sniffing and “war driving”
  - Identifying SSID in Wi-Fi networks
  - Traffic analysis
  - Useful when combined with other data
- Rogue access points
- Jamming (causing interference to an 802.11 network)
- Attacks on routing and packet forward infrastructure

Attacks on Mobile Ad hoc Networks

- Link layer attacks:
  - Vulnerability of 802.11 WEP to several types of cryptographic attacks
    - WEPCrack and AirSnort
  - DoS attacks on channel contention and reservation schemes
  - Exploiting binary exponential backoff to deny access to the wireless channel from its local neighbors
  - Backoffs at link layer incurring chain reaction in upper layer protocols such as TCP

- Network layer attacks in mobile ad hoc networks:
  - Routing attacks: advertising routing updates that do not follow specification … disrupt protocol operation and poison routing state at other nodes
  - JellyFish attacks: target closed-loop flows responsive to delay or loss – i.e. target end-to-end congestion control of TCP
    - Packet reordering
    - Periodic dropping
    - Delay-variance attacks
    - Duplicating packets
  - Blackhole attacks: target open-loop flows by dropping all packets after correctly receiving them at MAC layer

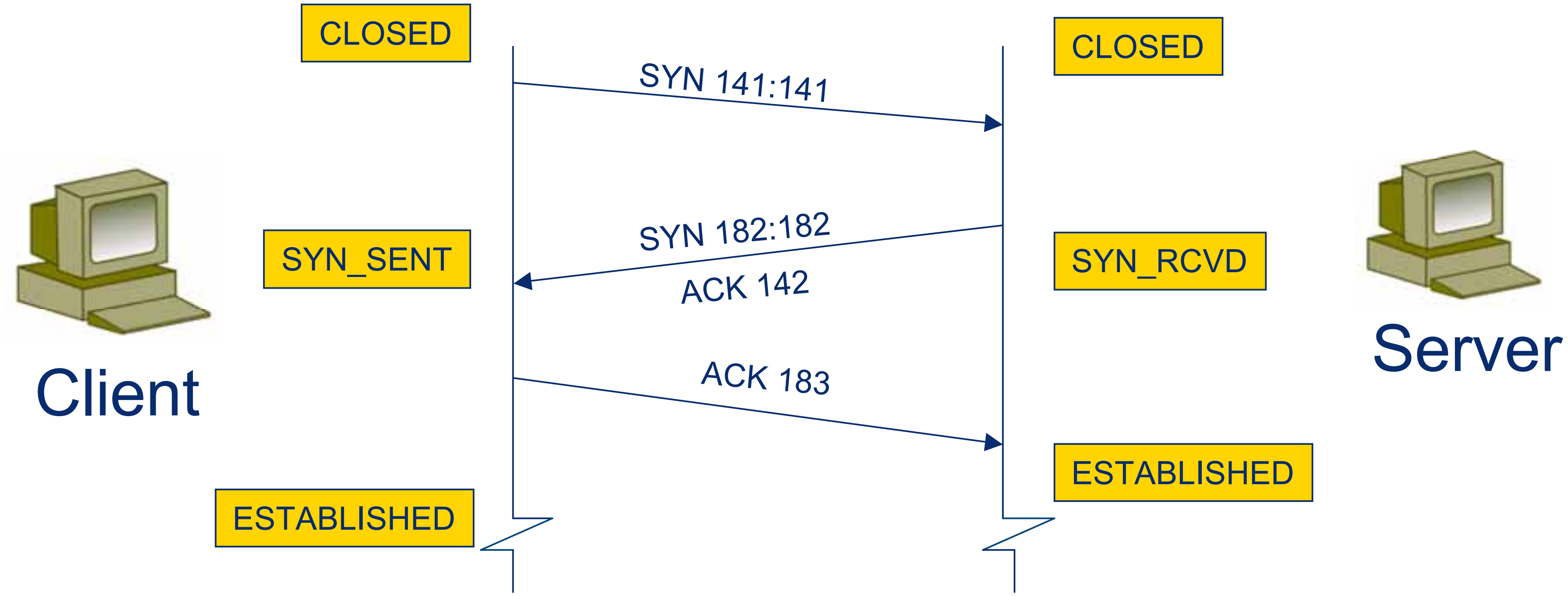
Classification of Attacks

- Snooping
  - Identifying SSID in Wi-Fi networks
  - Traffic analysis
  - Useful when combined with other data
- Man-in-the-middle attack
  - Replaying captured messages
- Bogus access points
- Attacks based on signal leakage

- Jumping-off point from which attacks are launched
- Attacks on keys in wireless networks:
  - Brute-force attacks
  - Dictionary attacks
  - Algorithmic attacks

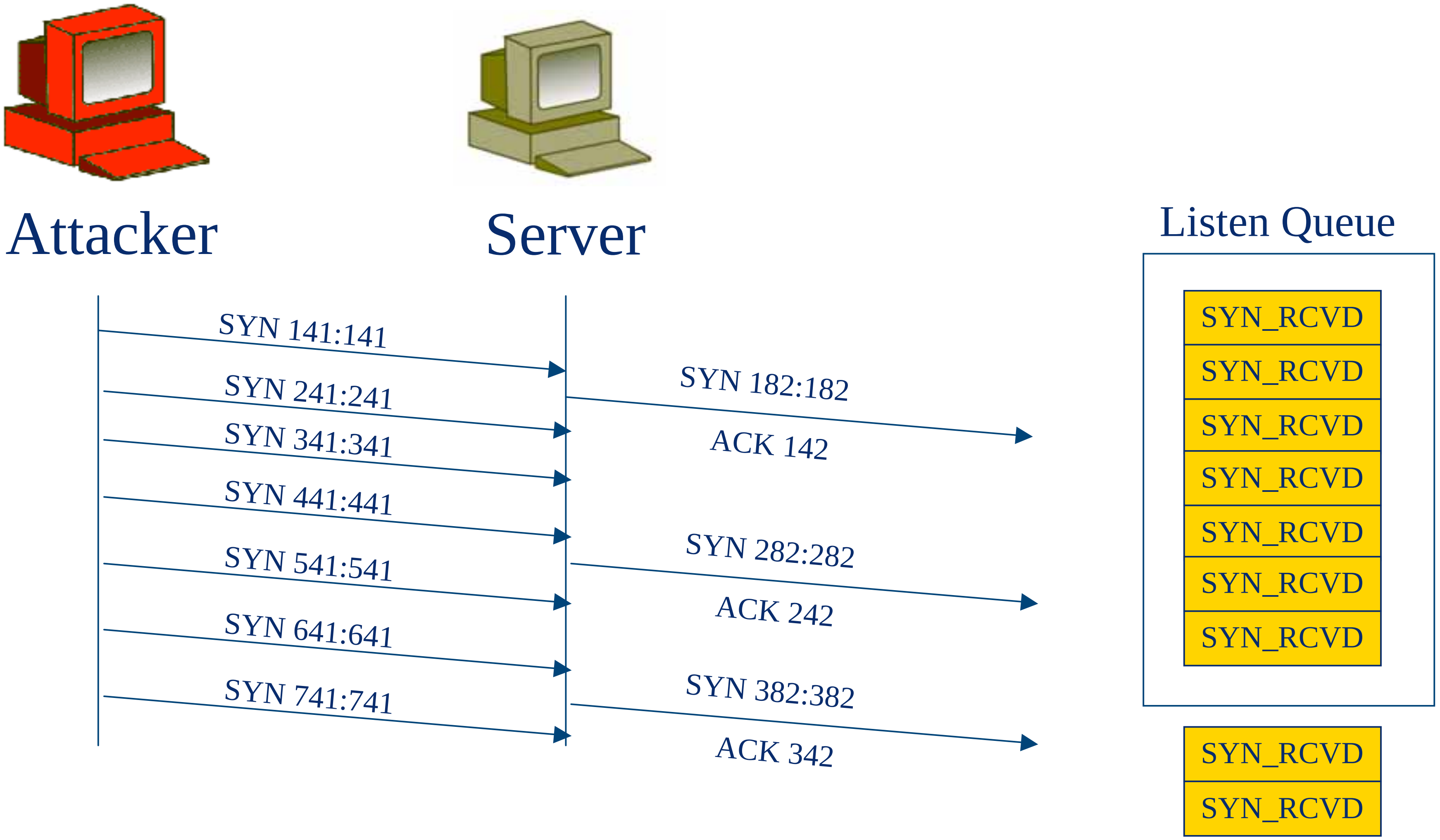
Denial-of-Service Attacks

Example: TCP SYN Flood



Normal sequence for TCP connection establishment (3-way handshake)

Example: TCP SYN Flood (cont.)



Attacker

Server

Listen Queue

SYN 141:141

SYN 241:241

SYN 341:341

SYN 441:441

SYN 541:541

SYN 641:641

SYN 741:741

SYN 182:182

ACK 142

SYN 282:282

ACK 242

SYN 382:382

ACK 342

SYN_RCVD

SYN_RCVD

SYN_RCVD

SYN_RCVD

SYN_RCVD

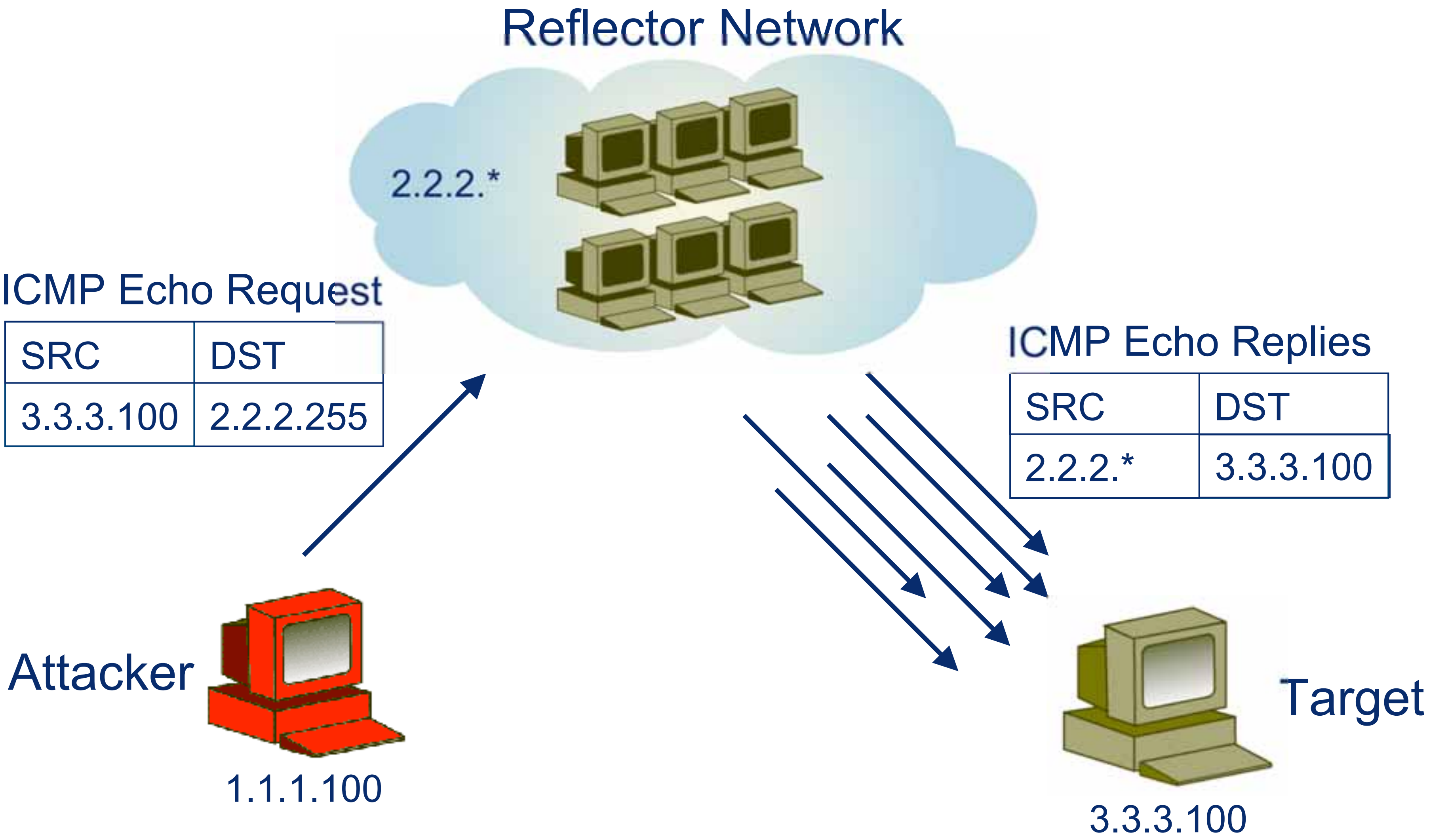
SYN_RCVD

SYN_RCVD

SYN_RCVD

SYN_RCVD

Example: Smurf Attack



ICMP Echo Request

SRC	DST
3.3.3.100	2.2.2.255

ICMP Echo Replies

SRC	DST
2.2.2.*	3.3.3.100

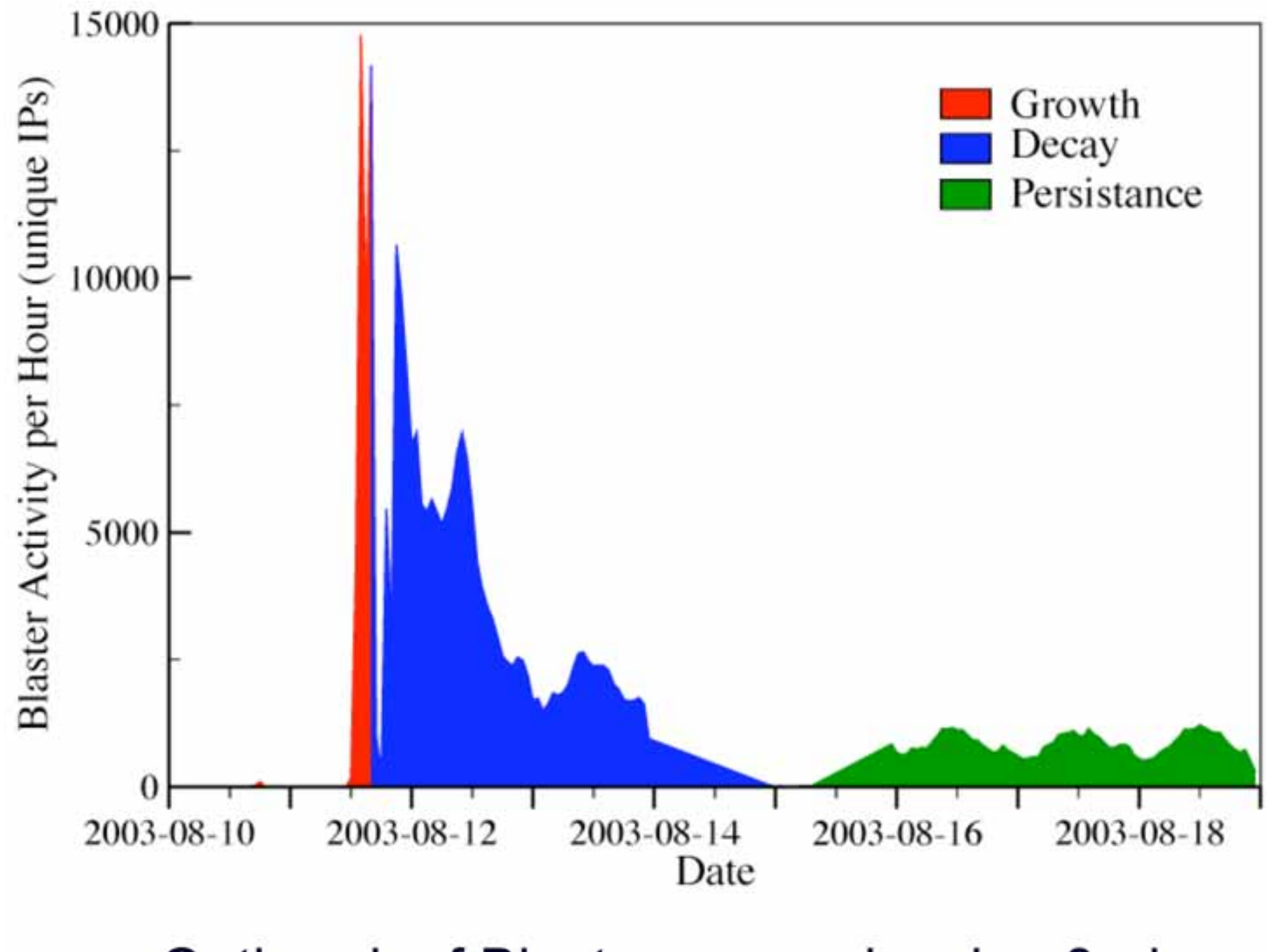
Denial-of-Service Attacks

- DoS attacks by gaining a foothold in the wired network
- DoS attacks using rouge wireless devices
- DoS attacks on wireless access points
- DoS attacks on services offered to mobile users
- DoS attacks by jamming frequency channels
- DoS attacks via network-layer packet blasting

Traffic analysis techniques employed by existing DDoS detection and mitigation solutions is not readily applicable to wireless networks with mobile nodes.

- What about malicious code, worms and viruses?
  - Implications for wireless networks and mobile devices

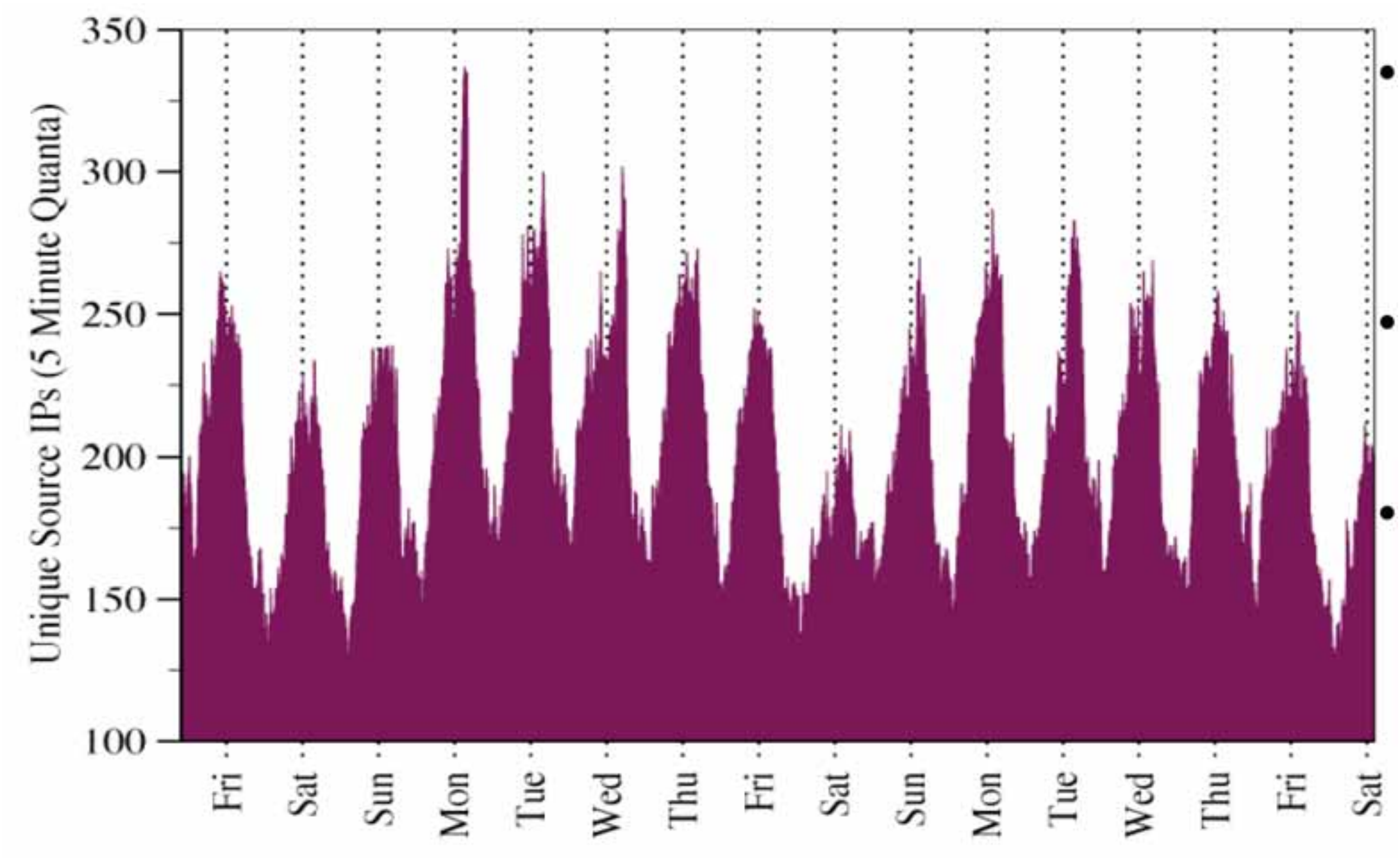
Internet Worms



- Blaster Worm released August 11, 2003
- IMS Observed 286,000 IPs
- Doubling every 2.3 hours
- 40,000 hosts/hour
- Half-life = 10.4 hours

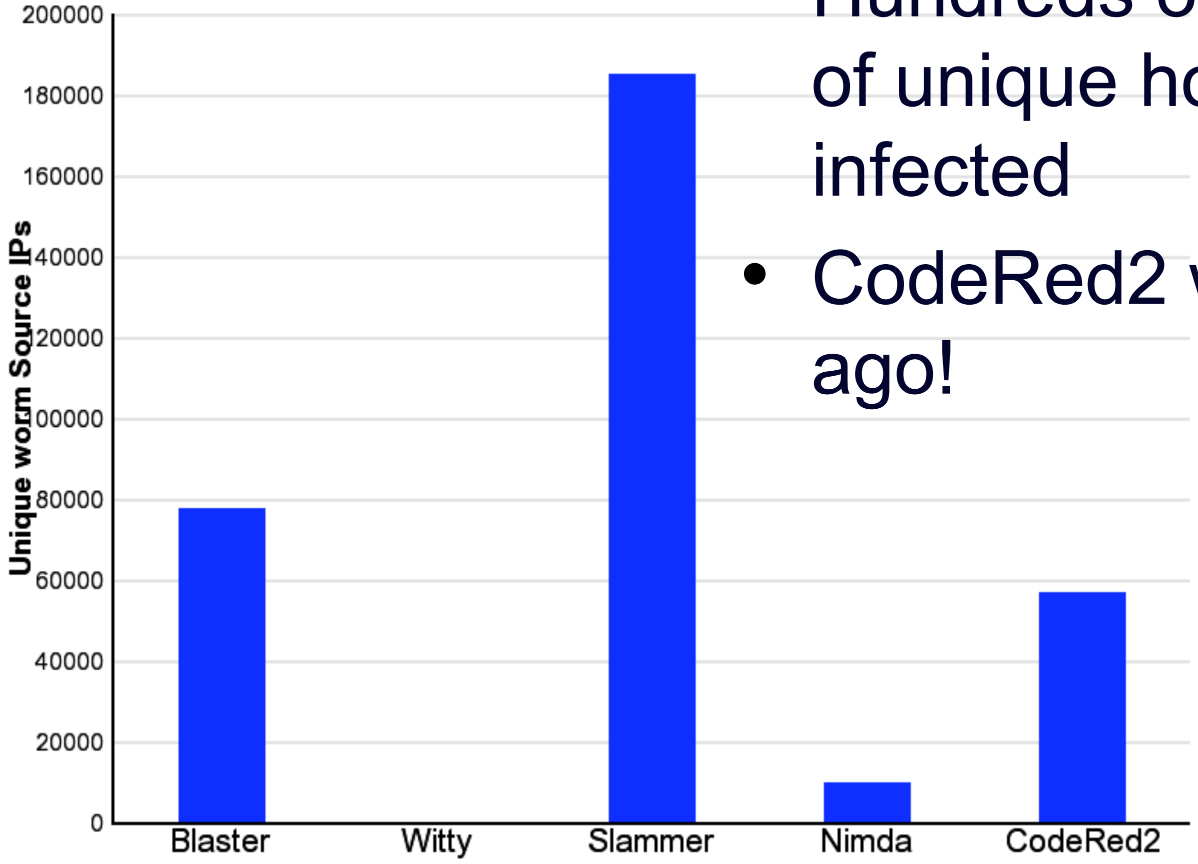
Outbreak of Blaster worm showing 3-phase life cycle

Blaster Circadian Pattern



- Cycles correspond with work week
- Saturday sees lowest activity
- Are infected hosts being rebooted?

Persistence of Internet Worms



- Hundreds of thousands of unique hosts still infected
- CodeRed2 was *years* ago!

Airborne Viruses

- As handheld devices become increasingly pervasive and interconnected, smart phones and PDAs will become increasingly susceptible to worms, viruses and Trojan horses.

- Broad range of applications: email, sms, web surfing, multi-player games, camera, e-transactions

- “The race to own a new platform!”
- Unlike desktop counterparts, security measures for these devices are relatively immature. Combined with unsecured wireless networks, the potential for fast propagating viral spread multiplies.
- Several methods of infection:
  - Synchronizing handheld with its desktop
  - Passing malicious code by infrared beam
  - Passing via unsecured wireless access

Airborne Viruses (cont.)

Palm OS Phage Virus:

- The first to successfully attack the Palm OS handheld platform in 2000. When executed, infects all third-party application program.
- When a carrier palm is synchronized with a clean palm, the clean palm could receive the virus in any infected file.
- This virus in turn copy itself to all other applications.

- Palm Security Update: Posted August 20, 2003

"This SecurityPatch.prc software will address a password security issue that was discovered on Palm Zire 71 and Tungsten T2 handhelds. The issue relates to a condition that may compromise the password lock out of the device."

Airborne Viruses (cont.)

- Windows CE PDAs have most of the ingredients for viral spread: fast processor, writeable memory, Pocket MS Word, Pocket outlook mail client.

- Potential is even greater if you combine a Microsoft mobile device OS with .NET distributed programming platform … small footprint, interconnected and running on a broad range of intelligent devices including cameras, Internet appliances, smart phones.

Airborne Viruses (cont.)

- Internet-based smart phones are increasingly vulnerable.

- Example:
  - SMS-based attack on Tokyo's emergency response system
  - Denial-of-service attack using SMS messages
  - The message hit 100,000 users inviting them to visit a web page
  - Activated a script to call 110, the emergency response number in Tokyo

Airborne Viruses (cont.)

Bluetooth Vulnerability: The Register June 2005

- Tel Aviv University in Israel - have come up with an exploit which allows hackers to pair with devices without alerting their owner.

- gets around limitations of a security attack first described by Ollie Whitehouse of security firm @Stake last year … needed to eavesdrop the initial connection process between two devices.

- a way to force this pairing process by masquerading as a device, already paired with a target, that has supposedly forgotten a link key used to secure communications.

Internal -v- Perimeter Environment

	PERIMETER	INTERNAL
NETWORK	**TENS** of targets **MEGABITS** of traffic	**THOUSANDS** of targets **GIGABITS** of traffic
APPLICATIONS	**TENS** of applications **WEB, MAIL, DNS**	**HUNDREDS** of applications **CUSTOM** protocols, **PEER-TO-PEER, COMMERCE**
POLICY	**INSIDE** and **OUTSIDE** groups **DEFAULT DENY**	**HUNDREDS** of groups **DEFAULT ALLOW**

Internal -v- Perimeter Protection

	PERIMETER	INTERNAL
THREATS	KNOWN EXPLOITS SCANNING	INSIDER MISUSE ZERO-DAY ATTACKS
IMPACT	INTERNET OUTAGE	DISRUPTION TO CONSUMER and BUSINESS ACCESS
DEPLOYMENT	ACCESS POINTS	DISFUSED THROUGHOUT NETWORK

Questions?

- What if we expand the pool of bots and botnets to include 2+Billion smart phone and PDAs?
- How do secure a broad rage of new mobile platforms and applications?
- What is the deployment model for security devices such as firewall, IDS, IPS? Where is the perimeter?
- How would convergence of networking and security devices in the wired world affect mobile computing?
- …

Timed Asynchronous System Models for Dependable Mobile/Pervasive/* Systems

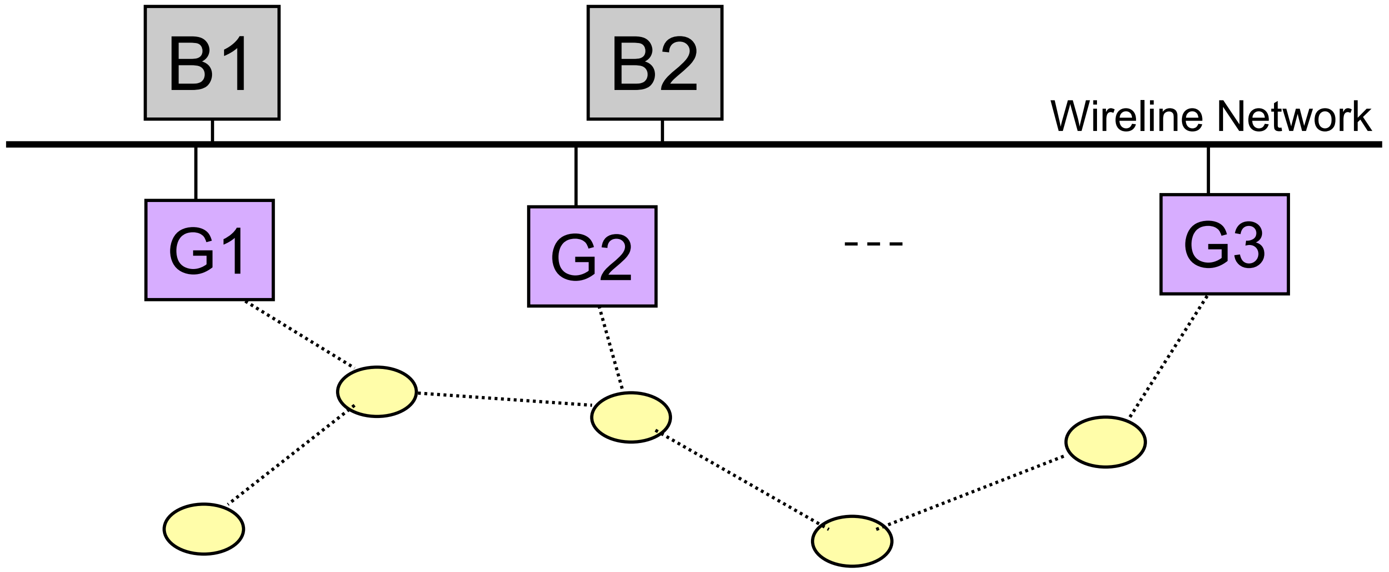
Christof Fetzer
Dresden University of Technology
Germany

Application Domain: **Technology Assisted Living**

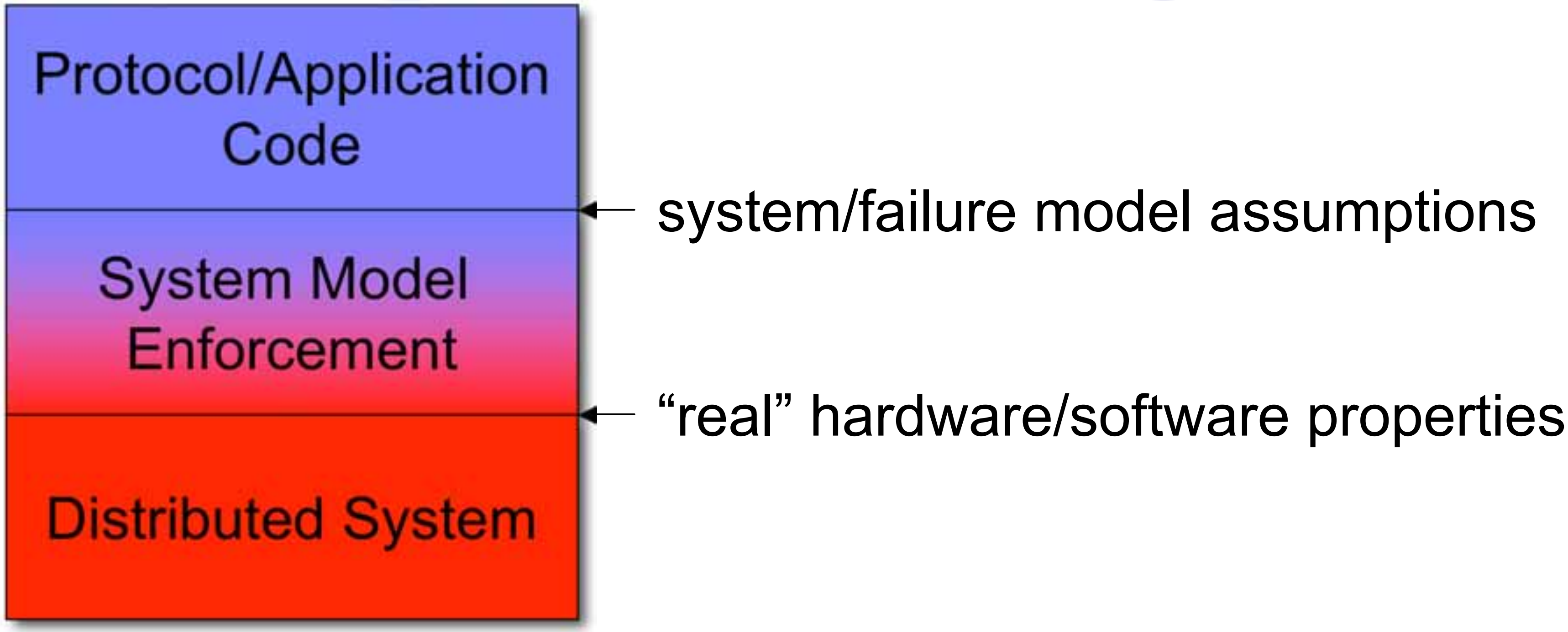
- Home/garden sensor network
  - e.g.: Intel uses motion sensors to check the health status of persons
- Need for dependability
  - application is safety critical...
- Some sort of physical security

Underlying Distributed System

- Mobile nodes
- Network technologies
  - Wireless and wired Ethernet



System Model Assumptions



Protocol/Application Code

System Model Enforcement

Distributed System

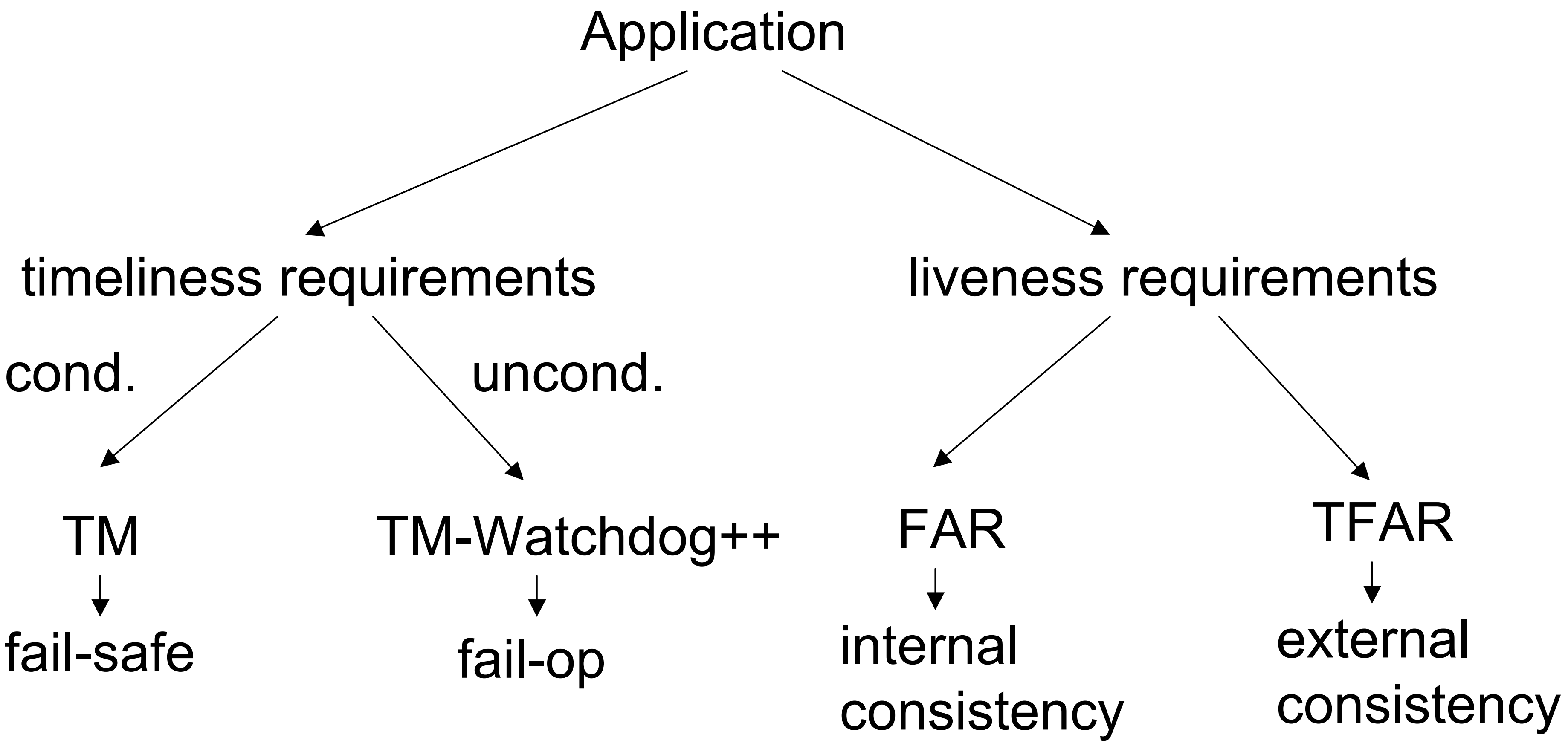
system/failure model assumptions

“real” hardware/software properties

Goals:

1) Simplify protocol development & permit correctness proofs
2) Probability that assumptions are violated are negligible

Application Dependency


Application

timeliness requirements

liveness requirements

cond.

uncond.

TM

fail-safe

TM-Watchdog++

fail-op

FAR

internal consistency

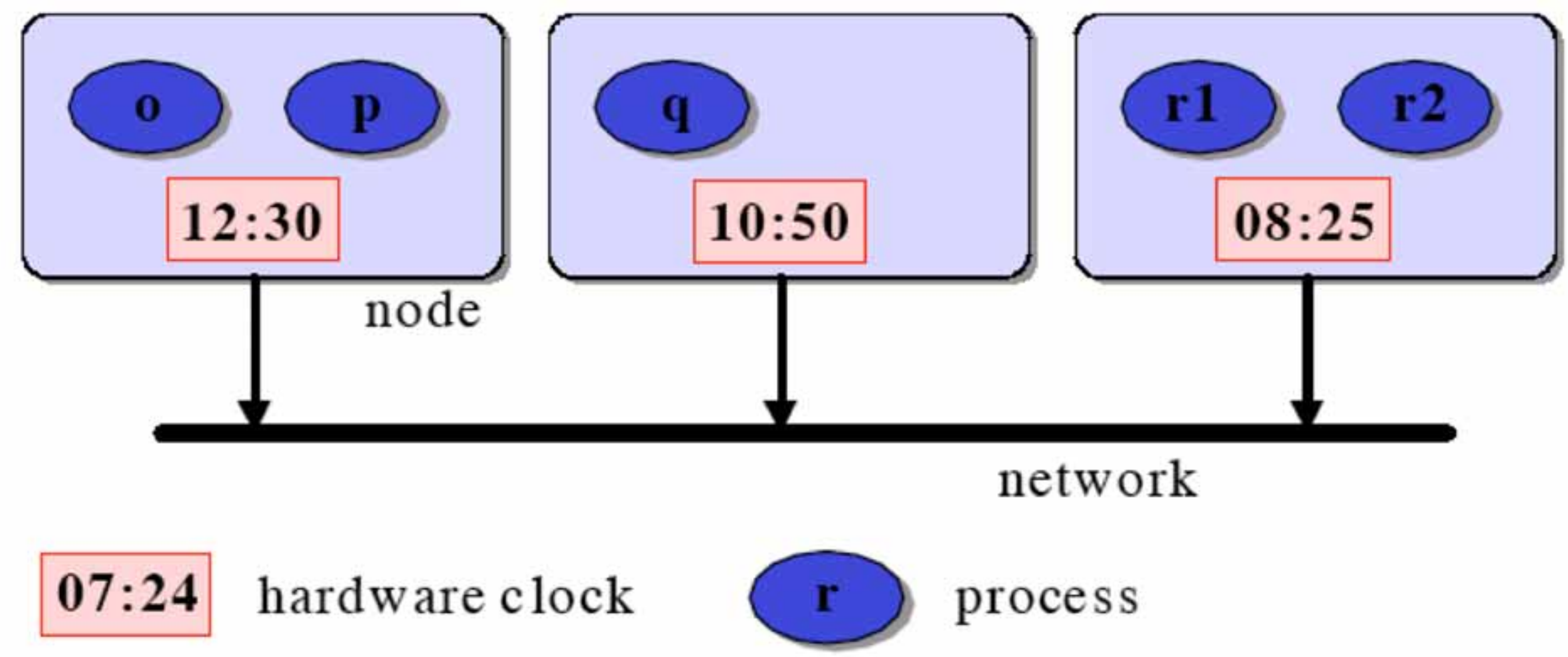
TFAR

external consistency

Timed Asynchronous System Model (TM)

[1]

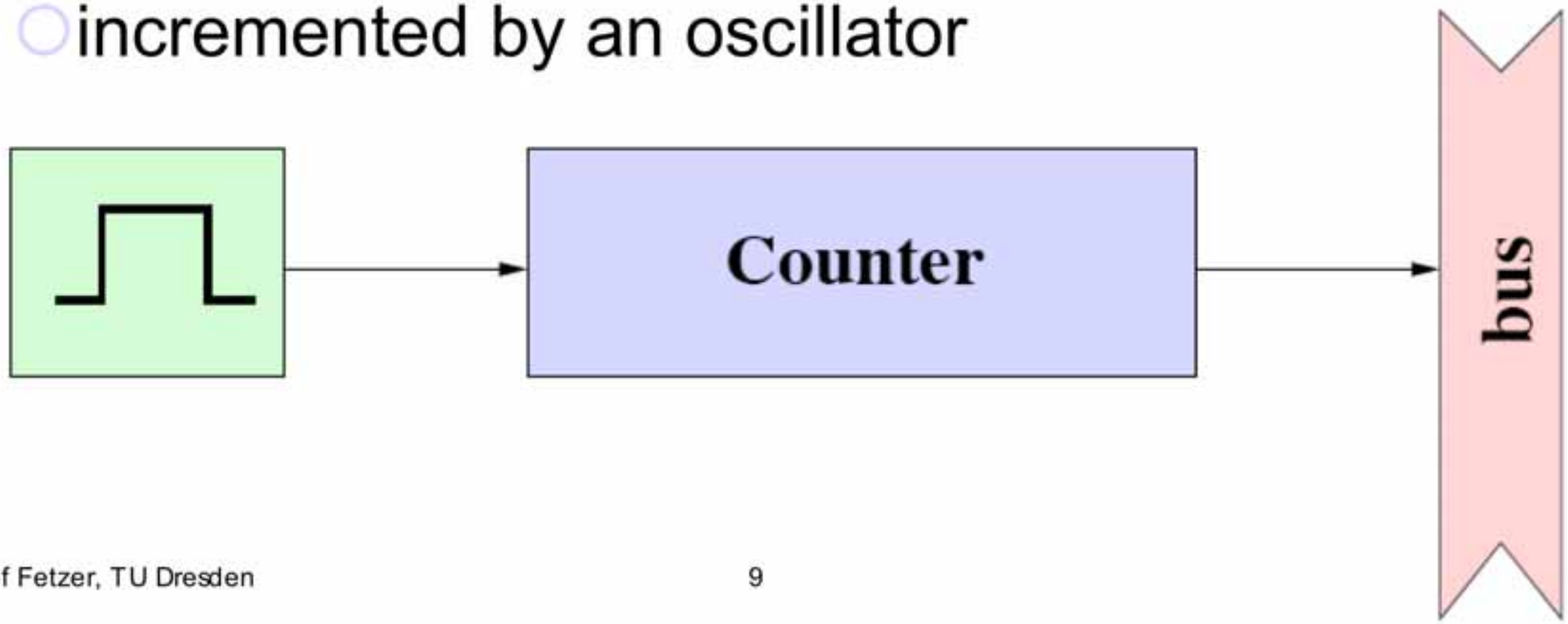
Services


o
p
12:30
node
q
10:50
r1
r2
08:25
network
07:24
hardware clock
r
process

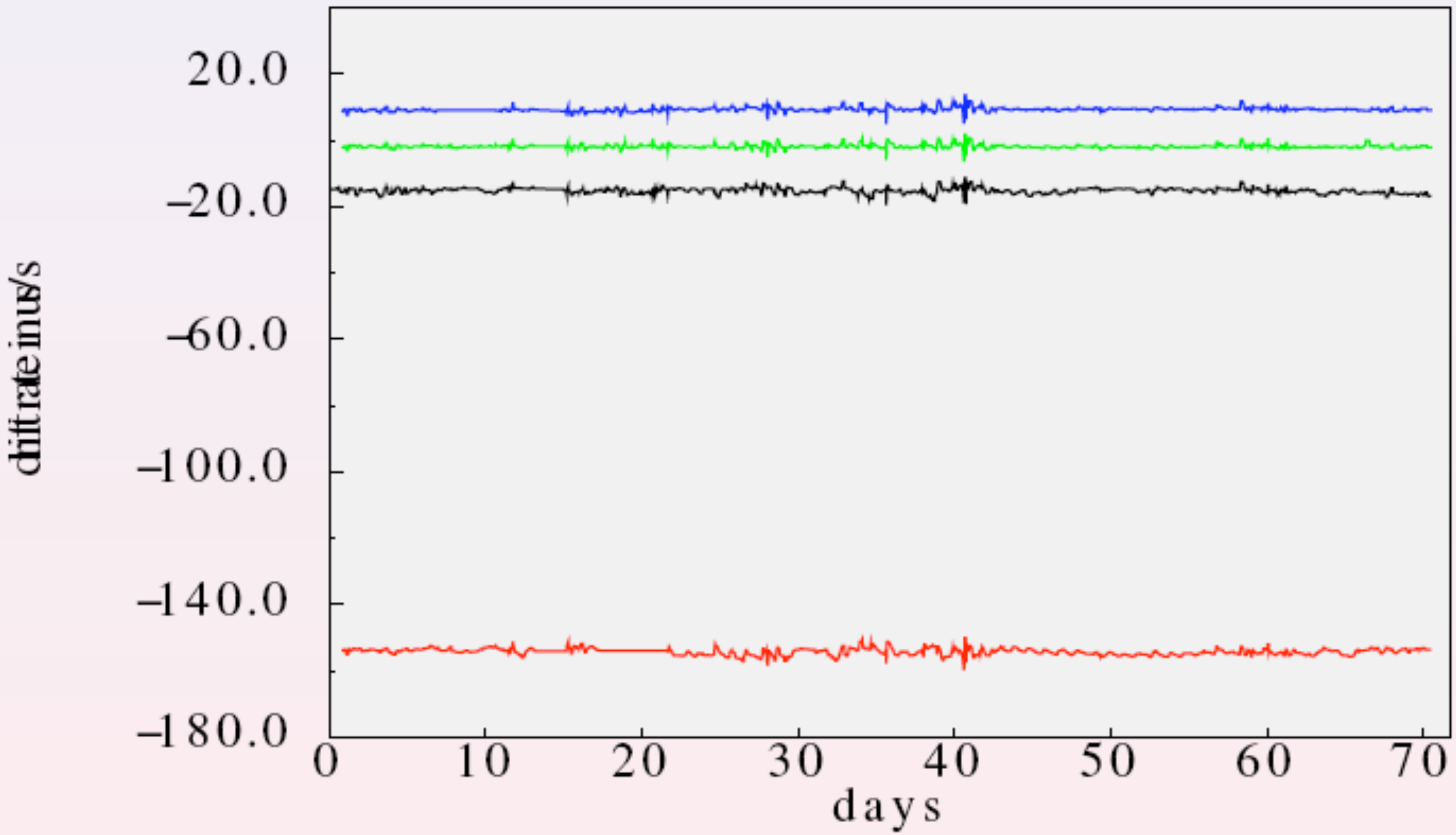
Local Hardware Clock Service

Local Hardware Clocks

- We assume that each computer p has a hardware clock Hp
- A hardware clock can be implemented by a hardware counter
  - incremented by an oscillator



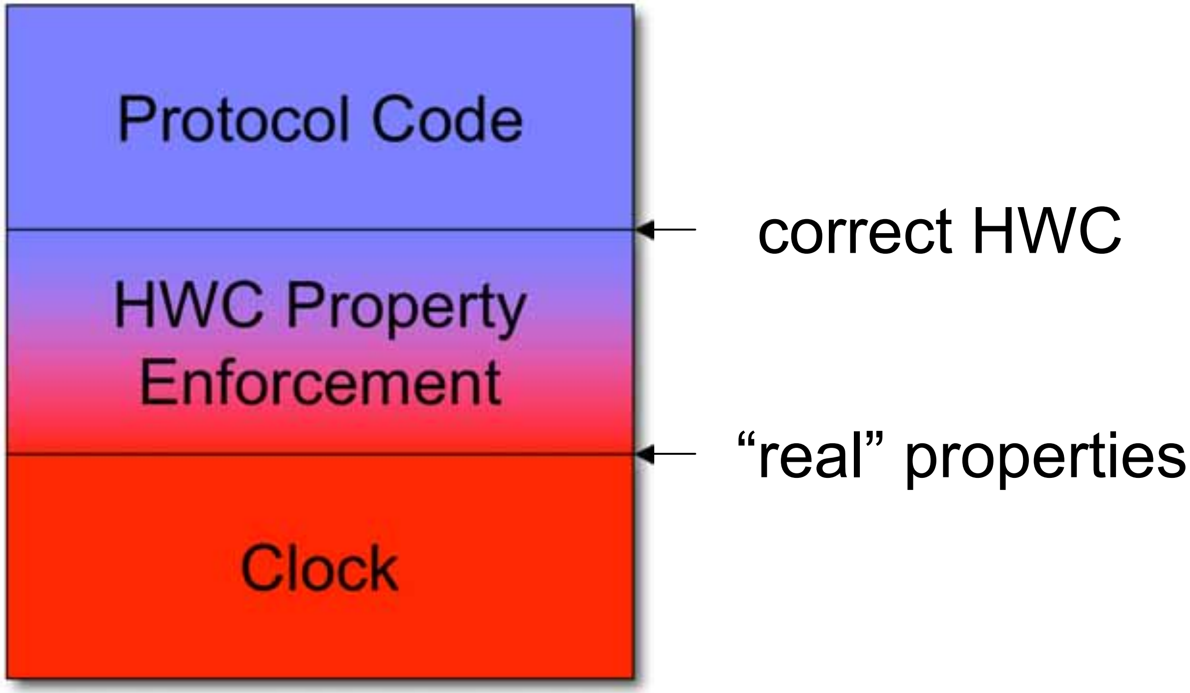
Measurements



Failure Assumption

- Failure Assumption:
  - Each correct process has a correct hardware clock, i.e., clock with a bounded drift rate.
- *Bounded drift rate*:
  - process can measure length of a time interval [s,t] with a max. error of ρ(t-s)

Hardware Clock Enforcement



Protocol Code

correct HWC

HWC Property Enforcement

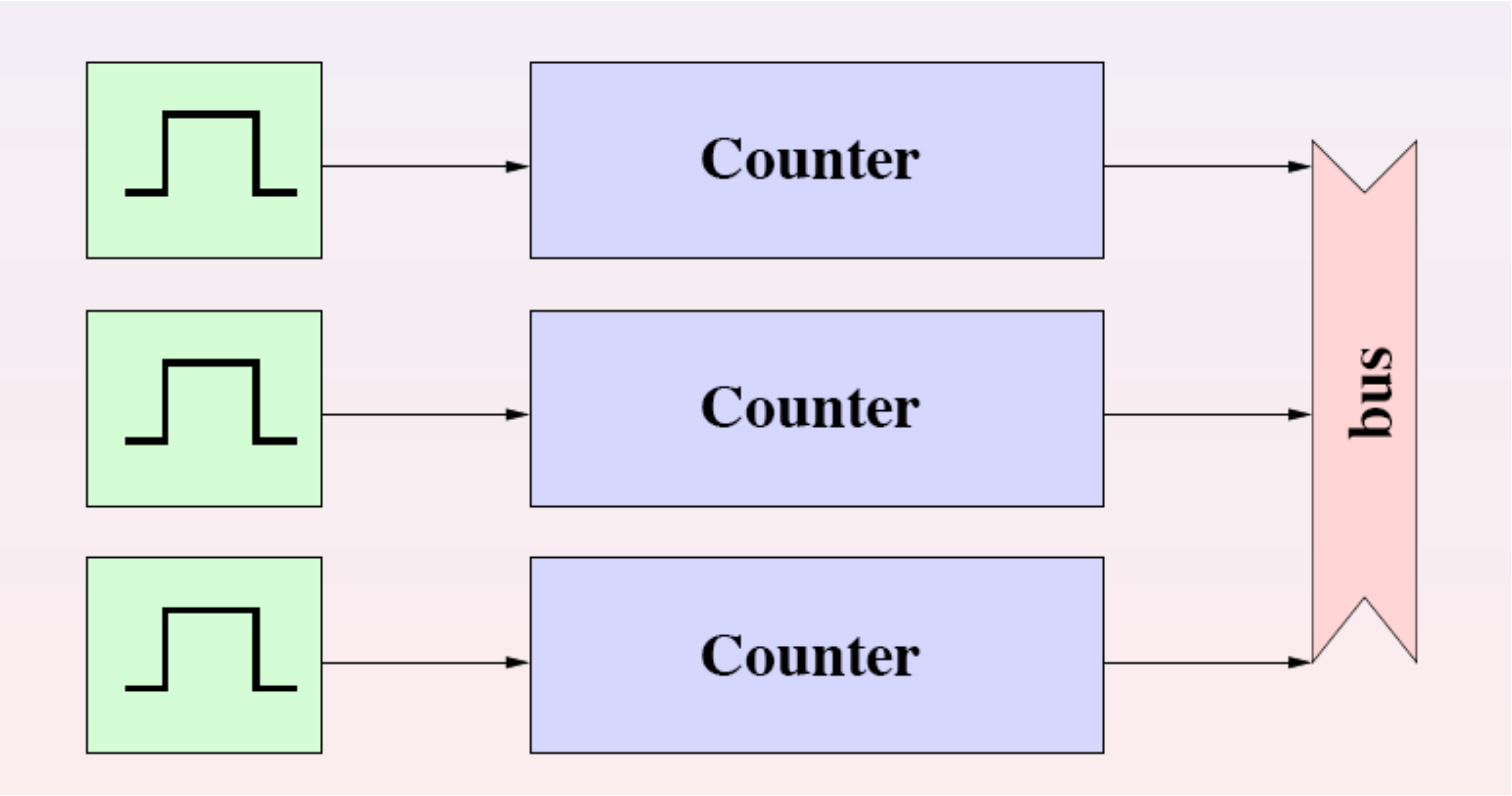
“real” properties

Clock

Clock Failure Semantics Enforcement

- We can try to detect clock failures and force a process to
  - crash if its hardware clock is faulty
- We can try to mask clock failures
- We can try to do both

Replicated Hardware Clock [2]



Replicated Hardware Clock

- Pentium processor has counter that is incremented in each cycle
  - Read counter with instruction: `rdtsc`
- Computers have hardware real-time clock
- Approach:
  - Can use different on-board clocks to enforce clock failure assumption

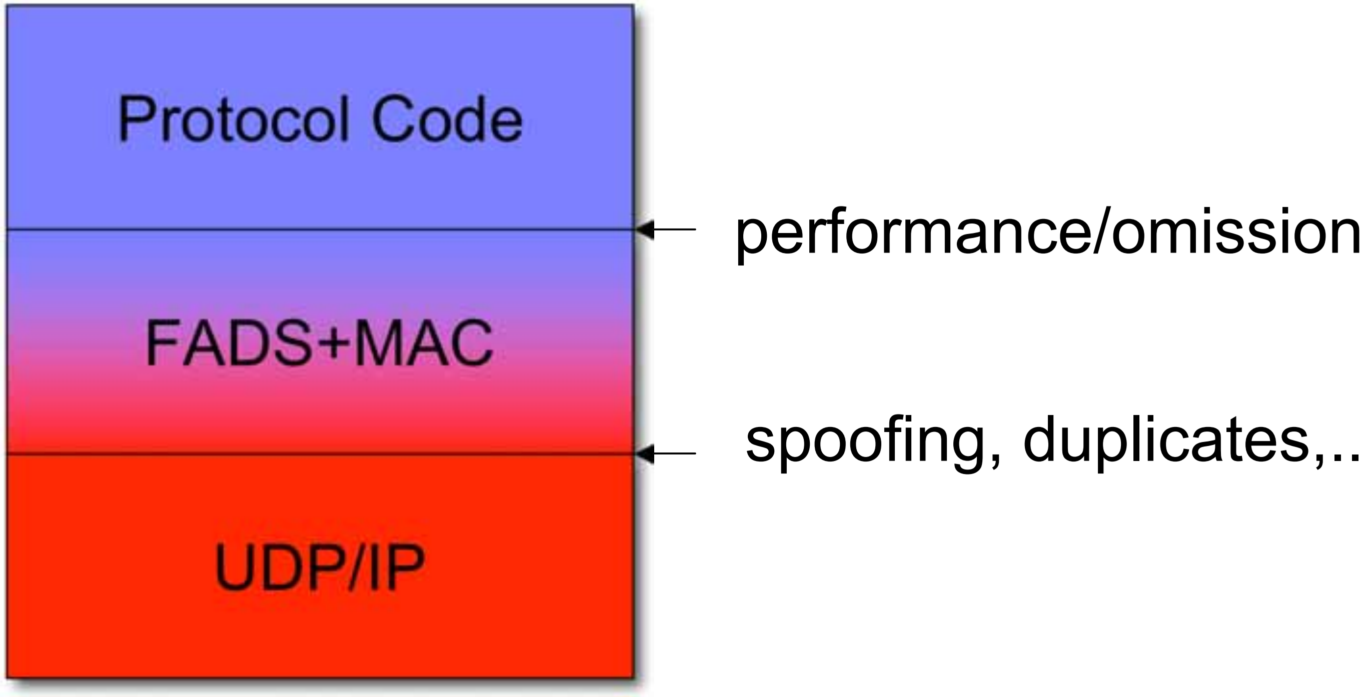
Datagram Service

Datagram Service

- Semantics:
  - At most once delivery of messages
- Performance failure:
  - message transmission delay > δ.
- Omission failure:
  - message transmission delay = ∞

- Note: No bound on the number of failures!

Datagram Failure Semantics Enforcement



Protocol Code

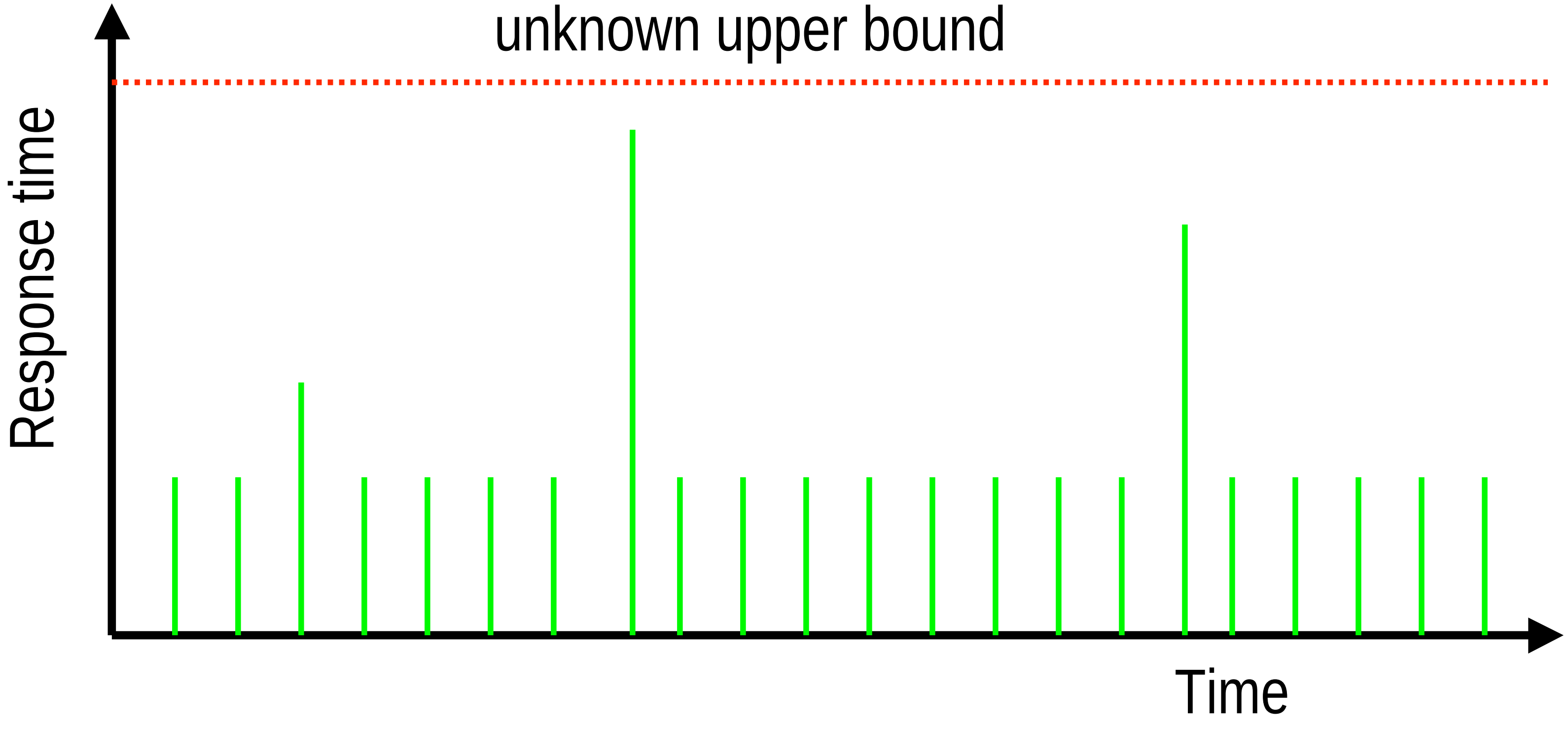
FADS+MAC

UDP/IP

performance/omission

spoofing, duplicates,..

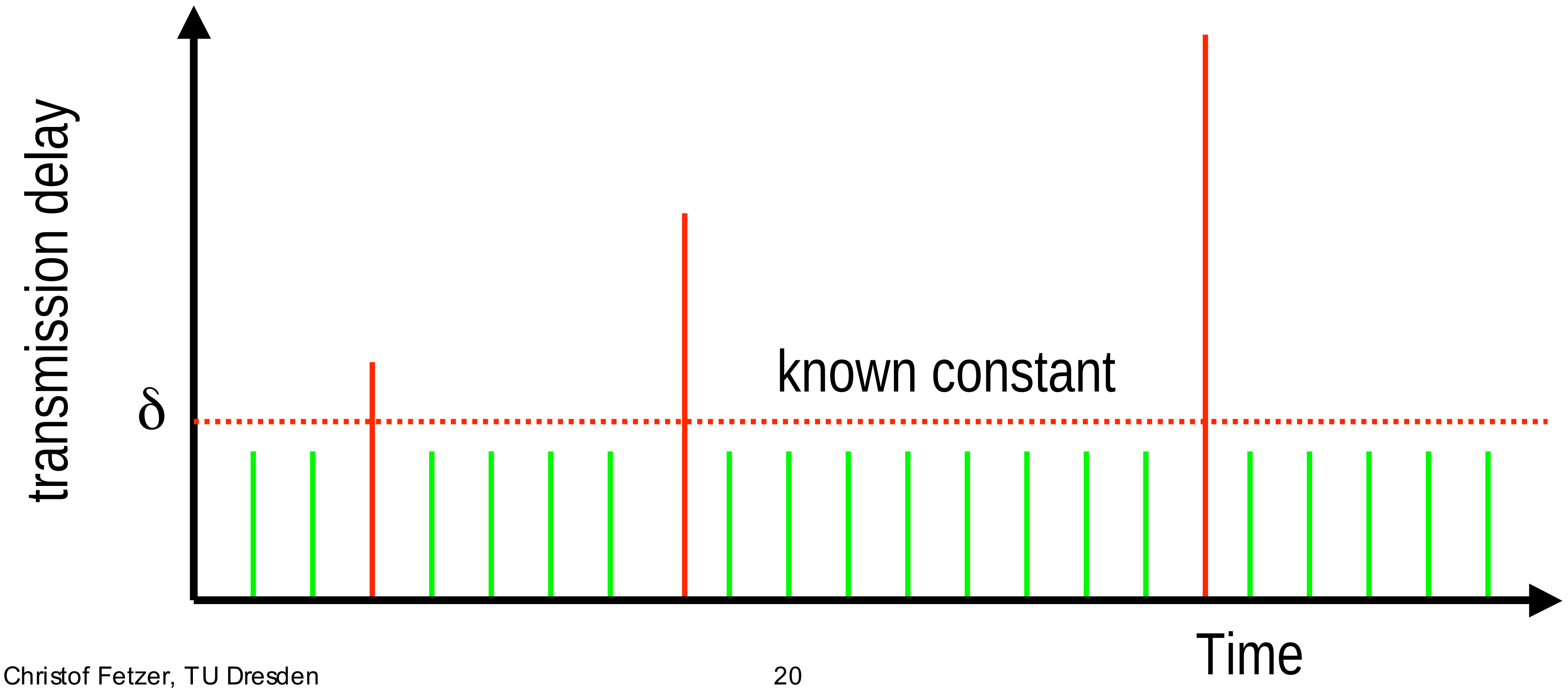
Partially Synchronous Systems



Timed Model: No Upper Bound

🟩 timely message

🟥 late message



Conditional Timeliness Requirements

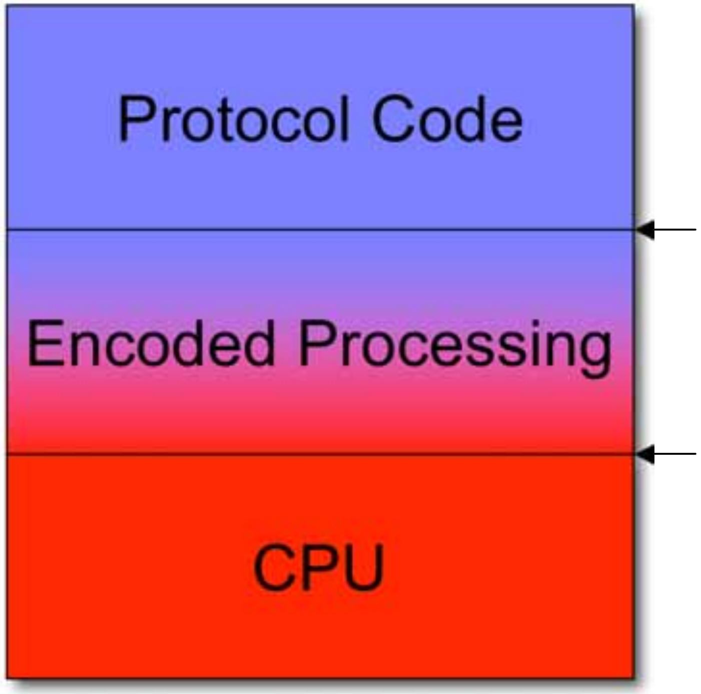
- Timeliness Requirement:
  - have to achieve something good in **D** seconds
- Conditional Timeliness Requirement:
  - have to achieve something good in **D** seconds if system is *stable*.

Process Service

Process Service

- Failure assumption:
  - **Processes have crash / performance failure semantics**

Process Failure Semantics Enforcement



Protocol Code

Encoded Processing

CPU

crash/performance failures

“arbitrary” failures

Possibilities and Impossibilities in the Timed Model

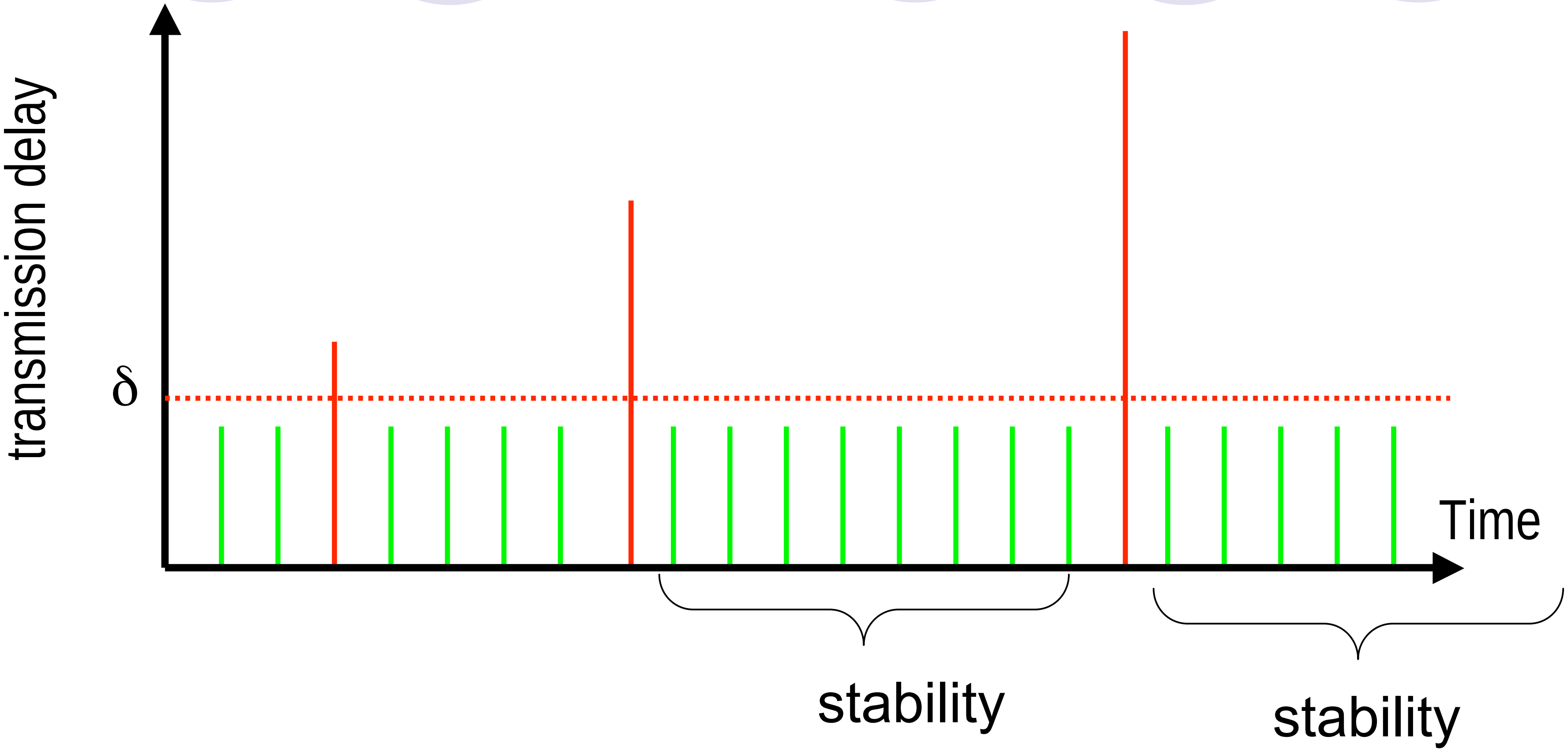
Most Standard Problems are impossible to solvable in TM

- For example, cannot solve
  - *consensus*,
  - *strong leader election*
  - eventually perfect failure detector
  - ...
- Reason:
  - Timed Model permits runs in which no message is delivered!

Two Approaches

- Change the problem:
  - enforce service properties whenever the underlying system is stable (synchronous)
  - if properties might be violated, signal to clients that properties are not guaranteed
    - we call that fail-awareness [3]
- *Add additional assumptions:*
  - *infinitely often the system is **stable***

Stability and instability periods



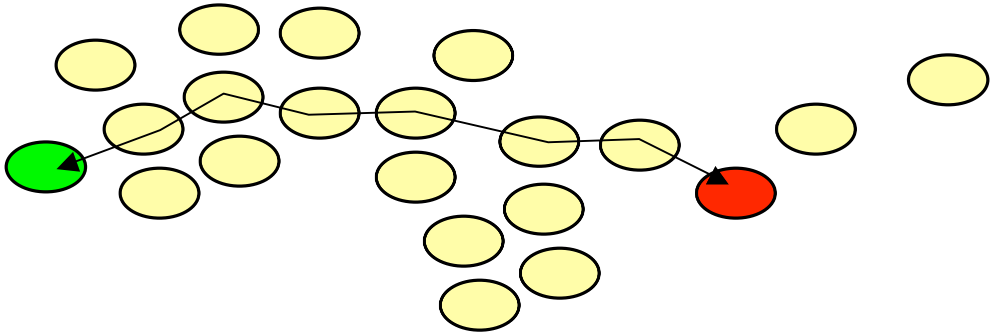
🟩 timely message 🟥 late message

Conditional Timeliness Requirements

- Timeliness Requirement:
  - ○ have to achieve something good in **D** seconds
- Conditional Timeliness Requirement:
  - ○ have to achieve something good in **D** seconds if system is *stable*.

Transmission delay...

- depends on diameter, density, ...
    - expect more variance in mobile/* systems
- How could nodes dynamically adjust δ?



Need to agree on a new δ

- Do we need the system to stabilize?
  - need to adjust δ when the system is unstable
- Do we really need a hardware clock?
  - e.g., change of clock frequency in mobile systems might complicate things...
  - use of minimal assumptions

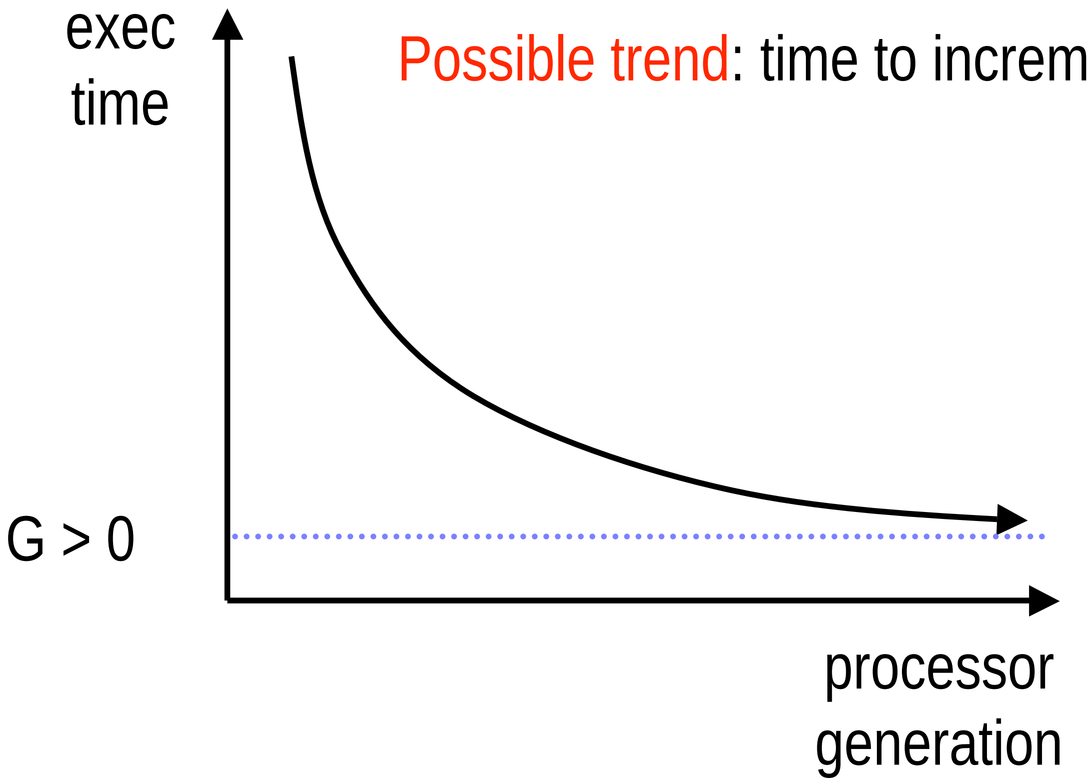
Finite Average Response Time Model (Far) Model

[5]

Observation 1:

Computers are not infinitely fast!

Max. Speed of ++ is bounded



Possible trend: time to increment integer

Weak Clock

- Clock with some max. unknown speed:

```
int tick = 0 ;
process Tick() {
    forever { tick++; }
}

int ReadClock() { return tick; }
```

Arbitrary Clock Failures

```
int tick = 0, last = 0; const int maxd = ...;
process Tick() { forever { tick++; } }

int ReadClock() {
        if (H() > tick) {
                tick = min(H(), tick+(tick-last)*maxd);
        } else { tick = max(H(), last); }
        last = ++tick;
        return last;
}
```

Weak Clock Semantics

- For each clock tick, at least some minimum unknown time G has passed
- What is it good for?
  - timeouts!

Observation 2:

In all well engineered systems(*), average transmission delay is finite.

(*) we need to take care of protocols without flow control

Communication System

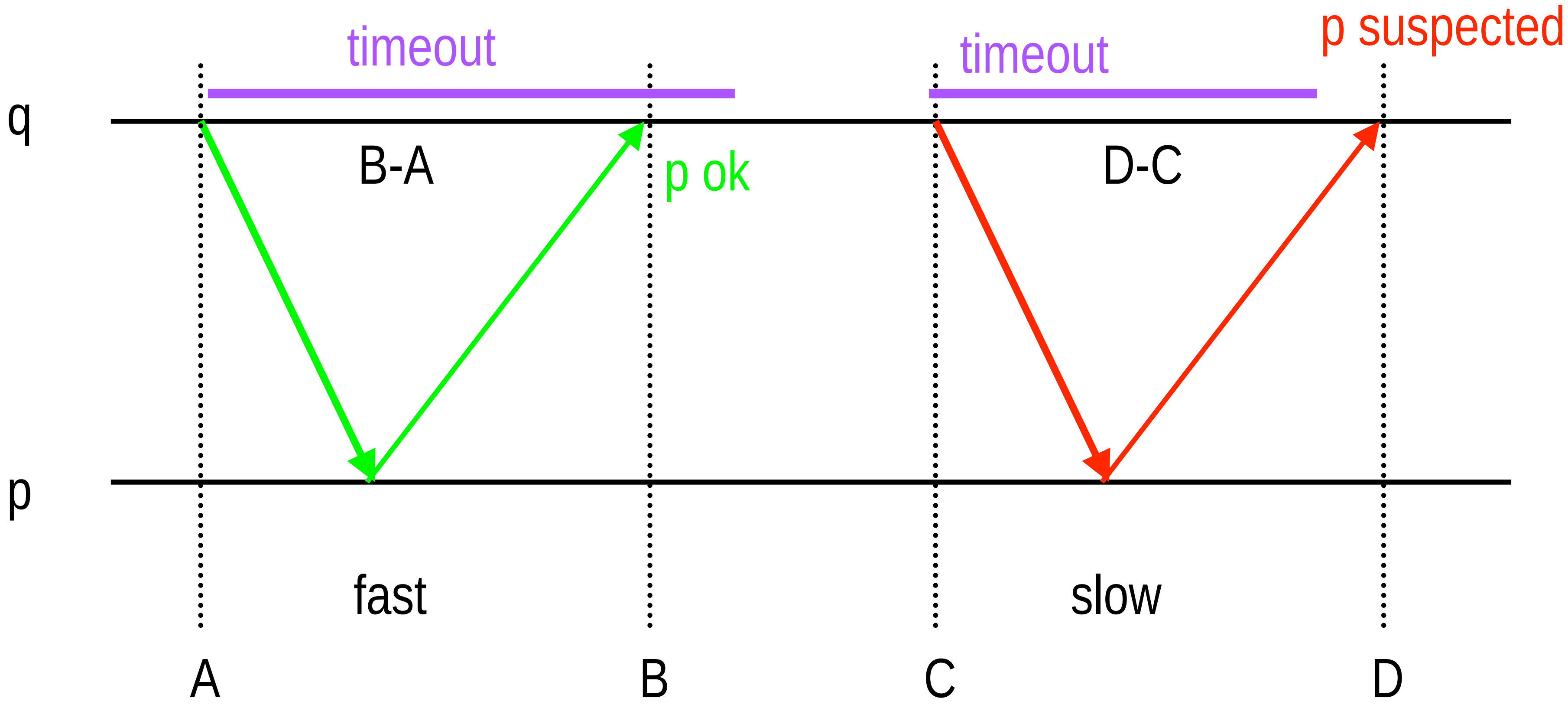
- We use **stubborn channels**
  - only reliable transmission of last message is guaranteed
  - need to wait for delivery of last message before transmitting new message

Finite Average Response Time

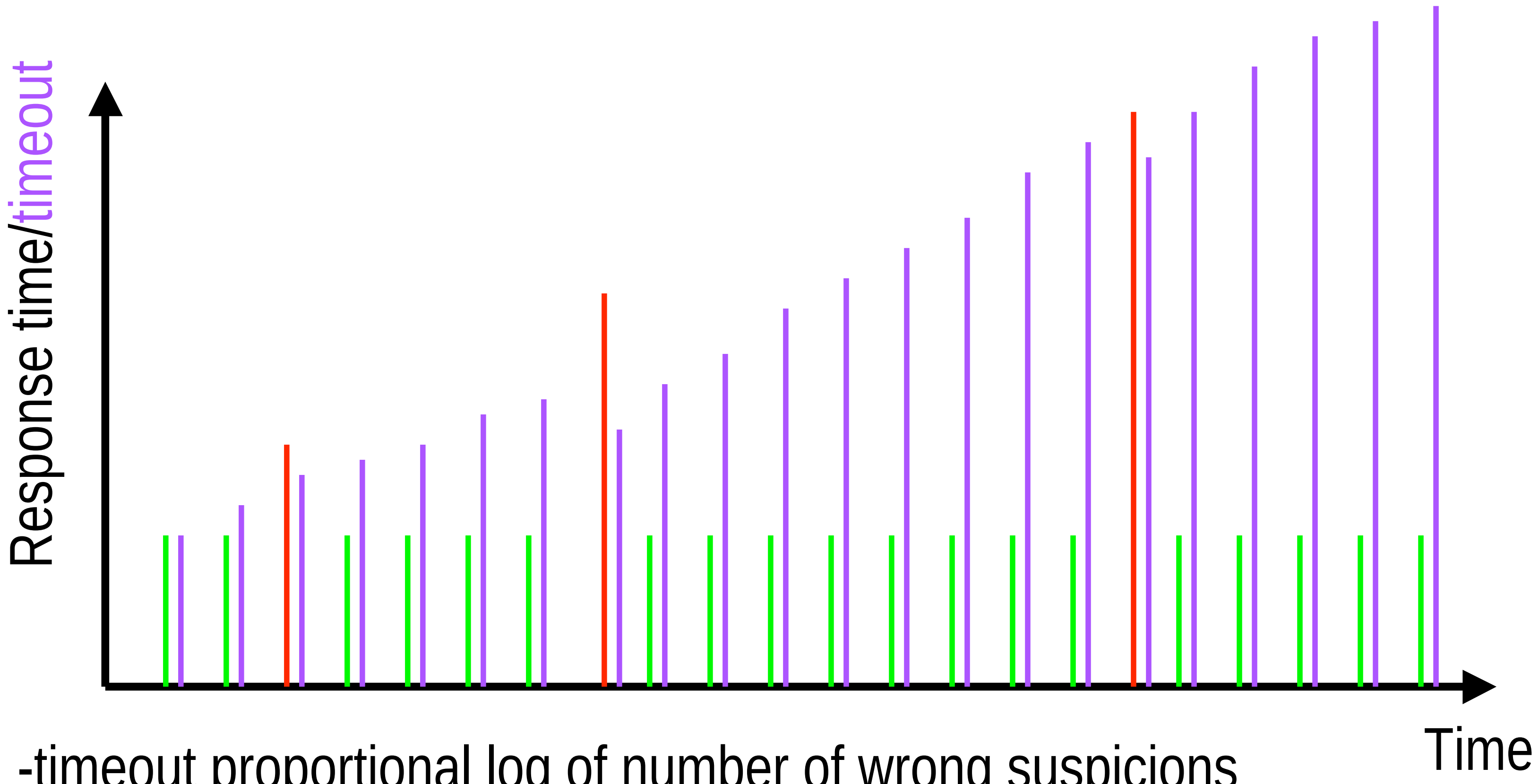
- **Assumption**:
  - average response time of link between any two correct processes is finite
  - average: $\lim_{k \to \infty}$(average of k first responses)

- Result:
  - Assumptions 1+2 sufficient to implement an eventually perfect failure detector [5]

Eventually Perfect Failure Detector



Timeout Adaptation



-timeout proportional log of number of wrong suspicions

-timeout proportional number fast messages since last slow message

Finite Average Response (FAR) Model [5]

- Eventually perfect failure detector (and hence consensus protocol) can be implemented in a system with
  - NO upper/relative bound on transmission delay
  - NO upper/relative bound on processing delay
  - NO assumption that system stabilizes
  - NO clocks, failure detectors, etc
- But
  - average response time must be finite
  - unknown min exec time for some operation

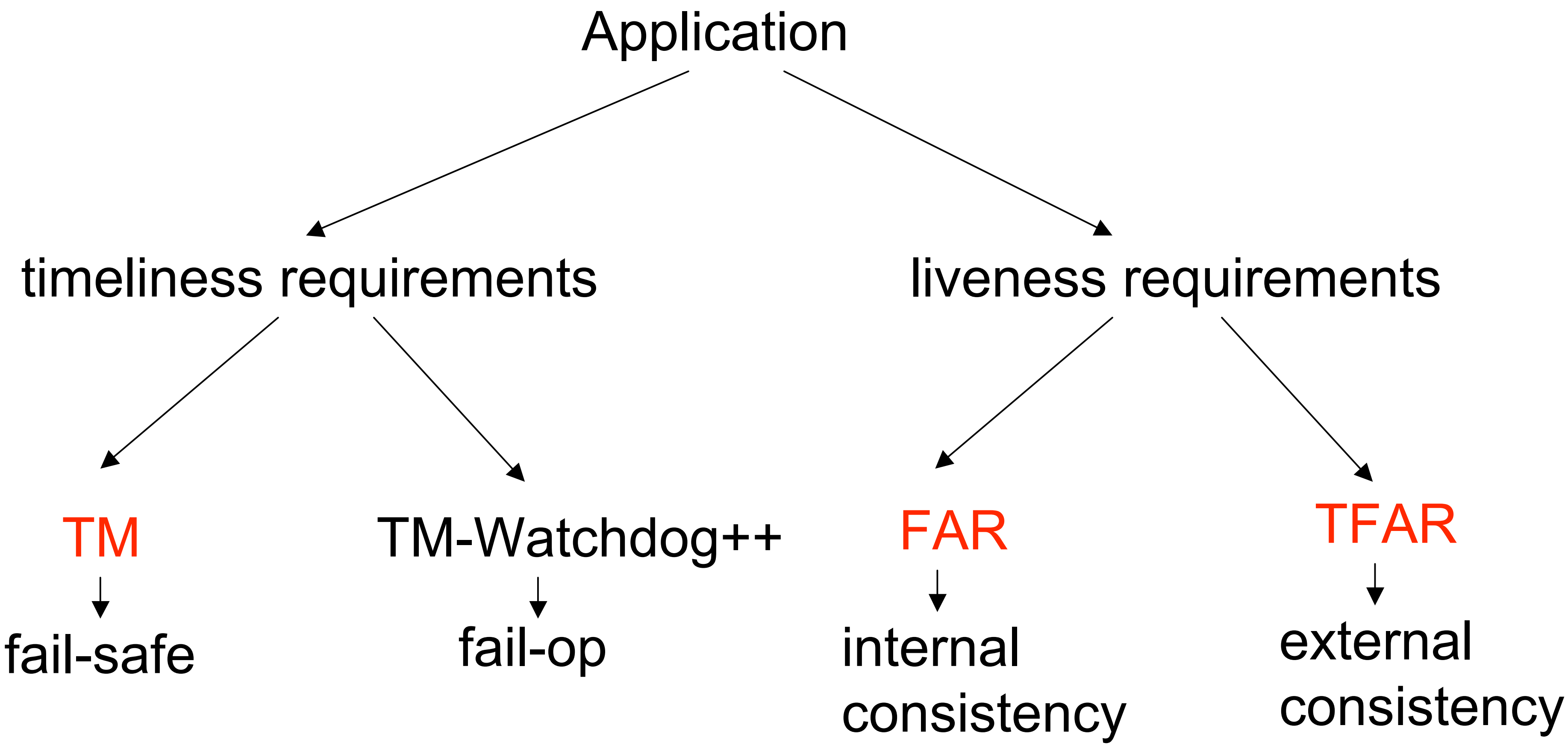
Timed Far Model

FAR Model [6]

Impossibility Result

- *Strong leader election* problem, i.e.,
  - infinitely often there is a leader
  - at any point in time there is at most one leader
- impossible to solve in FAR model [6]
  - adding a clock solves the problem
  - →Timed FAR model

Conclusion



Application

timeliness requirements

liveness requirements

TM → fail-safe

TM-Watchdog++ → fail-op

FAR → internal consistency

TFAR → external consistency

References

[1] F. Cristian and C. Fetzer, *The Timed Asynchronous Distributed System Model*, IEEE Transactions on Parallel and Distributed Systems

[2] C. Fetzer, F. Cristian, *Building Fault-Tolerant Hardware Clocks*, DCCA1999

[3] C. Fetzer and F. Cristian, *Fail-Awareness: An Approach to Construct Fail-Safe Applications*, Journal of Real-Time Systems, 2003

[4] C. Fetzer, F. Cristian, *A Fail-Aware Datagram Service*, IEE Proceedings - Software Engineering, 1999

[5] C. Fetzer, U. Schmid, M. Süßkraut, *On the Possibility of Consensus in Asynchronous Systems with Finite Average Response Times*, ICDCS 2005

[6] M. Süßkraut, C. Fetzer, *Leader Election in the Timed Finite Average Response Time Model*

Session 2.3

Mobility and Ubiquitous Computing

Moderator and Rapporteur

Henrique Madeira, University of Coimbra, Portugal

Sextant: A Comprehensive Localization Framework for Nomadic Computing

Emin Gün Sirer
Saikat Guha, Rohan Murty, Hongzhou Liu, Kevin Walsh

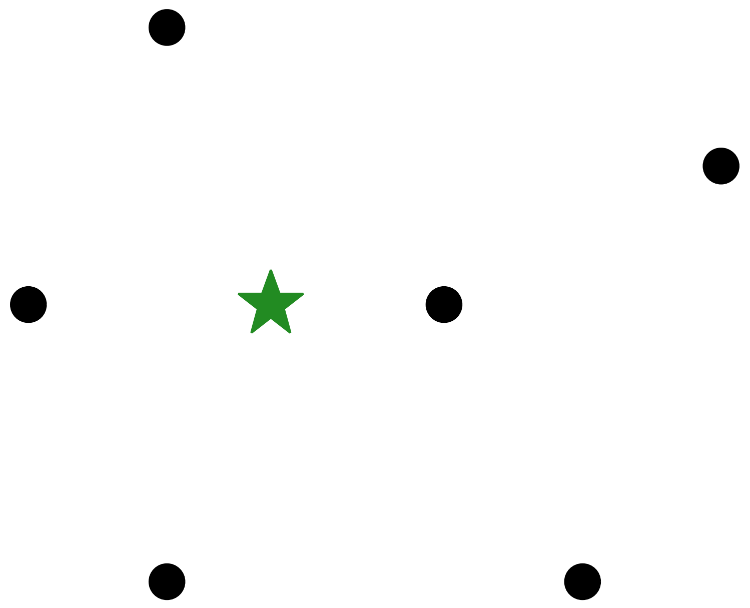
Cornell University

IFIP WG 10.4, July 4, 2005

Dependable Nomadic Systems

- Nomadic systems pose many problems
  - Localization (Sextant, [Mobihoc 2005])
  - Programming Model (MagnetOS, [MobiSys 2005])
  - Routing (SHARP, [Mobihoc 2002])
  - Path Selection (DPSP, [Mobihoc 2001])
  - Simulation (SNS, [WSC 2003, TOMACS 2004])
  - ...
- Need to figure out the location of nodes in order to provide novel location-based services
- Need a new programming model for performing long-lived computations in mobile networks

Challenges in Localization



Hardware

- Expensive
- Power Consuming

Infrastructure

- Initial setup required
- Not always available

Modeling

- Irregular wireless coverage area
- Introduces error

Sextant Approach

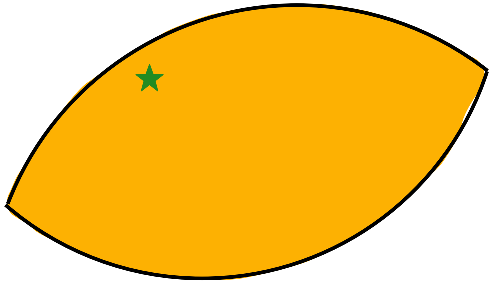
- Extract geometric constraints
- Disseminate them transitively
- Solve in a distributed manner

Sextant Approach

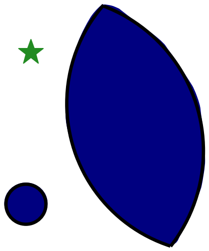
Contributions

- Unified Node and Event localization
- Accurate
  - Negative as well as positive information
  - Explicit representation
- Practical
  - Constraint extraction
  - Deployed on MICA-2 motes, laptops and PDAs

Sextant Approach



Positive constraint

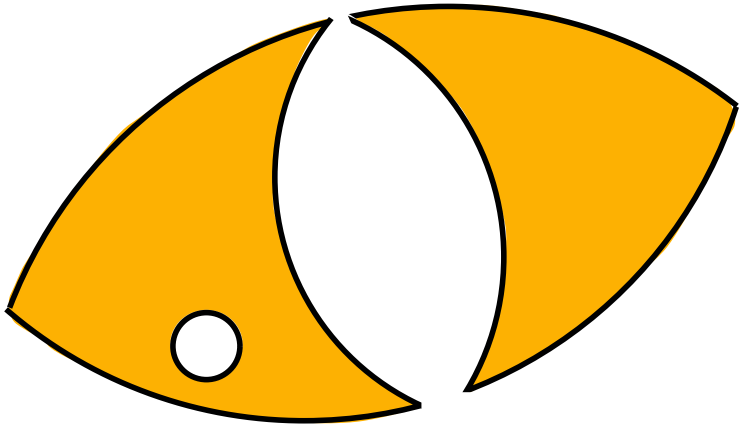


Negative constraint

Contributions

- Unified Node and Event localization
- Accurate
  - Negative as well as positive information
  - Explicit representation
- Practical
  - Constraint extraction
  - Deployed on MICA-2 motes, laptops and PDAs

Sextant Approach



- Need not be convex
- May have holes
- May have disconnected components

Contributions

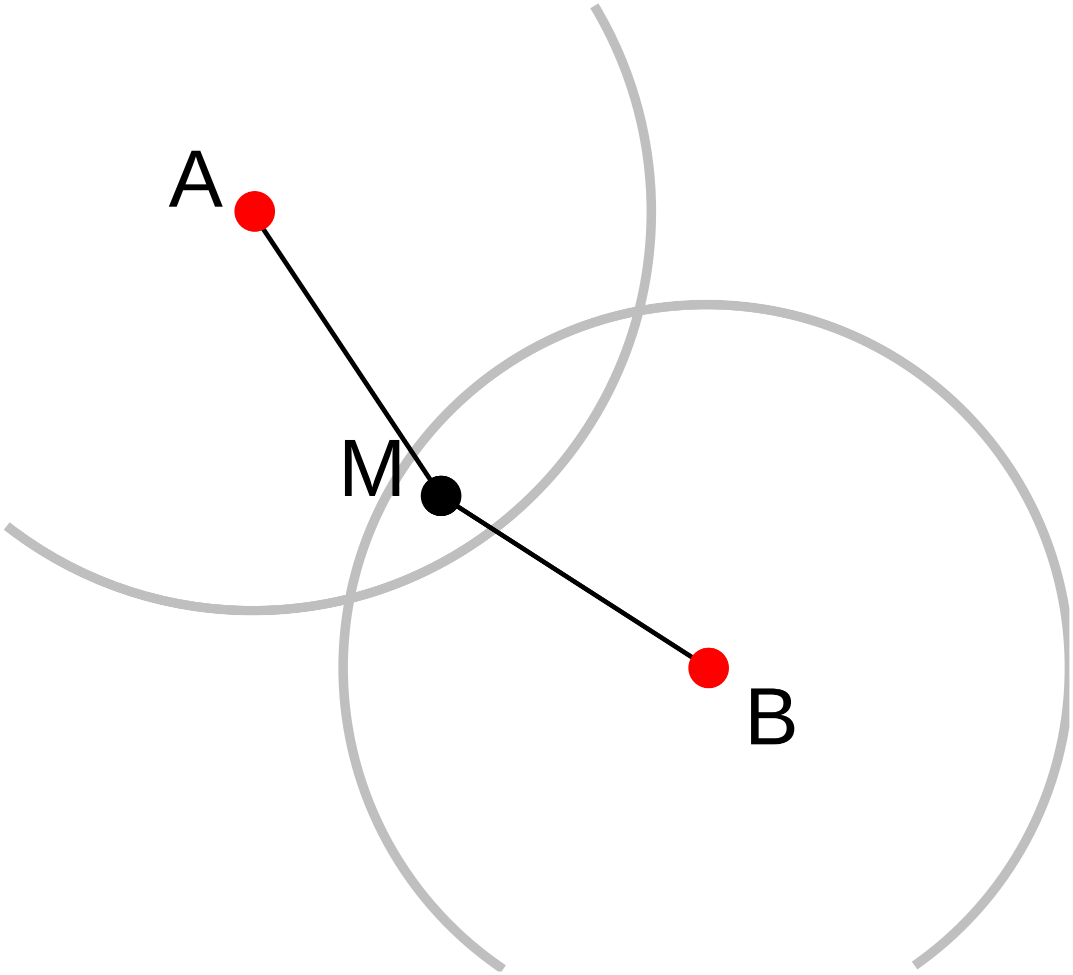
- Unified Node and Event localization
- Accurate
  - Negative as well as positive information
  - Explicit representation
- Practical
  - Constraint extraction
  - Deployed on MICA-2 motes, laptops and PDAs

Sextant Approach

Contributions

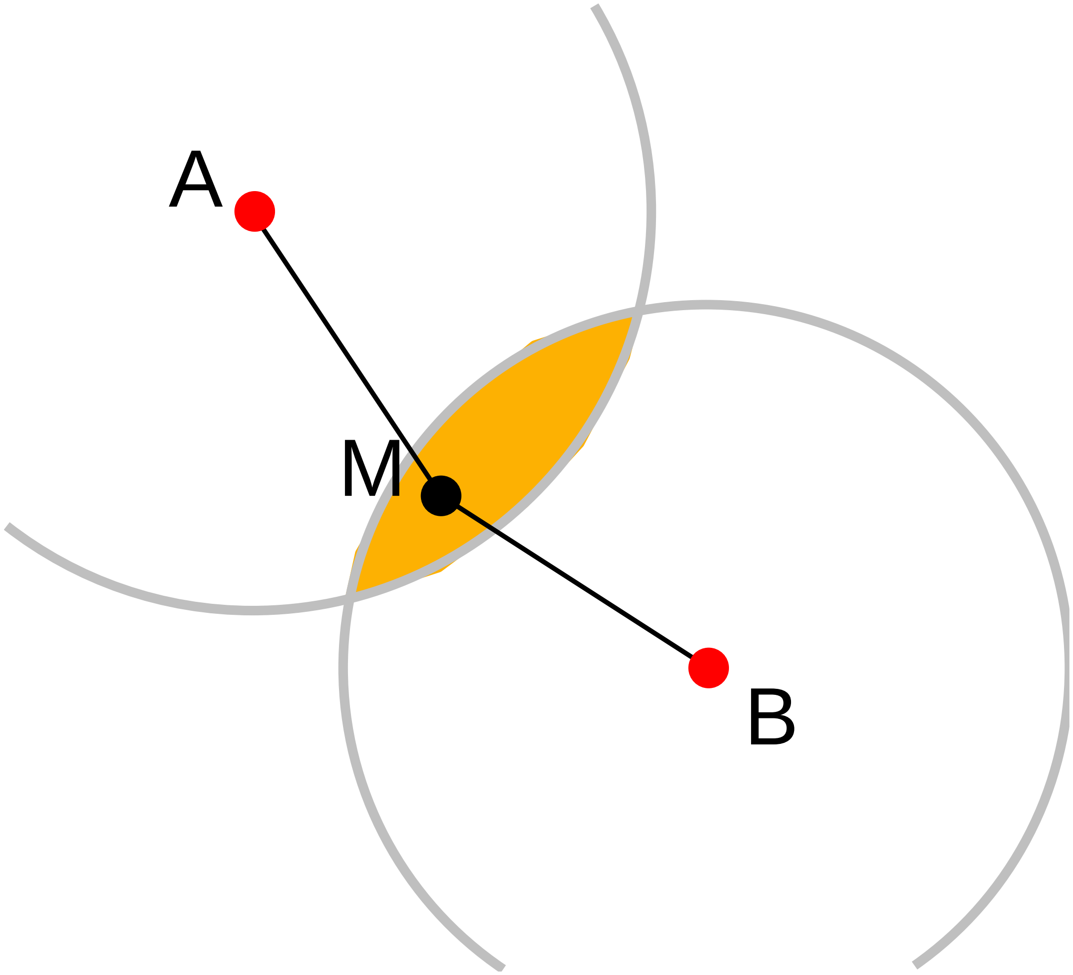
- Unified Node and Event localization
- Accurate
  - Negative as well as positive information
  - Explicit representation
- Practical
  - Constraint extraction
  - Deployed on MICA-2 motes, laptops and PDAs

Node Localization



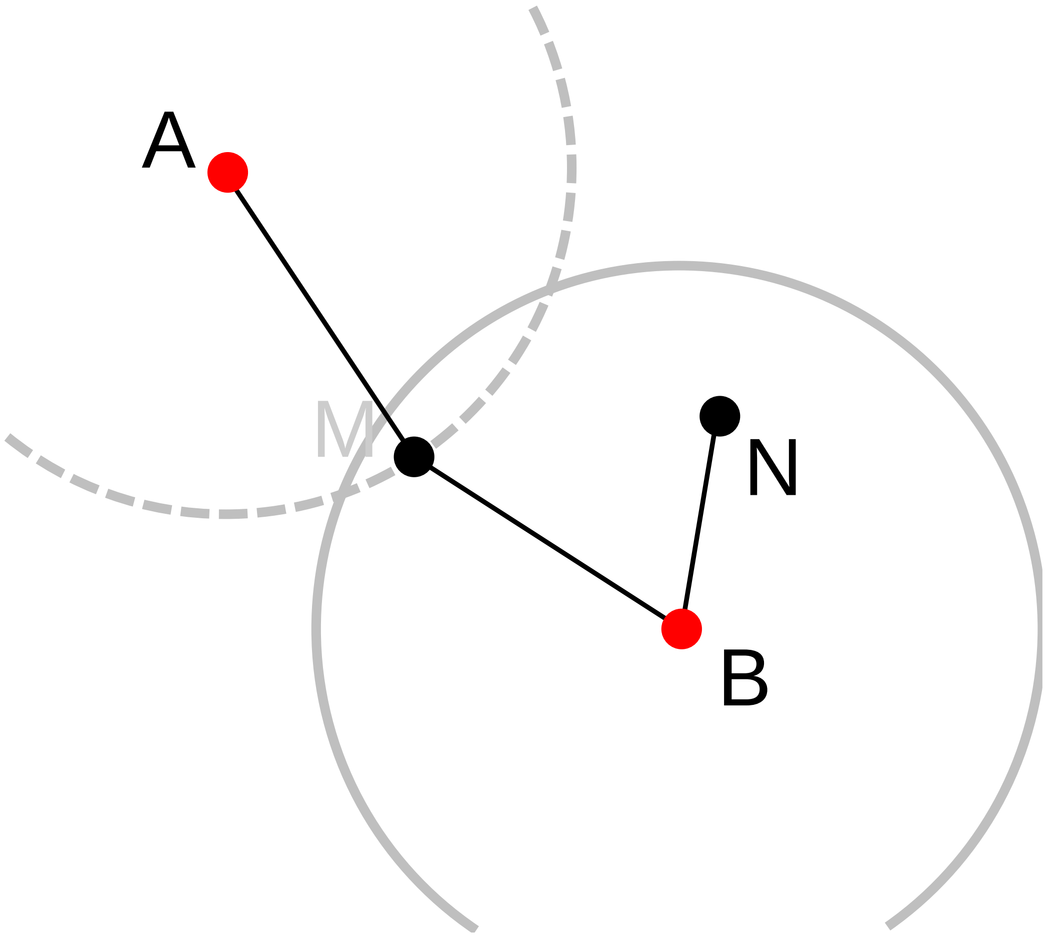
Positive Information

Node Localization



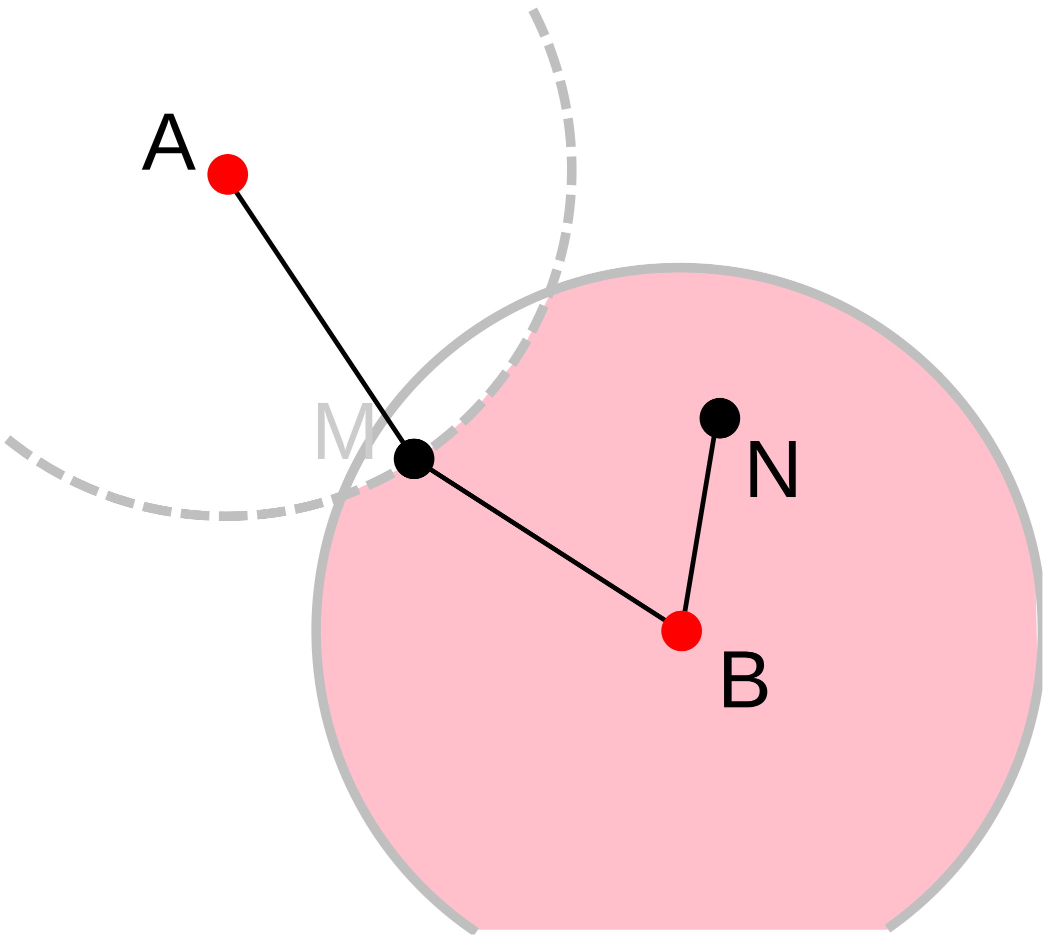
Intersection of Positive Information

Node Localization



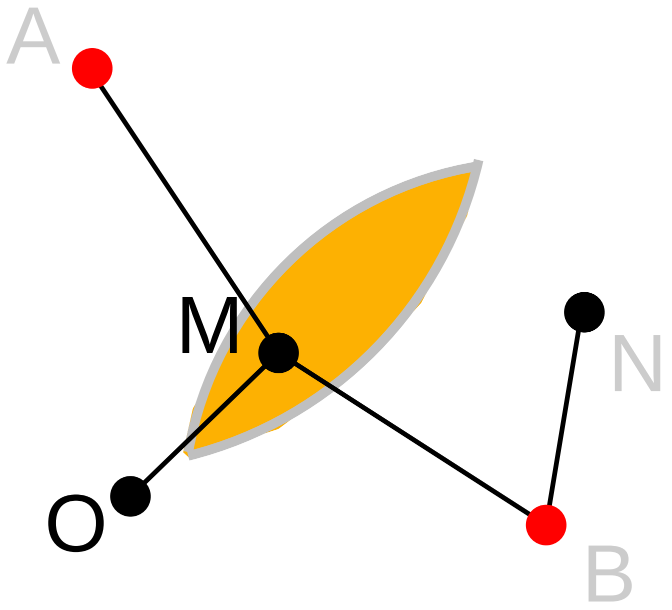
Negative Information

Node Localization



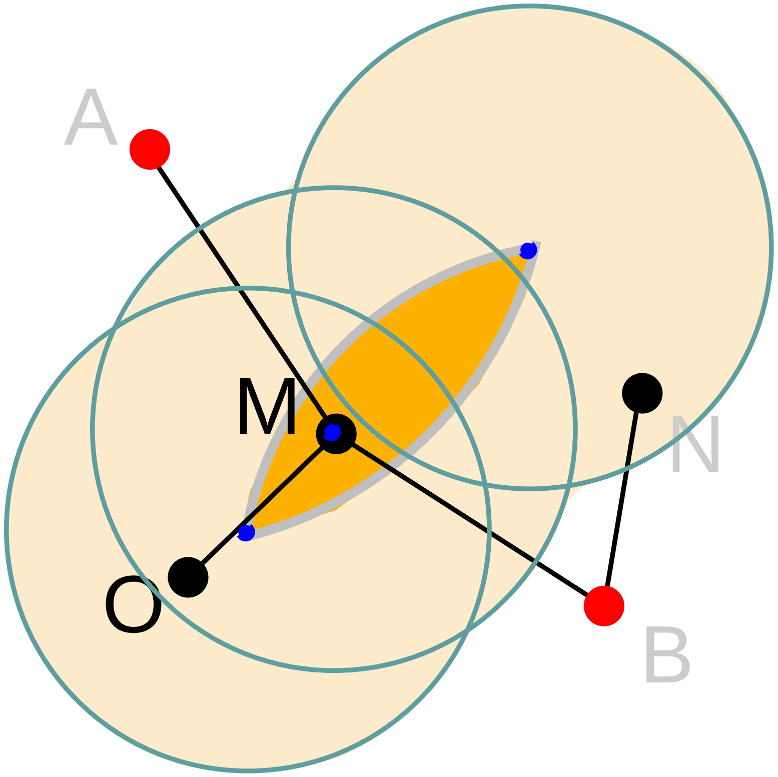
Subtraction of Negative Information

Node Localization



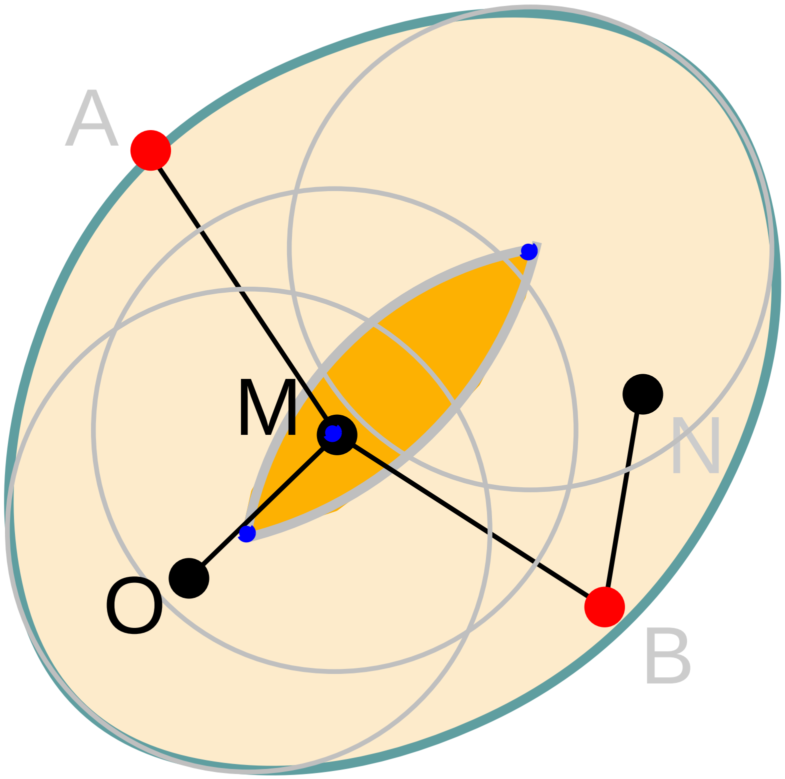
Transitive Dissemination of Positive Information

Node Localization



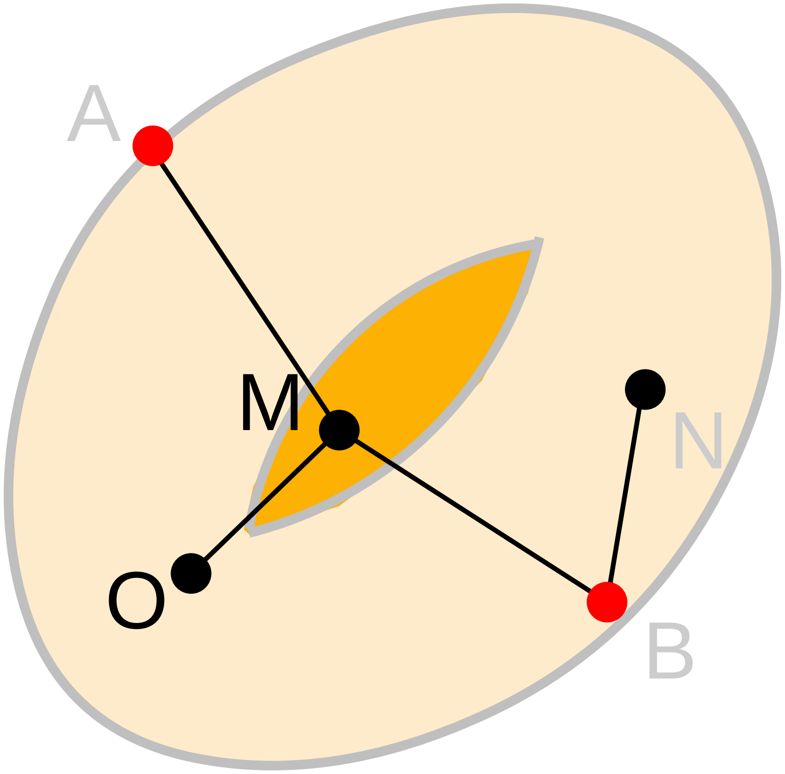
Transitive Dissemination of Positive Information

Node Localization



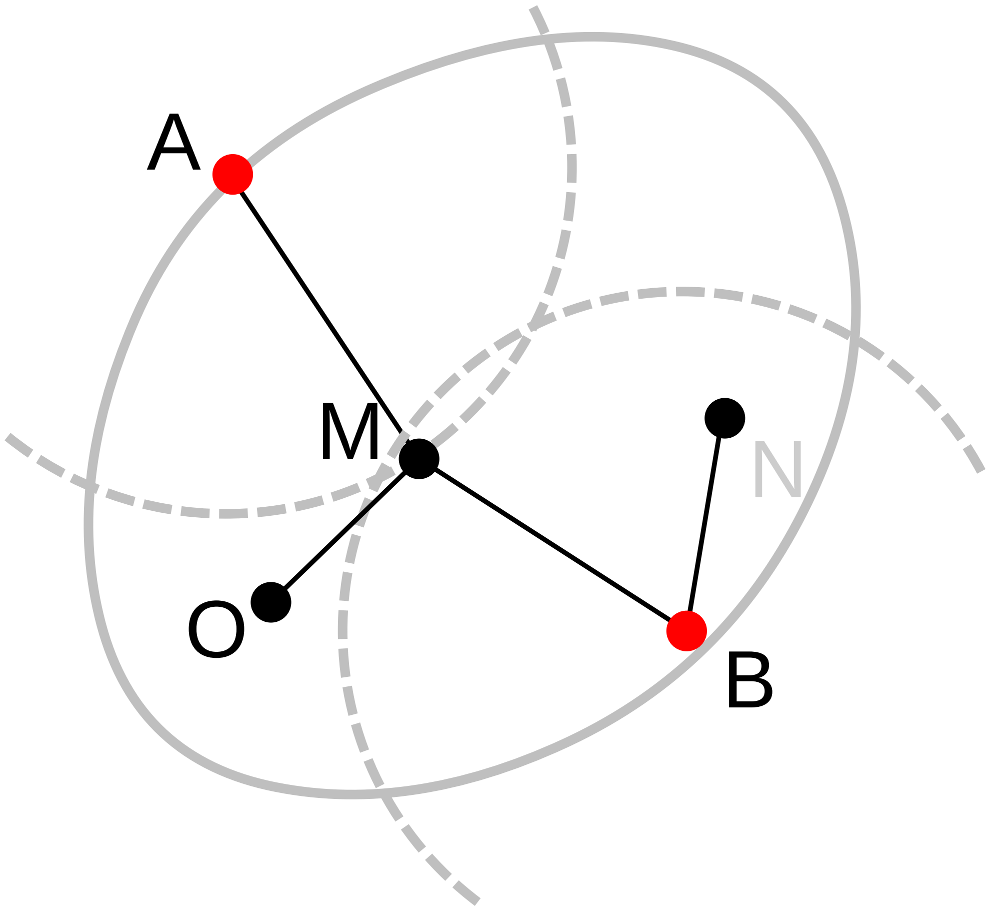
Transitive Dissemination of Positive Information

Node Localization



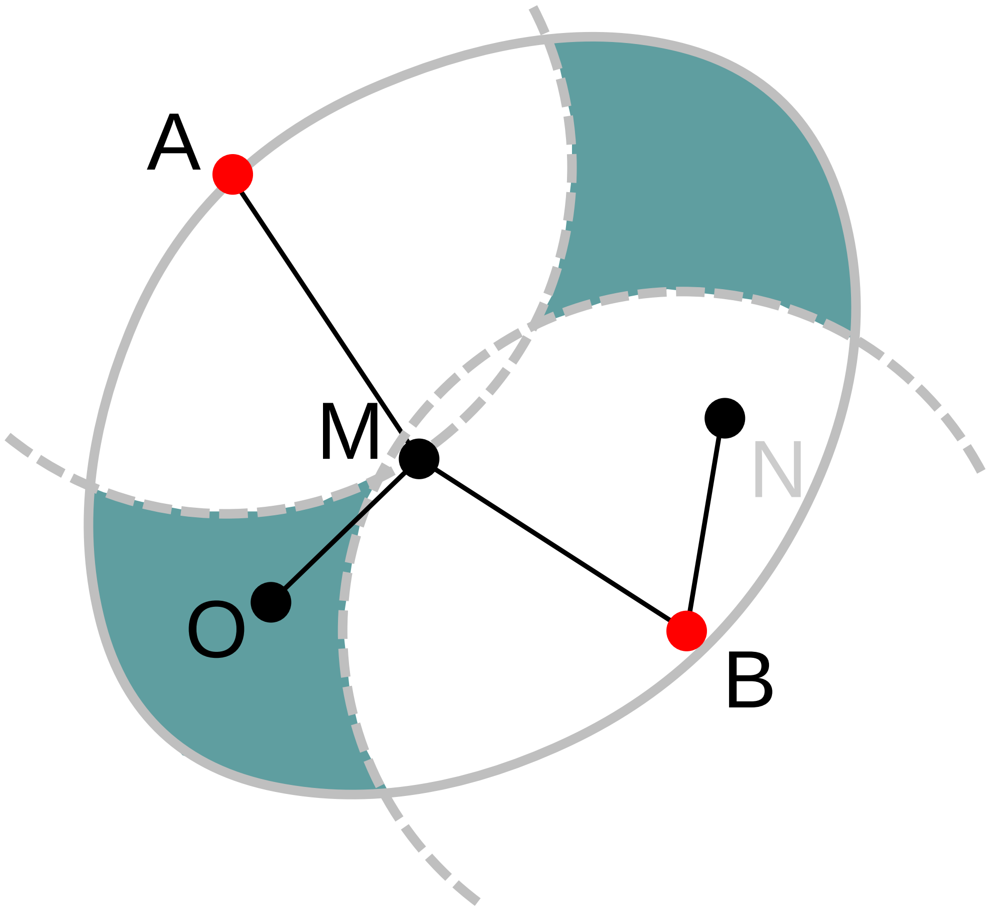
Transitive Dissemination of Positive Information

Node Localization



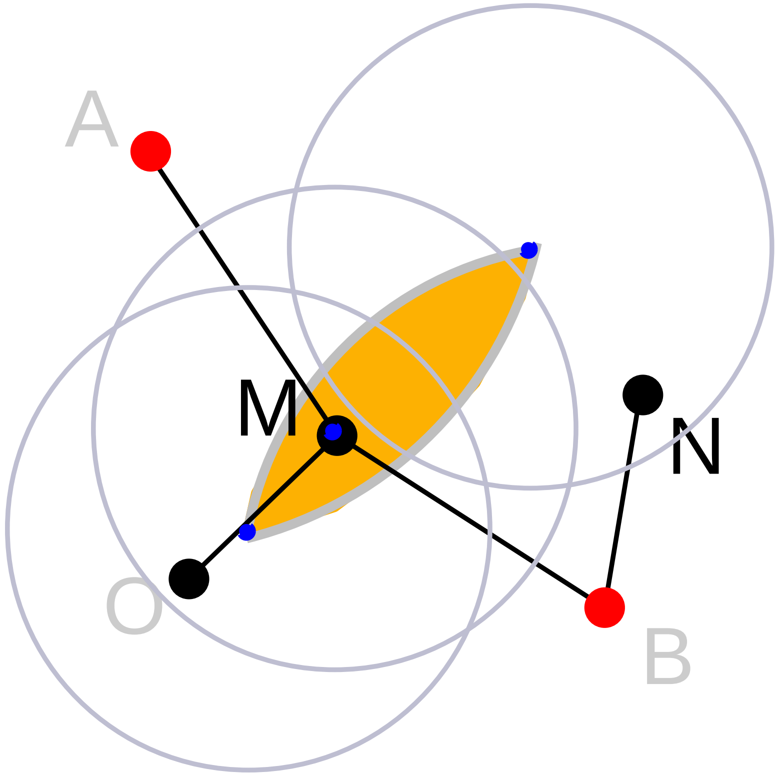
Combining Positive and Negative Information

Node Localization



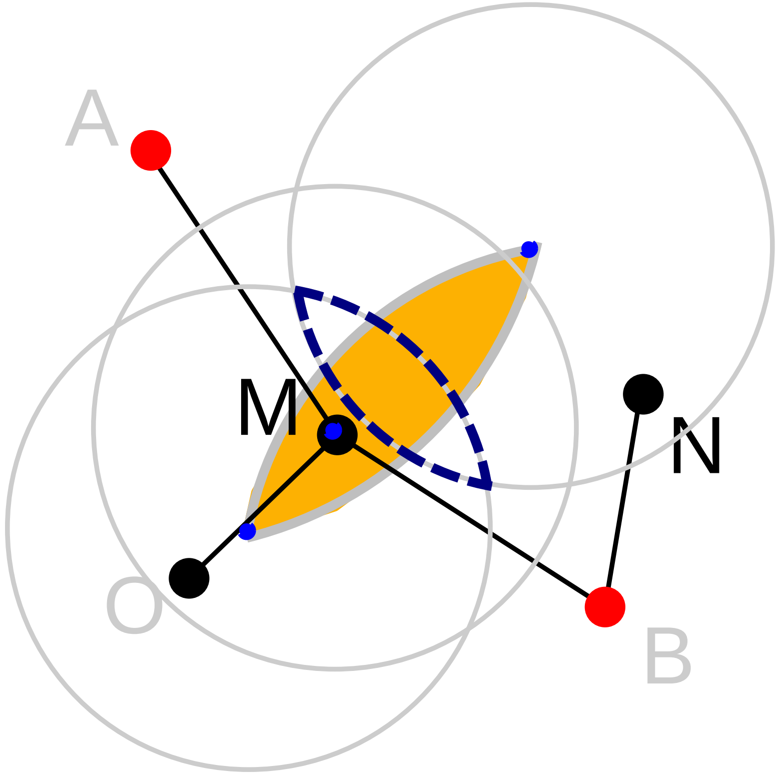
Combining Positive and Negative Information

Node Localization



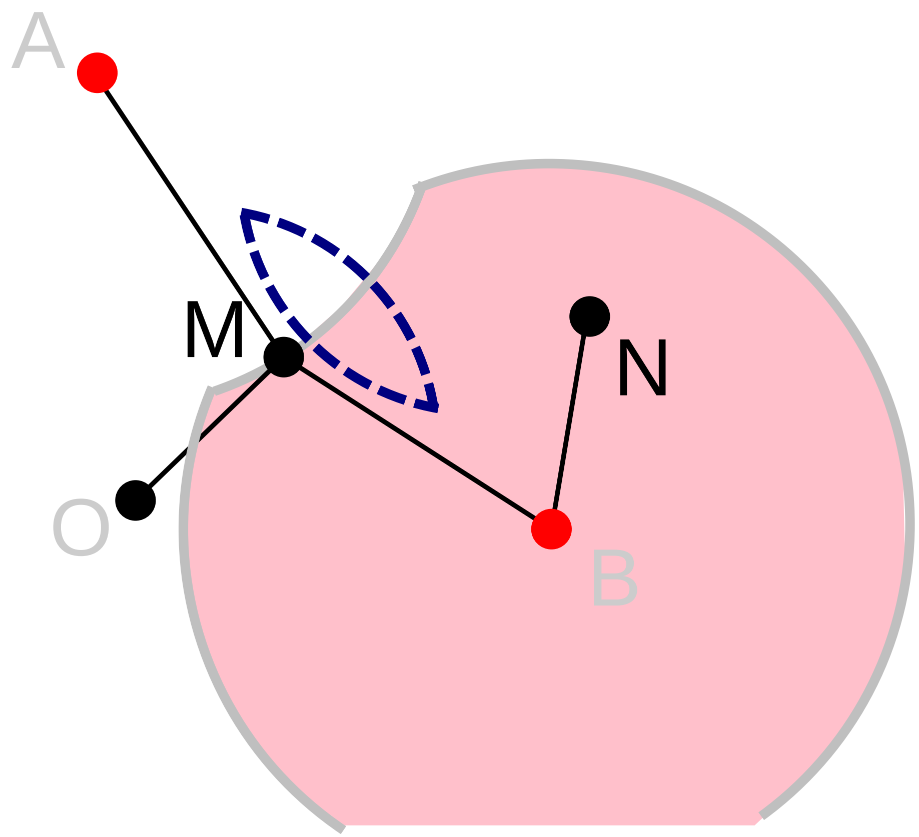
Transitive Dissemination of Negative Information

Node Localization



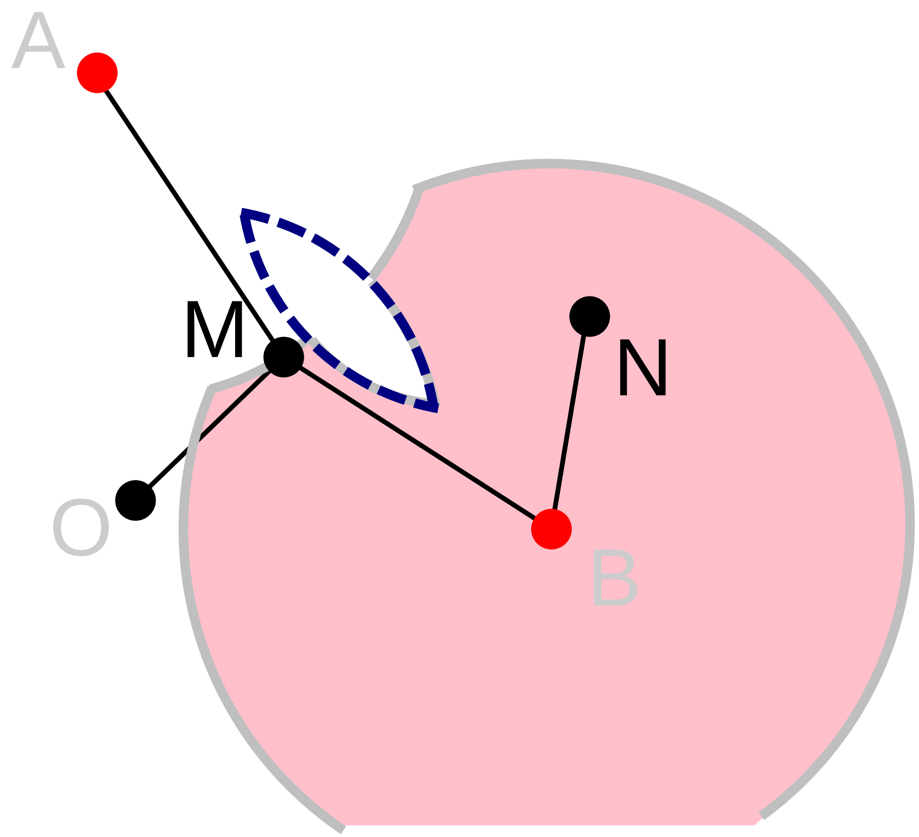
Transitive Dissemination of Negative Information

Node Localization



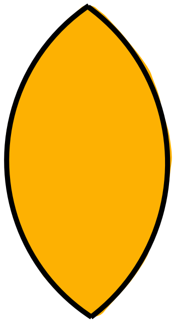
Refining Location Estimates

Node Localization



Refining Location Estimates

Sextant Approach



Polygons with Bézier boundaries

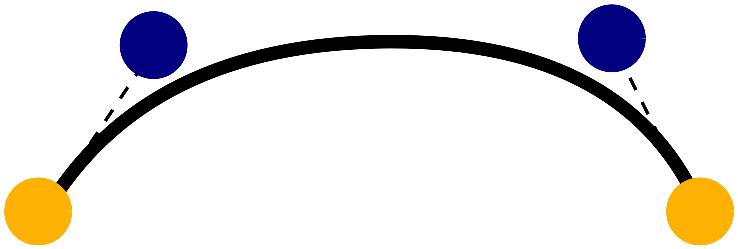
Each Node x

- Location Estimate: $\mathcal{E}_x$
- Positive Constraint: $\mathcal{P}_x$
- Negative Constraint: $\mathcal{N}_x$
- Set of positive constraints: Γ_x
- Set of negative constraints: Θ_x

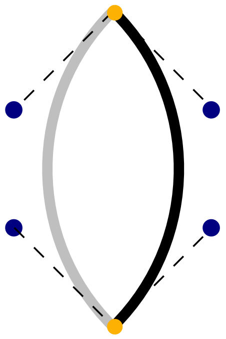
Invariant

$$\mathcal{E}_x = \bigcap_{p \in \Gamma_x} p \setminus \bigcup_{n \in \Theta_x} n$$

Sextant Approach



Bézier curve



Polygons with Bézier boundaries

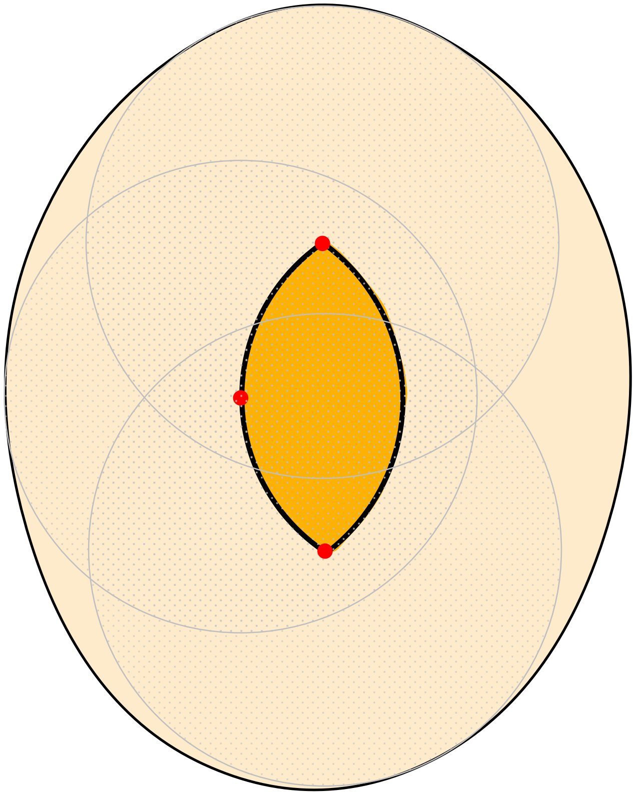
Each Node x

- Location Estimate: $\mathcal{E}_x$
- Positive Constraint: $\mathcal{P}_x$
- Negative Constraint: $\mathcal{N}_x$
- Set of positive constraints: Γ_x
- Set of negative constraints: Θ_x

Invariant

$$\mathcal{E}_x = \bigcap_{p \in \Gamma_x} p \setminus \bigcup_{n \in \Theta_x} n$$

Sextant Approach



Union of circles in $\mathcal{E}_x$

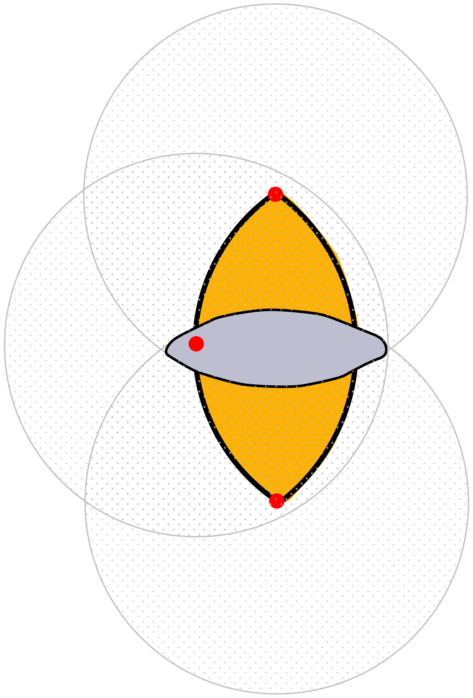
Each Node x

- Location Estimate: $\mathcal{E}_x$
- Positive Constraint: $\mathcal{P}_x$
- Negative Constraint: $\mathcal{N}_x$
- Set of positive constraints: Γ_x
- Set of negative constraints: Θ_x

Invariant

$$\mathcal{E}_x = \bigcap_{p \in \Gamma_x} p \setminus \bigcup_{n \in \Theta_x} n$$

Sextant Approach



Intersection of circles in $\mathcal{E}_x$

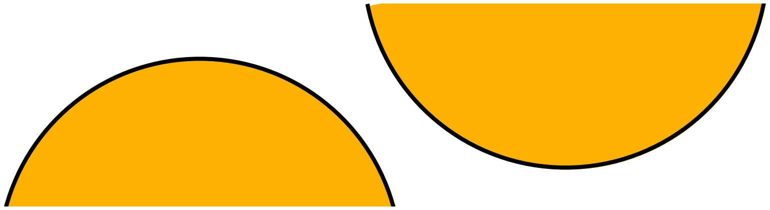
Each Node x

- Location Estimate: $\mathcal{E}_x$
- Positive Constraint: $\mathcal{P}_x$
- Negative Constraint: $\mathcal{N}_x$
- Set of positive constraints: Γ_x
- Set of negative constraints: Θ_x

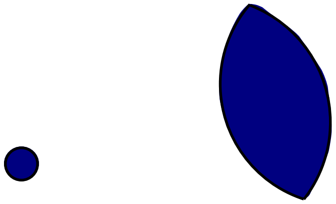
Invariant

$$\mathcal{E}_x = \bigcap_{p \in \Gamma_x} p \setminus \bigcup_{n \in \Theta_x} n$$

Sextant Approach



Γ_x: learned from wireless neighbors



Θ_x: learned from wireless non-neighbors

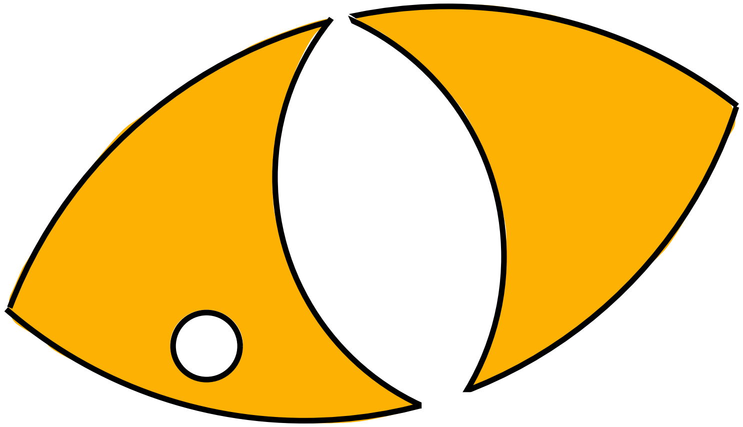
Each Node x

- Location Estimate: $\mathcal{E}_x$
- Positive Constraint: $\mathcal{P}_x$
- Negative Constraint: $\mathcal{N}_x$
- Set of positive constraints: Γ_x
- Set of negative constraints: Θ_x

Invariant

$$\mathcal{E}_x = \bigcap_{p \in \Gamma_x} p \setminus \bigcup_{n \in \Theta_x} n$$

Sextant Approach



Each Node x

- Location Estimate: $\mathcal{E}_x$
- Positive Constraint: $\mathcal{P}_x$
- Negative Constraint: $\mathcal{N}_x$
- Set of positive constraints: Γ_x
- Set of negative constraints: Θ_x

Invariant

$$\mathcal{E}_x = \bigcap_{p \in \Gamma_x} p \setminus \bigcup_{n \in \Theta_x} n$$

Event Localization

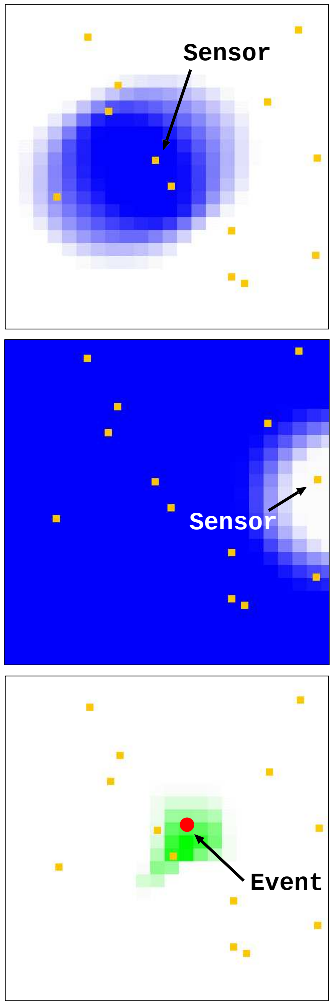
Similarity to Node Localization

- Constraints from sensing hardware vs. wireless radio
- Boolean sensed/not-sensed signal vs. boolean connectivity

Differences from Node Localization

- Annotate resultant areas with probabilities

Event Localization



Bayesian Probability

Positive Contribution

Sensor somewhere in $\mathcal{E}$ detects event; probability event in grid $\mathcal{G}_i$.

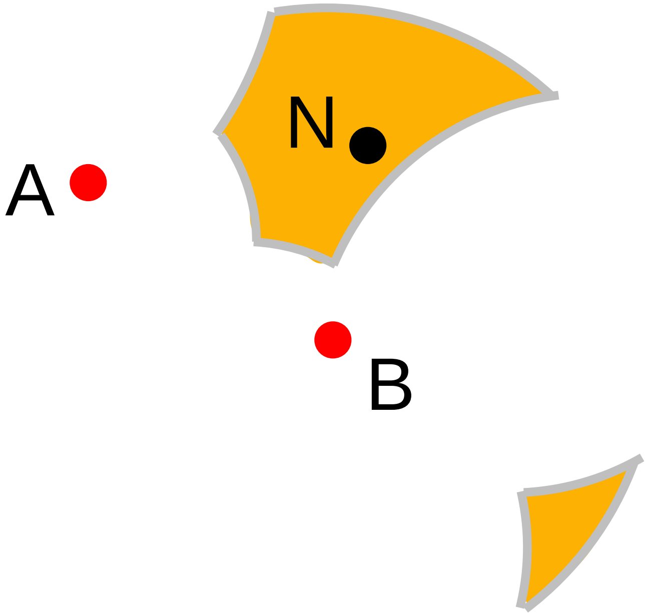
Negative Contribution

Sensor somewhere in $\mathcal{E}$ does not detect event; probability event in grid $\mathcal{G}_i$.

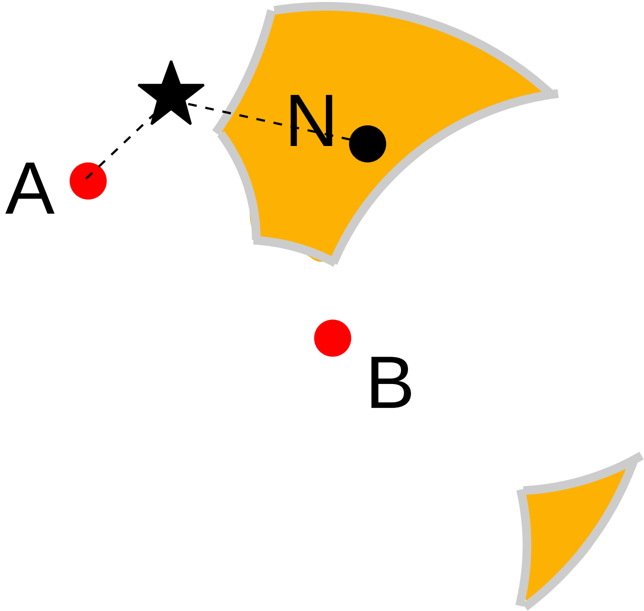
Solution

Product of positive and negative contributions from sensors sensing and not-sensing the event.

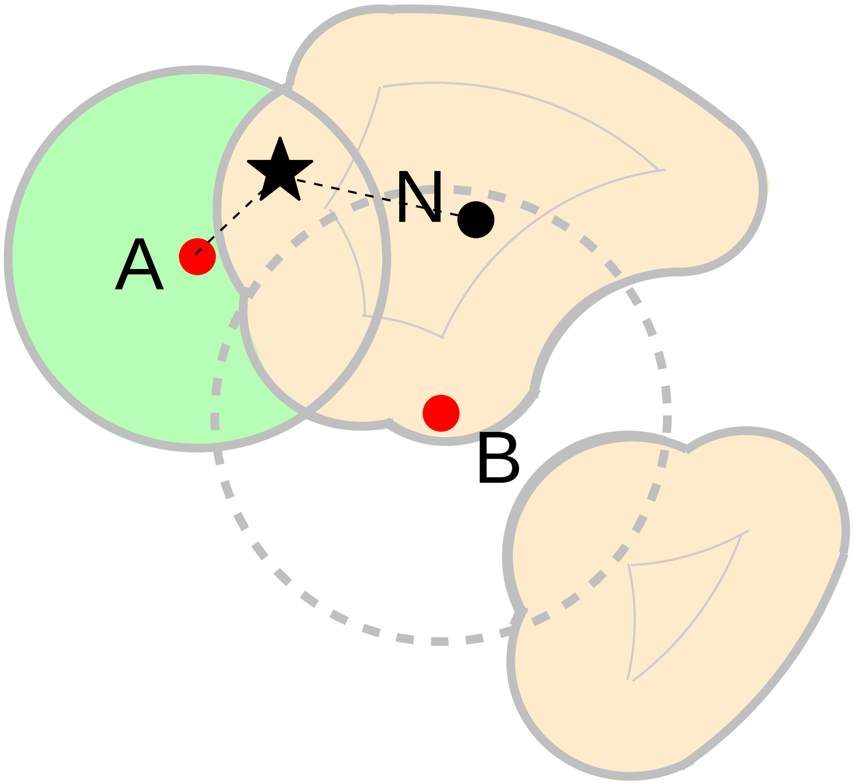
Feedback



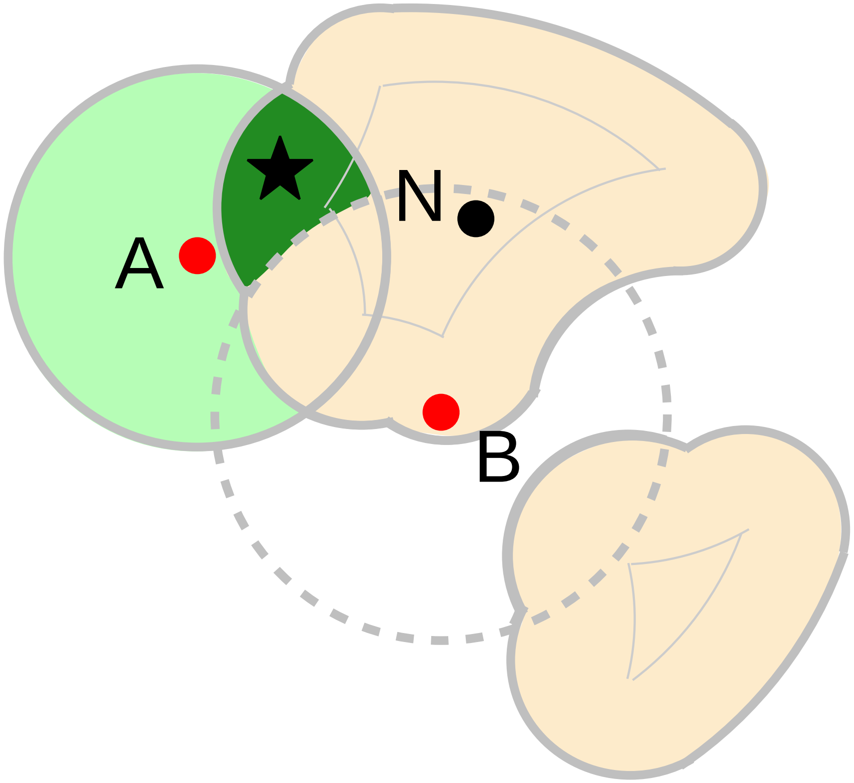
Feedback



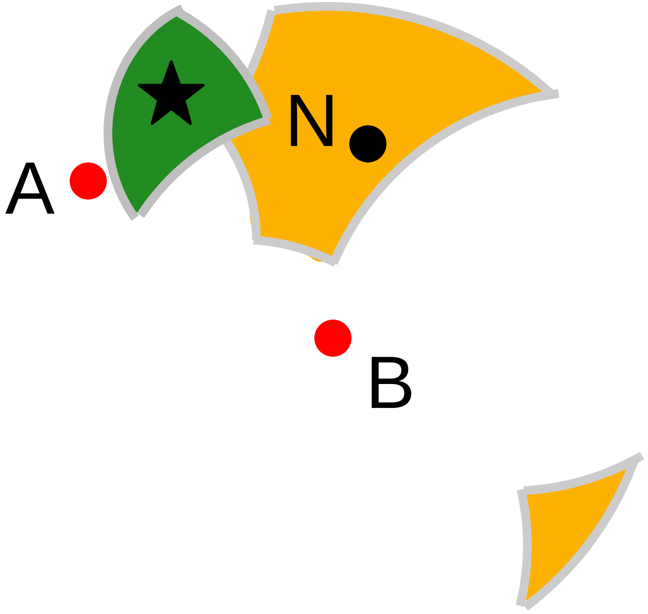
Feedback



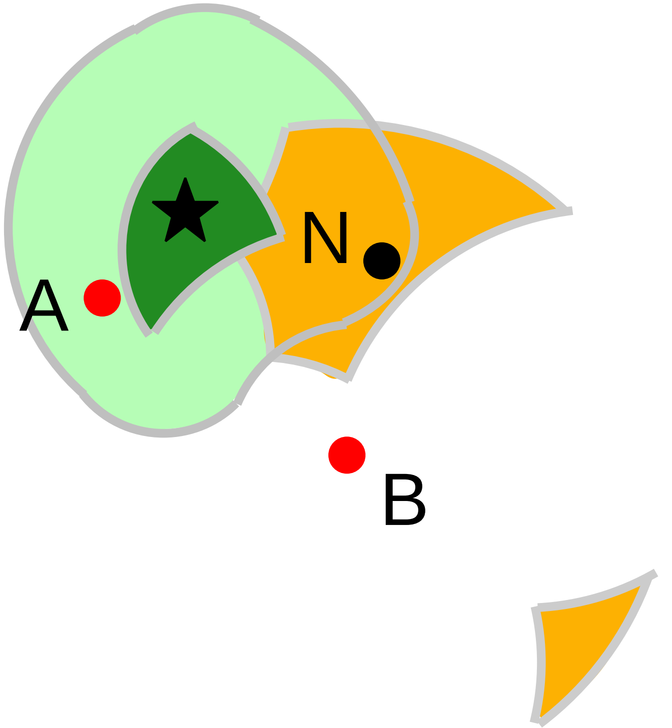
Feedback



Feedback

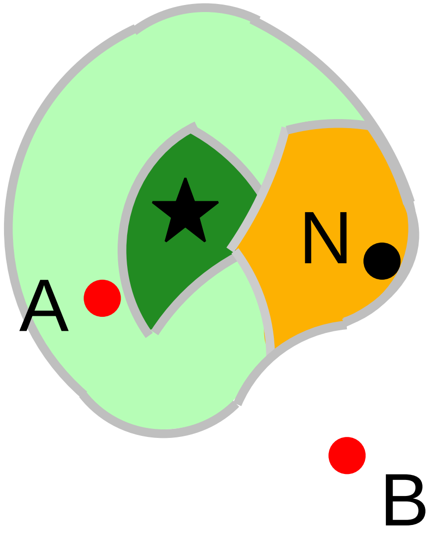


Feedback



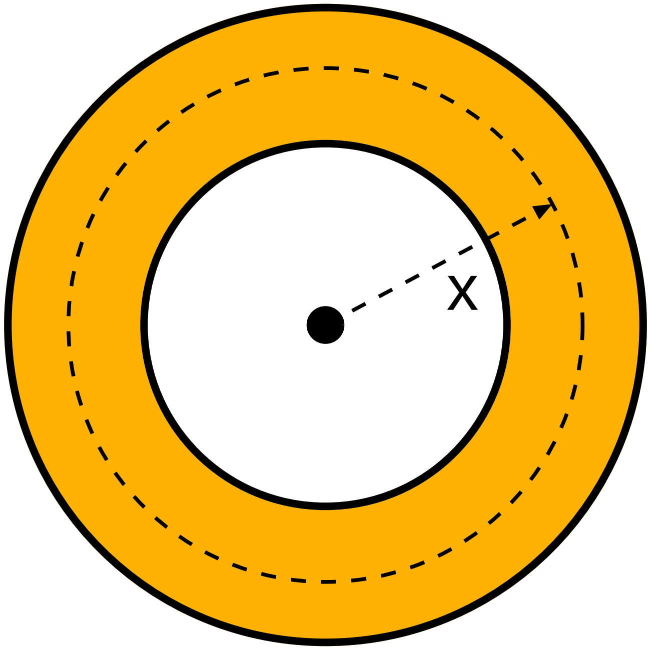
Events as a Source of Constraints

Feedback



Events as a Source of Constraints

Optimizations



Annulus for range x

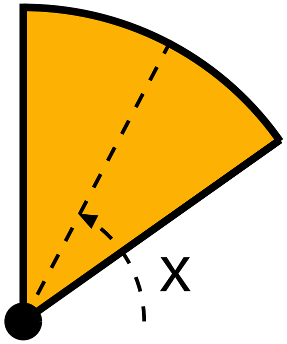
Wireless Hardware

- Range Measurements
- Angle of Arrival

Sensor Hardware

- Event Distance
- Directional Sensors

Optimizations



Wireless Hardware

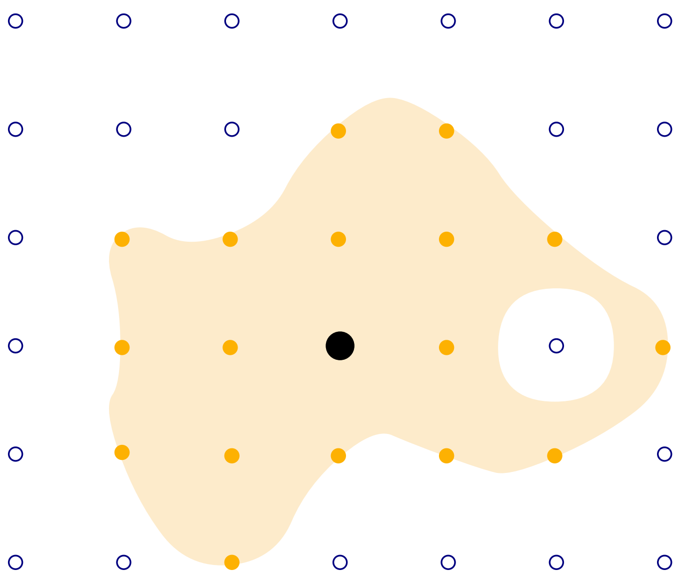
- Range Measurements
- Angle of Arrival

Sensor Hardware

- Event Distance
- Directional Sensors

Sector for angle x

Modeling



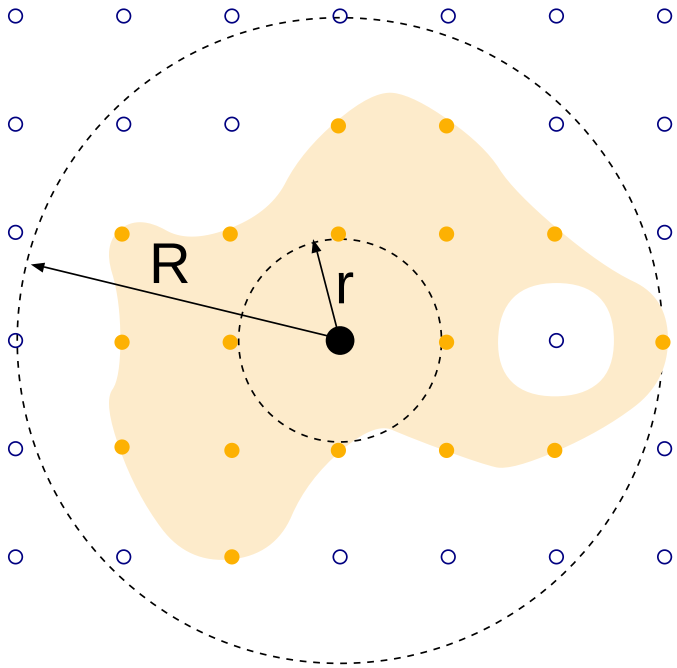
Wireless coverage area is non-convex and has holes

Wireless Radio

Boolean packet-received / packet-not-received.

- All reachable nodes $\leq R$ away
- All unreachable nodes $\geq r$ away

Modeling



Wireless Radio

Boolean packet-received / packet-not-received.

- All reachable nodes $\leq R$ away
- All unreachable nodes $\geq r$ away

Protocol

Neighborhood Discovery

- Nodes transmit periodic beacons
- Threshold beacon reception required for boolean connectivity

Gossip

Disseminate constraints as long as they are useful

- Positive information – used only at first hop
- Negative information – used within the first few hops

Validation of Node Localization

Implementation

- Implemented on MICA-2 motes, laptops and PDA
- About 2kB of storage per node
- About 80kB data transmitted per node until convergence

Setup

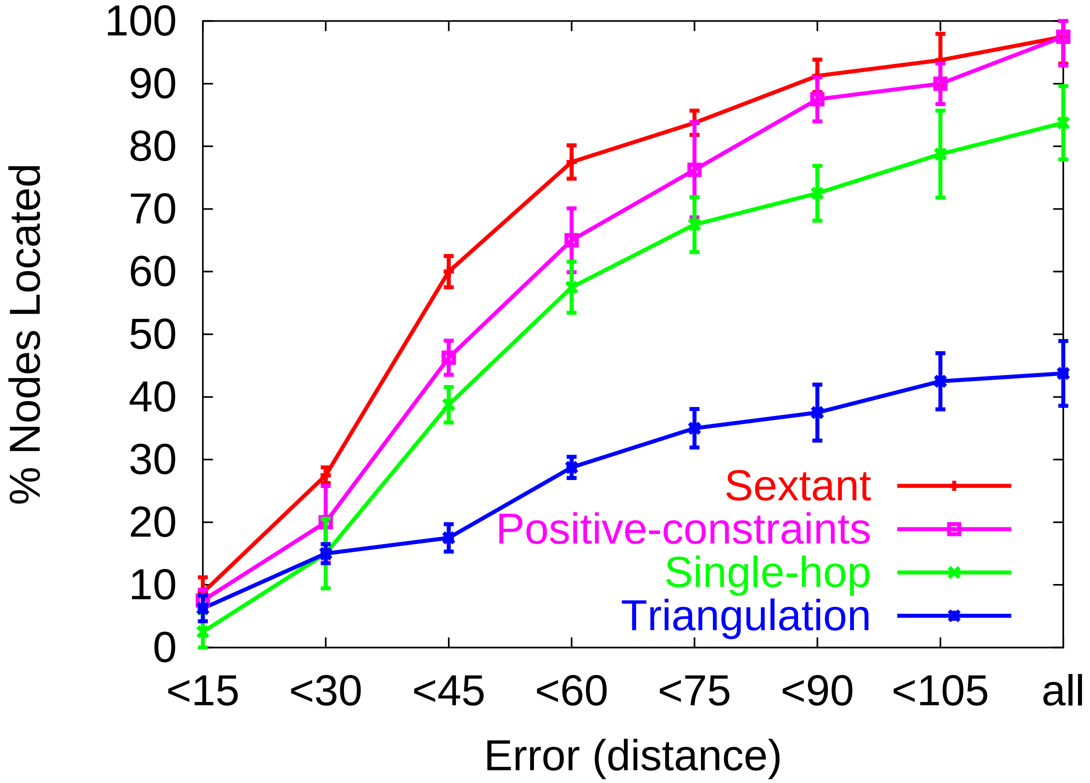
- 50 MICA2 motes placed in a grid pattern
- Landmarks chosen at random
- 80% packet reception threshold chosen for connectivity

Validation of Node Localization

Comparing Node Localization

- **Triangulation** – Centroid of neighbor nodes
  - GPSLess
- **Single-hop** – No transitive dissemination
  - Active Badge, Cricket, GPSLess, Localization Using Moving Target
- **Positive-constraints** – No negative information
  - APS, Convex position estimation, N-hop Multilateration, Robust Positioning
- **Sextant**

Validation of Node Localization

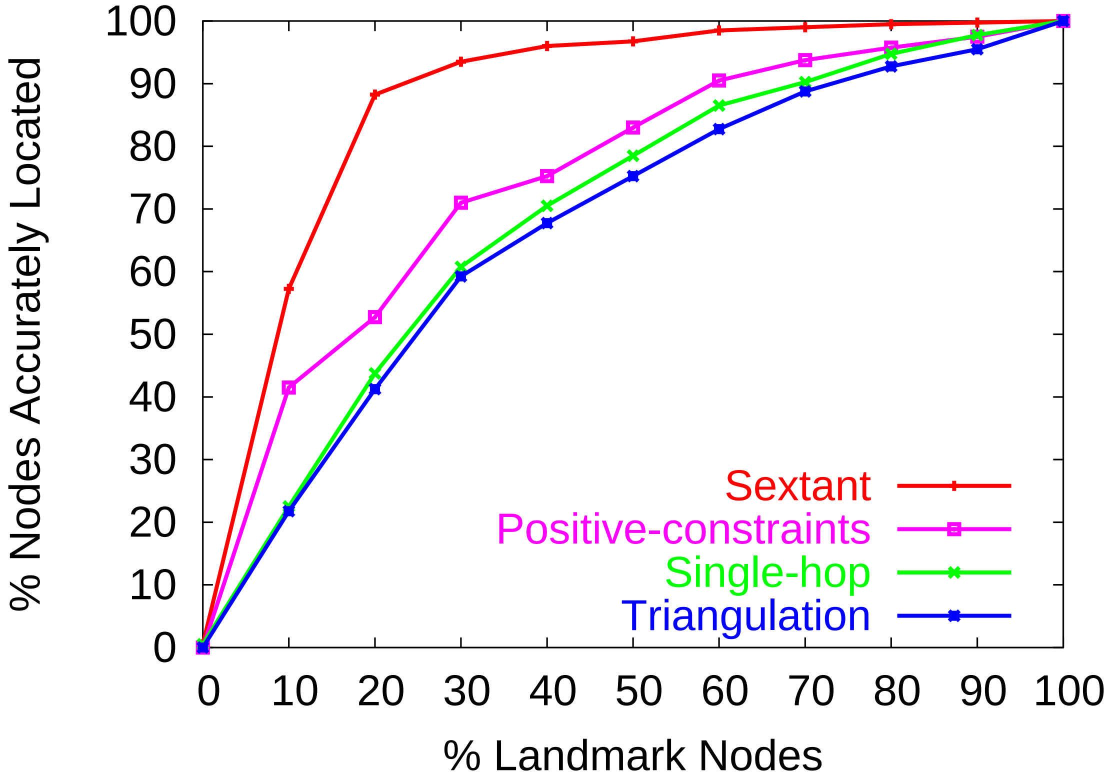


Node Localization

- Accurate
- Efficient
- Scalable

Sextant locates more nodes accurately

Validation of Node Localization

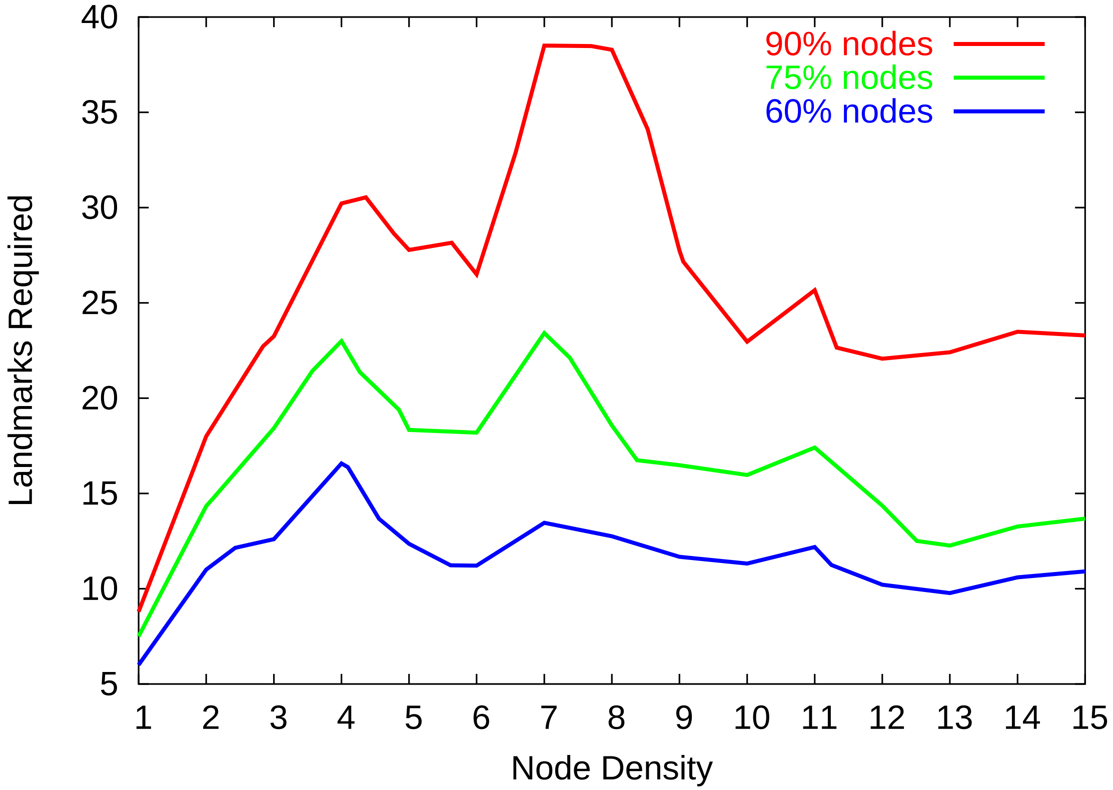


Node Localization

- ▶ Accurate
- ▶ Efficient
- ▶ Scalable

Sextant requires few landmarks

Validation of Node Localization



Node Localization

- Accurate
- Efficient
- Scalable

Sextant requires fixed landmark density

Validation of Event Localization

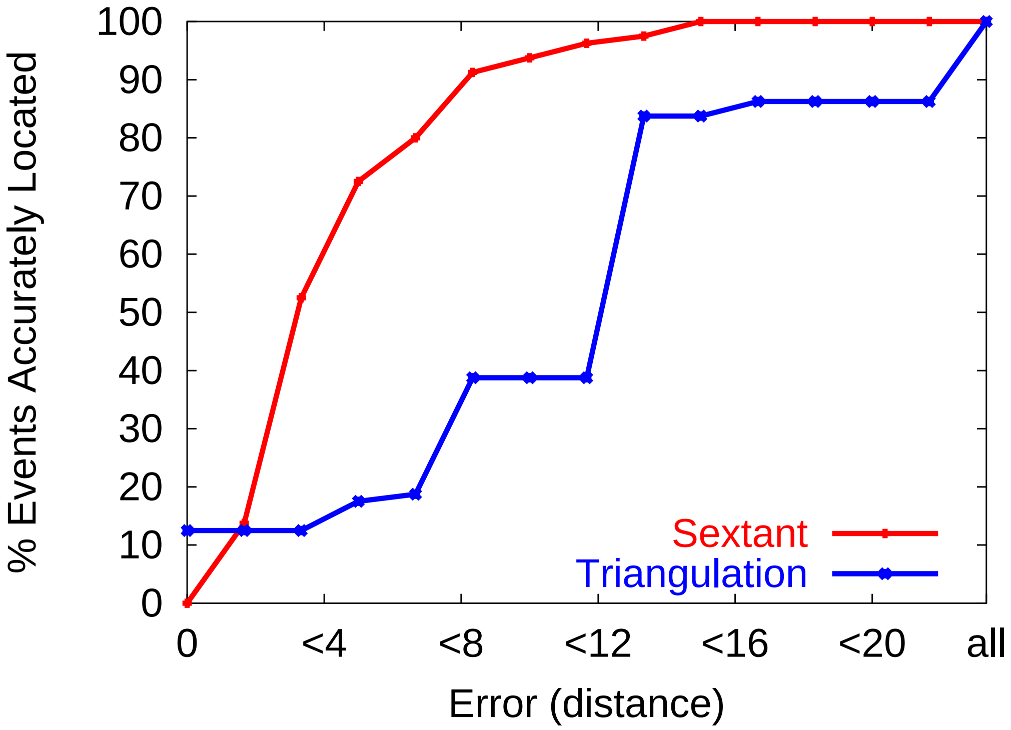
Setup

- 50 MICA2 motes placed in a grid pattern
- Event is a flash of light
- Appreciable change in analog value triggers sensor

Comparing Event Localization

- **Triangulation** – Centroid of sensors reporting the event
  - Acoustic Ranging
- **Sextant**

Validation of Event Localization

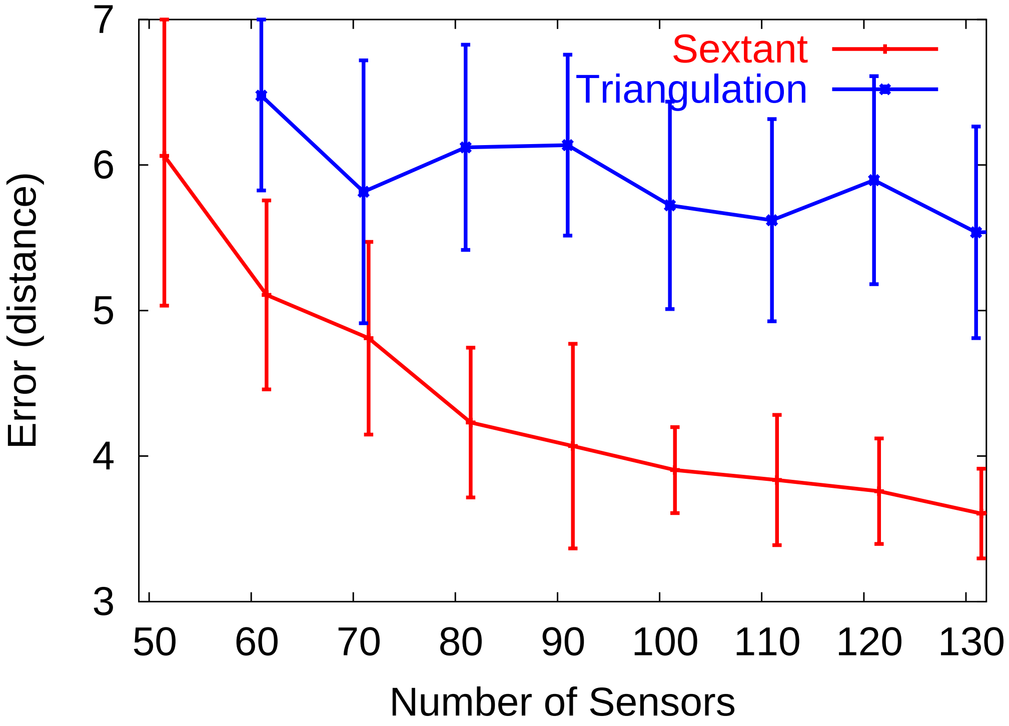


Event Localization

- ▶ Accurate
- ▶ Efficient
- ▶ Robust

Sextant locates more events accurately

Validation of Event Localization

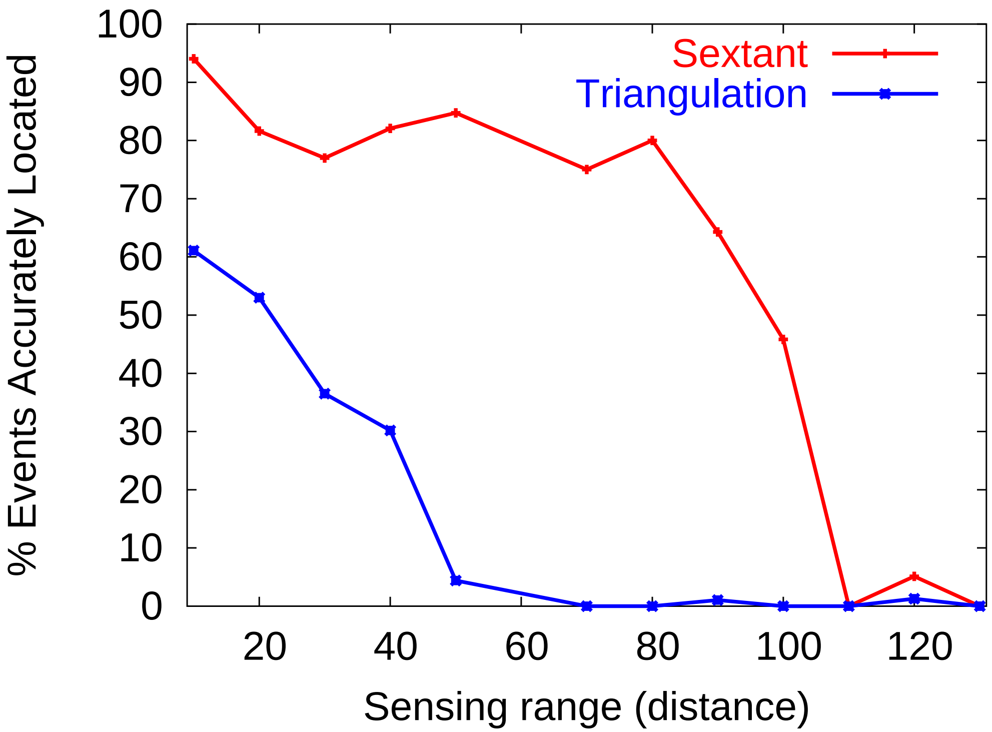


Event Localization

- Accurate
- Efficient
- Robust

Accuracy improves with nodes

Validation of Event Localization



Event Localization

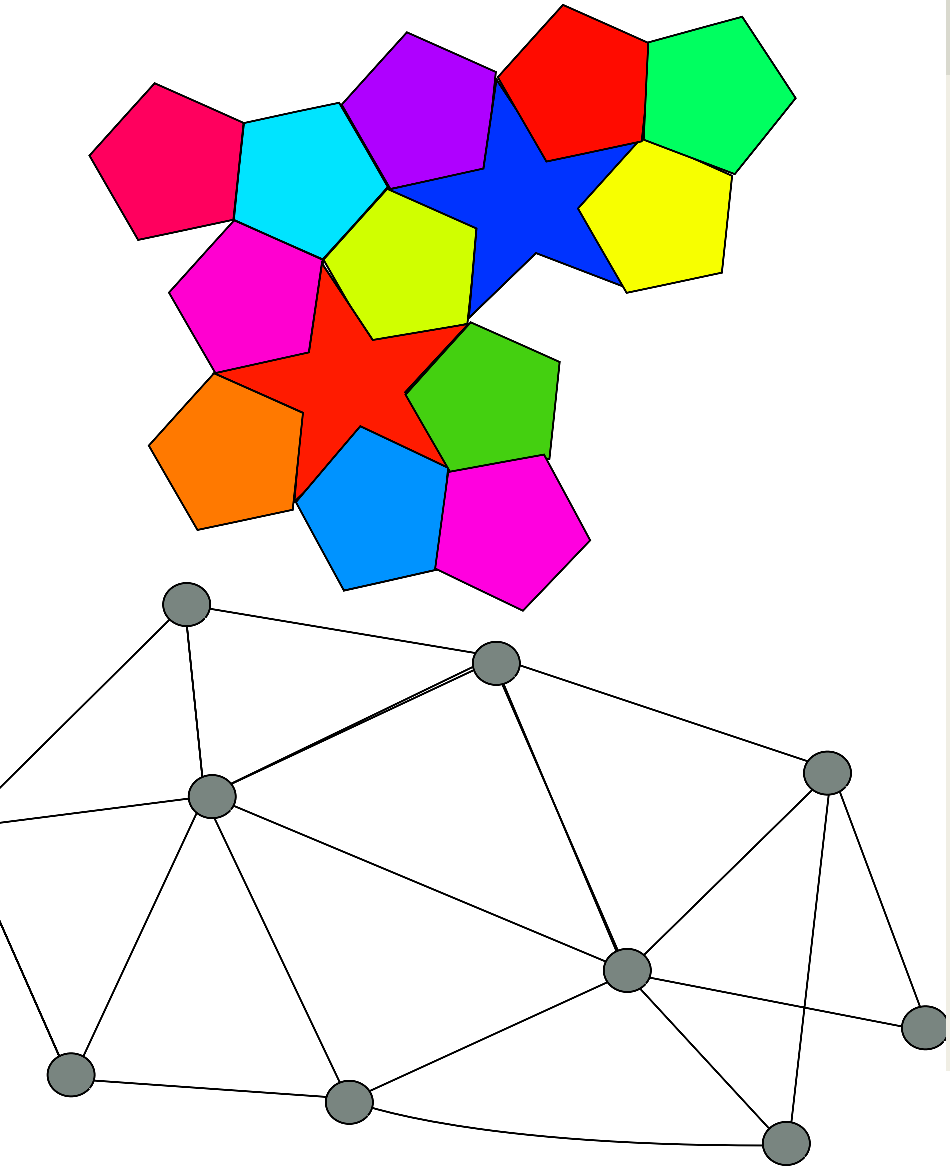
- Accurate
- Efficient
- Robust

Sextant independent of sensing range

Programming Model for Ad Hoc Networks

- Current state of the art is to view the network as a system of systems
  - Forces all applications to implement their own mechanisms for state migration
  - Tedious, error-prone
  - Multiple applications may conflict
- Fundamental problem stems from lack of an arbiter
  - Need a system layer to perform resource mediation

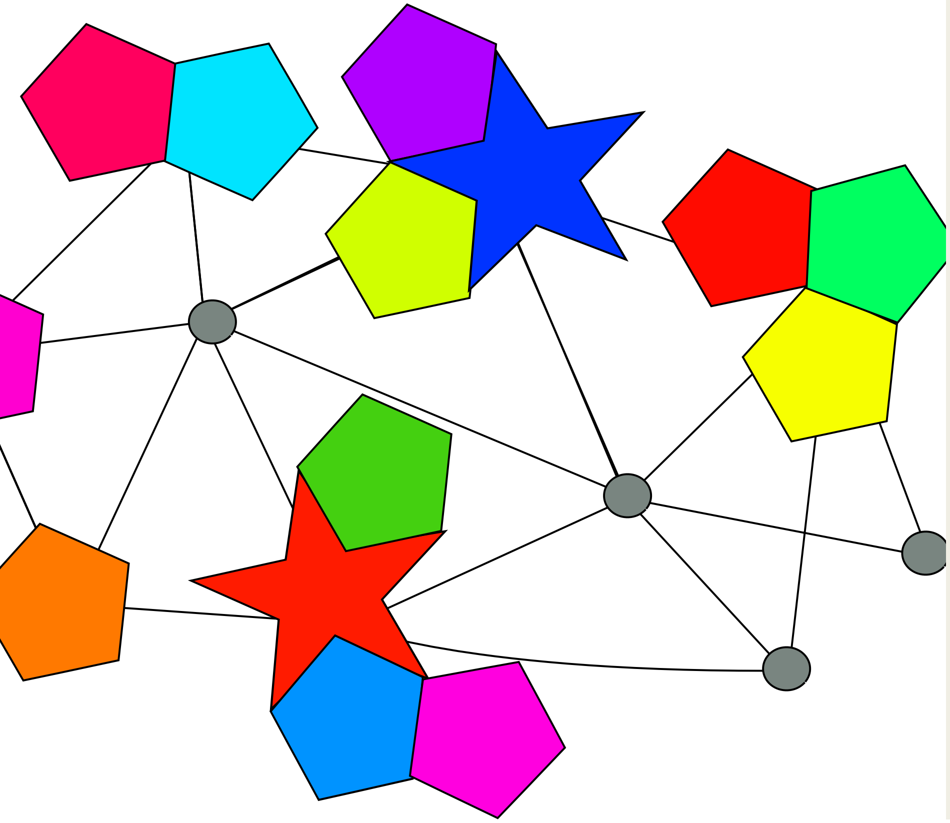
MagnetOS Approach



Contributions

- Programmer writes monolithic application for a single JVM
- MagnetOS statically partitions the application into communicating objects
  - Objects can reside anywhere in the network
- MagnetOS dynamically finds a good placement of objects on nodes in the network
  - Energy efficiency is the key goal

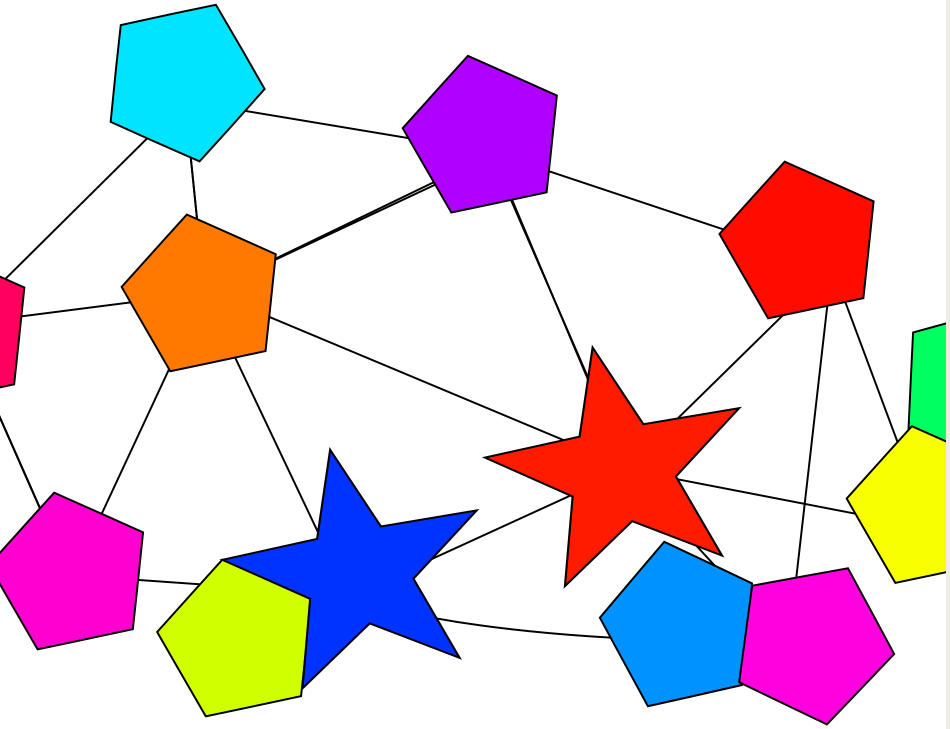
MagnetOS Approach



Contributions

- Programmer writes monolithic application for a single JVM
- MagnetOS statically partitions the application into communicating objects
  - Objects can reside anywhere in the network
- MagnetOS dynamically finds a good placement of objects on nodes in the network
  - Energy efficiency is the key goal

MagnetOS Approach



Contributions

- Programmer writes monolithic application for a single JVM
- MagnetOS statically partitions the application into communicating objects
  - Objects can reside anywhere in the network
- MagnetOS dynamically finds a good placement of objects on nodes in the network
  - Energy efficiency is the key goal

MagnetOS Implementation and Status

- Implemented most of the system
  - Static rewriter (50K loc)
  - Space-optimized JVM for x86 and StrongARM (30K loc)
  - Dynamic runtime (25K loc)
- Working on adding transparent replication
  - Based on message logging
  - Driven initially by programmer annotations

Summary

- Sextant is a localization framework that achieves high accuracy and scalability
  - Explicit representation of regions using Bézier curves
  - Conservative and comprehensive extraction of negative as well as positive constraints
  - Transitive dissemination of constraints
  - Use of events to refine node location
- Sextant is practical
- MagnetOS simplifies programming mobile systems
  - Many new directions based on transparent rewriting

http://www.cs.cornell.edu/People/egs/sextant/
http://www.cs.cornell.edu/People/egs/magnetos/

Related Work

Positive Information

- **GPS-Free** '01: Capkun, Hamdi and Hubaux
- **APS** '01: Niculescu and Nath
- **Convex Position Estimation** '01: Doherty, Pister and Ghaoui
- **Robust Positioning** '02: Savarese, Rabay and Langendoen
- **N-hop Multilateration** '02: Savvides, Park and Srivastava
- **APS-AoA** '03: Niculescu and Nath
- **Mere Connectivity Localization** '03: Shang, Ruml, Zhang and Fromherz
- **Connectivity-Based Positioning** '04: Bischoff and Wattenhofer
- **Unit Disk Approximation** '04: Kuhn, Moscibroda and Wattenhofer
- **Virtual Coordinates** '04: Moscibroda, O'Dell and Wattenhofer

Related Work

Single-Hop

- **Active Badge** ’92: Want, Hopper, Falcão and Gibbons
- **GPS-Less** ’00: Bulusu, Heidemann and Estrin
- **RADAR** ’00: Bahl and Padmanabhan
- **Cricket** ’00: Priyantha, Chakraborty and Balakrishnan
- **RF-Based Location Tracking** ’04: Lorincz and Welsh
- **VORBA** ’04: Niculescu and Nath
- **Localization Using a Moving Target** ’04: Galstyan, Krishnamachari, Lerman and Pattem

Related Work

Event Localization

- **Fine-grained Localization** ’01: Savvides, Han and Srivastava
- **Collaborative Processing** ’03: Zhao, Liu, Guibas and Reich
- **Acoustic Ranging** ’04: Sallai, Balogh, Maroti and Ledeczi
- **Countersniper** ’04: Simon, Maroti, Ledeczi et al.
- **Entity Tracking** ’02: Brooks, Griffin and Friedlander
- **Energy-Efficient Surveillance** ’04: He, Krishnamurthy, Stankovic et al.

A Network Service Provider's View of Ubiquitous Computing

Rick Schlichting

Director, Software Systems Research Department
AT&T Labs-Research
Florham Park, NJ 07932, USA

Answer to homework

- Ubiquitous computing and pervasive computing includes embedded devices, while nomadic computing does not.

Introduction

It's all about scale — ubiquitous computing means more endpoints and (much!) more data.

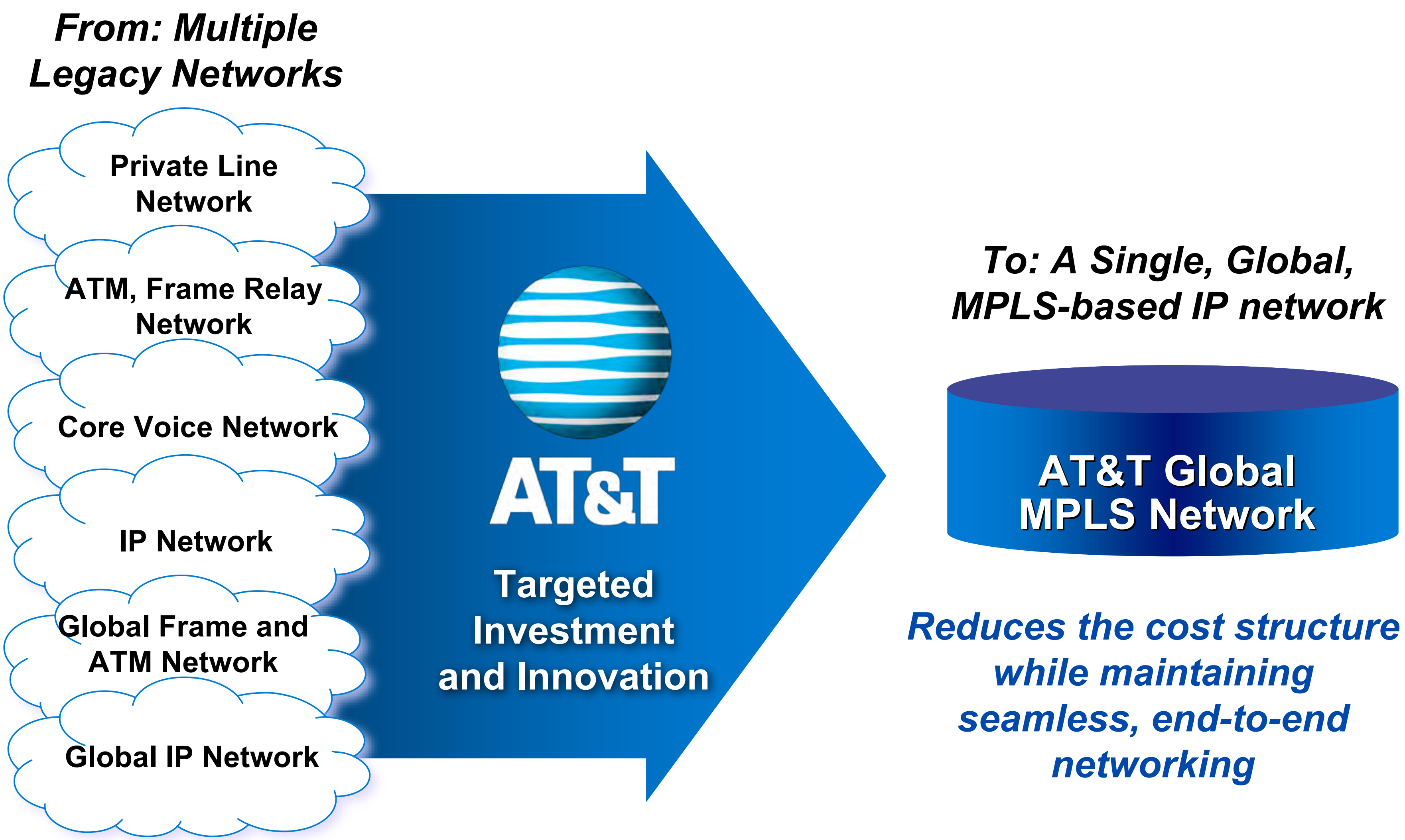
- **Talk Outline**
  - AT&T: Trends
  - Ubiquitous computing: Current business drivers.
  - Ubiquitous computing: Research in information and software systems.

AT&T: From Telephone Company to Network Service Provider

- **History**
  - 1876: Telephone invented by Alexander Graham Bell.
  - 1877: Bell Telephone Company founded; becomes parent of Bell System of local exchanges.
  - 1885: AT&T formed as subsidiary of Bell Telephone Company to build and operate long distance network.
  - 1899: AT&T becomes parent of Bell System.
  - 1925: Bell Telephone Laboratories established.
  - 1984: AT&T splits from 7 Regional Bell Operating Companies (RBOCs).
  - 1996: AT&T splits from NCR and Lucent (including Bell Labs); AT&T Labs formed.
  - 2005: SBC proposes to acquire AT&T.
- **Everything is now about IP, converged networks, and serving the enterprise space**
  - Operate largest IP backbone in the U.S.
  - 1000 MPLS switching nodes worldwide.
  - 76K miles of route fiber in the U.S.
  - First to provide coast-to-coast OC-192 (10 Gbits/sec).
  - Operate 22 IDCs
- **“The World’s Networking Company”**

RETURN ON COMMUNICATIONS

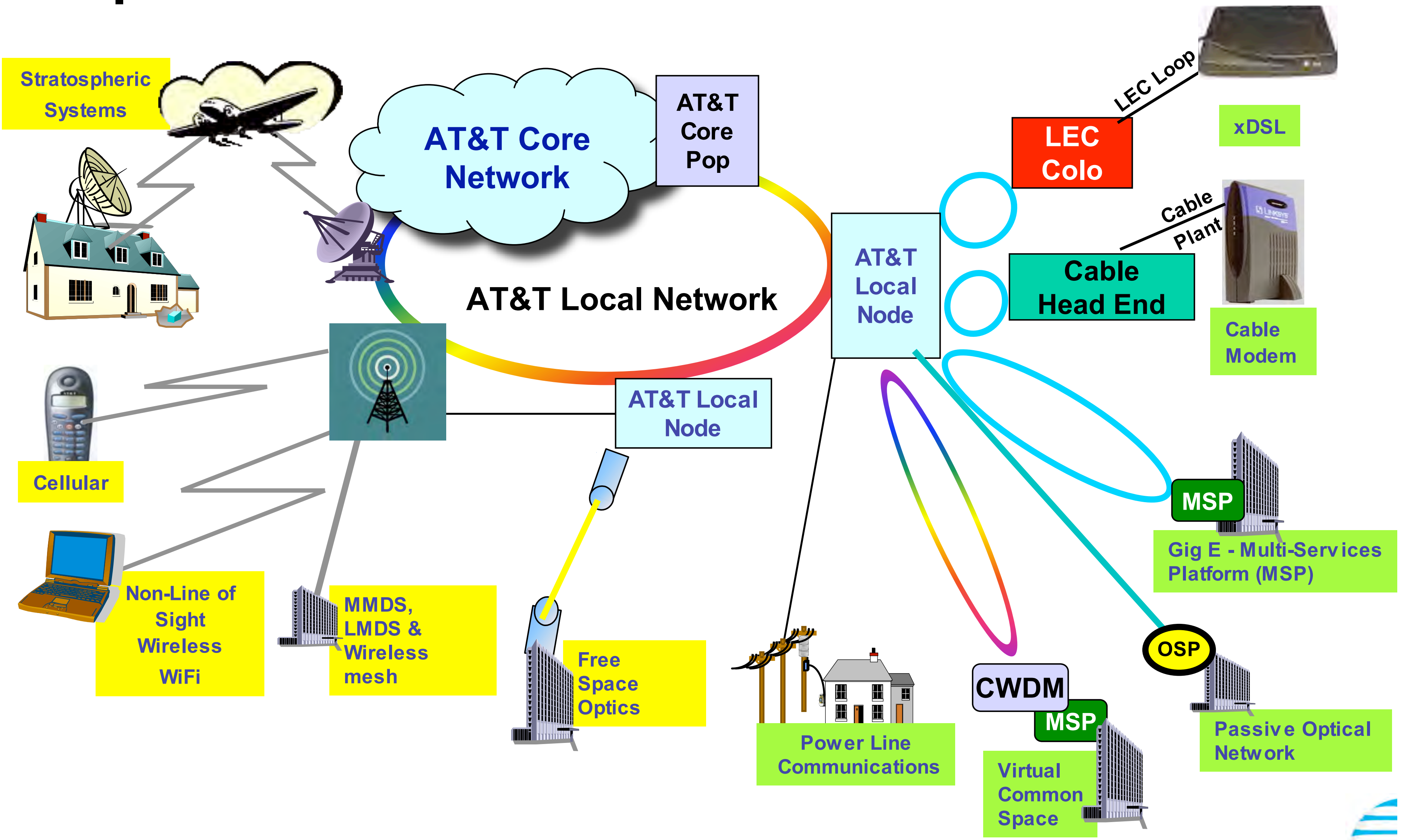
AT&T's Network Evolution



From: Multiple Legacy Networks
Private Line Network
ATM, Frame Relay Network
Core Voice Network
IP Network
Global Frame and ATM Network
Global IP Network
AT&T
Targeted Investment and Innovation
To: A Single, Global, MPLS-based IP network
AT&T Global MPLS Network

Reduces the cost structure while maintaining seamless, end-to-end networking

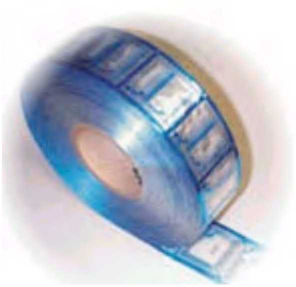
Heterogeneity: Access Technologies and Endpoints



RETURN ON COMMUNICATIONS

Ubiquitous Computing: Current Business Driver is RFID

- RFID (Radio Frequency Identification) is being used and will become more prevalent in inventory and asset tracking systems
- Goal is to have RFID on every item in the supply chain
- EPC - Electronic Product Code
  - Electronic Product Code (ePC) is a new product numbering standard under development by the Uniform Code Council that can be used to detect, track, and control a variety of items using radio frequency identification (RFID) technology. The 96-bit ePC code links to an online database, providing a secure way of sharing product-specific information along the supply chain.
- Small-size and low-cost (near-term goal of $0.05 per tag, moving to <$0.01) would drive to virtually all types of consumer goods



Roll of RFID Tags

RFID Tags for Pallets and Boxes

RFID Tags

Companies Evaluating / Implementing RFID Solutions According to EPC & Industry Sources

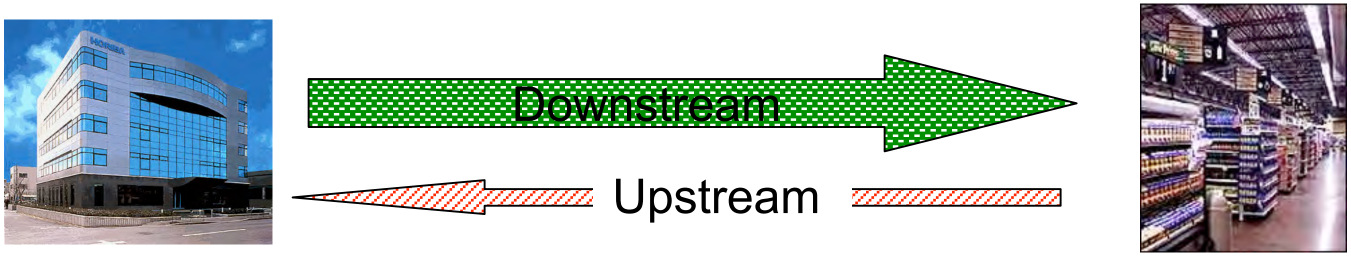
Client	Industry Vertical	Application
American Express	- Financial	- Contactless Payment System (ExpressPay)
Best Buy	- Retail	- Track & Trace / Asset Management
Coca-Cola	- Retail	- Track & Trace / Asset Management
CVS	- Retail	- Payment System / Track & Trace
Department of Defense	- Government	- Track & Trace / Chain of Custody
DHL	- Transportation	- Track & Trace*
Federal Express	- Transportation	- Track & Trace*
General Mills	- Manufacturing	- Track & Trace
HP	- Manufacturing	- Track & Trace / Asset Management
Johnson & Johnson	- Manufacturing	- Track & Trace
Home Depot	- Retail	- Track & Trace / Asset Management
Kelloggs	- Manufacturing	- Track & Trace
Kimberly-Clark	- Manufacturing	- Track & Trace
Kodak	- Manufacturing	- Track & Trace
Merck	- Pharma	- Track & Trace / Chain of Custody
Micro Beef Technologies	- Ranching	- Track & Trace / Chain of Custody
Mobil Speedpass	- Retail	- Contactless Payment System
Novartis	- Pharma	- Track & Trace / Chain of Custody
Pfizer	- Pharma	- Track & Trace / Chain of Custody
Roche	- Pharma	- Track & Trace / Chain of Custody
Schering – Plough	- Pharma	- Track & Trace / Chain of Custody
Target	- Retail	- Track & Trace / Asset Management
Tesco	- Retail	- Track & Trace / Asset management
The Gillette Company	- Manufacturing	- Track & Trace
Tyson	- Manufacturing	- Track & Trace / 300M cases per year
UPS	- Transportation	- Track & Trace*
Visy Paper	- Manufacturing	- Track & Trace
…		

RETURN ON COMMUNICATIONS

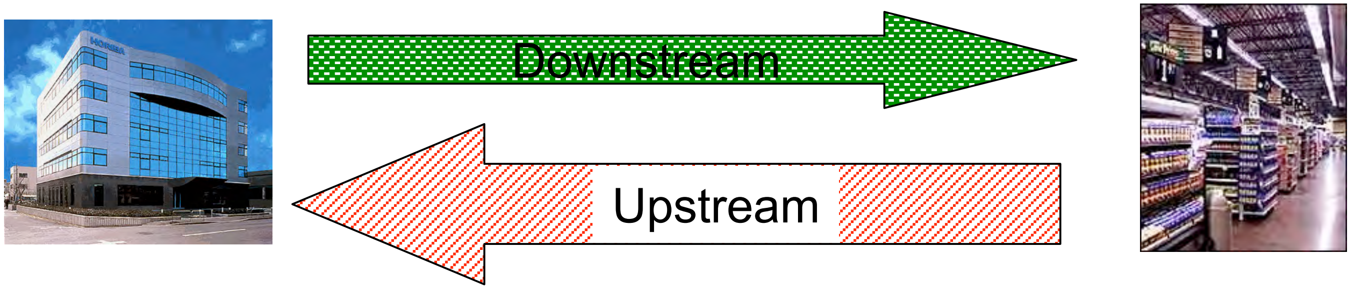
RFID - Network Implications

- With RFID, more information is at the network edge and will feedback to central sites (like distribution centers and corporate headquarters)
- Existing information exchange could reverse (i.e. more coming from the edge back to the central site rather than a pushed down)
- Would make existing network access systems, such as VSAT terminals and ADSL, inadequate for the new task
- Would drive SYMMETRICAL broadband deployment further to the edge

Today - More Information Flow from Corporate to Edge



Tomorrow - More Information Flow from Edge to Corporate



Object Naming Service (ONS)

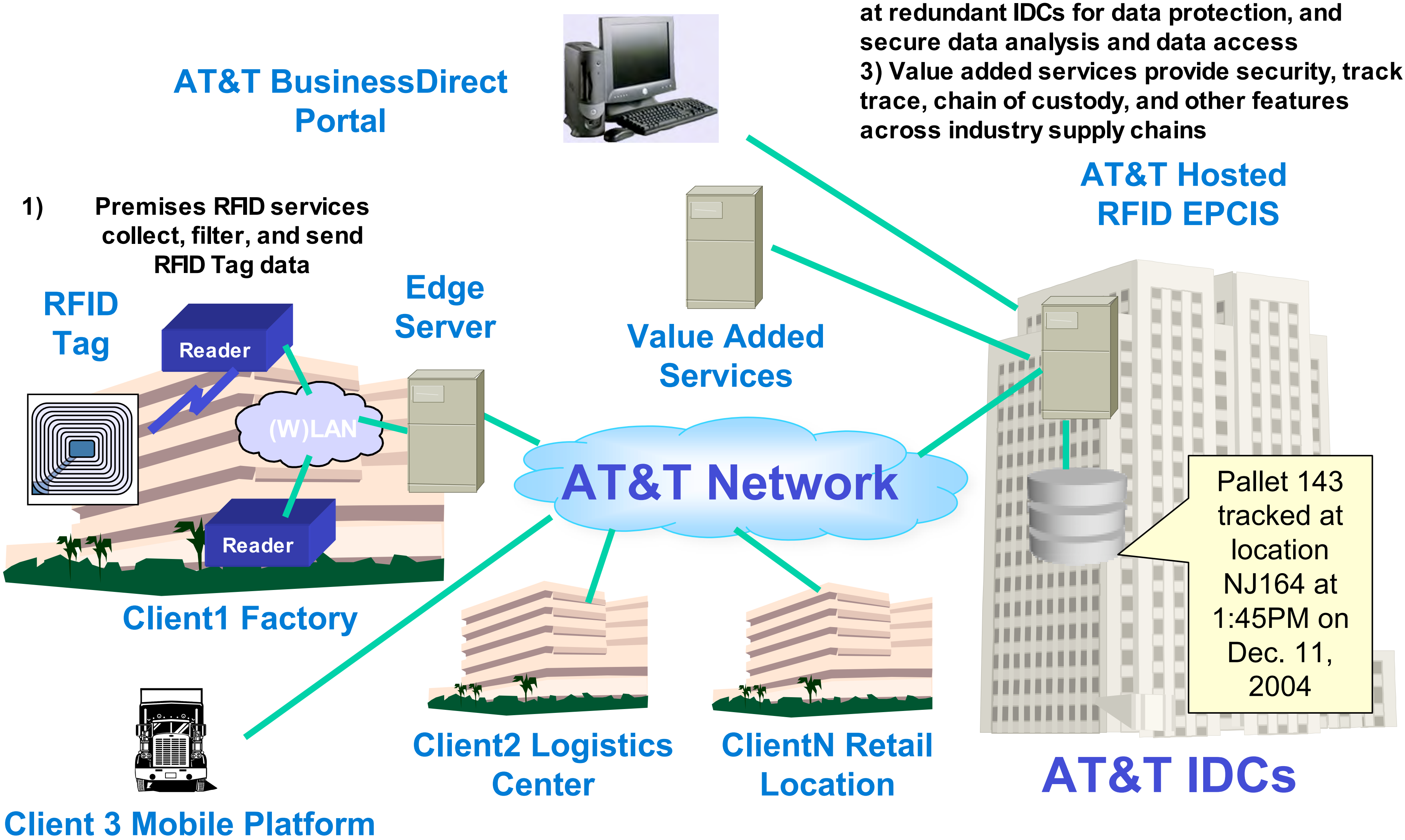
- ONS tells computer systems where to find information about any object with an electronic product code (EPC) for RFID applications.
- Designed in a similar concept like a URL for the internet. Based in part on the Internet Domain Name System (DNS) – routes information to appropriate network endpoints
- The EPC means nothing without the ONS information about the actual product instance carrying the EPC.
- The ONS is accessed via IP networking in a distributed fashion
- The amount of data transactions for ONS service is expected to grow at a phenomenal rate.
  - Today the worldwide Internet handles 17 billion messages a day.
  - Several industry sources have estimated that the worldwide ONS network will need to handle approximately <u>4 quadrillion message</u> a day by 2012 (note: item level tagging is assumed).

RFID Services

2) Managed Hosted Transaction-Based and/or client dedicated Electronic Product Code Information Service (EPCIS) store the read data at redundant IDCs for data protection, and secure data analysis and data access

3) Value added services provide security, track, trace, chain of custody, and other features across industry supply chains

1) Premises RFID services collect, filter, and send RFID Tag data

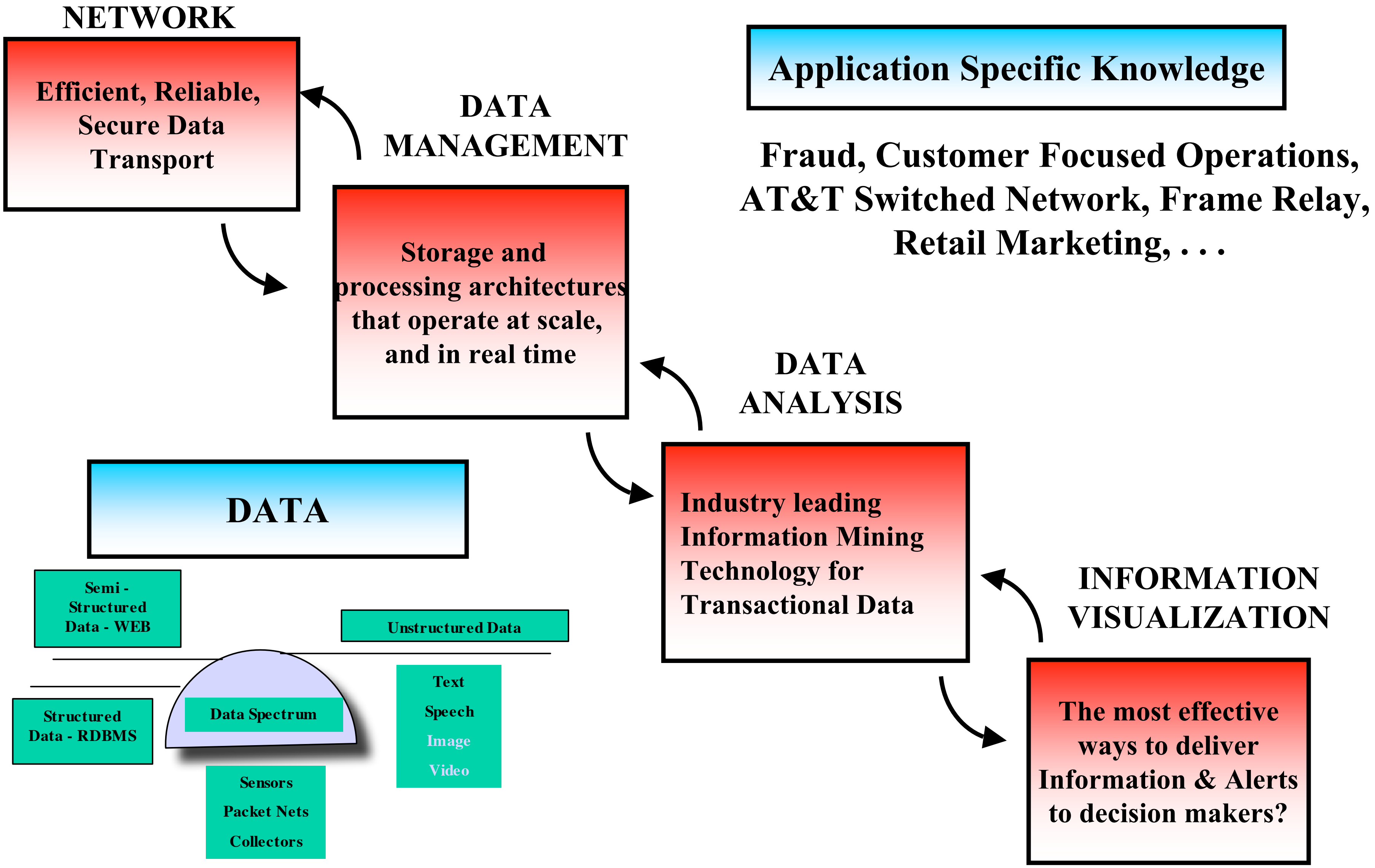


Ubiquitous Computing: Research in Information and Software Systems

The Next Bottleneck - Information

- We are no longer CPU constrained, e.g. 5 GHz CPUs
- We are no longer memory constrained, e.g. multi-GB memories
- We are no longer disk constrained, e.g. 160 GB disk
- We are becoming less bandwidth constrained, e.g. cable, DSL, FSO, WiFi
- We could easily be constrained by our ability to extract useful information from massive amounts of data
- Ubiquitous computing means lots of data, and data of different types!

AT&T Data Mining Approach



NETWORK

Efficient, Reliable, Secure Data Transport

DATA MANAGEMENT

Storage and processing architectures that operate at scale, and in real time

DATA ANALYSIS

Industry leading Information Mining Technology for Transactional Data

INFORMATION VISUALIZATION

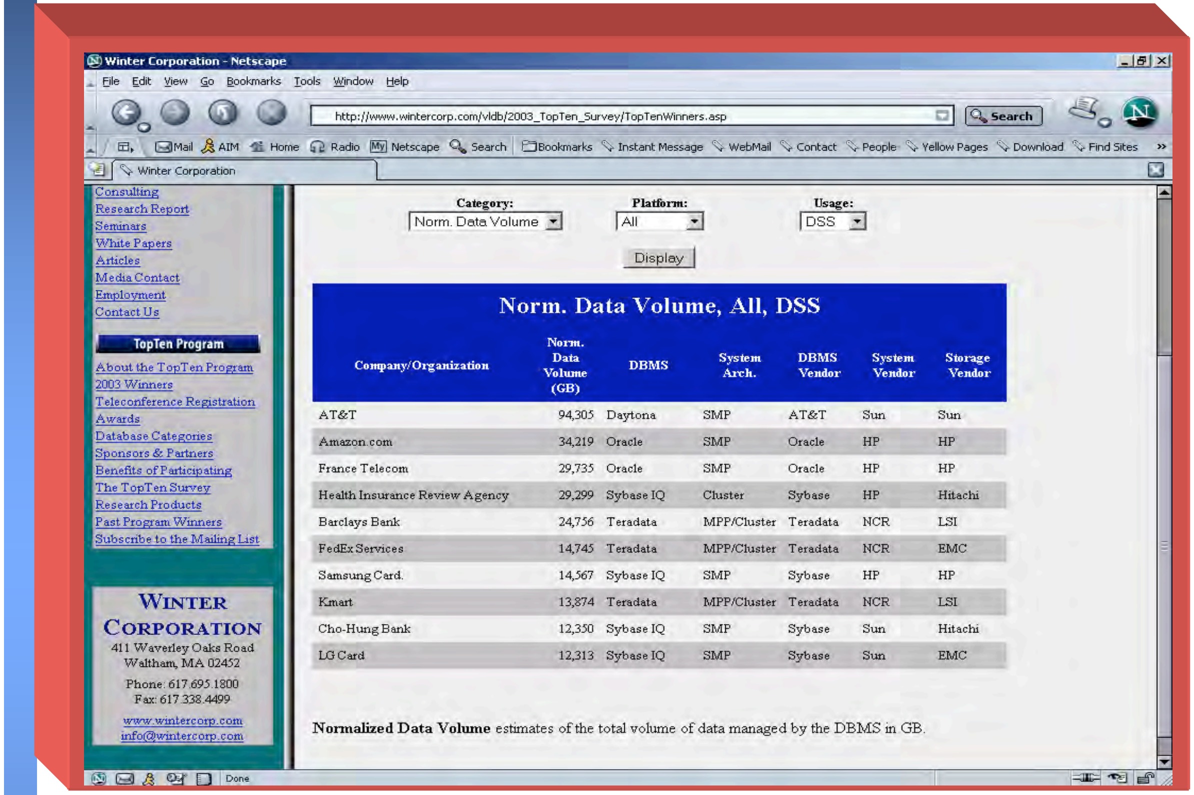
The most effective ways to deliver Information & Alerts to decision makers?

Application Specific Knowledge

Fraud, Customer Focused Operations, AT&T Switched Network, Frame Relay, Retail Marketing, . . .

DATA

Daytona: Managing Data at AT&T Scale



Category: Norm. Data Volume
Platform: All
Usage: DSS

Norm. Data Volume, All, DSS

Company/Organization	Norm. Data Volume (GB)	DBMS	System Arch.	DBMS Vendor	System Vendor	Storage Vendor
AT&T	94,305	Daytona	SMP	AT&T	Sun	Sun
Amazon.com	34,219	Oracle	SMP	Oracle	HP	HP
France Telecom	29,735	Oracle	SMP	Oracle	HP	HP
Health Insurance Review Agency	29,299	Sybase IQ	Cluster	Sybase	HP	Hitachi
Barclays Bank	24,756	Teradata	MPP/Cluster	Teradata	NCR	LSI
FedEx Services	14,745	Teradata	MPP/Cluster	Teradata	NCR	EMC
Samsung Card.	14,567	Sybase IQ	SMP	Sybase	HP	HP
Kmart	13,874	Teradata	MPP/Cluster	Teradata	NCR	LSI
Cho-Hung Bank	12,350	Sybase IQ	SMP	Sybase	Sun	Hitachi
LG Card	12,313	Sybase IQ	SMP	Sybase	Sun	EMC

Normalized Data Volume estimates of the total volume of data managed by the DBMS in GB.

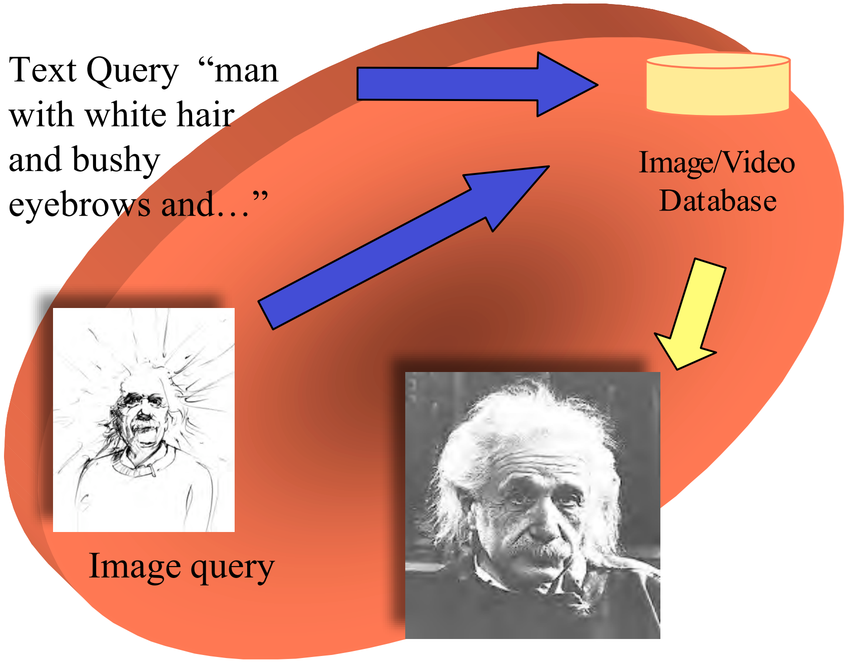
- **Massive amounts of data can be collected, but hard to manage in commercial DBs**
- **Daytona enables scalable data management**
  - organizes and stores massive amounts of data on disk, supported by indices and a data dictionary
  - permits concise expression of sophisticated queries
  - provides answers to those queries quickly
  - data in a concurrent, crash-proof environment
  - proven reliability

Applications across AT&T:

- SCAMP – AT&T Call Detail Data Base of Record
  - largest publicly known data warehouse
- Global Fraud Mgt. System – All AT&T Call Fraud
- Traffic Analysis System (TAS) – IP Traffic Analysis
- STORM/FLOOD – Network Security Monitors
- Gigascope – IP Packet Monitoring & Analysis (OC48)

RETURN ON COMMUNICATIONS

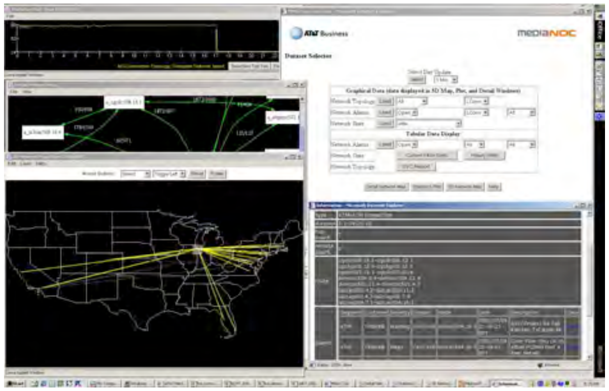
Analysis: Video and Image Data Mining



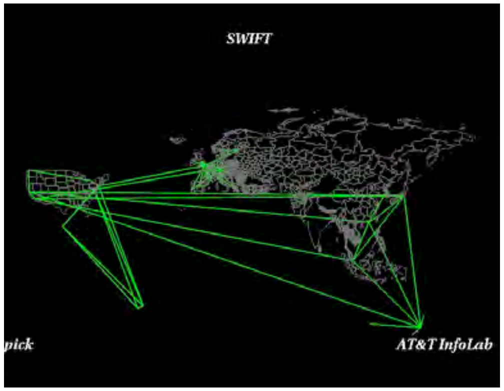
- Strengthen AT&T's hosting offers in the image/video space with higher value-added services
- Enhance AT&T's video conferencing portfolio with automatic indexing.
- Provide summarization services to broadcast video customers.

- Automatic annotation of large image and video databases for better content-based retrieval.
- Techniques for automatically labeling image and video content with descriptive text.
- Flexibility to support consumer-grade digital cameras, and compressed-domain processing tolerant to multiple compression formats.

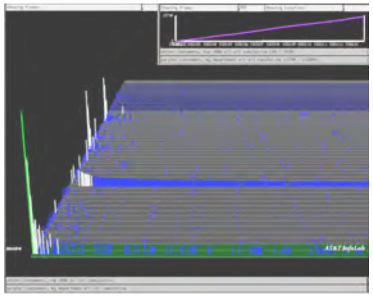
SWIFT: Visualizing Large-Scale Services



customer end-to-end views

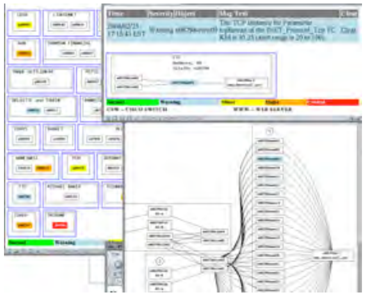


EMEA global network

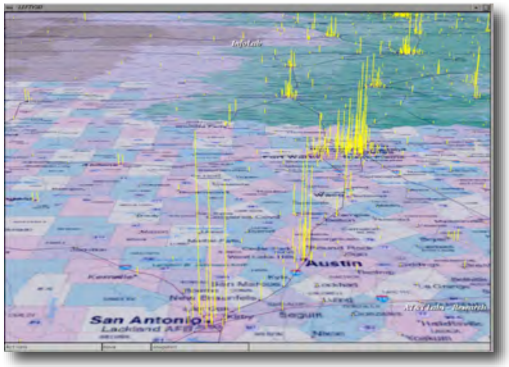


retail transactions

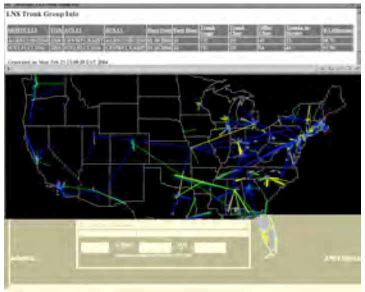
high-res aerial images



IDC customer view



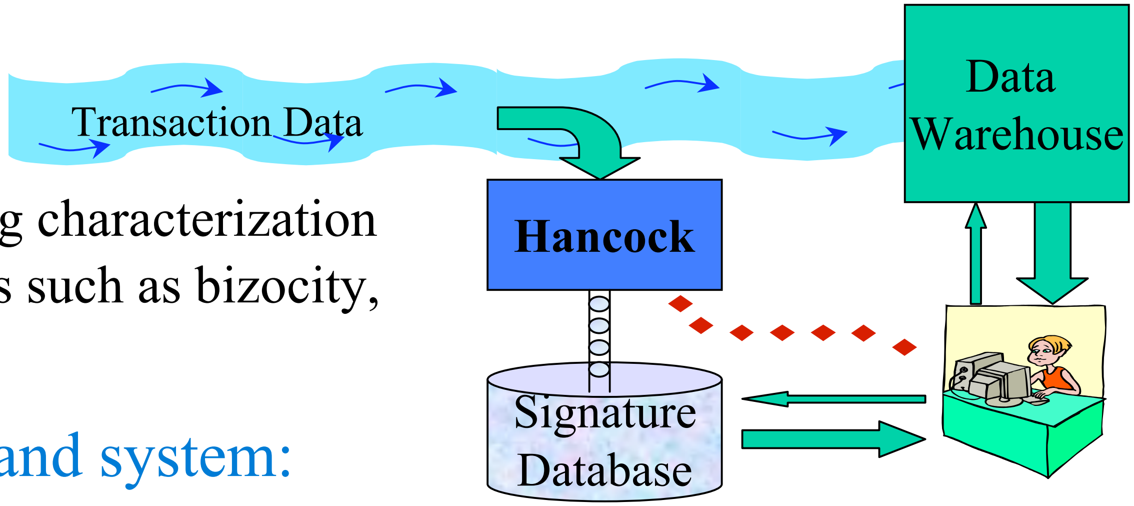
a wireless service with multiscale geopolitical map overlays



Project X-Ray

- Swift runs at full scale on data sets with hundreds of millions of items.
- It enables data integration in the human interface.
- It offers 3D graphics and animation for visual querying, navigating from a global view of the entire data set down to individual records.
- It works with both live data feeds and stored historical state simultaneously.
- Swift runs on anything from desktop clients up to large Powerwalls.
- Current work is to generalize using ODBC, JDBC, XML.

Large Scale Data Stream Processing



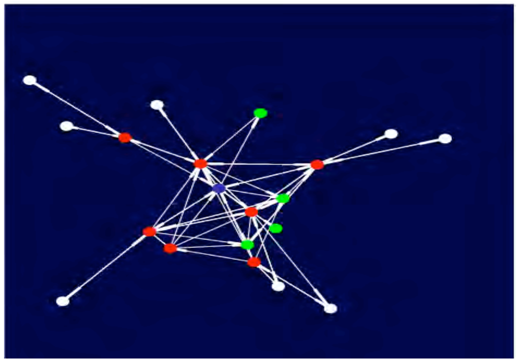
Signature: an evolving characterization of customers’ behaviors such as bizocity, fraudicity, usage, etc.

Hancock language and system:

- Succinct specification of signatures.
- Data streams processed and stored with compression.

Community of Interest:

- Fraud detection, record linkage, etc.
- 228M phone #’s, 120 bytes per #.
- 7GB collection.
- Update daily in 2 hours.



- Inbound calls
- Known fraudster
- Outbound calls

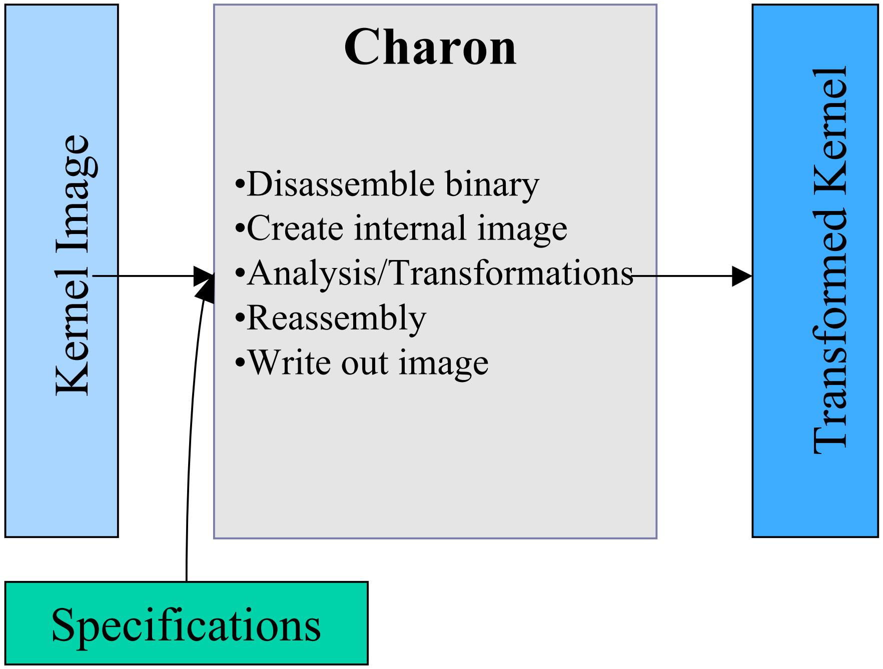
Cassyopia: Software System Optimization

Compiler-assisted holistic system optimization.

- **Goals**
  - Optimize across address spaces and different types of address spaces (e.g, user processes+kernel).
  - Optimize for different metrics, including performance, memory footprint, fault tolerance, security.
  - Optimize across address spaces that execute on separate machines.
  - Both static and dynamic optimizations.
- **Use compiler optimization techniques in novel ways**
  - Most of the work based on the PLTO, a binary rewriting tool for the IA-32 architecture.

Charon: Automated Kernel Specialization

Perform automated kernel transformations.



- **Uses**
  - Kernel specialization for small or specialized devices such as sensors, motes, routers, cell phones, etc. (*kernel compaction*).
  - To expose OS state to application or middleware to enable, e.g, adaptation.
- **Tool being built by modifying PLTO.**

Conclusions

Ubiquitous computing means more endpoints and more data.

- **Challenges**
  - – Network architectures and management.
  - – Information handling and mining.
  - – Software and systems.

Session 2.4

Synthesis and Wrap Up

Moderator

W. Kent Fuchs

Challenges in Mobile Distributed Systems

- *Autonomous Clustering and Hierarchical Routing for Mobile Ad Hoc Networks*

  Yoshiaki Kakuda

- *The Crumbling Perimeter: Mobile Networking and Internal Security Issues*

  Farnam Jahanian

- *Timed Asynchronous System Models for Dependable Mobile/Pervasive Systems*

  Christof Fetzer

Yoshiaki Kakuda

Challenging Issues in Routing for Mobile Ad Hoc Networks

- Routing for large-scale networks

  → protocols based on hierarchical routing and autonomous clustering

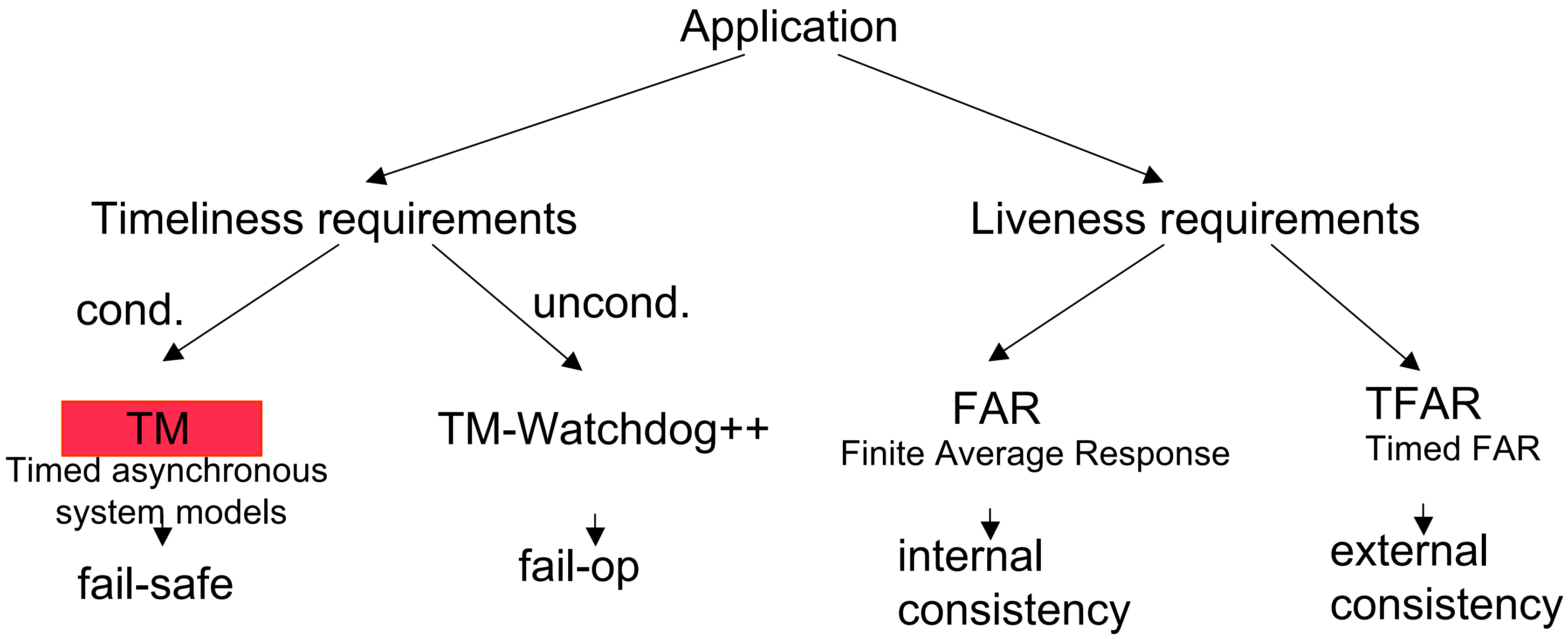
- Routing for asymmetric (heterogeneous) networks
- Location-based routing
- Energy efficient routing
- Secure routing
- QoS routing (several levels)

Farnam Jahanian

- Trends in threats / attacks (nature / rates / impact / protection solutions)
  - Nature :
    - Past: availability attacks (infrastructure disruption)
    - New attacks directly target people (ID theft, phishing)
- Rethink the protection solutions
- Internal Security Challenge
  - Evolving Threat models
  - Evolving Trust model
  - Evolving Business model
- No solution proposed ... But several questions!

- What if we expand the pool of bots and botnets to include 2+Billion smart phones and PDAs?
- How to protect against DoS attacks from a massive number of widely-distributed wireless devices?
- How to protect against millions of persistent infected mobile devices?
- Can traffic analysis techniques be applied to wireless networks and mobile applications?
- How to apply anomaly detection?
- How to secure a broad rage of new mobile platforms and applications? How to protect sensitive data on mobile devices?
- Where is the perimeter? What is the deployment model for security devices such as firewall, IDS, IPS? Where do you analyze, detect and stop potentially malicious traffic?
- Can the routing infrastructure be secured?
- How would convergence of networking platforms and security devices in the wired world affect mobile computing?

Christof Fetzer



- Mobile / pervasive: TM
- Model assumptions: simplify protocol development & correctness proof
- Mobile: communication assumptions are very weak

  Assumption about response time

  Average transmission delay is finite

Gün Sirer: A comprehensive localizing framework for self-organizing systems

- **Localization problem**: knowing where nodes are is a difficult problem if assuming realistic assumptions.
- Sextant
  - Uses both **positive and negative information** to localize nodes.
    - **Positive constraints**
    - **Negative constraints**
    - **Bézier curves to represent regions**
  - Nodes constantly disseminates information on their location.
  - Event localization interacts with node localization: events helps node localization and vice-versa.
- Mobility and malicious behaviors introduce new problems
- Programming model see mobile networks as a system of systems. Resource mediation layer is needed.
- Replicated objects can be used to provide some redundancy (may have node identification problems)

Session 3 - Summary report from Henrique Madeira

Rick Schlichting: A network service provider view of ubiquitous nomadic computing

- **Scale matters**: ubiquitous computing means more endpoints and more data. **The huge amount data is the problem**!
- Heterogeneity is there.
- RFID (Radio Frequency Identification) is essential for ubiquitous computing
- RFID services will change information exchange volumes.
- Object Naming Services (ONS), a kind of DNS for ubiquitous computing.
- Research issues:
  - CPU speed, memory,.. constraints are not the problem
  - The amount of data is the problem. How to manage, analyze and visualize all that data? Traditional DB cannot handle this amount of data.
  - Data reduction techniques?
  - Data → Information → Knowledge

Question: no research issues on dependability?

Session 3 - Summary report from Henrique Madeira

Our Assignments

- Carl Landwehr (based on discussion with Joel Birnbaum of HP who reports having long discussions with Mark Weiser)
  - Nomadic computing: people move around, computers may or may not
  - Mobile computing: computers move around, people may or may not
  - Ubiquitous computing: computers are all around, but you are aware of their presence; you may use them explicitly.
  - Pervasive computing: computers are everywhere, but have disappeared into the woodwork. You aren't aware you are using them.
- What are the top 5 problems that need to be solved to enable dependable nomadic computing?

- mobile *adj.* **1. Moving or capable of moving readily (*especially from place to place); SYN:* nomadic, peregrine, roving, wandering.**

- nomadic *adj.* **1. Of or pertaining to nomads, or their way of life; wandering; moving from place to place for subsistence; "a nomadic tribe."**

Computers that move from place to place

- ubiquitous *adj.* **1. Existing or being everywhere, or in all places, at the same time; omnipresent.**

- pervasive *adj.* **1. Tending to pervade, or having power to spread throughout; of a pervading quality.**

Computers embedded in physical objects
(that may or may not move from place to place)

HOMEWORK: Find your way in the jungle of perviquitous systems

Paulo Esteves Veríssimo

Navigators Group,
LaSIGe, Laboratory for Large-Scale Informatic Systems
Univ. Lisboa
pjv@di.fc.ul.pt

http://www.di.fc.ul.pt/~pjv

``X axis'':

- Nomadic
  - you go from place to place, but you are not quite on-line in between
- Mobile
  - you go from place to place, *and* you are on-line in between

``Y axis'':

- Ubiquitous
  - you compute wherever you are, desirably with seamless power and connectivity.
    - e.g. GLOBAL COMPUTING Initiative of the EU.

- Pervasive
  - computers exist everywhere, they *permeate* the environment, the objects you use, you yourself.
    - (i) may be an enabler of 'ubiquitous';
    - (ii) generates considerable amount of information, picture as metaphor ``event sprays'', we have to learn how to cope with.
    - e.g. DISAPPEARING COMPUTER Initiative of the EU

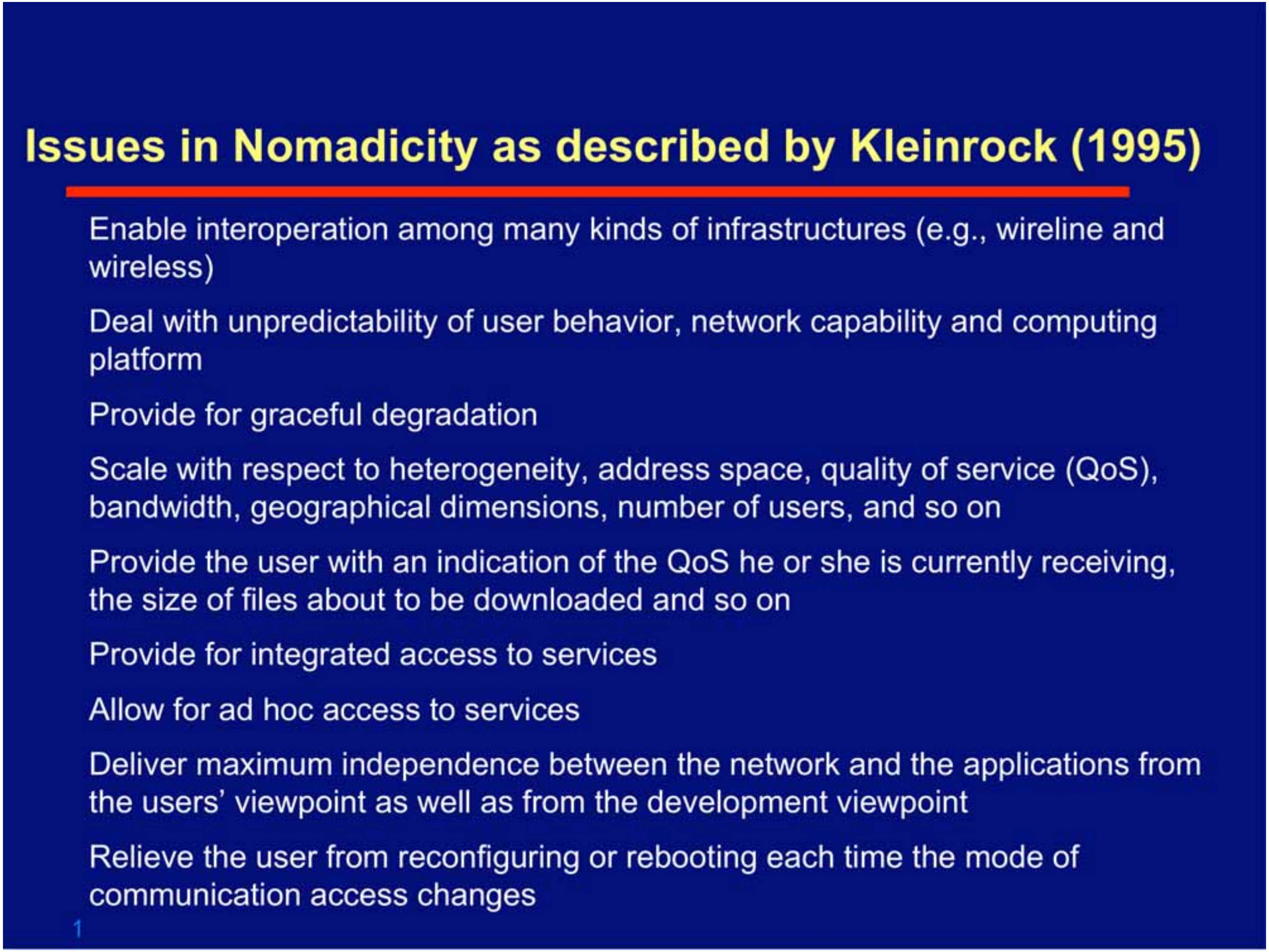
``Y axis'':

- Ubiquitous
  - you compute wherever you are, desirably with seamless power and connectivity.
  - Orthogonal to nomad/mobile. Gives a dimension of scale to the latter (many places to migrate to, many paths where I can move through and be on-line)
  - e.g. GLOBAL COMPUTING Initiative of the EU.
- Pervasive
  - computers exist everywhere, they *permeate* the environment, the objects you use, you yourself.
  - Essentially, the effects, seen from the same level of abstraction as ubiquitous was mentioned, are:
  - (i) it may be an enabler of 'ubiquitous';
  - (ii) it generates considerable amount of information, picture as metaphor ``event sprays'', that we have to learn how to cope with.
  - e.g. DISAPPEARING COMPUTER Initiative of the EU.

Relation to embedded systems:

- This world will become what may called ``complex embedded systems'' or more appropriately **``systems of embedded systems'':**
    - ad-hoc collections of largely wireless and mobile entities
    - active environments of pervasive and inconspicuous devices, that can also be moved as we move furniture
    - will be formed by recursive collections of small-scale embedded systems as we know them today
    - e.g. Embedded CO-OPERATING OBJECTS Initiative of the EU.

- Navigators group:
  - http://www.navigators.di.fc.ul.pt/



Issues in Nomadicity as described by Kleinrock (1995)

Enable interoperation among many kinds of infrastructures (e.g., wireline and wireless)

Deal with unpredictability of user behavior, network capability and computing platform

Provide for graceful degradation

Scale with respect to heterogeneity, address space, quality of service (QoS), bandwidth, geographical dimensions, number of users, and so on

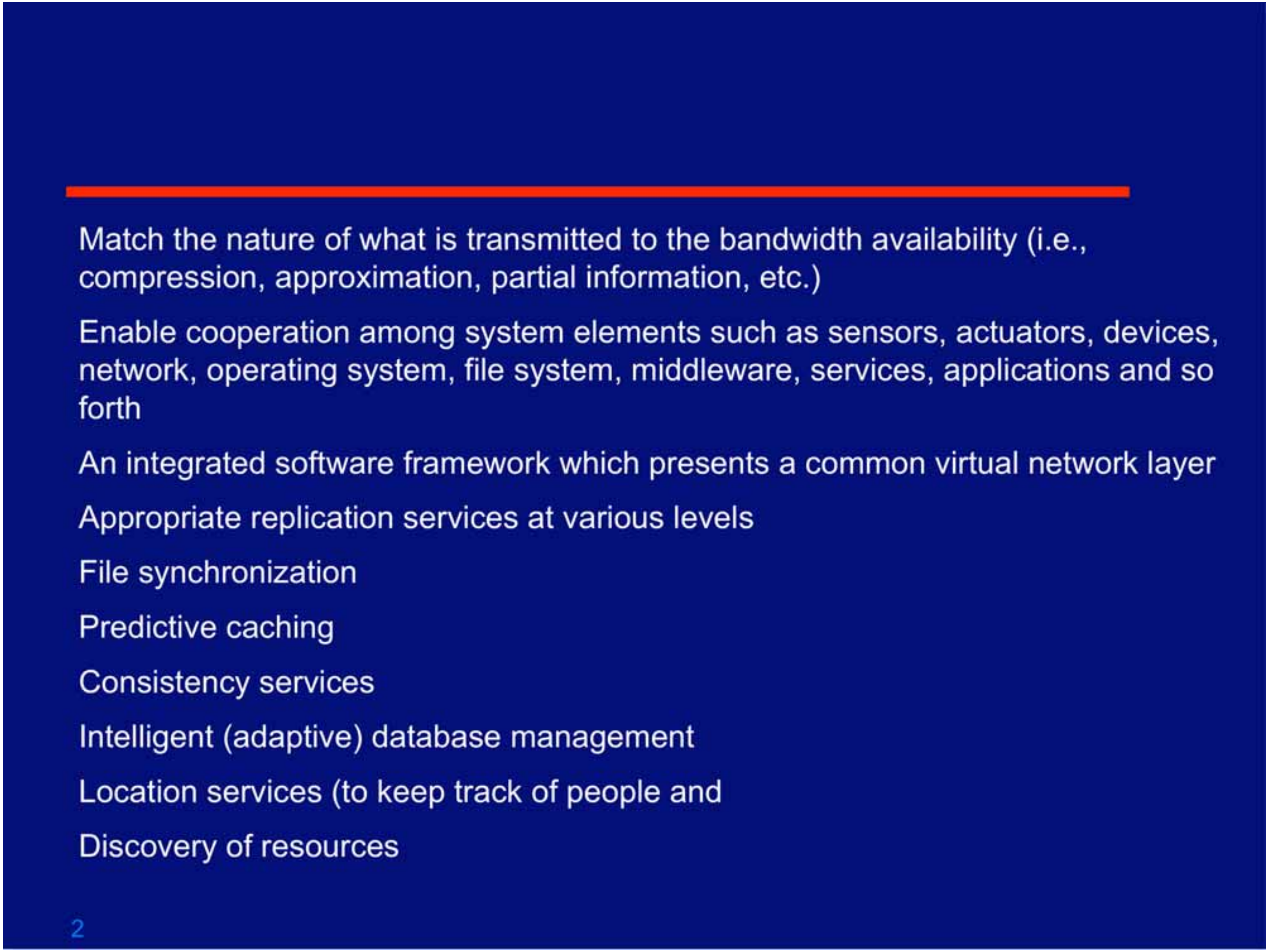
Provide the user with an indication of the QoS he or she is currently receiving, the size of files about to be downloaded and so on

Provide for integrated access to services

Allow for ad hoc access to services

Deliver maximum independence between the network and the applications from the users’ viewpoint as well as from the development viewpoint

Relieve the user from reconfiguring or rebooting each time the mode of communication access changes



Match the nature of what is transmitted to the bandwidth availability (i.e., compression, approximation, partial information, etc.)

Enable cooperation among system elements such as sensors, actuators, devices, network, operating system, file system, middleware, services, applications and so forth

An integrated software framework which presents a common virtual network layer

Appropriate replication services at various levels

File synchronization

Predictive caching

Consistency services

Intelligent (adaptive) database management

Location services (to keep track of people and

Discovery of resources

2

IFIP WG 10.4

Business Meeting

Chair
Jean Arlat, LAAS-CNRS, Toulouse France

48th IFIP WG 10.4 Meeting

Hôtel de Yama, Hakone, Japan

Friday July 1 — Tuesday July 5, 2005

Business Meeting

Monday July 4, 2005

Agenda

- IEEE/IFIP DSNs - DSN-2005, DSN-2006, DSN-2007
- IEEE Trans. on Dependable and Secure Computing
- Future WG Meetings — 49, 50, ...
- SIG on Dependability Benchmarking
- TC-10 Conference at WCC'2006
- Other Supported Events
- Part restricted to WG members

IEEE/IFIP International Conference on Dependable Systems and Networks

Yokohama, Japan (June 28 - July 1, 2005) - 353 Attendees!

Philadelphia, PA, USA (June 25-28, 2006)

- General Chair: Chandra Kintala (Stevens Inst. of Technology, Hoboken, NJ, USA)
- Conference Coordinator: David Taylor (Univ. of Waterloo, Canada)
- DCCS Program Chair: Lorenzo Alvisi (University of Texas, Austin, USA)
- PDS Program Chair: Aad Van Moorsel (University of Newcastle Upon Tyne, UK)

Edinburgh, Scotland (June 25-28, 2007)

- General Chair: Tom Anderson (University of Newcastle Upon Tyne, UK)
- Conference Coordinator: Mohamed Kaâniche (LAAS-CNRS, Toulouse, France)
- DCCS Program Chair: Zbigniew Kalbarzyck (Univ. of Illinois Urbana-Champaign, USA)
- PDS Program Chair: Peter Buchholz (TU Dortmund, Germany)

Anchorage, AL, USA (TBD, 2008)

- General Chair: Phil Koopman (Carnegie Mellon University, Pittsburgh, PA, USA)

IEEE Transactions on Dependable and Secure Computing

htpp://computer.org/tdsc

- Second Year
- Two Special Issues:
  - "Oakland 2005" (IEEE Symp on Security & Privacy, May 2005)
  - DSN-2005 (DCCS & PDS)
- Think of submitting a paper!

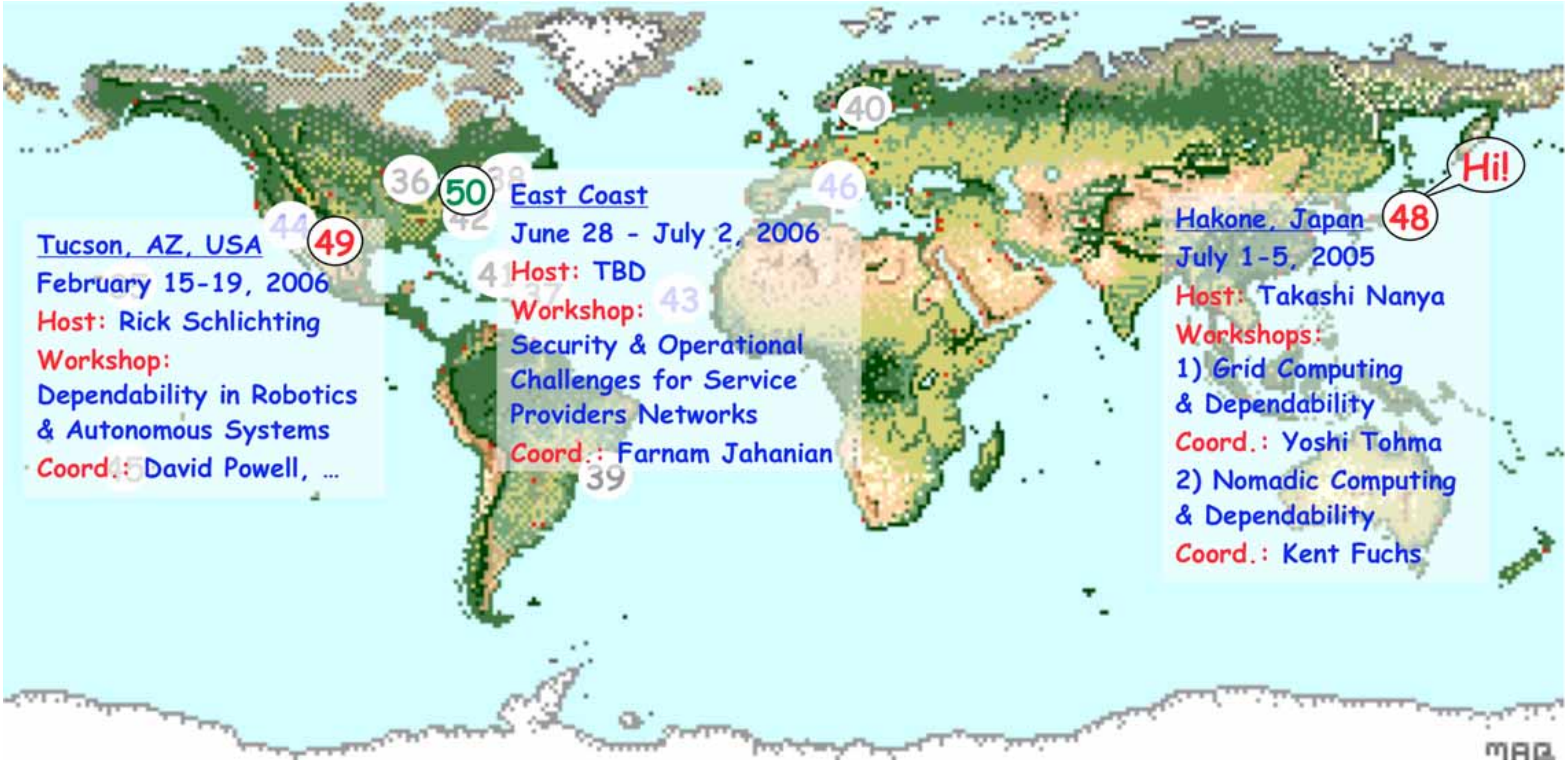
(Some) Proposals for Workshop Topics

- Grid Computing and Dependability (Yoshi)
- Nomadic Computing and Dependability (Kent)
- Dependability in Robotics and Autonomous Systems (David,...)
  [Possibly in connection with Int. Advanced Robotics Programme WG on Robot Dependability]
  —> 49th meeting - linked to DCCS-2006 PC Meeting
- Security and Operational Challenges for Service Providers Networks (Farnam)
  —> 50th meeting - linked to DSN-2006
- Software Dependability (Karama,...)
- Critical Infrastructures (Carl, Bill,...)
- ...

Major Workshop Topics

distributed computing, parallel computing, real-time systems, certification of dependable systems, specification methods, design diversity, specification and validation of hard dependability requirements, methodologies for experiments, VLSI testing and fault tolerance, hardware- and-software testing and validation, fault tolerance in new architectures, communication networks, algorithms for distributed agreement, cars and computers, accidental *vs.* intentional faults, robotics and dependability, limits in dependability, avionics and dependability, dependability issues in medical computing, security and dependability, tools for dependable system design and evaluation, railway safety, safety cases, dependability in automotive electronics, computer systems benchmarking with applications to dependability, time and dependability, dependability, survivability, and integrity in e-commerce transactions and infrastructure, dependability benchmarking, utilization of formal methods in dependable systems, challenges and directions for dependable computing, dependability and survivability, middleware for adaptivity and dependability, measuring assurance in cyberspace + hardware design and dependability, open source and dependability, human computer interaction and dependability, autonomic web computing, grid computing and dependabity + nomadic computing and dependability, …

Future Meetings



Tucson, AZ, USA
February 15-19, 2006
Host: Rick Schlichting
Workshop:
Dependability in Robotics & Autonomous Systems
Coord.: David Powell, …

East Coast
June 28 - July 2, 2006
Host: TBD
Workshop:
Security & Operational Challenges for Service Providers Networks
Coord.: Farnam Jahanian

Hakone, Japan
July 1-5, 2005
Host: Takashi Nanya
Workshops:
1) Grid Computing & Dependability
Coord.: Yoshi Tohma
2) Nomadic Computing & Dependability
Coord.: Kent Fuchs

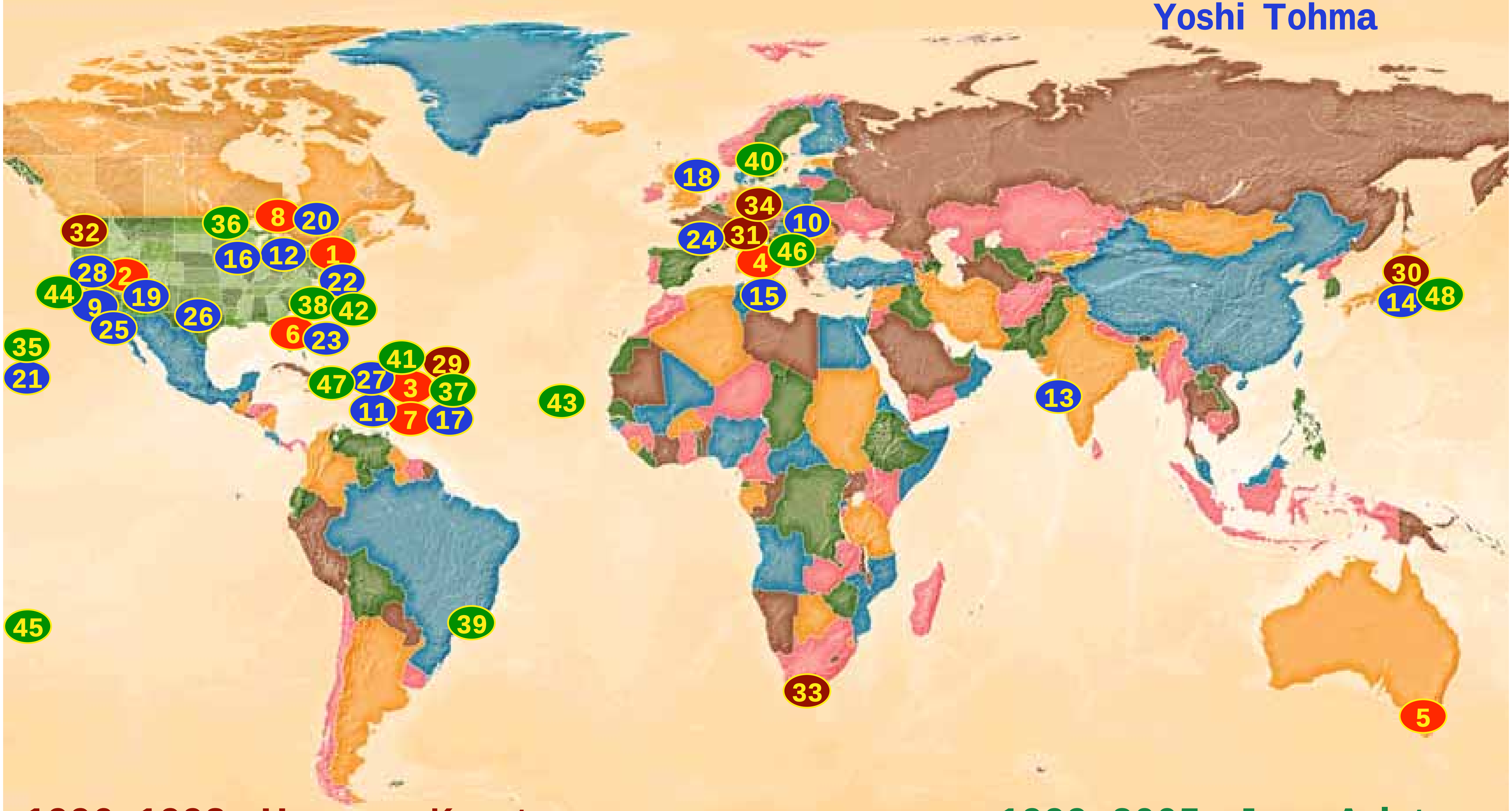
51 Open (Mexico, India ?)

52 Scotland: Shore of Loch Lomond

Next ?

1981-1985: Al Avižienis
Alain Costes

1986-1995: Jean-Claude Laprie
John Meyer
Yoshi Tohma



1996-1998: Hermann Kopetz
Jacob Abraham
Hirokazu Ihara

1999-2005: Jean Arlat
Takashi Nanya
Bill Sanders

SIG'DeB

http://www.ece.cmu.edu/~koopman/ifip_wg_10_4_sigdeb

- Panel held at DSN-2005

- Next SIG Meeting : November 8, 2005
  -> Workshop on Dependability Benchmarking organized jointly with ISSRE-2005, Chicago, IL (8-11 Nov., 2005)

TC-10 Conference at IFIP WCC-2006
Biologically Inspired Cooperative Computing

- Chairs: Franz Rammig (U. Paderborn-Chair TC10) & Mauricio Solar (U. Santiago Chile)
- Program Chairs: Yi Pan (U. Georgia) & Hartmut Schmek (U. Karlsruhe)
- Not bio-informatics -> Four Streams:
  - (1) Modelling and Reasoning about Collabarative Self-Organizing Systems (10.1)
  - (2) Collaborative Sensing and Processing Systems (10.3)
  - (3) Dependability of Collaborative Self-Organizing Systems (10.4)
  - (4) Design and Technology of Collaborative Self-Organizing Systems (10.5)
- PC (to include)

Wolfgang Nebel	Freiburg	Germany	Deborah Estrin*	Los Angeles	USA
Henk Sips	Delft	The Netherlands	Bernhard Sendhoff*	Offenbach	Germany
Albert Y. Zomaya	Sydney	Australia	Jean Arlat	Toulouse	France
Stephan Olariu	Norfolk	USA	Kim Kane	Irvine	USA
Ivan Stojmenovic	Ottawa	Canada	Eliane Martins	Campionas	Brasil
Johnnie Baker	Kent	USA	Roy A Maxion	Pittsburgh	USA
Ricardo Reis	Porto Alegre,	Brasil	Takashi Nanya	Tokyo	Japan
Marco Dorigo*	Brussels	Belgium	William H. Sanders	Urbana	USA
Xiaodong Li*	Melbourne	Australia	Richard D. Schilchting	Florham Park	USA
Luca M. Gambardella*	Manno-Lugano	Switzerland	Charles Rattray	Stirling	UK
Daniel Polani*	Hatfield	UK	Jochen Pfalzgraf	Salzburg	Austria
Christian Müller-Schloer*	Hannover	Germany	Leslie S. Smith	Stirling	UK

* To be confirmed

Other (in cooperation) Events

- **WORDS-2005** (10th Int. Workshop on Object-oriented Real-time Dependable Systems), **Sedona, AZ, USA, February 2-4, 2005 —** http://asusrl.eas.asu.edu/srlab/activities/words05/words05.htm
- **EDCC-2005** (5th European Dependable Computing Conference), Budapest , Hungary, **April 20-22, 2005** — http://sauron.inf.mit.bme.hu/EDCC5.nsf
- **4th IARP/IEEE-RAS/EURON** Workshop on Technical Challenges for Dependable Robots in Human Environments**, Nagoya, Japan, June 16-18, 2005**
- **SAFECOMP-2005** (24th International Conference on Computer Safety, Reliability and Security, Fredrikstad, Norway, **September 28-30, 2005** — http://www.safecomp.org
- **LADC-2005** (2nd Latin-American Symposium on Dependable Computing), Salvador, Bahia, Brazil, **October 25-28, 2005** — http://www.lasid.ufba.br/ladc2005
- **PRDC-2005** (11th Int. Symp. Pacific Rim Dependable Computing), Changsha, China, **December 12-14, 2005 —** http://sc.hnu.cn/newweb/communion/prdc2005/presentation.htm
- **SAFECOMP-2006** (25th International Conference on Computer Safety, Reliability and Security, **Gdansk, Poland, September, 27-29 2006**
- **EDCC-2006** (6th European Dependable Computing Conference), Coimbra, Portugal, **October 14-17, 2006** — http://edcc.dependability.org

DSN2005 Summary

Registration

DSN Registrants : 350

- Paid(full registration) : 332
- Free invitees : 18
  - Sponsor representatives: 16
  - Honorary general chair: 1
  - Keynote speaker: 1

Tutorial-only: 3

DSN show-up (1)

- Academia: 234
- Corporation:82
- Government:11
- Unknown:5

- Total: 332

DSN Show-up : 21 countries

Japan:	132
USA:	104
France:	11
Italy:	11
Korea:	10
Portugal:	10
Germany:	9
UK:	7
Sweden:	6
China, Taiwan, Israel, Brazil :	4
Netherlands, Spain, Swiss:	3
Canada, Russia:	2
Mexico, Norway, Singapore:	1

Tutorials registrants

- A : 41
- C : 23
- E : 28

Registration Income

• DSN conference:	17,022,000 Yen
• Tutorials:	1,365,000 Yen
• Total Income:	18,387,000 Yen

Award and Grant

- IEEE CS: 9600 US$
- IFIP TC10: 3000 Euro

- Carter Award 750 US$ x 2
- Student Grant 500 US$ x 26

DSN2005 Sponsors

Telecommunications Advancement Foundation

HITACHI Inspire the Next — Hitachi, Ltd

FUJITSU — FUJITSU LIMITED

JR Railway Technical Research Institute — Railway Technical Research Institute

NEC — NEC Corporation

RCAST Univ. Tokyo — RCAST, University of Tokyo

ICF — International Communications Foundation

DAIDO — Daido Signal Co.,Ltd.

MITSUBISHI ELECTRIC — MITSUBISHI ELECTRIC CORPORATION

OKI — OKI Electric Industry Co.,Ltd.

NISSAN — Nissan Motor Co.,Ltd.

NIPPON SIGNAL THE NIPPON SIGNAL CO., LTD. — The Nippon Signal Co.,Ltd.

JR-EAST — East JapanRailway Company

NTT — Nippon Telegraph and Telephone Corporation

Panasonic — Matsushita Electric Industrial Co.,Ltd.

IBM — IBM Corporation

Kyosan Electric Mfg. Co., Ltd. — Kyosan Electric Mfg.Co.,Ltd.

Sun microsystems — Sun Microsystems

Inoue Foundation for Science

Thank you !

DSN 2006 Update - June05

Organizing Committee:

General Chair:	Chandra Kintala
Conf. Coordinator:	David Taylor
PC Chair for DCCS:	Lorenzo Alvisi
PC Chair for PDS:	Aad van Moorsel
Finance:	Sachin Garg
Local Arrangements:	Navjot Singh (Chair), Bengi Karacali
Publicity:	Timothy Tsai
Registration:	Rick Buskens (Chair), Yennun Huang
Publications:	Priya Narasimhan
Workshops:	Neeraj Suri
Tutorials:	Joanne Dugan
Student Forum:	Christof Fetzer
Fast Abstracts:	Saurabh Bagchi

DSN 2006 Update - June05

- **Hotel** contract signed by IEEE with Sheraton Society Hill, Philadelphia, PA, USA
  - 1 mile from downtown, 10 miles from airport
  - Total room block 630 nights, Room rate $159/night (single/double)
  - 10 meeting rooms, largest meeting room can hold 950
  - Internet, a/v, etc. costs may now be $12K
- **Social Event:** One of 2 possibilities
  - Exclusive tour of Constitution Center and dinner in the center
  - Banquet on a famous docked ship and a tour of something in Philadelphia

DSN 2006 Update - June05

- **Publicity**
  - 1-page ad in DSN2005 proceedings
  - 1-sheet, 2-sided, 3-paneled hard-copy CFC printed; 8000 copies
    - Distributed mailing through volunteers
  - Larger-size posters; 100 copies; please take and post them
  - Web-site with details – to be ready in July05
- **Publications**
  - Have a quote from IEEE for the 1st Volume of Proceedings – 850 pages, 350 hard copies, 375 CD; costs about $25K for Vol. 1
- **Program**
  - DCCS, PDS, Workshops, Tutorials, Student Forum and Fast Abstracts, Panel(s)
  - DCCS and PDS committees formed

DSN 2006 Update - June05

- Financials:
  - Working on IEEE TMRF and IFIP Event form for budget approvals
  - Industry Support/Funding - None yet
- Registration
  - IEEE or 3rd party (non-IEEE) services?
- Estimated charges as of today; preliminary numbers only
  - Advanced registration fees for
    - Members: $565, non-members: $710
    - Student members: $270, non-members: $340
  - Late/On-site registration fees for
    - Members: $680, non-members: $850
    - Student Members: $330, student non-members: $415
  - Social Event: $100

DSN 2006 Update - June05

Schedule for the next DSN-Fiscal Year:

- 1Q: July – Sept 05
  - Social event finalization, TMRF approval, advance loan from IEEE/IFIP, update web-site, registration services vendor selection, setup paper submission process, select publication vendor, fund-raising, CFP advertising, …
- 2Q: Oct – Dec 05
  - Registration website design, publications submission web-site and process, a/v vendor selection, keynote speaker search, CFP advertising, Carter Award process, …
- 3Q: Jan – Mar 06
  - Program committee meetings and program decisions, local arrangement logistics, souvenirs, advance program design and printing, …
- 4Q: Apr – Jun 06
  - Call for Participation advertising, final program, solicitation for Proceedings publication, Registration tasks, Meals and social details, …

DSN 07
Edinburgh, Scotland

Tom Anderson

Centre for Software Reliability

School of Computing Science

University of Newcastle upon Tyne, UK

Date and Location

Monday 25 - Thursday 28, June 2007

Edinburgh, Scotland

Weather

Normal climate for Southern Scotland
in June is:

bright, sunny, dry, and pleasantly warm.

Venue

EICC:

Edinburgh International Conference Centre

Purpose built - excellent audio visual capabilities

Modern facilities - completed 2001

Multiple seminar rooms - flexible configuration

Display areas

All housed in a single attractive building

Prime City centre location

EICC

DSN 05, Yokohama, 1st July 2005 5

Accommodation

Edinburgh offers a huge range of hotels, with four at 5* standard in the city centre, and many others at a full range of cost/quality levels.

Adjacent to the EICC is the Sheraton Grand (5*, finest spa facilities in Europe according to Condé Nast). Only a few minutes walking to the Hilton Caledonian, Novotel, Travelodge.

Sheraton Grand Hotel

Costs

UK is traditionally expensive, but:

- Registration can be kept to a similar level to DSN 04
- Hotel charges currently range upwards from about £80 for 3* hotels, £99 for 4* and £145 for 5*
- We will negotiate a discounted conference rate for rooms at a range of recommended hotels

Transportation

Airport: 8 miles from City Centre, with flights to 16 European destinations and more than 30 non-budget flights from London to Edinburgh every day. There is also a daily flight to New York. Glasgow airport is 40 minutes away.

Excellent rail links to rest of Britain, including Glasgow, Newcastle and London

Attractions

Capital city of Scotland

Malt Whisky, excellent restaurants

Edinburgh Castle, Holyrood Palace, Royal Yacht “Brittania”

Museums: National, Royal, Flight, War, Costume, Country Life

Usher Hall, Portrait Gallery of Scotland

Forth railway bridge (3 x double cantilever)

Arthur’s Seat and Salisbury Crag

Yacht Malt

DSN 05, Yokohama, 1st July 2005

Excursion (one option)

Short journey to the historic city of Stirling, location of

- the battle of Stirling Bridge (1297)
- the Wallace memorial
- and Stirling Castle.

Conference dinner at the castle

(Not) William Wallace

Stirling Castle

Sponsors

We have the benefit of close contacts with a large number of major industrial players, and also with a very large number of smaller organisations. Prospects for DSN donations are distinctly encouraging.

Evening Light in June

DSN 05, Yokohama, 1st July 2005 16

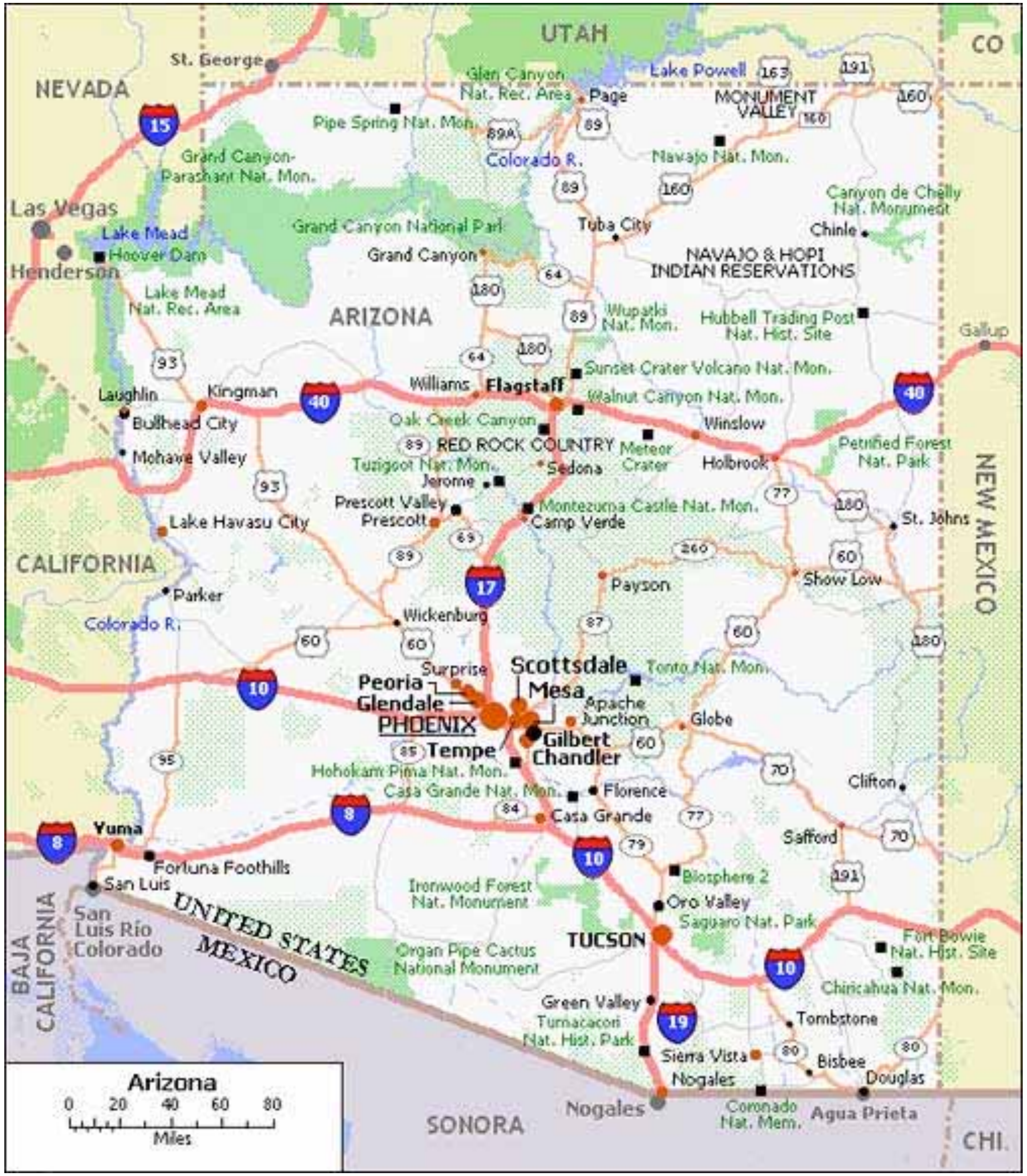
WG10.4 2006 Winter Meeting

- Arizona!

+ Unique, easily accessible and….. warm!

- Not a French island.

- Dates: Feb 15 (W) evening through Feb 19 (Su).



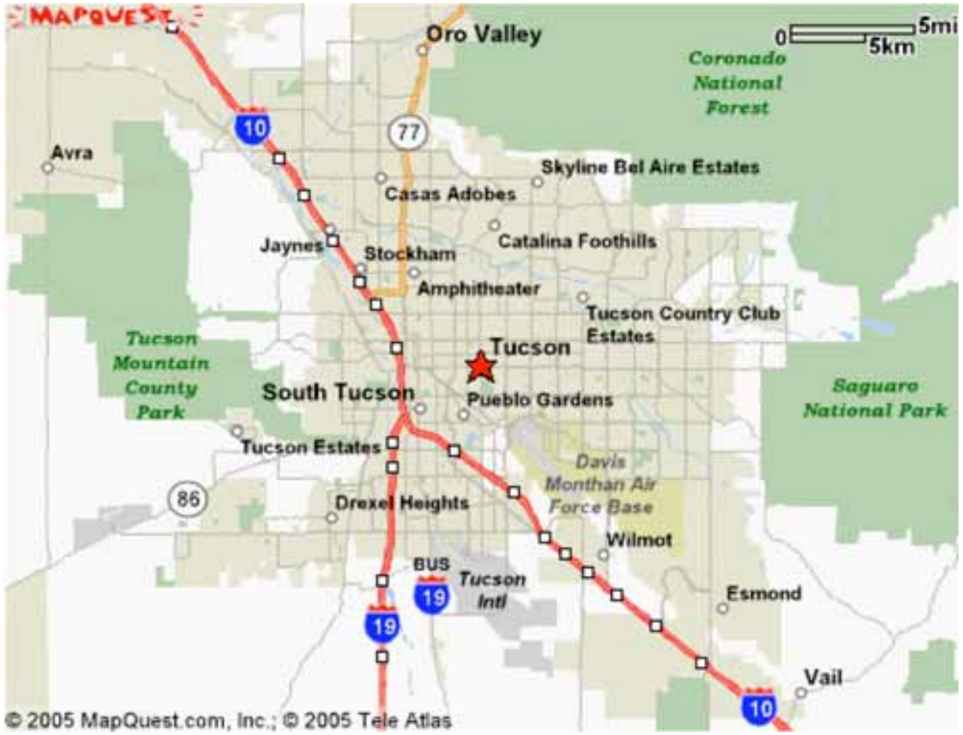
Location

- TBD, but likely in Tucson area
- Phoenix/Scottsdale also a possibility
- Sorry, Grand Canyon National Park not possible. :-(
- Status: Currently running approvals through AT&T.

Tucson

- Lots of tourist attractions and excursion options.

- AZ-Sonora Desert Museum, San Xavier mission, Mt. Lemmon, Sabino Canyon, Pima Air & Space Museum, Saguaro National Park, Old Tucson, ….



Tucson

- Lots of tourist attractions and excursion options.
- AZ-Sonora Desert Museum, San Xavier mission, Mt. Lemmon, Sabino Canyon, Pima Air & Space Museum, Saguaro National Park, Old Tucson, ….
- Tombstone, Bisbee, Tubac, Kartchner Caverns State Park, Nogales, Kitt Peak National Observatory, Organpipe National Monument, Madeira Canyon.
- Good restaurants!

Accessibility

- Medium size airport, conveniently located.
- Non-stops: Albuquerque, Atlanta, Chicago, Dallas, Denver, Houston, Las Vegas, Los Angeles, Minneapolis, Phoenix, Salt Lake City, San Diego, Seattle.
- Can also fly to Phoenix and drive 2 to 2.5 hrs.
- From DCCS PC meeting in Austin, TX (Feb 15):
  - Lv Austin 6:00p, Ar Tucson 9:02p (American, stop in DFW)
  - Lv Austin 6:00p, Ar Tucson 9:09p (America West, change in Phx)
  - Lv Austin 6:00p, Ar Tucson 11:09p (Delta, change in SLC)
  - Lv Austin 7:05p, Ar Tucson 10:27p (Continental, change in Houston)
  - Lv Austin 7:35p, Ar Tucson 11:25p (Southwest, change in Las Vegas)

EDCC-6

Coimbra, Portugal

18-20 October 2006

General Chair
João Gabriel Silva
University of Coimbra

Program Chair
Johan Karlsson
Chalmers University

EDCC-6

Submission deadline

2 April 2006

http://edcc.dependability.org/

Research Reports

Session 1

Moderator

William H. Sanders, UIUC, USA

Research Report
Dependable TCP/IP Networking

Elias P. Duarte Jr.

Federal University of Parana

Curitiba, Brazil

The 48th Meeting of IFIP WG 10.4

Hakone, Japan

July 1-5 2005

Outline

- Why work on TCP/IP dependability?
- An Overview of Dependable Network Management
- Current work: WAN Monitoring
  - DNR: Distributed Network Reachability
- GigaMan P2P: A Management Framework for the Brazilian Gigabit Backbone
  - Fault-Tolerant Routing
- Monitoring Dynamic Networks
- Distributed Integrity Checking
- Other Projects

We All Know That...

- The estimated number of Internet users has grown to 800 million persons worldwide
- Applications are increasingly critical for individuals & organizations
- How can one *monitor* such connected sets of heterogenous networks?
- What about re-configuration & control?

Integrated Network Management

- Monitoring & Control (Configuration)
- Independent of Operating System

- The 5 original management functions include:
  - Security Management
  - Performance Management
  - Configuration Management
  - Accounting Management
  - and.......

Fault Management

- Perhaps the most important function
  - At the very least you want to know what is working and what has crashed...
  - FAULT MANAGEMENT MUST BE FAULT-TOLERANT
- Several approaches have been proposed:
  - Use of Management Proxies for reaching managed objects
  - Management by Replication: replicating objects so that they are available post-mortem (IETF Draft)
  - The application of Distributed System-Level Diagnosis for LAN Management

Testing Is An Issue

- Several heterogeneous units are monitored
- For each unit, a test procedure must be defined
  - e.g. check whether the toner is too low, which virtually represents a faulty printer

We Are Currently Working on WAN Monitoring

- DNR: a Distributed algorithm for computing Network Reachability
- An algorithm to determine which portions of the network are reachable & unreachable
- The network may get partitioned & heal later
- Implementation: SNMP-based, allowing a reliable map to be drawn
- Reliable in the sense that even if part of the system is faulty, fault-free nodes are able to get reachability information

GigaMAN-P2P: Managing the Brazilian Gigabit Backbone

- The Brazilian RNP (Academic-Research Network) is currently upgrading links
- There are several challenges for managing high-speed networks
- Nodes are Autonomous Systems, in the sense that they are administered independently
- A Peer-To-Peer (P2P) Management System is being proposed
- Specific research project: Fault-Tolerant Routing

Monitoring Dynamic Networks

- It is difficult to model and map dynamic decentralized networks
- Information might be stale
- We have been working on an intelligent approach based on swarm intelligence
- IAgents migrate throughout the network collecting topology information

Distributed Integrity Checking

- Consider a choice of peers from which you can download a program
- Can you trust all of them?
- Remember: this is the Internet!
- How can a set of peers build a web of trust?

Comparison-Based Diagnosis

- Nodes run comparisons and report comparison results
- A Generalized Model of Distributed Diagnosis has been proposed
- After receiving a file/an output:
  - The tester compares files/outputs
  - If the comparison results in a match, nodes are classified in the same set
  - If a mismatch results, nodes are classified in different sets, according to the result

The New Model

- Allows nodes to be trusted according to the set they belong to
- A large number of comparisons may be executed in a distributed fashion

Other Projects

- An Architecture for IP Packet Tracing
- HyperGrid: a Dependable Grid Infrastructure
- SLA Contract Checking Based on MultiDimensional Search
- JXTA SNMP Peer

Megascale Project
A Low-Power and Compact Cluster for High-Performance Computing

Hiroshi Nakamura (U. Tokyo)

Masaaki Kondo (U. Tokyo)

Hiroshi Nakashima (Toyohashi UT)

Mitsuhisa Sato (U. Tsukuba)

Taisuke Boku (U. Tsukuba)

Satoshi Matsuoka (TITECH)

http://www.para.tutics.tut.ac.jp/megascale/

Background: Mega-Scale Project (1/2)

- Many applications need Peta-Flops.
  - Computational Genetics/Biology
  - Simulation of Environment/Crimate/Disaster
  - Computational Chemistry/Phisics/...
- Can we achieve Peta-Flops by extending traditional MPP/clusters? → NO!!
  - Huge space requirement (Gym @ 10^4 PE)
  - Huge power requirement (10MW @ 10^4 PE)

→ We need a new approach!!

＝Peta-Flops with Commodity Technology

Background: Mega-Scale Project (2/2)

- Our Mega-Scale project aims to establish fundamental technologies for 10^6 scale parallel systems focusing on;
  - Feasibility to build them with realistic cost and space ➔ **low-power for smaller footprint/volumn**
  - Dependability to operate them with high reliability and fault-tolerance
  - Programmability to obtain maximum performance with minimum effort

  based on commodity technologies.
- about €3M for 5 years, supported by JST (Japan Science and Technology Agency)

MegaProto : Prototype

- Objective : Proof of our claims
  - commodity technology > HPC dedicated
  - low-power/high-density > high-end/low-dens.
- Platform for our software development

  still under development, but...
  - power-aware compilation
  - high-performance/dependable NW: RI2N (Redundant Interconnection with Inexpensive Network)
    - network trunking for performance
    - network redundancy for reliability
  - fault-tolerant cluster management
    - Skewed Checkpointing for Multiple Failures (SRDS'04)

Conceptual Design (1/2)

performance/power perspective

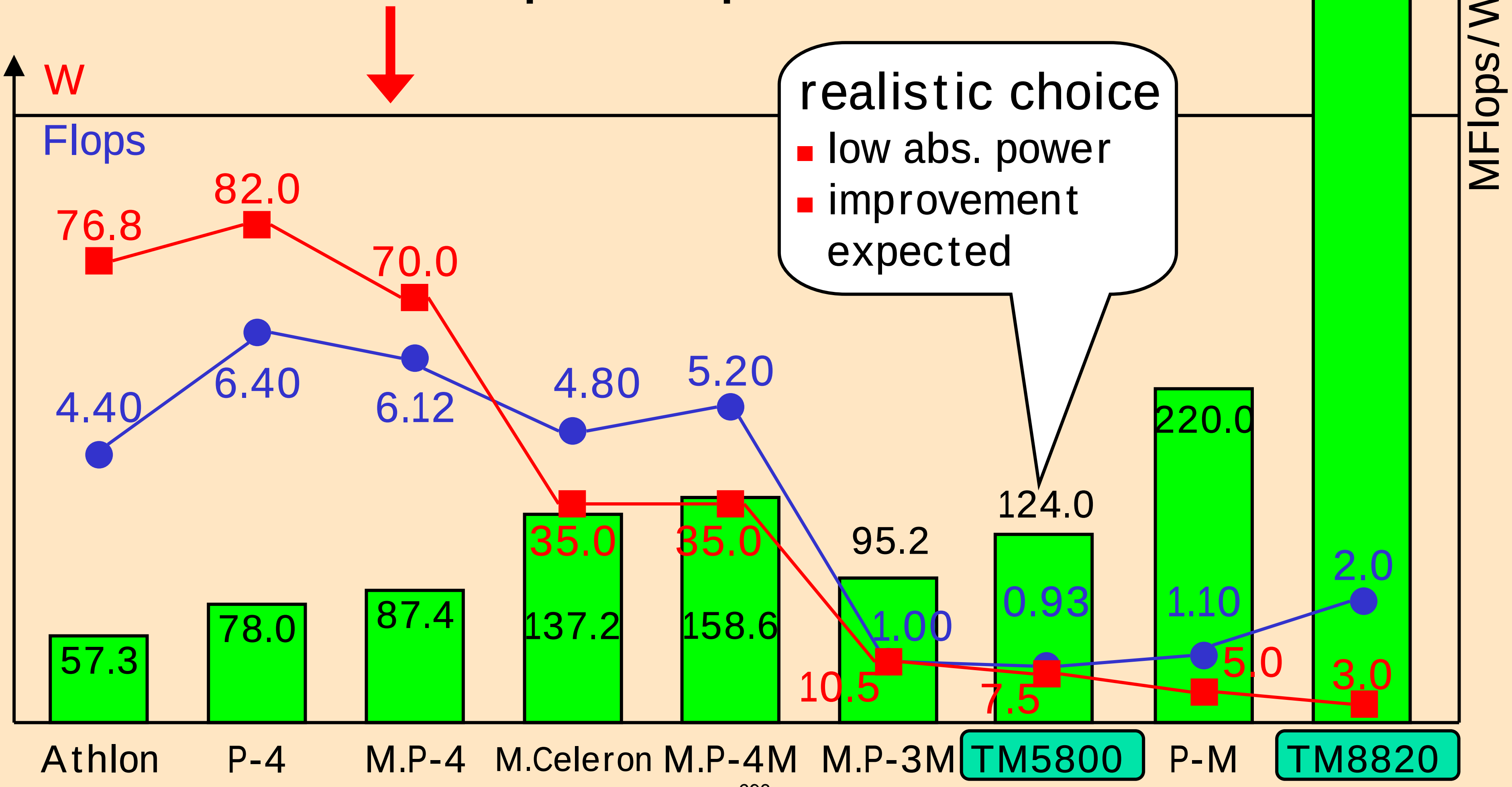
- Target power & perf./ (19" x 42U: 1rack)
  - peak perf. = 1TFlops
  - power = 10kW (300W/1U cooled by air)
  - → perf/power = 100MFlops/W
- Breakdown of power budget
  - processors = 1/4
    ⇒ 400MFlops/W
  - proc peripheral (mem. etc) = 1/4
  - network = 1/2

MegaScale

Conceptual Design (2/2)

now comes true!!

- Is 400MFlops/W processor available



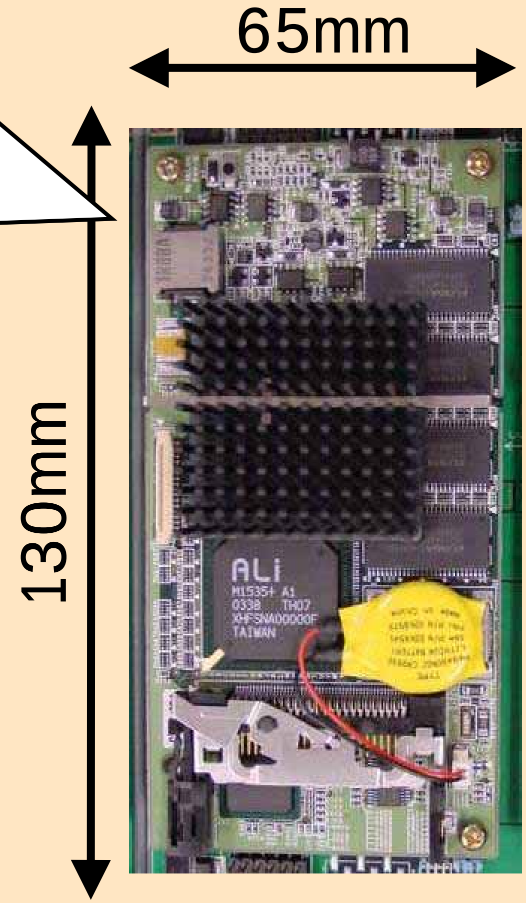
realistic choice

- low abs. power
- improvement expected

System Configuration (1/4)

version 1

- TM5800 (Crusoe)
- 0.93GFlops
- L1C =64KB L2C =512KB
- 256MB SDR



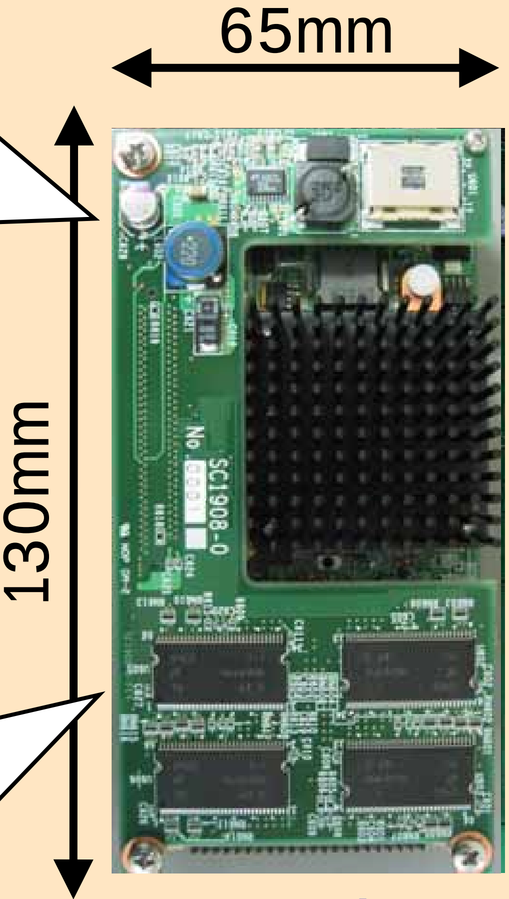
System Configuration (1/4)

version 1

- TM5800 (Crusoe)
- 0.93GFlops
- L1C =64KB L2C =512KB
- 256MB SDR

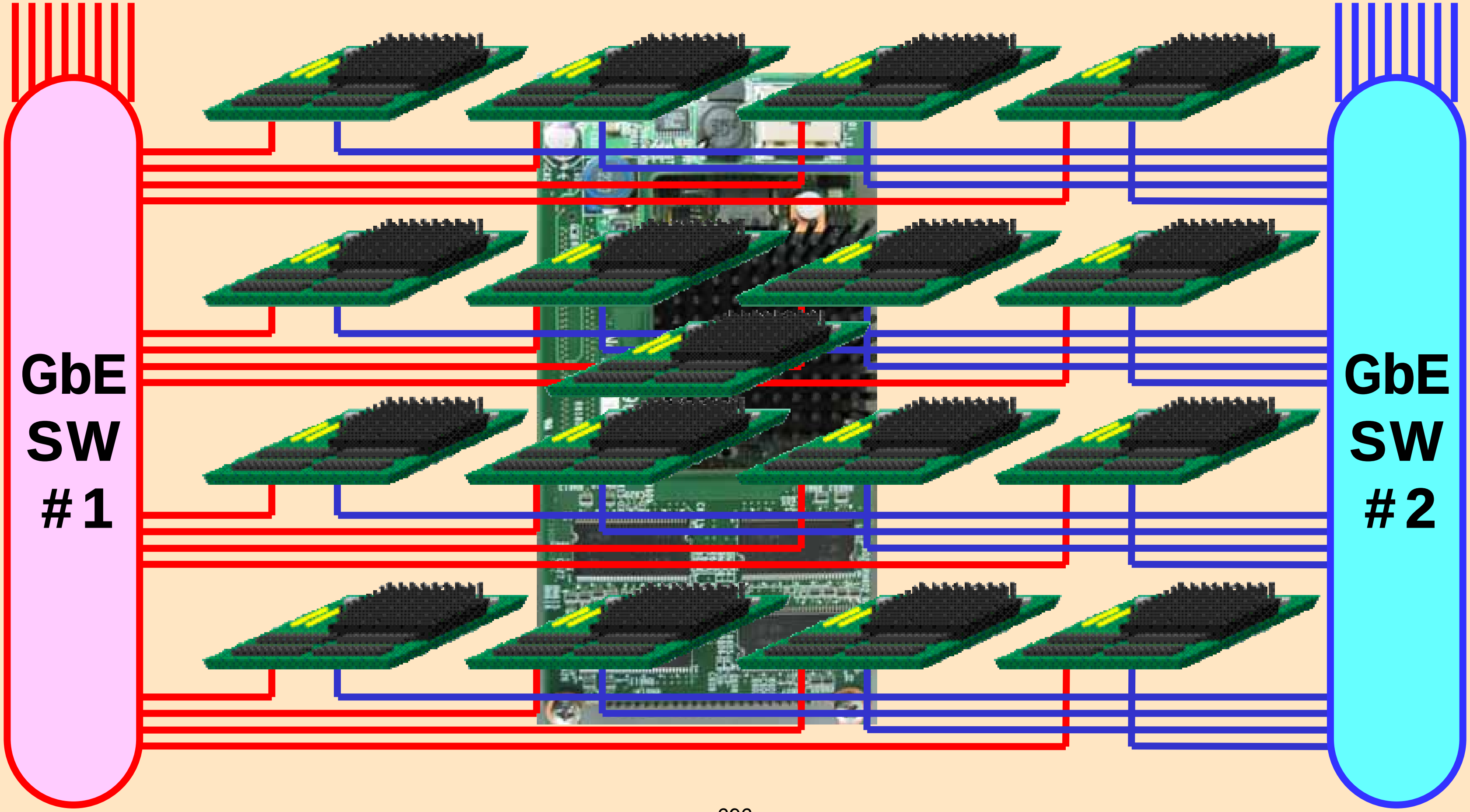
version 2

- TM8820 (Efficeon)
- 2.0GFlops
- L1C =192KB L2C =1MB
- 512MB DDR

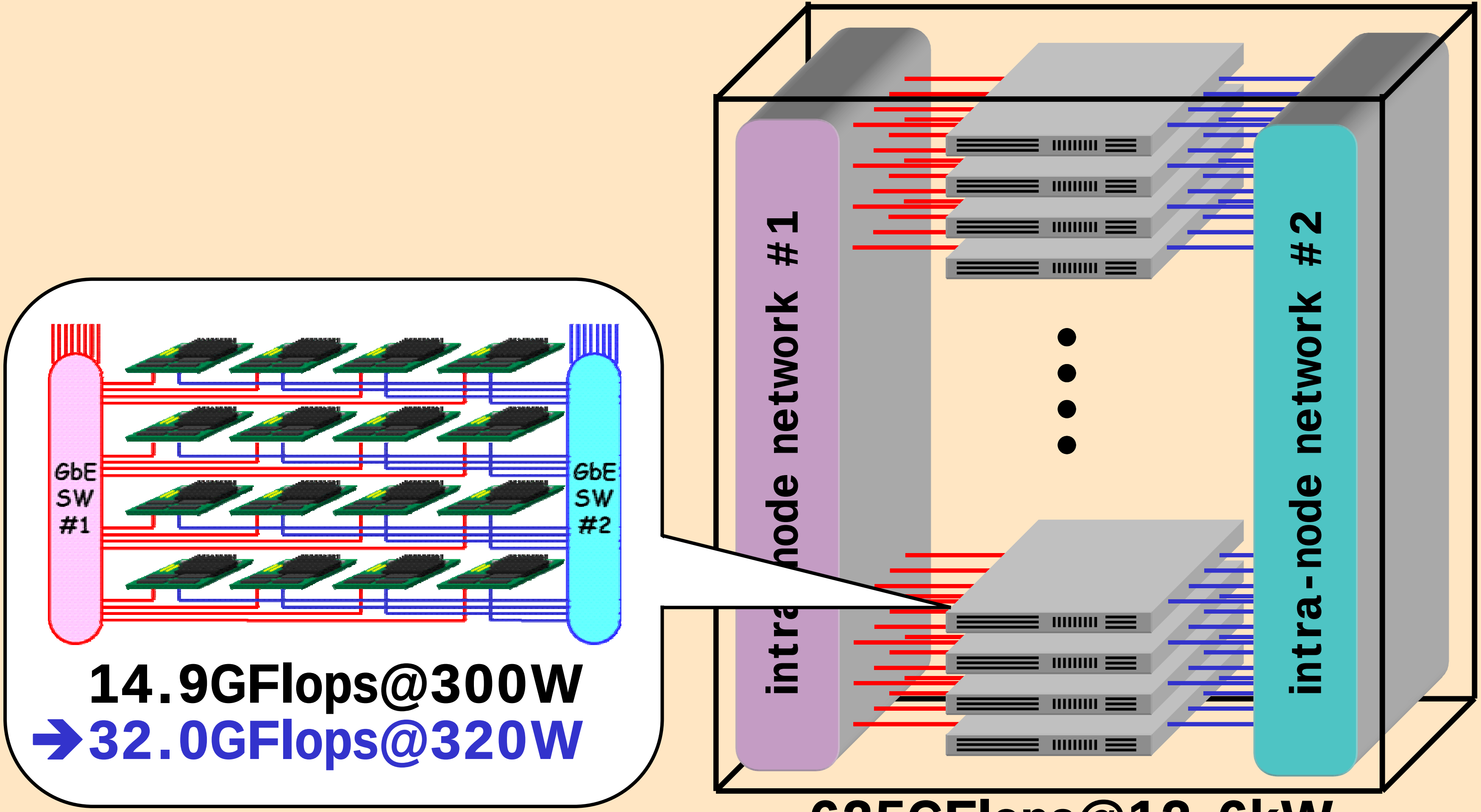


2-stage rocket !!

System Configuration (2/4)



System Configuration (3/4)



14.9GFlops@300W

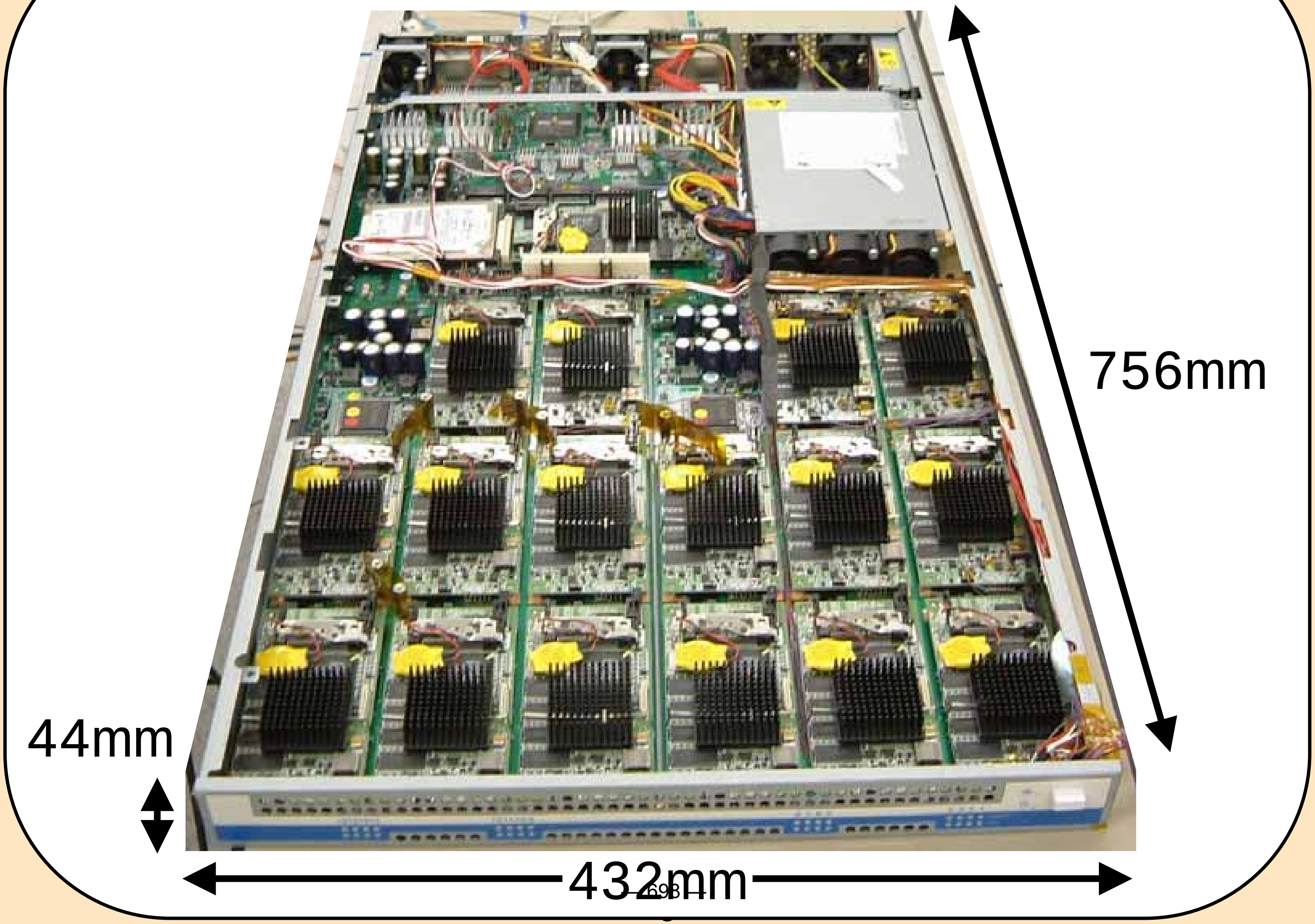
→32.0GFlops@320W

625GFlops@12.6kW

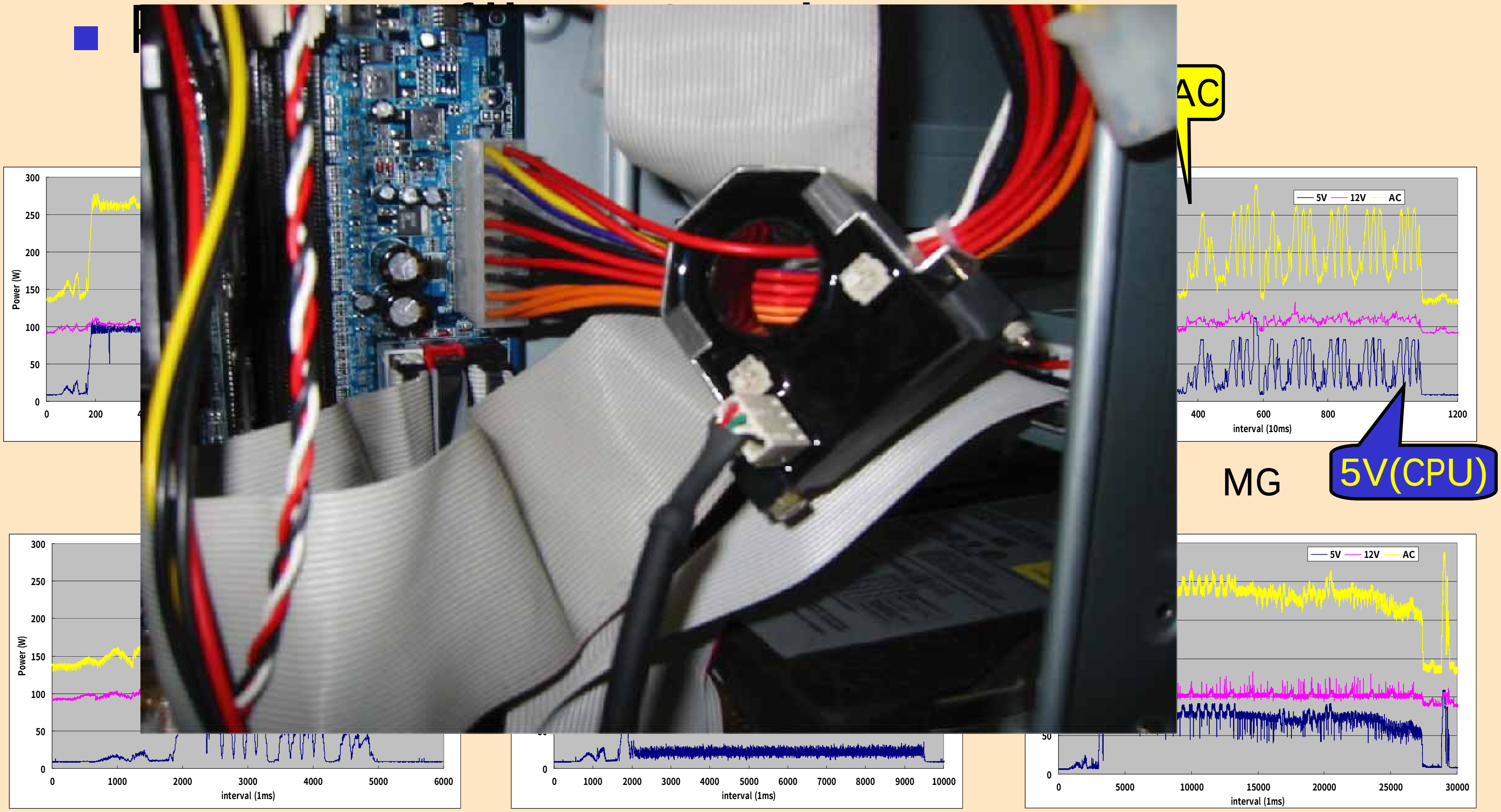
→1344GFlops@13.4kW

System Configuration (4/4)

version 1, delivered March, 2004



Performance Evaluation of MegaProto/Crusoe (1/3)



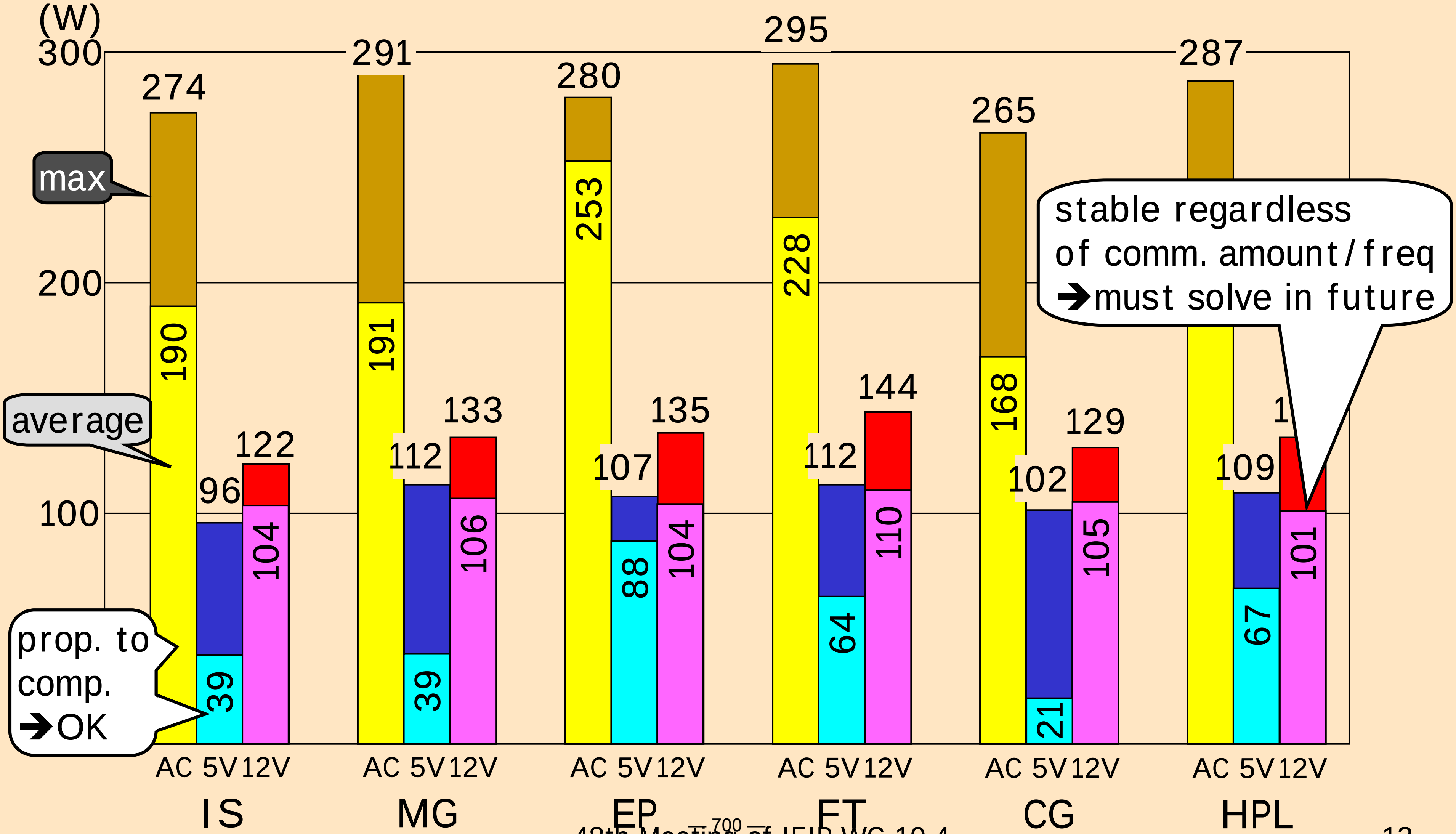
MG

IS

CG

HPL

Performance Evaluation of MegaProto/Crusoe (2/3)



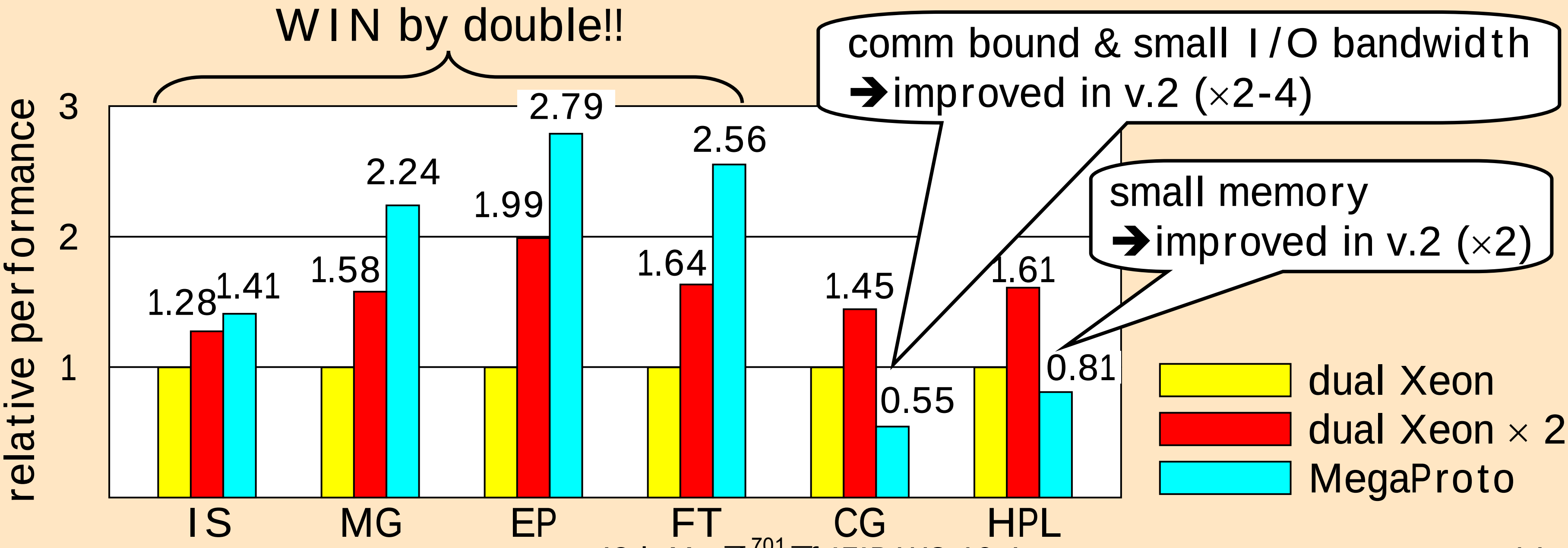
prop. to comp.
➔OK

stable regardless of comm. amount / freq
➔must solve in future

Performance Evaluation of MegaProto/Crusoe (3/3)

- v.s. 1U server (dual Xeon 3.06GHz, 1GB)

	dual Xeon 🟨	MegaProto 🟦
power / 1U	400W	300W
processor TDP	170W	120W
peak perf.	12.24 GFLOPS	14.88 GFLOPS



WIN by double!!

comm bound & small I/O bandwidth
➔improved in v.2 (×2-4)

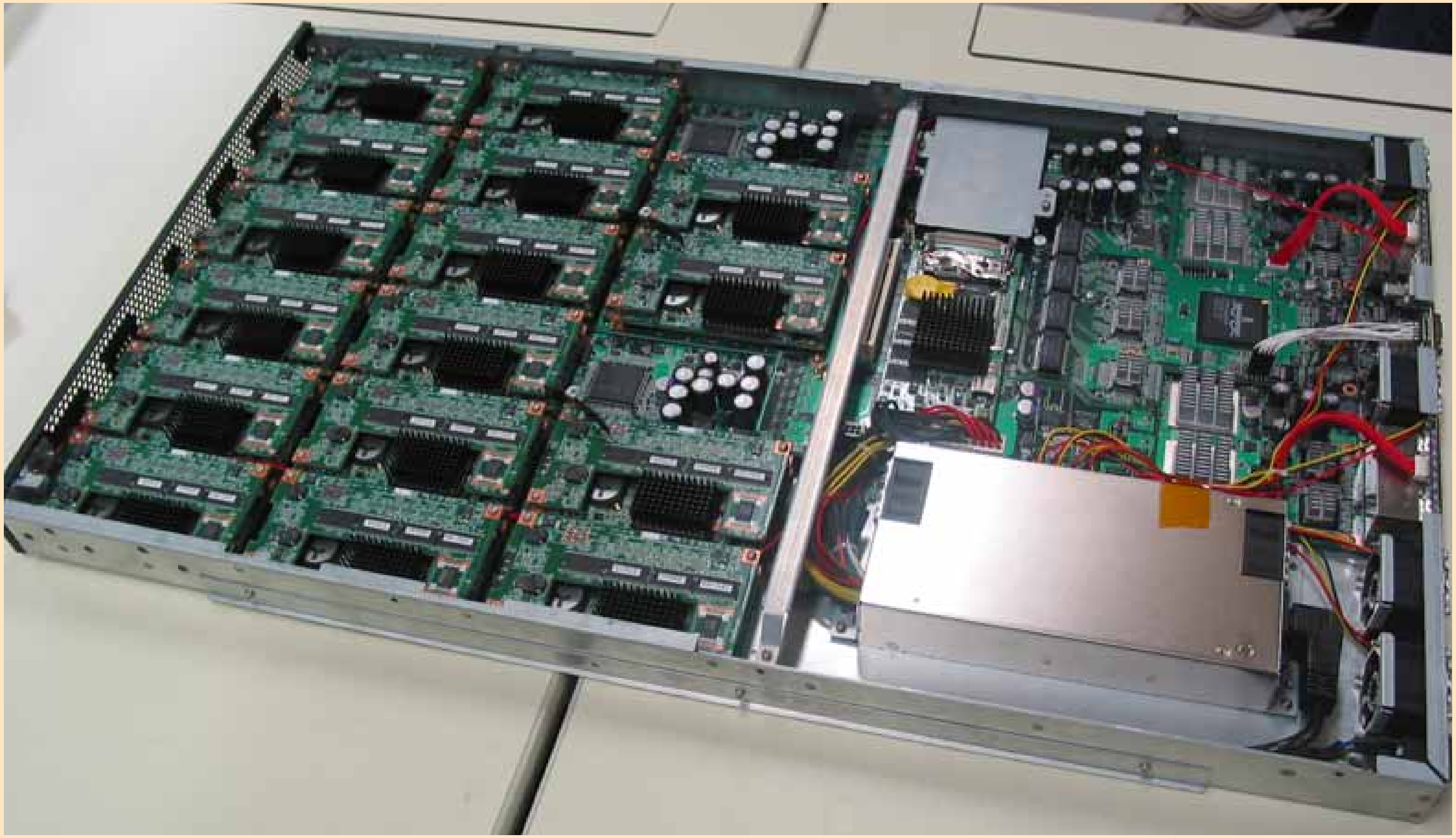
small memory
➔improved in v.2 (×2)

Summary

- Megascale Project : A Low-Power and Compact Cluster for High-Performance Computing
  - megascale high-performance low-power computing based on commodity technology
- MegaProto/Crusoe (version 1)
  - (TM5800@933MHz ＋ 2 x 1GbE) x 16 ＝14.9GFlops@300W (50MFlops/W)
  - 1.4-2.8 x dual-Xeon (IS,MG,EP,FT)
  - March, 2004 : 2 Unit (32 PE)
  - good performance/power
- MegaProto/Efficeon (version2)
  - (TM8820@1.0GHz ＋ 2 x 1GbE) x 16
  - June, 2005 : 20 Unit (320 PE)

MegaProto/Efficeon (version 2)

- delivered yesterday!



THE WHITE ROSE GRID
e-Science Centre of Excellence

Provenance-Aware Fault Tolerance for Grid Computing

Professor Jie Xu (*jxu@comp.leeds.ac.uk*)

Director of the WRG e-Science Centre of Excellence

University of Leeds & University of Newcastle upon Tyne, UK

THE WHITE ROSE GRID
e-Science Centre of Excellence

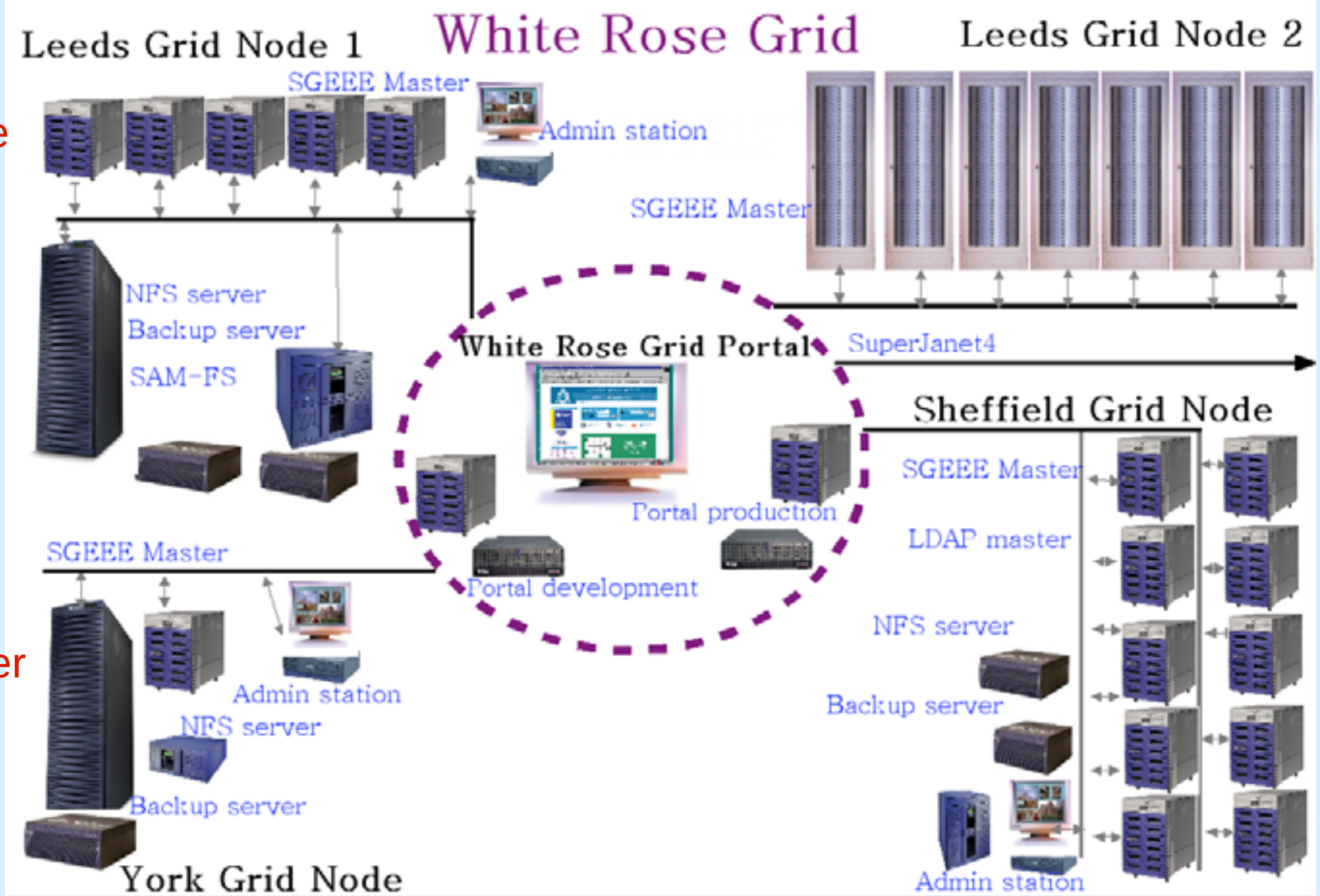
The White Rose Grid Project

- The **three Yorkshire Universities' project** (started in 2001, over £10M investment and research projects) http://www.wrgrid.org.uk/
- **Involves** Leeds (Profs K Brodlie, P M Dew & **J Xu**), York (Prof J Austin), and Sheffield (Profs G Tomlinson & P Fleming); under the guidance of the Chief Executive of WRUC (Dr Julian White – CEO of WRUC)
- White Rose University Consortium – a strategic partnership of the three Universities - http://www.whiterose.ac.uk
- Excellent partnership with **Computing Services** & Comp Science (Dr S Chidlow, C Cartledge, Dr A Turner)
- **Partners:** Esteem Systems in conjunction with Sun Microsystems & Streamline Computing
- Supported by Yorkshire Forward, Y&H Reg Dev Agency

THE WHITE ROSE GRID
e-Science Centre of Excellence

The WRG Architecture

General Purpose HPC node



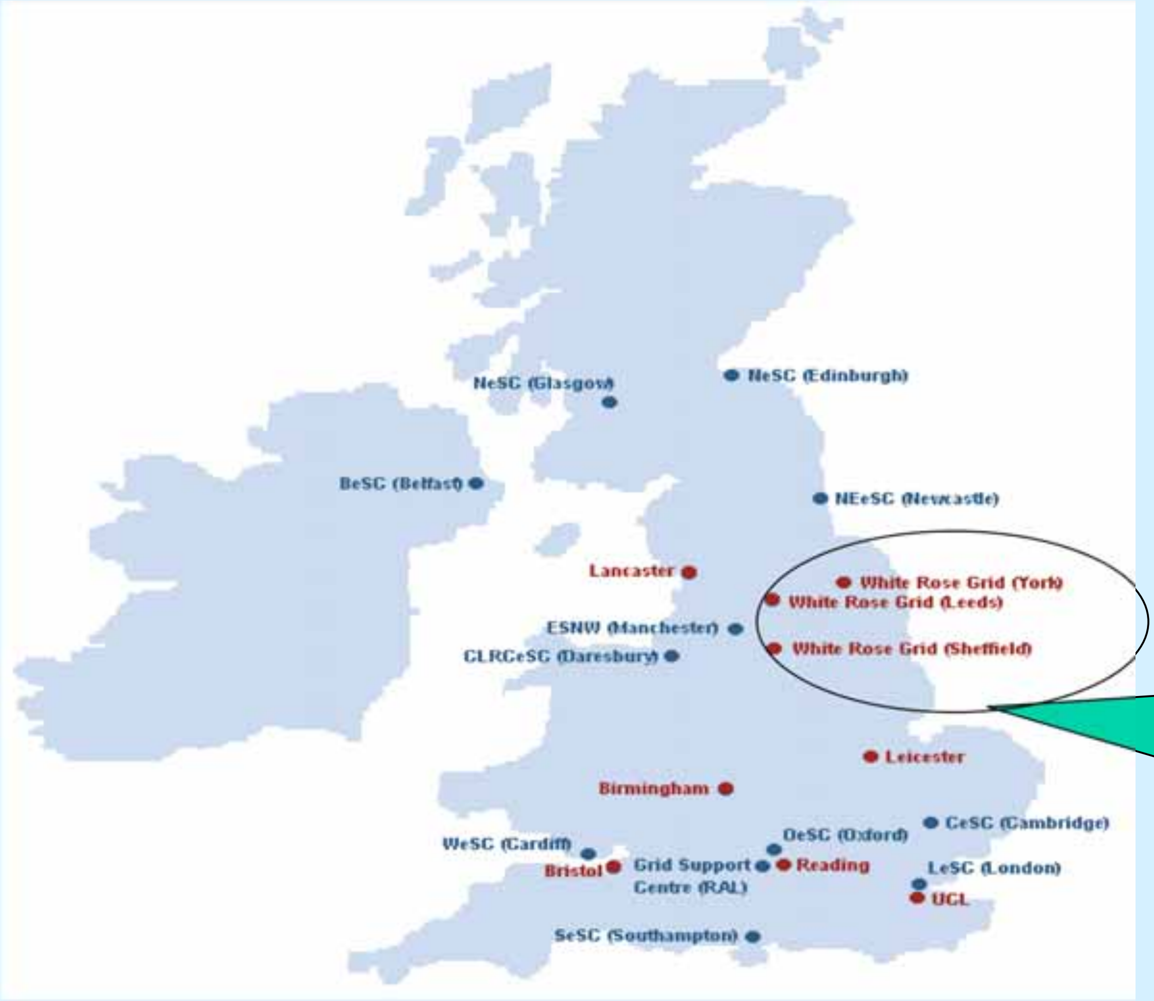
CFD node

Computer Science node

Engineering Application node

THE WHITE ROSE GRID
e-Science Centre of Excellence

UK e-Science Centres



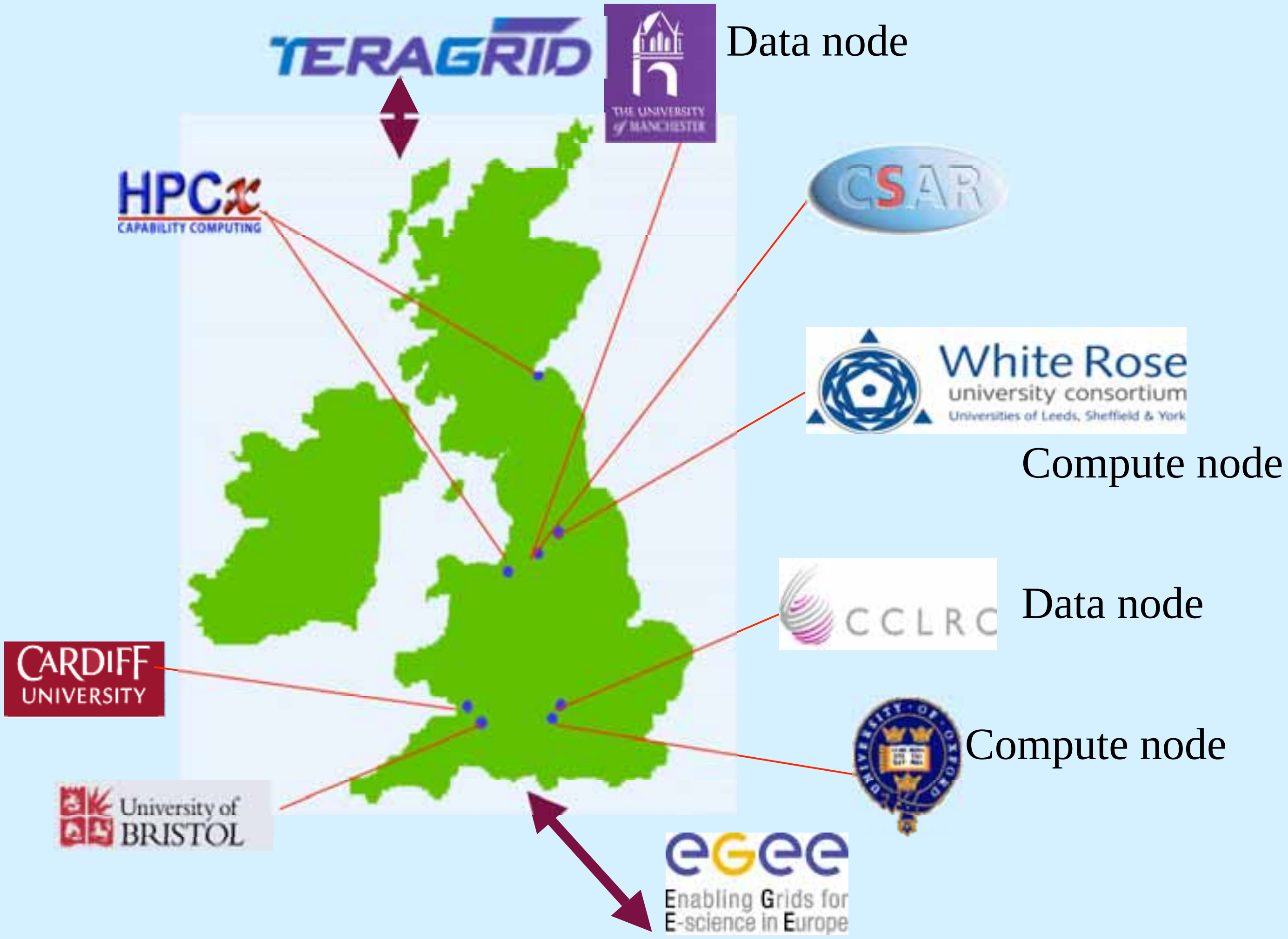
Our Centre:

- To offer focus for a variety of e-science issues and activities in our region
- To develop close links with the UK e-Science CP
- To develop a particular specialism: visualisation, distributed diagnostics and system dependability

The White Rose Grid e-Science Centre of Excellence

UK e-Science Centres (courtesy of NeSC)

National Grid Service



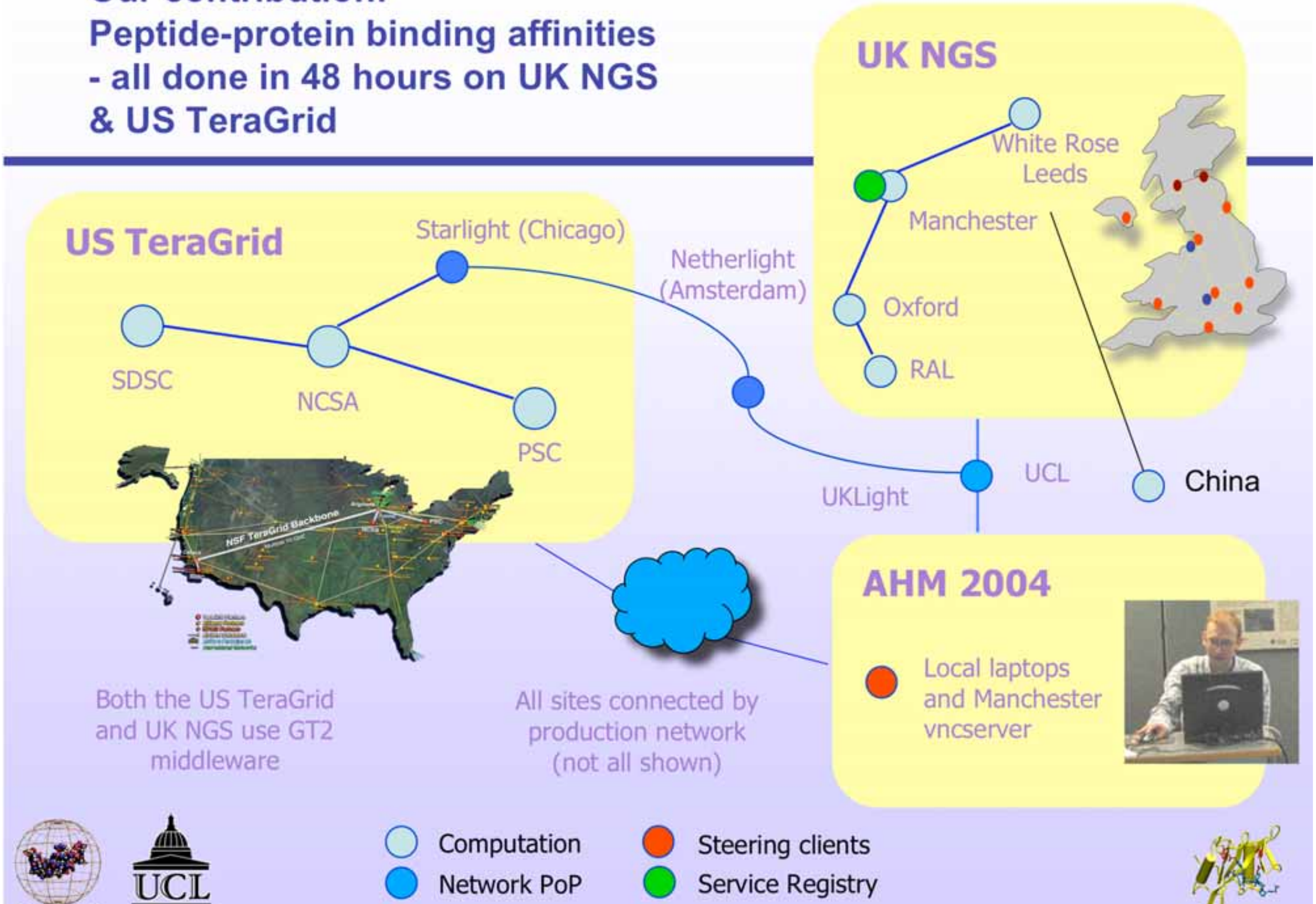
Data node

Compute node

Data node

Compute node

Our contribution: Peptide-protein binding affinities - all done in 48 hours on UK NGS & US TeraGrid



Both the US TeraGrid and UK NGS use GT2 middleware

All sites connected by production network (not all shown)

Local laptops and Manchester vncserver

- ○ Computation
- ● Network PoP
- 🟥 Steering clients
- 🟩 Service Registry

The ‘Shared Service’ Problem (1)

- A potential approach for achieving fault tolerance in a Grid/Web services environment is to invoke multiple functionally-equivalent services and to act upon the results returned from them, e.g. by comparison or voting.

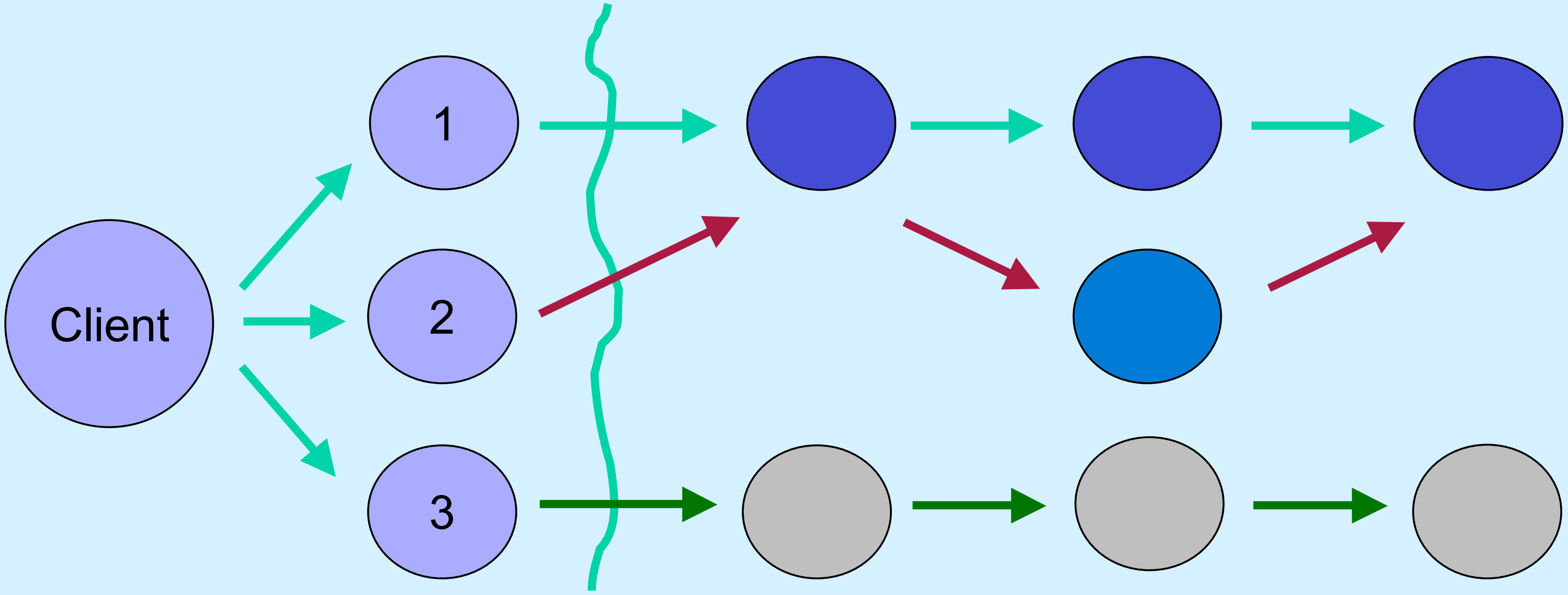
- A problem for this fault tolerance approach, however, *is* that in most SOA models, the implementational details of a service are hidden from a client of the service.

- The only information available to a client is the service’s interface and – possibly – some QoS metadata.

- This is an issue as services that initially appear disparate may – during the course of their execution – invoke one or more identical, “shared” services.

The 'Shared Service' Problem (2)

- The result is that different services may use the same shared services behind the scenes, which may make common mode failure (CMF) much more likely.



A Solution to This Problem

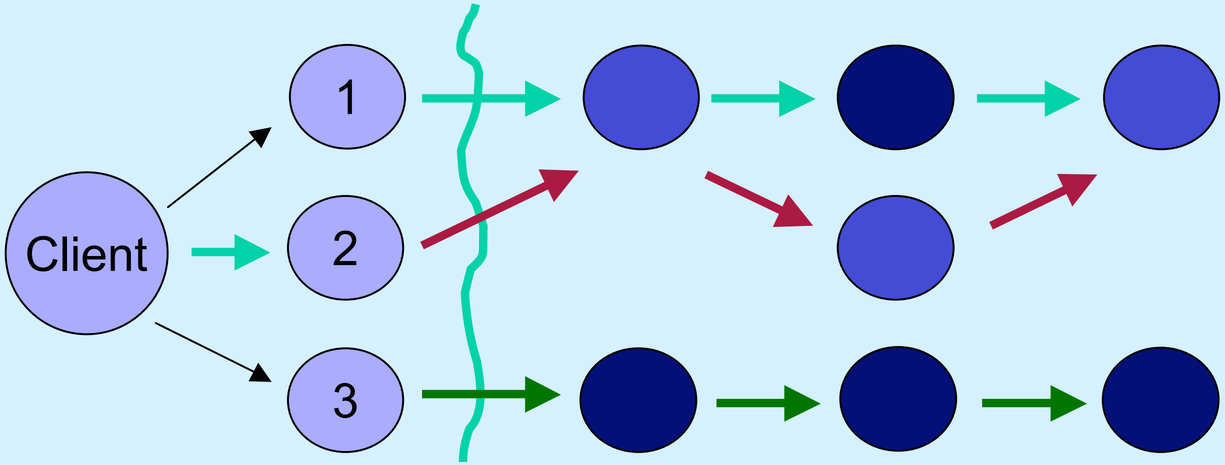
- One possible way of resolving this problem is to incorporate the technique of **provenance** in the fault tolerance approach used.
- Provenance is the documentation of the process that leads to a result.
- If we assume that data provenance is recorded, it will allow a fault tolerance scheme to build up a “view” of how each result it receives has been constructed.
- By possessing this view, a number of actions can be taken upon the results returned, e.g. weightings can be assigned to each service based upon how closely related it is to another service; services that have many common-dependencies can therefore have less “sway” in the voting algorithm used.

Weighted Voting

- In this “view”, s1 and s2 have 2 common dependencies, whilst s3 has no common dependencies.

- As a trivial example, we could therefore assign weightings of 0.5 to s1 and s2, and 1.0 to s3.

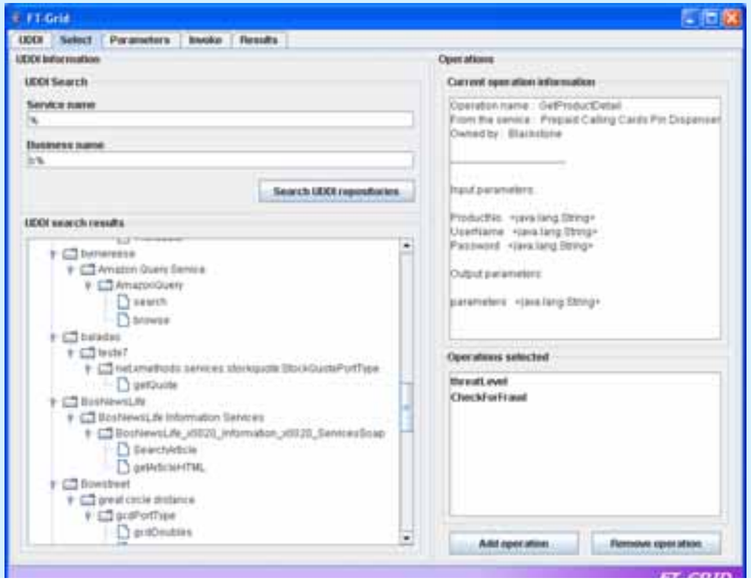
- In this case, should s1 and s2 agree, but s3 disagree with a result, then no overall “trusted” result will emerge.



FT-Grid: A framework for achieving fault tolerance

- We have implemented a java-based framework that facilitates the creation of fault tolerance schemes based on diverse services. This is called ***FT-Grid*.**

- The current implementation consists of both an API allowing developers to easily search for, invoke, and vote on services at run-time, and also a GUI to demonstrate the system.



UDDI information

UDDI Search

Service name: %

Business name: b%

UDDI search results

```
▾ bvnereese
  ▾ Amazon Query Service
    ▾ AmazonQuery
        search
        browse
▾ baladas
  ▾ teste7
    ▾ net.xmethods.services.stockquote.StockQuotePortType
        getQuote
▾ BosNewsLife
  ▾ BosNewsLife Information Services
    ▾ BosNewsLife_x0020_Information_x0020_ServicesSoap
        SearchArticle
        getArticleHTML
▾ Bowstreet
  ▾ great circle distance
    ▾ gcdPortType
        gcdDoubles
```

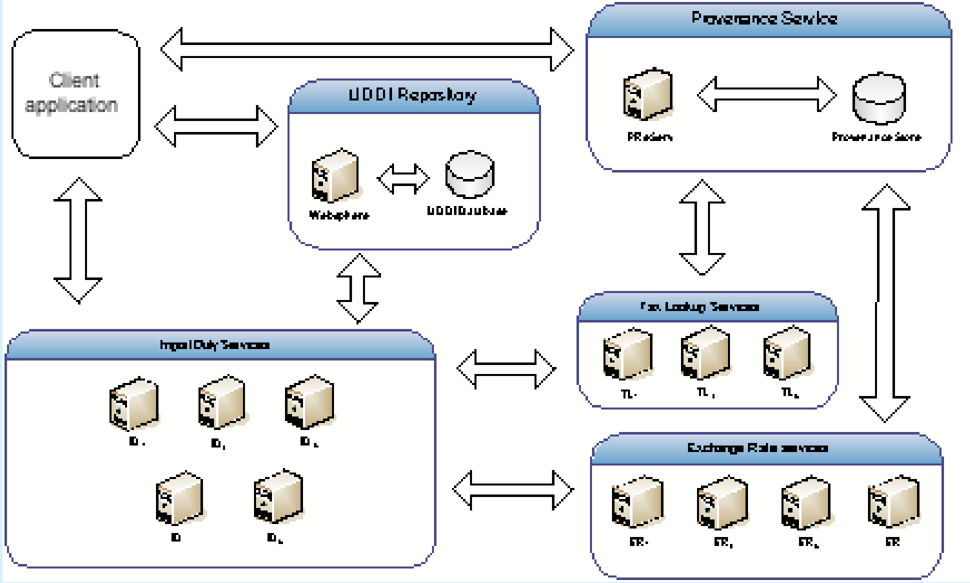
Operations

Current operation information

```
Operation name : GetProductDetail
From the service : Prepaid Calling Cards Pin Dispenser
Owned by : Blackstone

____________________

Input parameters:

ProductNo  <java.lang.String>
UserName   <java.lang.String>
Password   <java.lang.String>

Output parameters:

parameters  <java.lang.String>
```

Operations selected

threatLevel
CheckForFraud



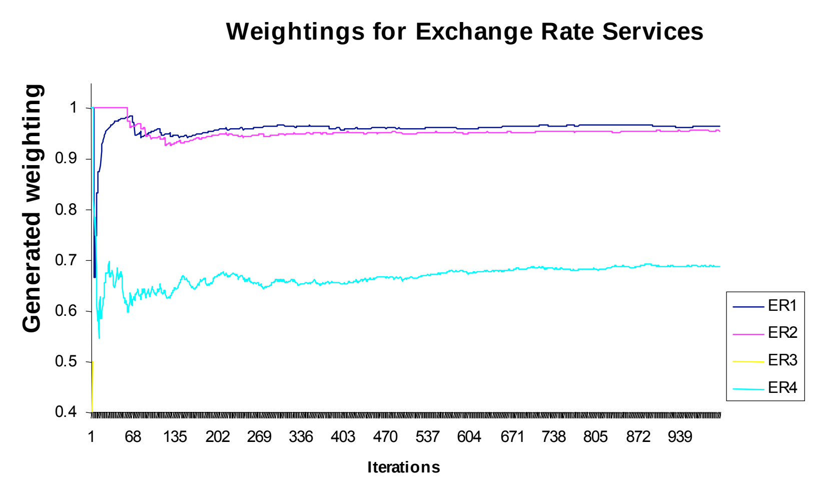
Comparison of Three Schemes

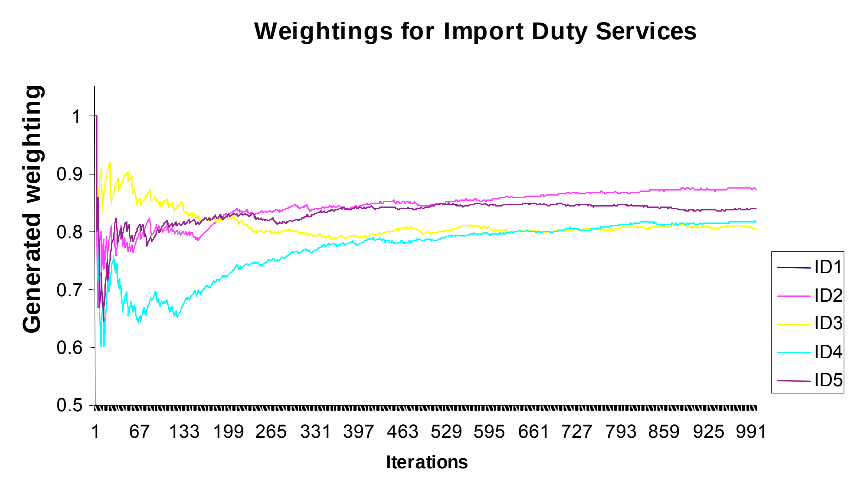
- Using FT-Grid, we developed a system that built up weightings based on the historical results of each service (the frequency with which a service's results agreed with the consensus). 15 Web services were involved and a Grid provenance system, called PASOA (developed at Southampton), was employed.
- We developed three systems in total:
  - A system without fault tolerance
  - A 'traditional' MVS system
  - A provenance-aware MVS system
- The traditional MVS system discarded results from services that had a weighting below a user-specified value, whilst the provenance-aware scheme discarded results where any service in a workflow fell below a user-specified value.
- This experiment yielded a large set of empirical data, and stress-tested both FT-Grid and the underlying infrastructure.

Some Experimental Results

- We performed 3 runs of 1000 tests on each scheme:

	Correct result	No result	CMF
Experiment 1 Run 1	828	172	-
Experiment 1 Run 2	858	142	-
Experiment 1 Run 3	822	178	-
Average	**836**	**164**	**-**
Experiment 2 Run 1	928	9	63
Experiment 2 Run 2	921	14	65
Experiment 2 Run 3	921	7	72
Average	**923.33**	**10**	**66.66**
Experiment 3 Run 1	996	4	0
Experiment 3 Run 2	990	10	0
Experiment 3 Run 3	996	4	0
Average	**994**	**6**	**0**





Brief Result Analysis

- The scheme without fault tolerance obtained a correct result in **83.6%** of all tests performed.
- The traditional MVS scheme obtained a correct result in **92.3%** of all tests performed, and a common-mode failure (CMF) occurred in **6.6%** of results.
- The provenance-aware MVS scheme obtained a correct result in **99.4%** of tests performed, and had no CMF.
- These results are encouraging, but it must be remembered that the test scenario was very simple, and in a more realistic environment (with more reliable services), the advantage of the provenance-aware scheme is likely to be reduced.
- We are making progress…

THE WHITE ROSE GRID
e-Science Centre of Excellence

Questions?

A new Programming Model for Dependable Adaptive Real-Time Applications

Presented by
António Casimiro

48th Meeting of IFIP Working Group 10.4
Hakone, Japan, July 1-5, 2005

Context

- Work developed in CORTEX, in which the concept of sentient objects was introduced
    - Autonomous entities with sentience (e.g. robots)
    - Geographical dispersion
    - Real-time & safety requirements
    - Availability
- Several issues addressed in CORTEX
    - Programming model for sentient applications
    - Interaction model
    - WAN-of-CANs architecture (systems-of-systems)

Dealing with uncertainty

- We defined a generic approach to reconcile uncertainty with the need for predictability
- This could be (and was) applied in CORTEX, for sentient applications
- Make the application behave [safely, timely, securely, etc] in the measure of what can be expected from the environment
- Provide guarantees in the way that is done

→ Dependable adaptation

Back to the roots

- Initial idea proposed in 1999
  - Formal definition of the relevant properties:
    - No-contamination
    - Coverage stability
  - Definition of approaches for dependable application programming:
    - Fail-safe approach (fail-safe applications)
    - Reconfiguration & adaptation (time-elastic, t-safe apps)
    - Replication

Meanwhile…

During the course of CORTEX

Programming principle

- General and systematic approach:
  - QoS coverage service
    - The user simply provides the needed coverage
    - The service indicates the bound that must be used
    - For applications with time-safety and time-elasticity
  - Timing failure detection service
    - The user provides a bound for some action
    - The service will execute an handler upon failure detection

Making it dependable

- To adapt the QoS it is necessary to:
  - monitor the actual QoS being provided
  - decide if adaptation is necessary

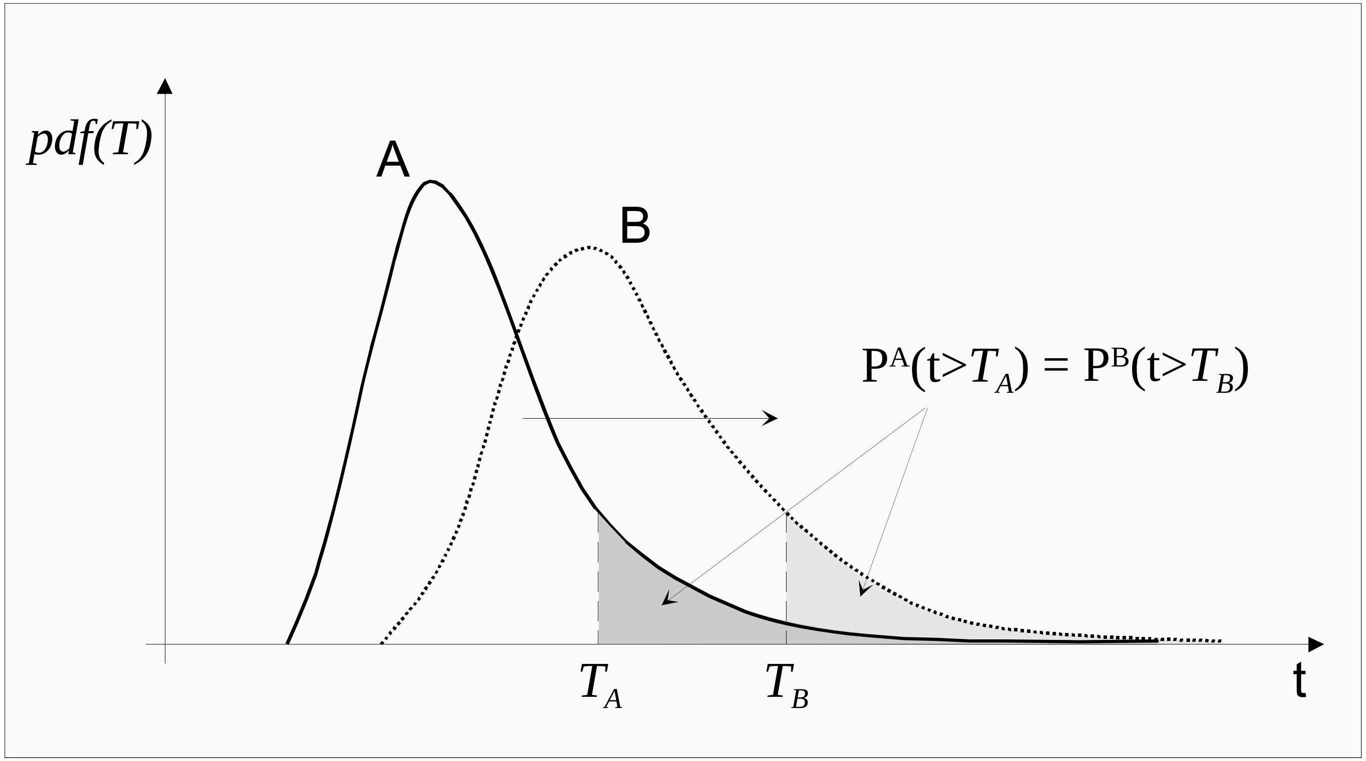
- To dependably adapt the QoS we must:
  - observe the environment in a dependable way
  - apply a rigorous strategy to decide when and how to adapt

Dependable adaptation

- First, it is necessary to trust the service that provides the measurements (durations)
  - in the value domain (correct measurements)…
  - …and in the time domain (timely measurements)

Dependable adaptation

- Then, decide when and how to adapt

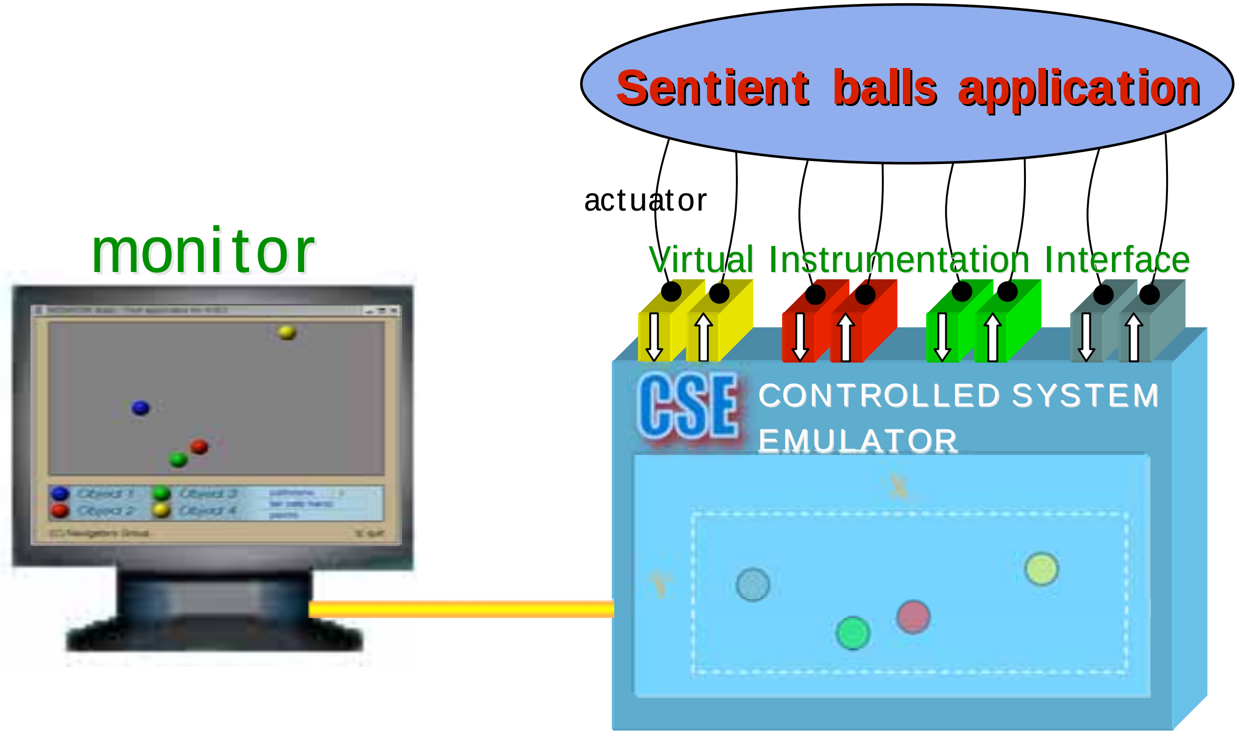


Finally…

We applied the programming model

Sentient balls application

- Physical environment is emulated


Sentient balls application
actuator
Virtual Instrumentation Interface
monitor
CSE
CONTROLLED SYSTEM EMULATOR

Emulator

- Emulated environment: four entities shaped as colored balls move in a space with a certain speed and direction
- A Virtual Instrumentation Interface allows to:
  - acquire ball positions, directions and speeds;
  - change ball movement (speed and direction)
- The sentient application (ball controllers) uses the TCB for the underlying services:
  - QoS Adaptation
  - Timing Failure Detection

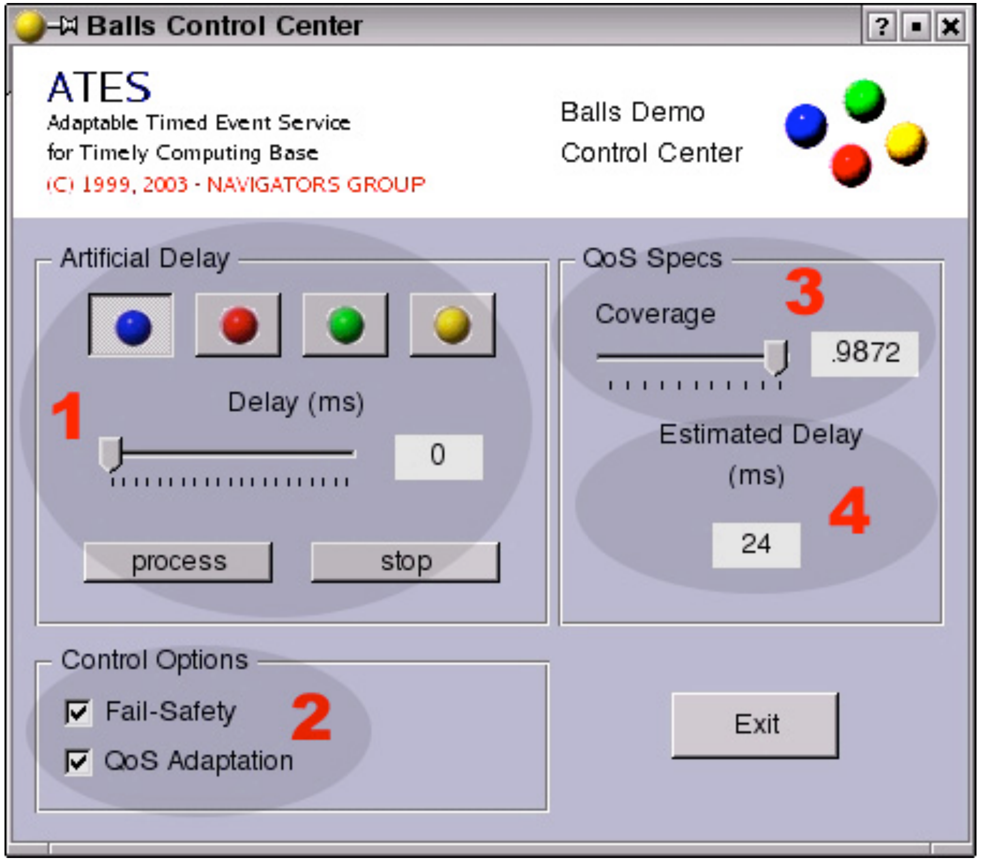
Fail-Safety Demo

- When Fail-Safety is ON:
  - Delivery delay of events is controlled using the TCB distributed TFD
  - Timing failure detected ➔ stop balls in timely way
- When Fail-Safety is OFF:
  - Timing failures can cause balls to crash!

QoS-Adaptation Demo

- When QoS-Adaptation is ON:
  - The service indicates the estimated delay that corresponds to requested coverage value
  - This value is used to determine and set ball speed that preserves safety
  - Coverage stability is achieved
- When QoS-Adaptation is OFF:
  - No speed adaptation takes place
  - Assumed delay keeps constant, possibly leading to coverage degradation due to timing failures

A small taste of it…



Balls Control Center

ATES

Adaptable Timed Event Service
for Timely Computing Base

(C) 1999, 2003 - NAVIGATORS GROUP

Balls Demo
Control Center

Artificial Delay

Delay (ms) 0

QoS Specs

Coverage .9872

Estimated Delay (ms) 24

Control Options

☑ Fail-Safety

☑ QoS Adaptation

Where is the paper?

- MAIN FEATURE of May 2005 issue of IEEE Distributed Systems On-Line Journal:
  - http://dsonline.computer.org
  - http://dsonline.computer.org/portal/site/dsonline/menuitem.9ed3d9924aeb0dcd82ccc6716bbe36ec/index.jsp?&pName=dso_level1&path=dsonline/0505&file=o5001.xml&xsl=article.xsl&
- *A New Programming Model for Dependable Adaptive Real-Time Applications*
  Pedro Martins, Paulo Sousa, António Casimiro, Paulo Veríssimo
  IEEE Distributed Systems Online, vol. 6, no. 5, 2005.

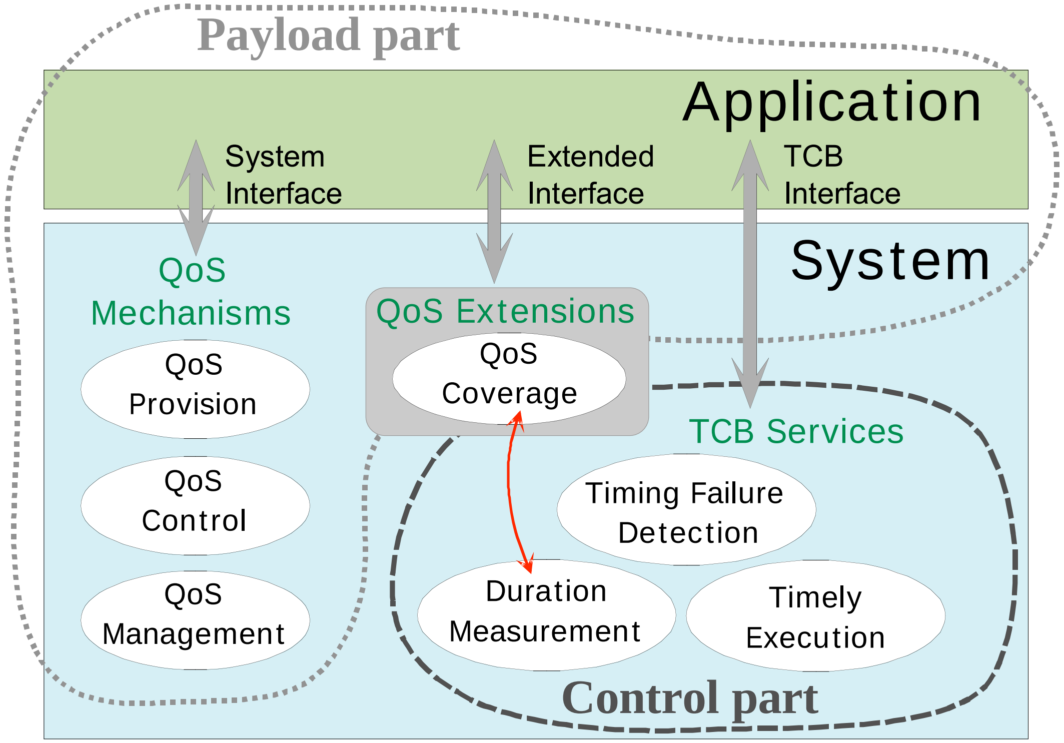
you may also get there from our web site, www.navigators.di.fc.ul.pt under "Recent Documents“.

…a small movie

Extra slides

QoS coverage service

In a system with a TCB



Implementation

- We use a known result from prob. theory:

$$\boxed{P(D>t) \leq \frac{V(D)}{V(D)+(t-E(D))^2},\ \text{for all } t > E(D)}$$

  - which allows the calculation of an upper bound for the probability of a time bound t being violated

- Given the coverage $C_{\min}$, t is obtained with:

$$\boxed{t = \frac{2E(D)+\sqrt{4E(D)^2-4(E(D)^2+V(D)-\frac{V(D)}{1-C_{\min}}})}{2}}$$

Implementation issues

- Estimation of Expected value and Variance
  - E(D) and V(D) correspond to the average and variance of a set of values obtained during an interval of mission
  - The size of the set depends on the application

- Contributing factors for accuracy loss:
  - Error associated to the measured durations
  - Error introduced by the estimation (finite number of samples)
  - Error that results from using an upper bound for the probability

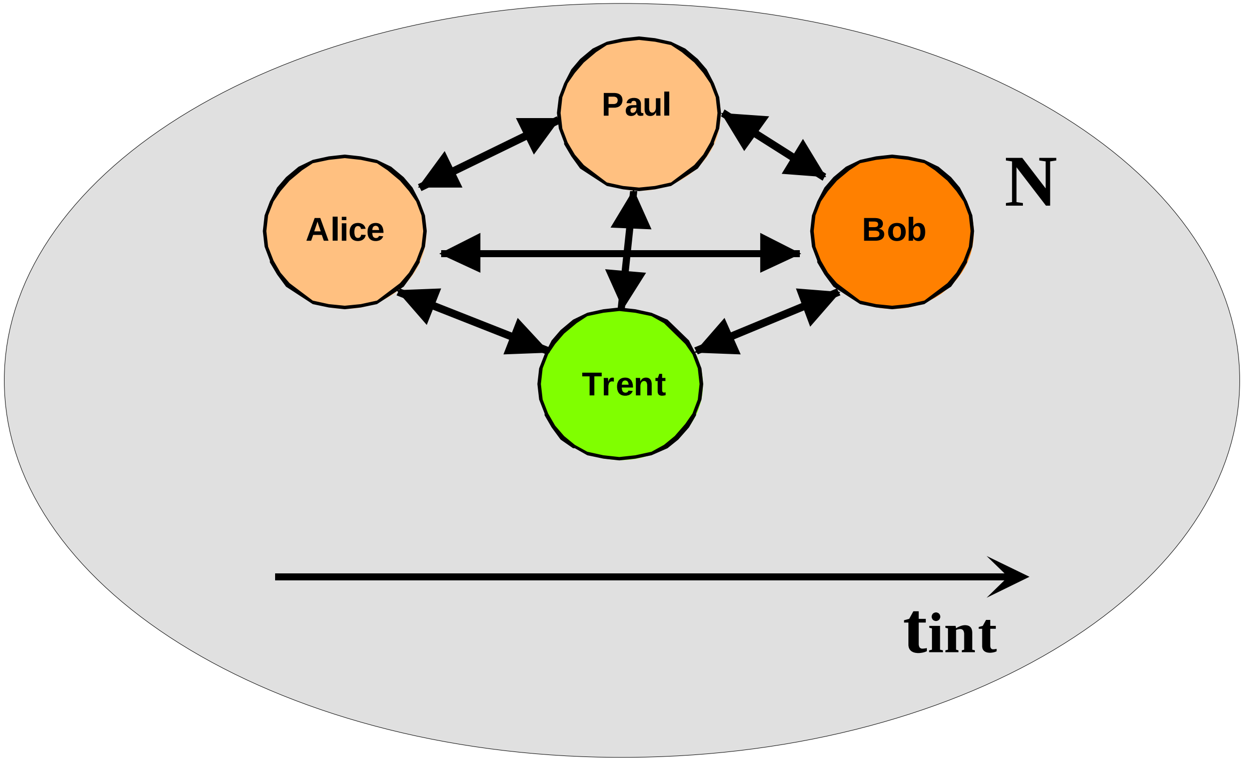
- Results can be improved by reducing errors:
  - Measure durations with smaller errors
  - Get rid of pessimistic assumptions (e.g. no recognition abilities)

FLP is back!
or
A forgotten dimension of time in distributed systems problems

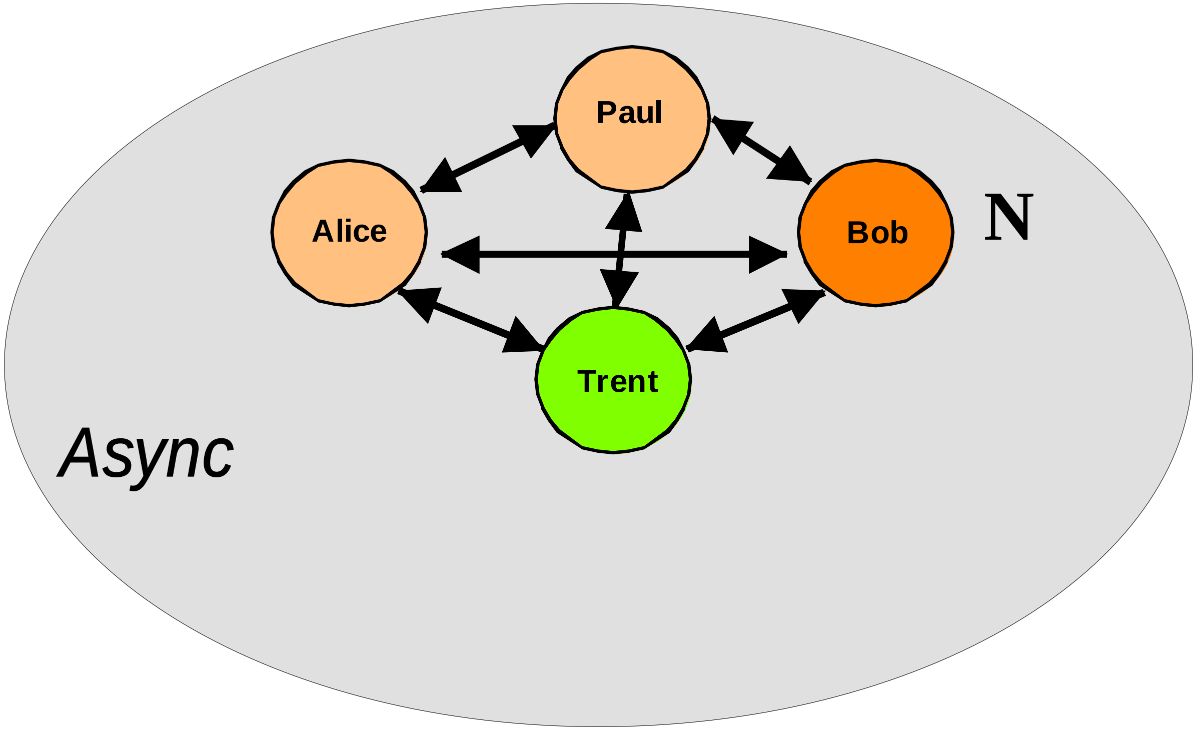
Paulo Esteves Veríssimo

Navigators Group,
LaSIGe, Laboratory for Large-Scale Informatic Systems
Univ. Lisboa
pjv@di.fc.ul.pt
http://www.di.fc.ul.pt/~pjv

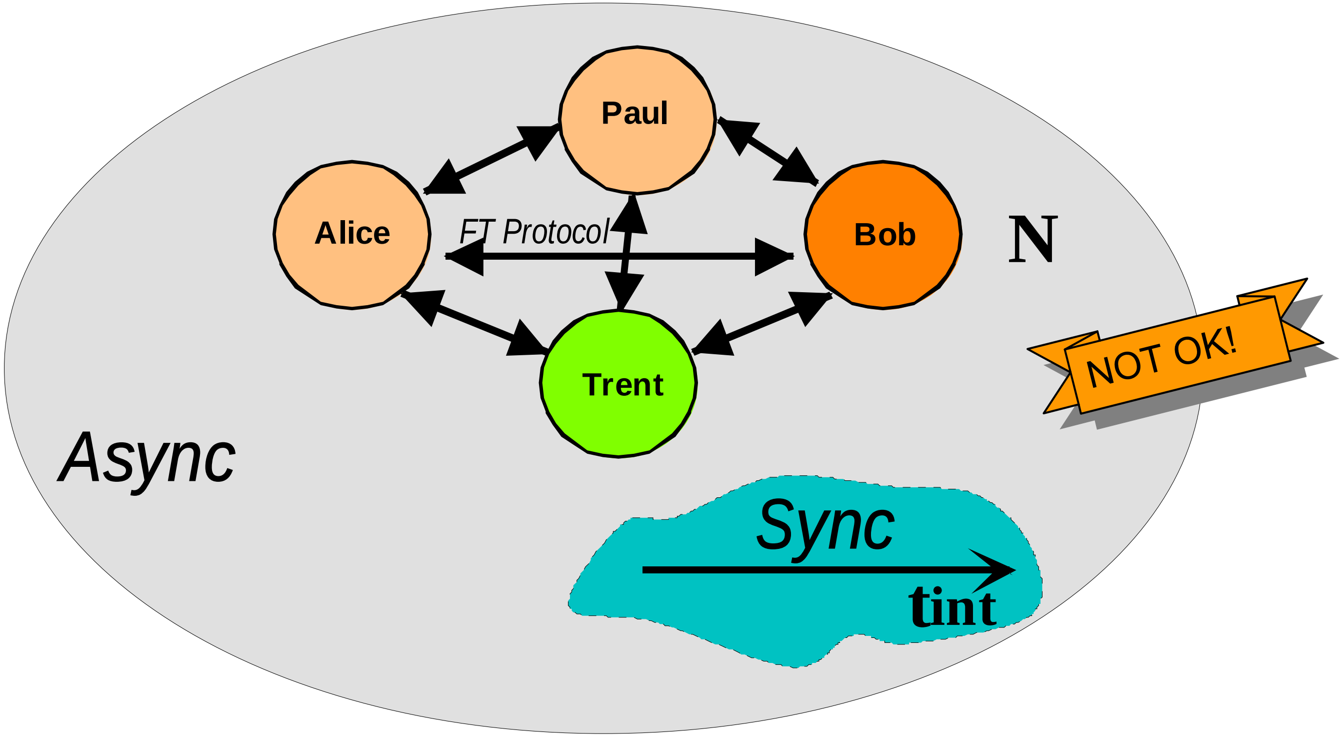
Classical Model



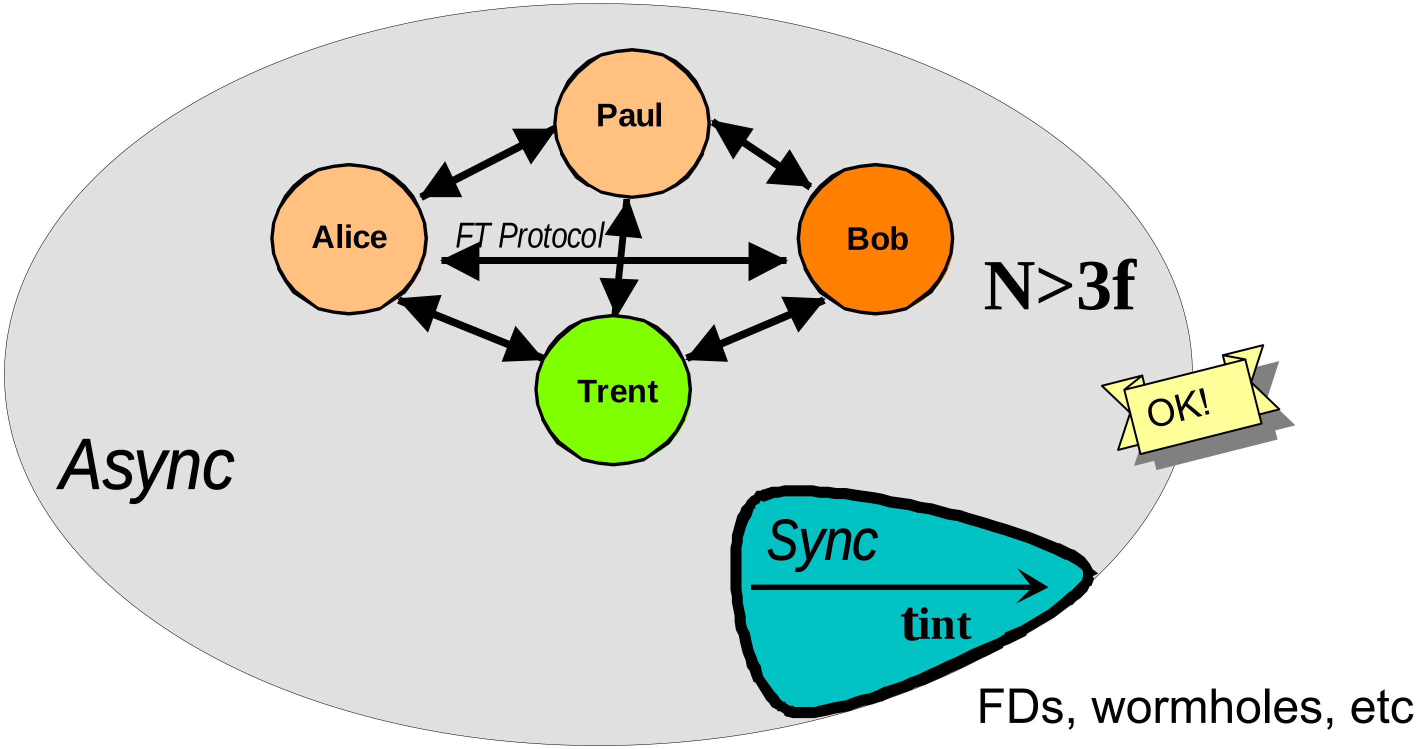
Classical Model - Async System



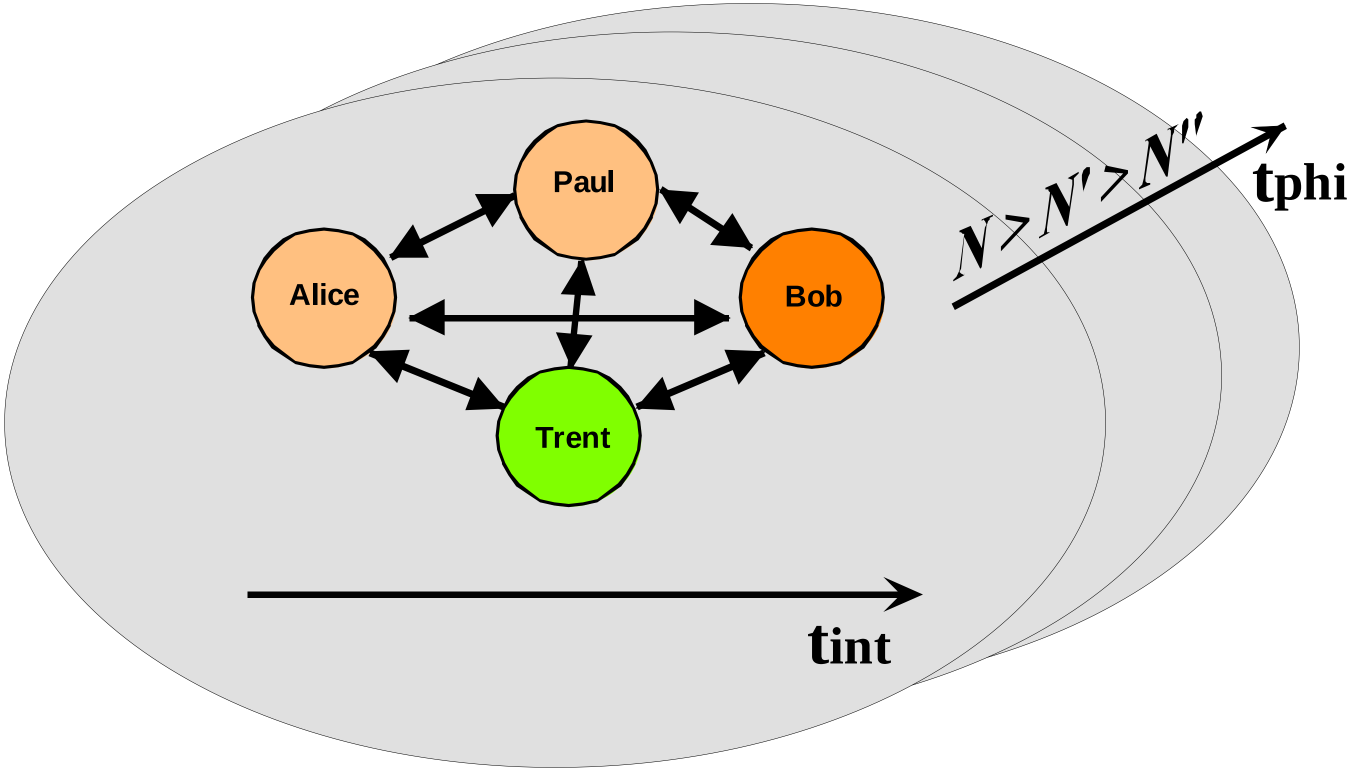
Classical Model - Async System with hidden sync assumptions



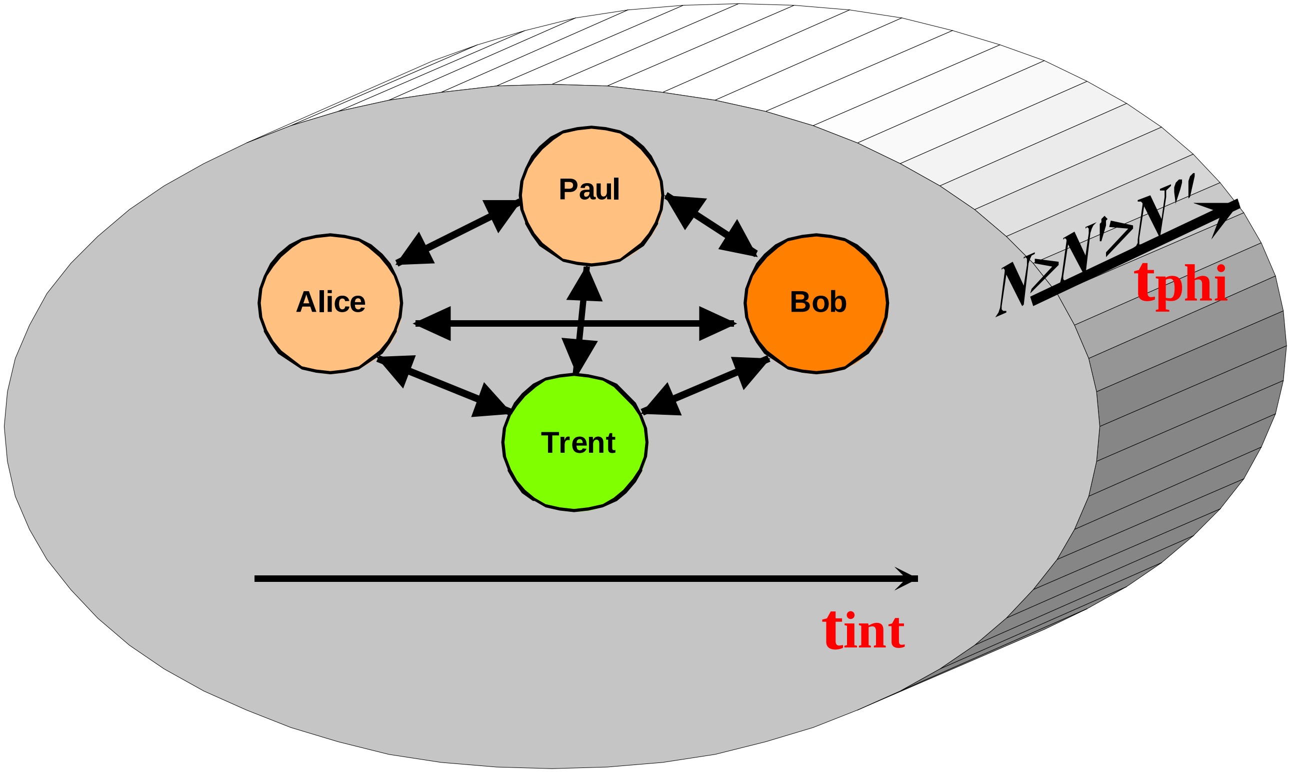
Classical Model - Correct FT Async system



Classical Model vs. Reality



Physical Model



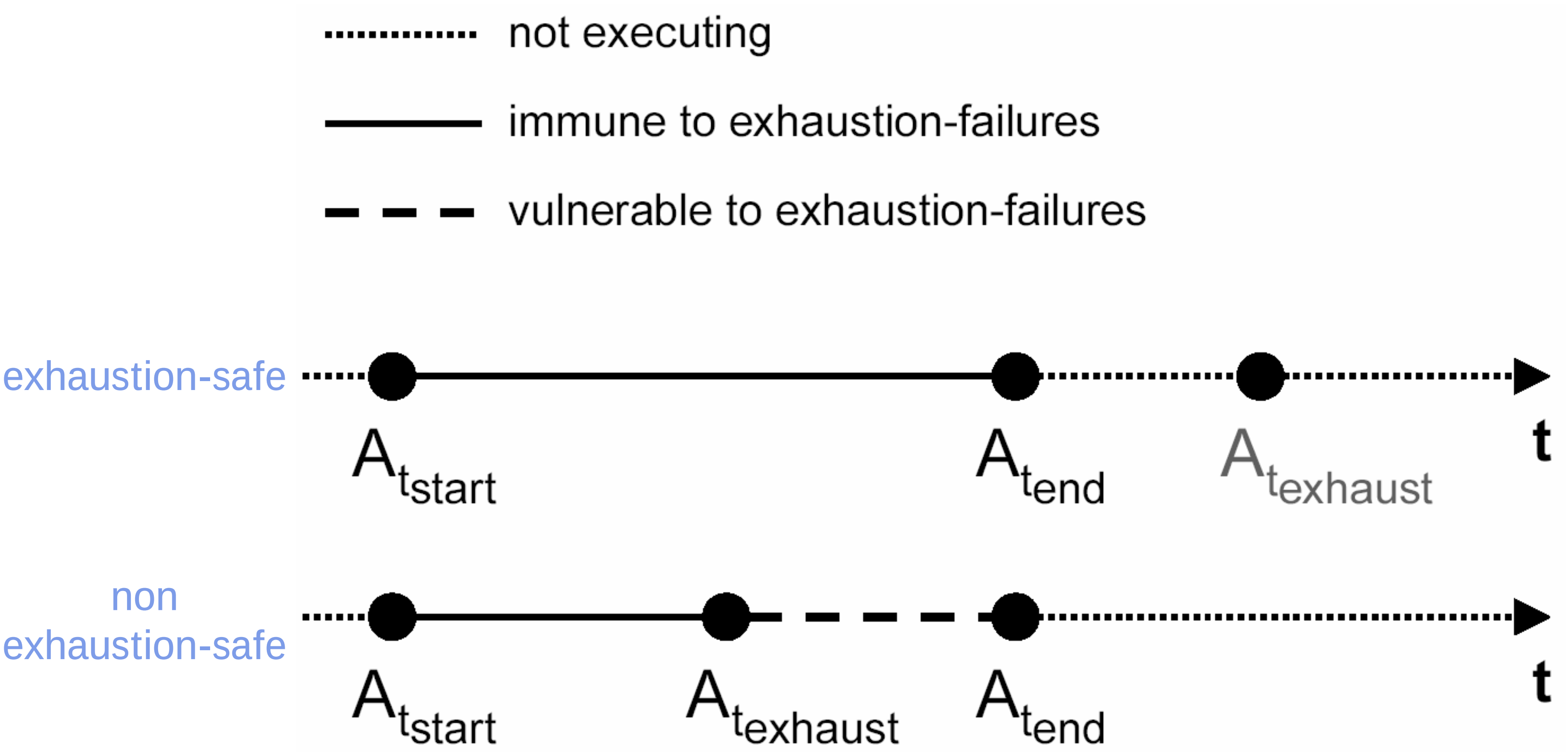
Focusing on Resources

- Fault and timing assumptions are an abstraction of the required resources.
  - e.g., f fault-tolerance means (n-f) correct nodes are required.
- Resource exhaustion: violation of a resource assumption.
  - e.g., f+1 nodes fail.
- Definition: An exhaustion-failure is a failure that results from resource exhaustion.
- Definition: A system is exhaustion-safe if it ensures that exhaustion-failures never happen.

Physical System Model (PSM)

- Allows to formally reason about how exhaustion-safety is affected by different combinations of timing and fault assumptions.
- A system execution is defined by
  - t_{start}: the RT start instant.
  - t_{end}: the RT termination instant.
  - $t_{exhaust}$: the RT instant when exhaustion occurs.
- Definition: A system is exhaustion-safe iff $t_{end} < t_{exhaust}$, for all executions.
  - e.g., a f fault-tolerant distributed system is exhaustion-safe if it terminates before f+1 failures being produced.

To Be or Not to Be Exhaustion-Safe



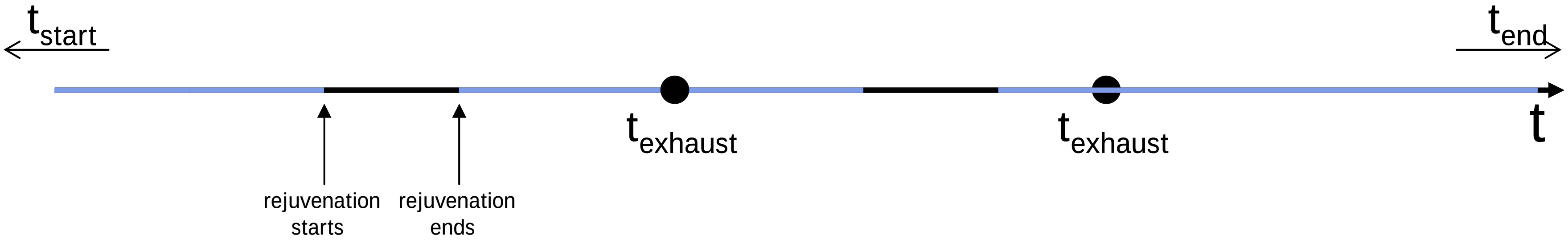
not executing

immune to exhaustion-failures

vulnerable to exhaustion-failures

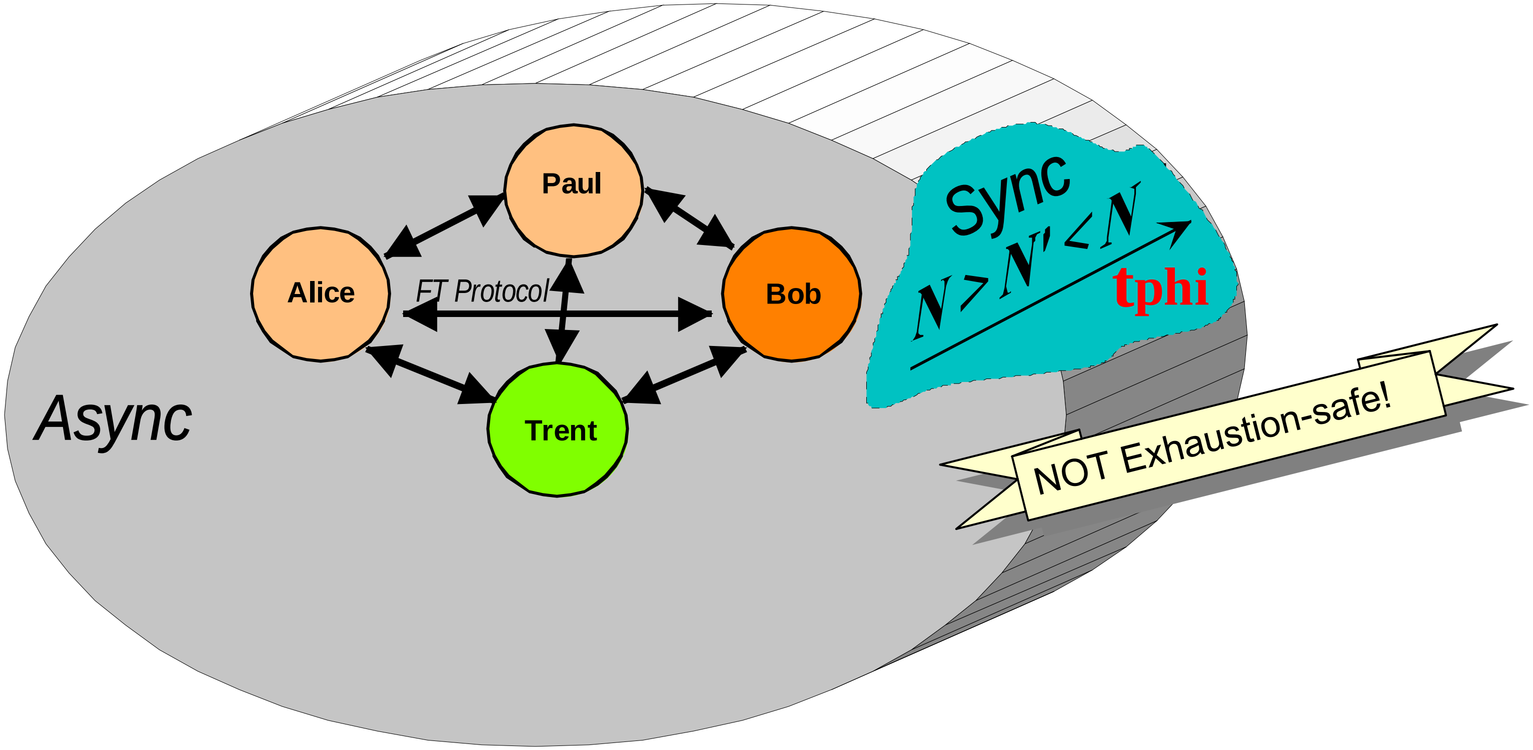
Proactive Recovery

- Goal: to constantly postpone $t_{exhaust}$ through periodic rejuvenation.
  - e.g., periodic rejuvenation of OS code .



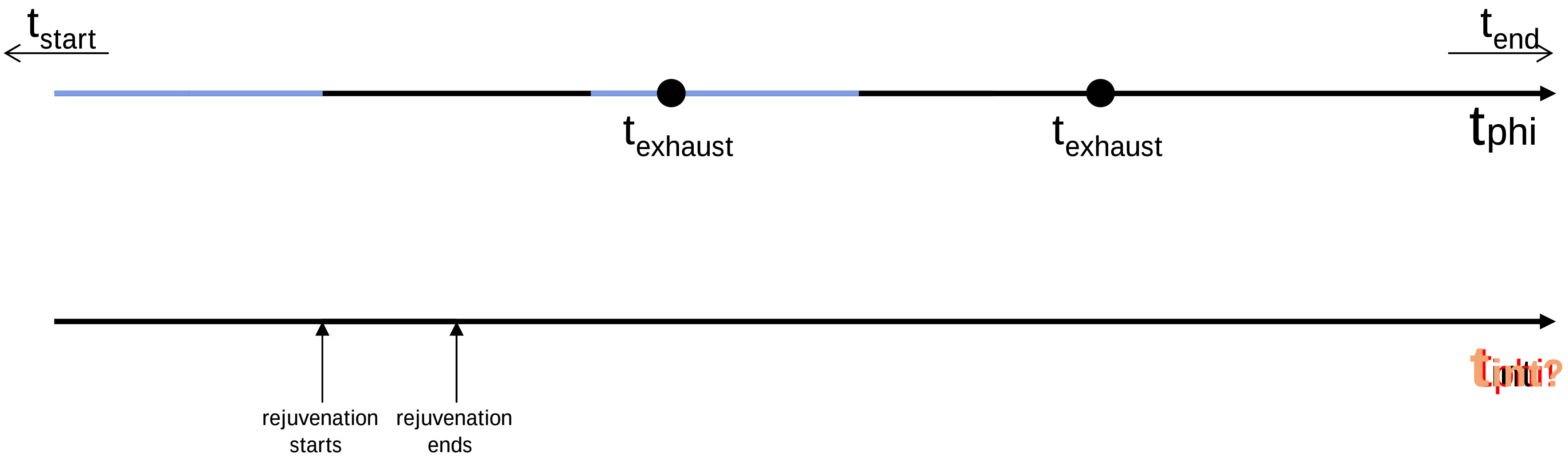
- A system is exhaustion-safe only if rejuvenations are always terminated before exhaustion.

Physical Model - Async system with hidden sync assumptions

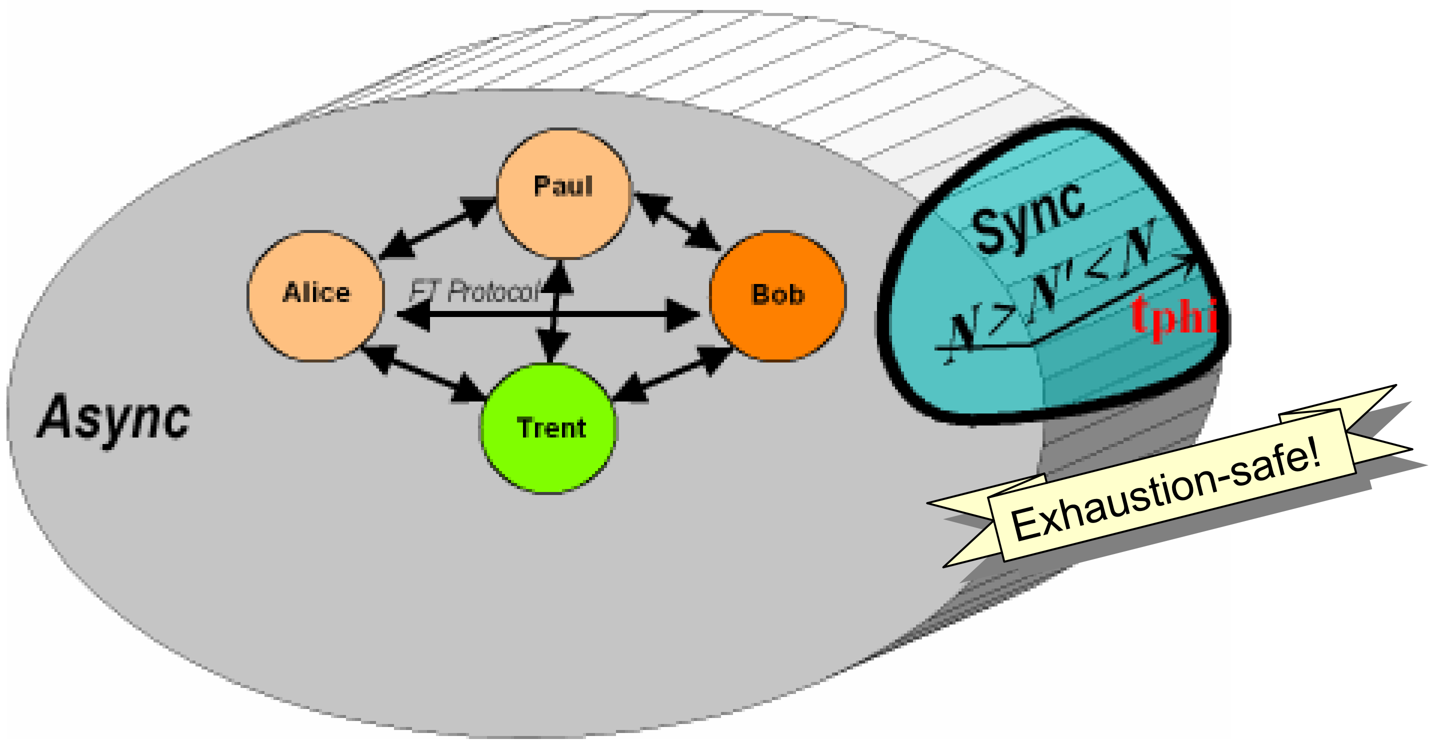


Proactive Recovery

- Goal: to constantly postpone $t_{exhaust}$ through periodic rejuvenation.
  - e.g., periodic rejuvenation of OS code .



Classical Model - Correct FT Async system



Conclusions

- Current state-of-the-art does not allow to construct exhaustion-safe distributed systems, specially in face of arbitrary faults:
  - Sync systems are vulnerable:
    - timing failures.
  - Async systems are vulnerable:
    - max number of faults + unbounded execution time.
  - Async systems with async proactive recovery are vulnerable:
    - max number of faults + unbounded rejuvenation period.

Future/Ongoing Work

- Combining proactive recovery and wormholes
  - Proactive recovery is useful to postpone $t_{exhaust}$ as long as it has timeliness guarantees.
  - Proposal: combine async payload system with sync proactive recovery subsystem.
  - See our recent tech report:
    - Proactive Resilience through Architectural Hybridization DI/FCUL TR 05-8, May 2005.
  - http://www.navigators.di.fc.ul.pt/

Human Expertise in Fault Detection and Adjustment

An Empirical Case Study

Rainer Knauf

Technical University of Ilmenau
School of Computer Science and Automation
Ilmenau, Germany

Setsuo Tsuruta

Tokyo Denki University
School of Information Environment
Tokyo, Japan

Avelino J.Gonzalez

University of Central Florida
Dept. of Electrical and Computer Engineering
Orlando, FL, USA

Content

1. System Evaluation and Refinement – An Issue of this WG?

2. Our Concept – An Overview

3. The Problem with Human Experience

🗐✍ Incorporating a Validation Knowledge Base (VKB) as a Model of Collective Experience

🗄✍ Incorporating Validation Expert Software Agents (VESA) as Models of Individual Experiences

6. A Prototype Test

   ➢ *Knowledge Base*

   ➢ *Test Cases*

   ➢ *Application Conditions*

7. Test Results

   ➢ *On the Usefulness of Modeling the experience*

   ➢ *Lessons Learnt*

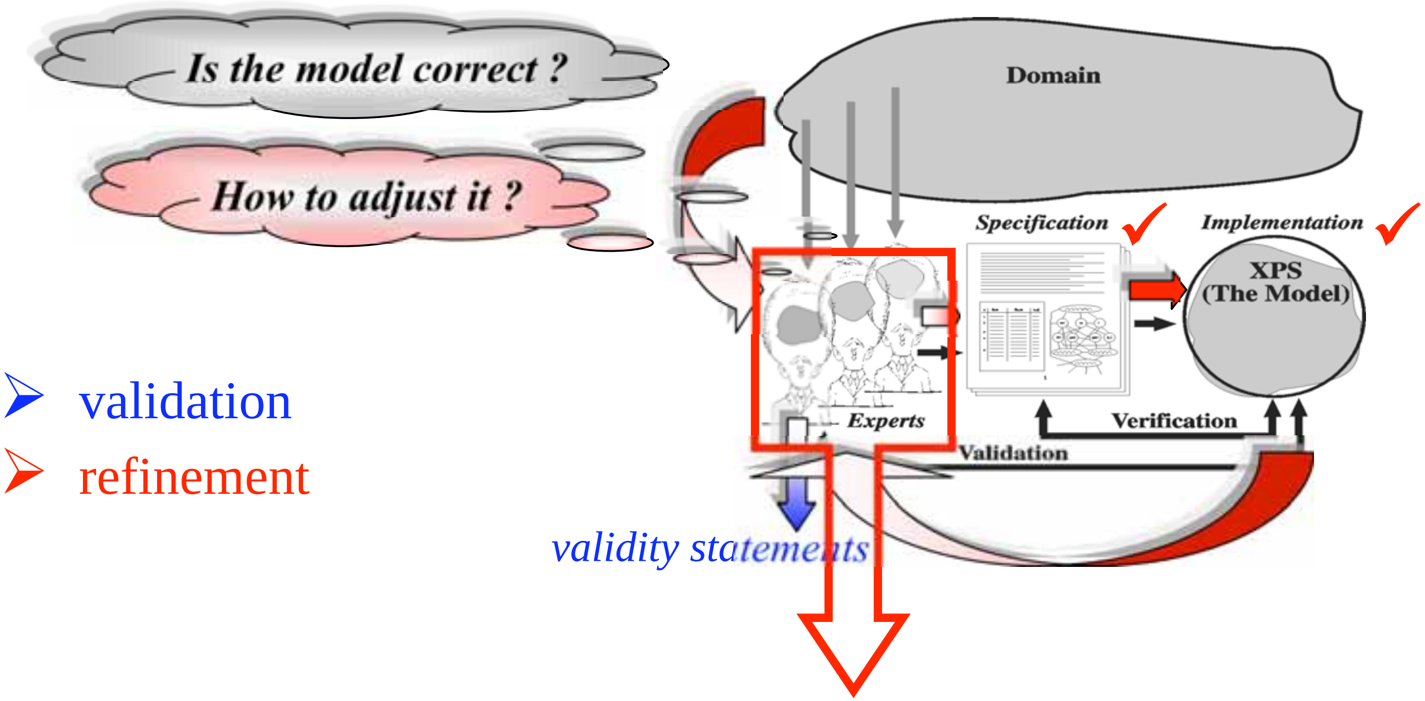
8. Summary and Conclusion

1. System Evaluation and Refinement – An Issue of this WG?

- ➢ *Today's opportunities to design and employ complex systems rise the question, whether or not we are able to control what we are able to build*
- ➢ *The impact of invalidity increases with the with the number today‘s systems‘ application fields and their sensibility to malfunctions*
- ➢ *Today‘s IT-systems may become a real threat without ensuring their validity*
- ➢ *Moreover, many interesting applications are characterized by some dynamics in their topical background.*
- ➢ *Thus, these systems need to be refined based on both, revealed invalidities and new topical insights.*

- ▪ In fact, these concerns are issues of **dependable computing**.
- ▪ Maybe they are not an issue of fault tolerance, but of **fault detection and adjustment** instead.

Verification, validation, and refinement – what‘s it?



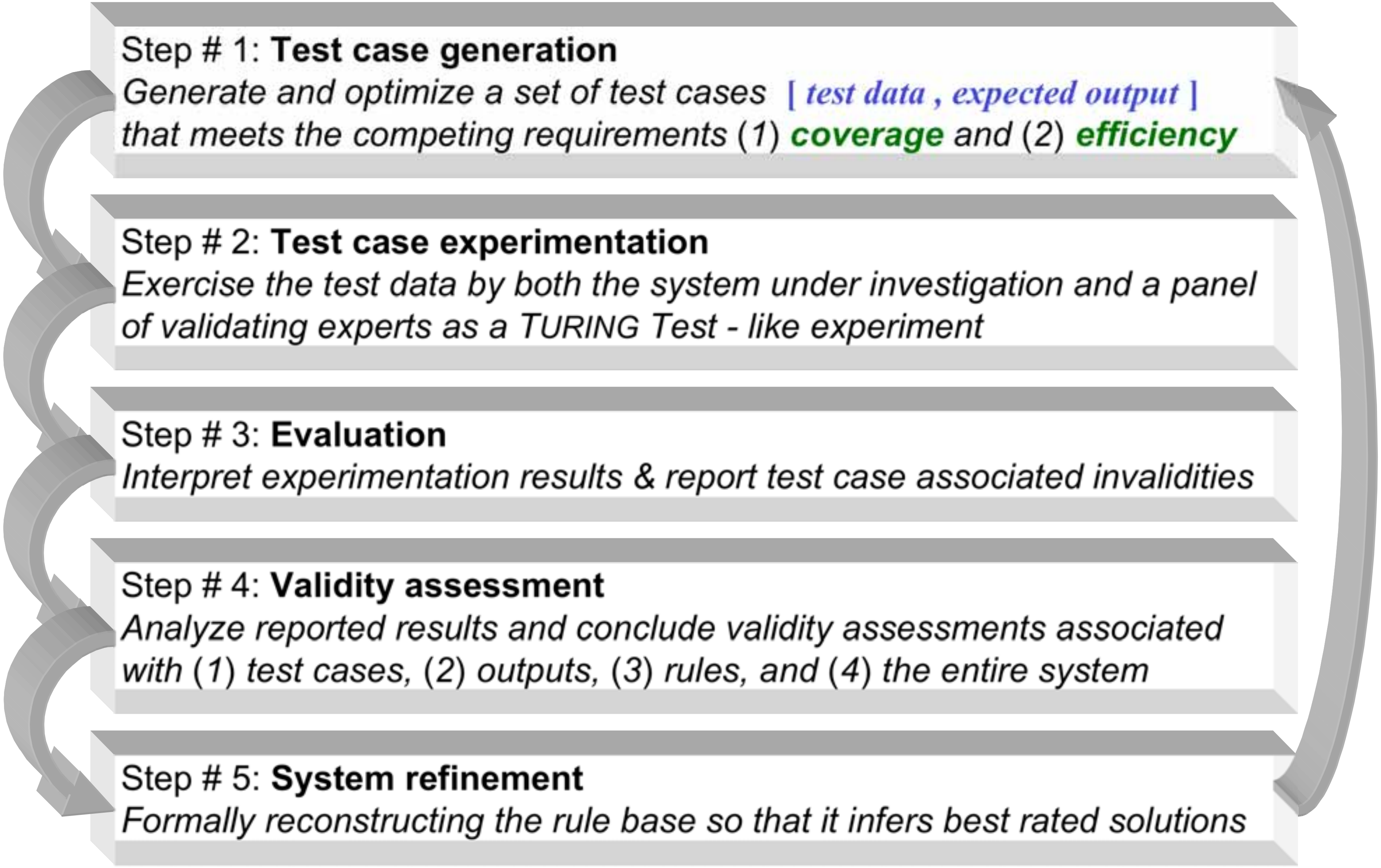
- validation
- refinement

Humans in the loop – a problem?

Yes, indeed!

But is there any alternative ?

2. Our Conceept – An Overview



Step # 1: **Test case generation**
Generate and optimize a set of test cases ***[test data , expected output]*** *that meets the competing requirements* (1) ***coverage*** *and* (2) ***efficiency***

Step # 2: **Test case experimentation**
Exercise the test data by both the system under investigation and a panel of validating experts as a TURING Test - like experiment

Step # 3: **Evaluation**
Interpret experimentation results & report test case associated invalidities

Step # 4: **Validity assessment**
Analyze reported results and conclude validity assessments associated with (1) *test cases,* (2) *outputs,* (3) *rules, and* (4) *the entire system*

Step # 5: **System refinement**
Formally reconstructing the rule base so that it infers best rated solutions

3 The Problem with Human Experience

What‘s the problem with employing human expertise for system validation?

☹ Experts have different beliefs, experiences and learning capabilities.

☹ Experts are not free of mistakes.

☹ Experts’ opinions about the desired system’s behavior

- differ from each other
- change over time as a result of misinterpretations, mistakes or new insights

☹ Experts are often too busy and/or too expensive to hire them for system validation and refinement.

How to get out of this misery ?

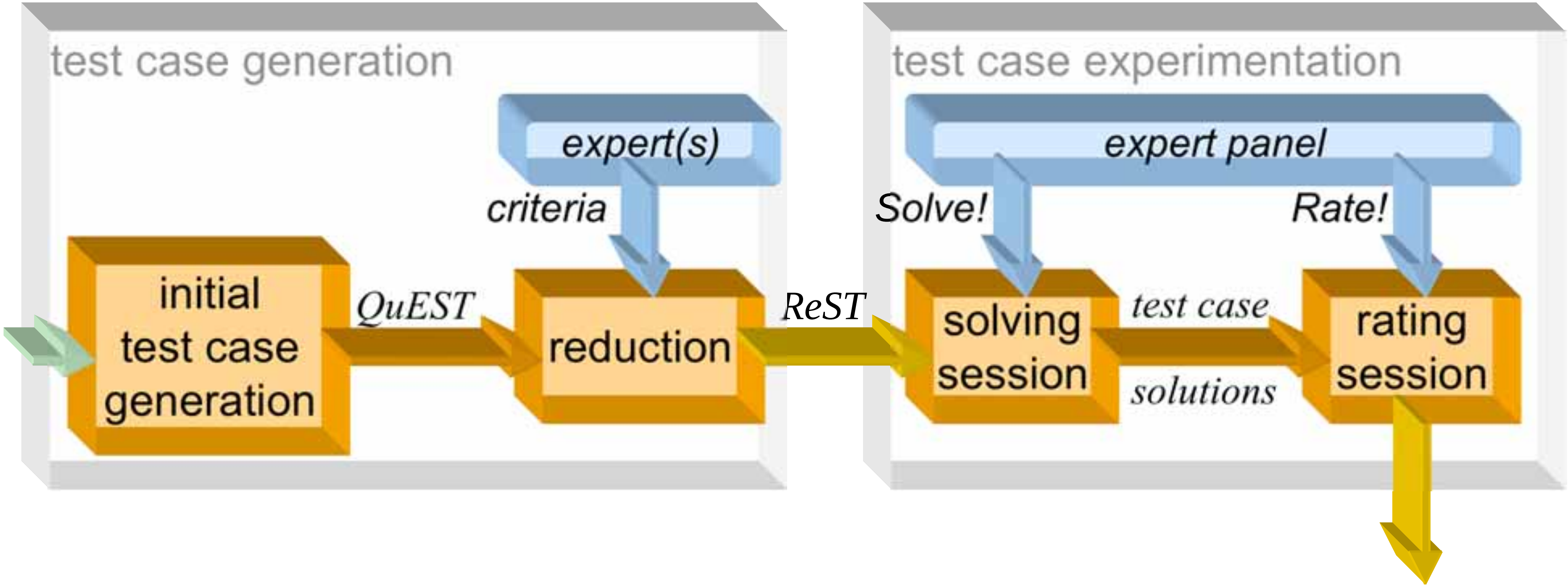
By

(1) **modeling their experience**

(2) **compensating some human weaknesses with this model**

The Involvement of Humans so far

Where is the human input into our validation technology ?



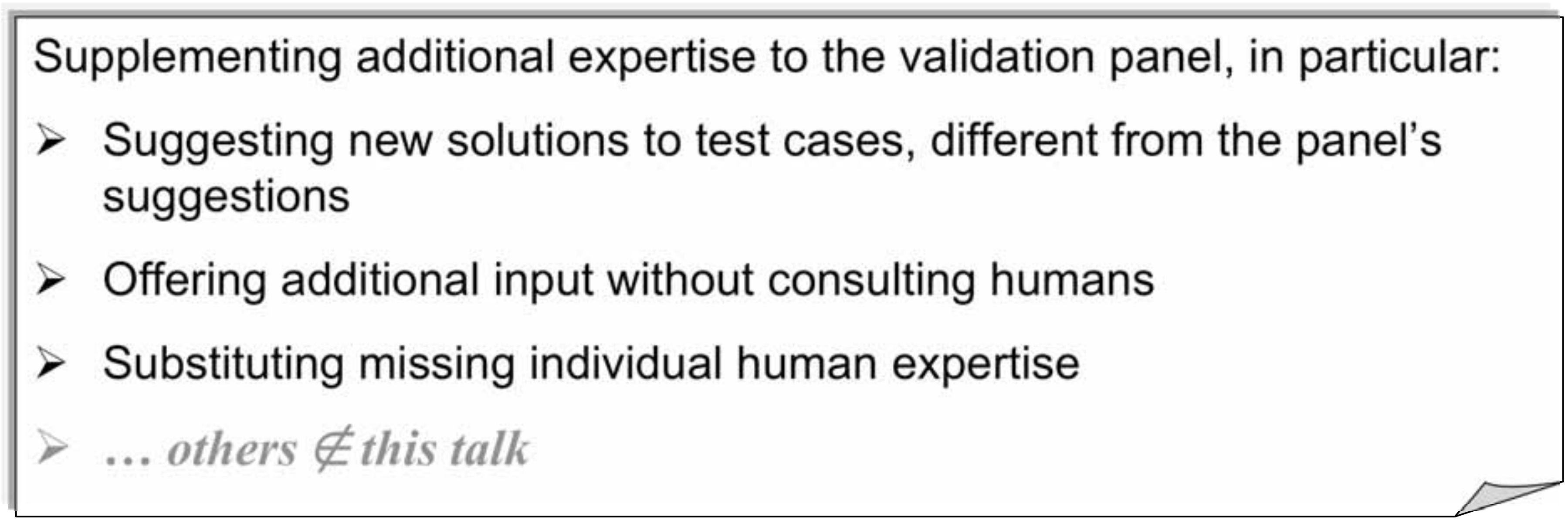
QuEST Quasi Exhaustive Set of Test Cases

- *a well-designed set that ensures coverage by formally analyzing the input space*

ReST Reasonable Set of Test Cases

- *a subset of **QuEST** that ensures the requirement efficiency by using validation criteria*

Objectives of modeling human experience

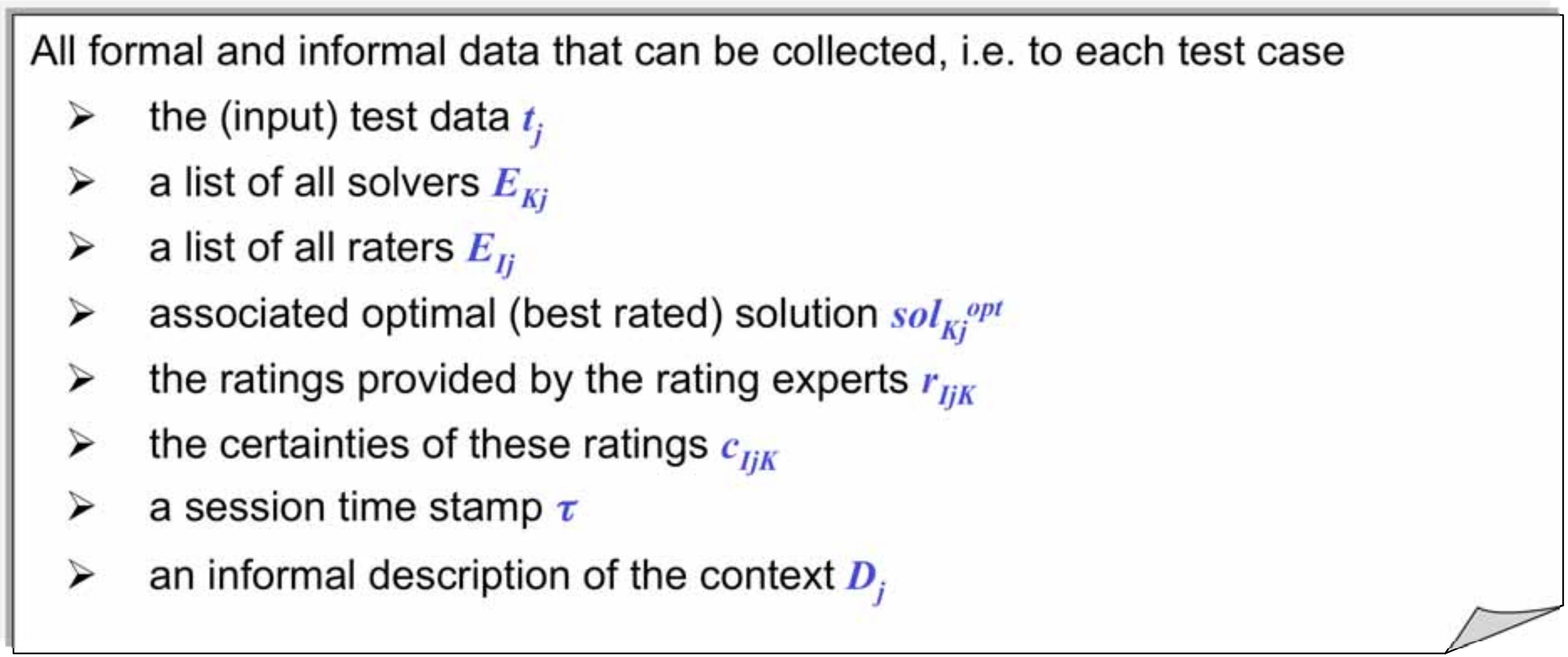


Supplementing additional expertise to the validation panel, in particular:

- Suggesting new solutions to test cases, different from the panel's suggestions
- Offering additional input without consulting humans
- Substituting missing individual human expertise
- *… others ∉ this talk*

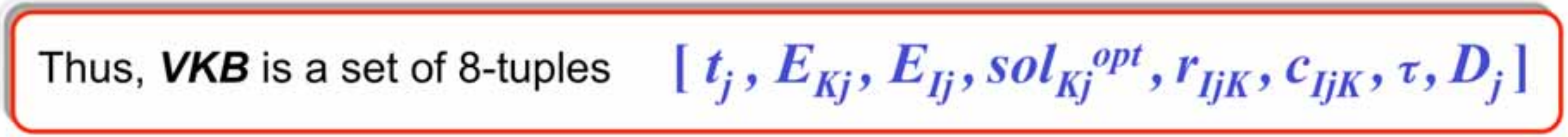
4 Incorporating a Validation Knowledge Base (VKB) as a Model of Collective Experience

4.1 The Content of VKB



All formal and informal data that can be collected, i.e. to each test case

- the (input) test data t_j
- a list of all solvers E_{Kj}
- a list of all raters E_{Ij}
- associated optimal (best rated) solution sol_{Kj}^{opt}
- the ratings provided by the rating experts r_{IjK}
- the certainties of these ratings c_{IjK}
- a session time stamp τ
- an informal description of the context D_j



Thus, ***VKB*** is a set of 8-tuples $[\, t_j\,, E_{Kj}, E_{Ij}, sol_{Kj}^{opt}, r_{IjK}, c_{IjK}, \tau, D_j \,]$

A part of *VKB* in the prototype test experiment

e_1, e_2, e_3
human experts

$t_1, t_2, \ldots$
test case inputs

$o_1, o_2, \ldots$
solutions (outputs)

τ
session #

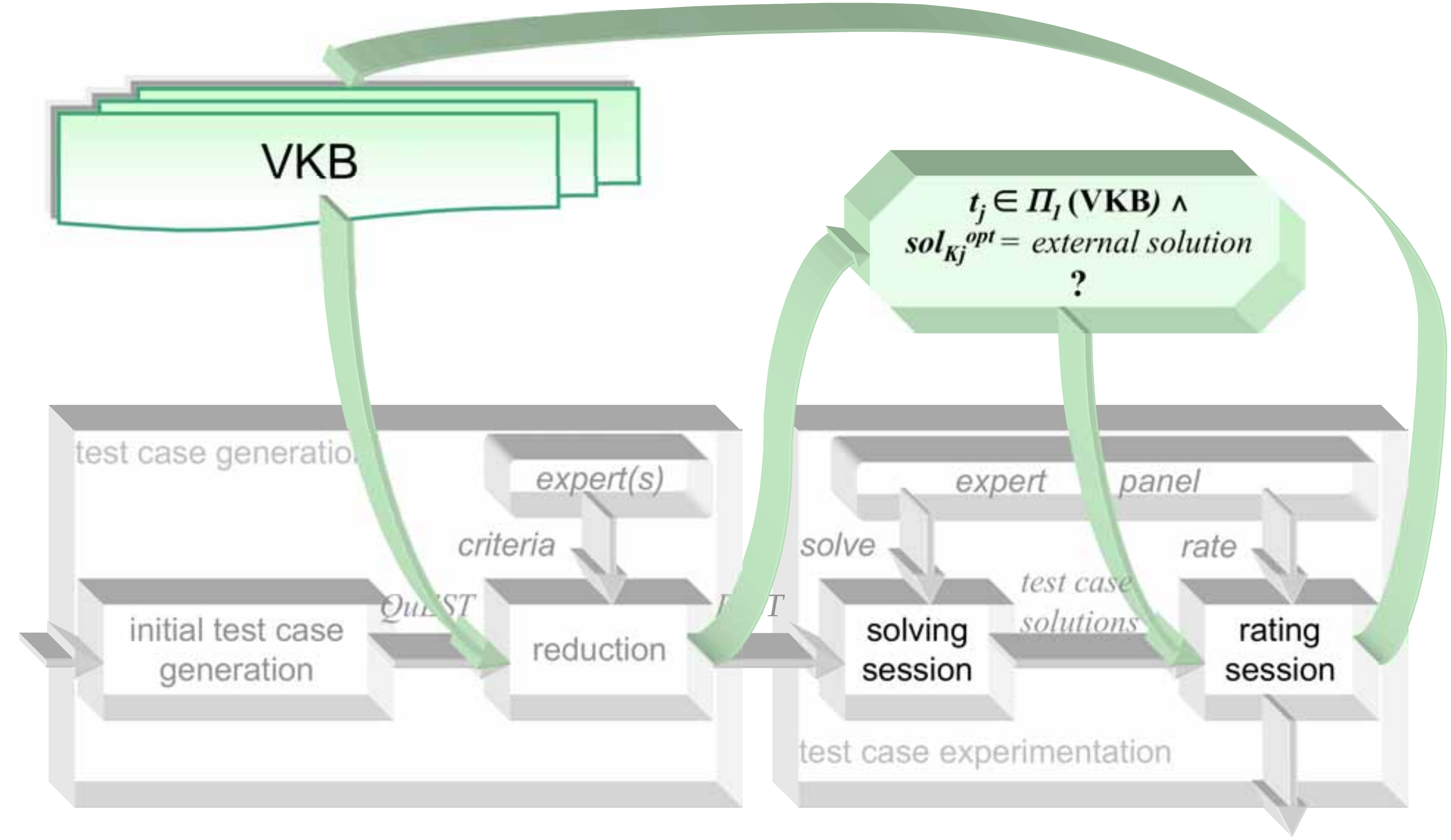
r
rating: 1 for correct, 0 for incorrect

c
certainty: 1 for certain, 0 for uncertain

t_j	E_{Kj}	E_{Ij}	sol_{Kj}^{opt}	r_{IjK}	c_{IjK}	τ	D_j
t_1	$[e_1, e_3]$	$[e_1, e_2, e_3]$	o_6	[1, 0, 1]	[0, 1, 1]	1	
t_1	$[e_3]$	$[e_1, e_2, e_3]$	o_4	[1, 0, 1]	[1, 1, 1]	3	
t_1	$[e_2]$	$[e_1, e_2, e_3]$	o_{17}	[0, 1, 0]	[1, 1, 1]	4	
t_2	$[e_1, e_3]$	$[e_1, e_2, e_3]$	o_7	[0, 0, 1]	[0, 0, 1]	1	
t_2	$[e_3]$	$[e_1, e_2, e_3]$	o_2	[1, 0, 1]	[1, 1, 1]	3	
t_2	[ ]	$[e_1, e_2, e_3]$	o_2	[1, 0, 1]	[1, 1, 1]	4	
t_3	$[e_2]$	$[e_1, e_2, e_3]$	o_{20}	[0, 1, 0]	[0, 1, 1]	1	
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
t_{42}	$[e_1, e_2, e_3]$	$[e_1, e_2, e_3]$	o_{23}	[1, 1, 1]	[1, 1, 1]	2	
t_{42}	$[e_1, e_2, e_3]$	$[e_1, e_2, e_3]$	o_{23}	[1, 1, 1]	[1, 1, 1]	3	

4.2 The Usage of VKB

External collective experience: $sol \in$ VKB, but not provided by the panel

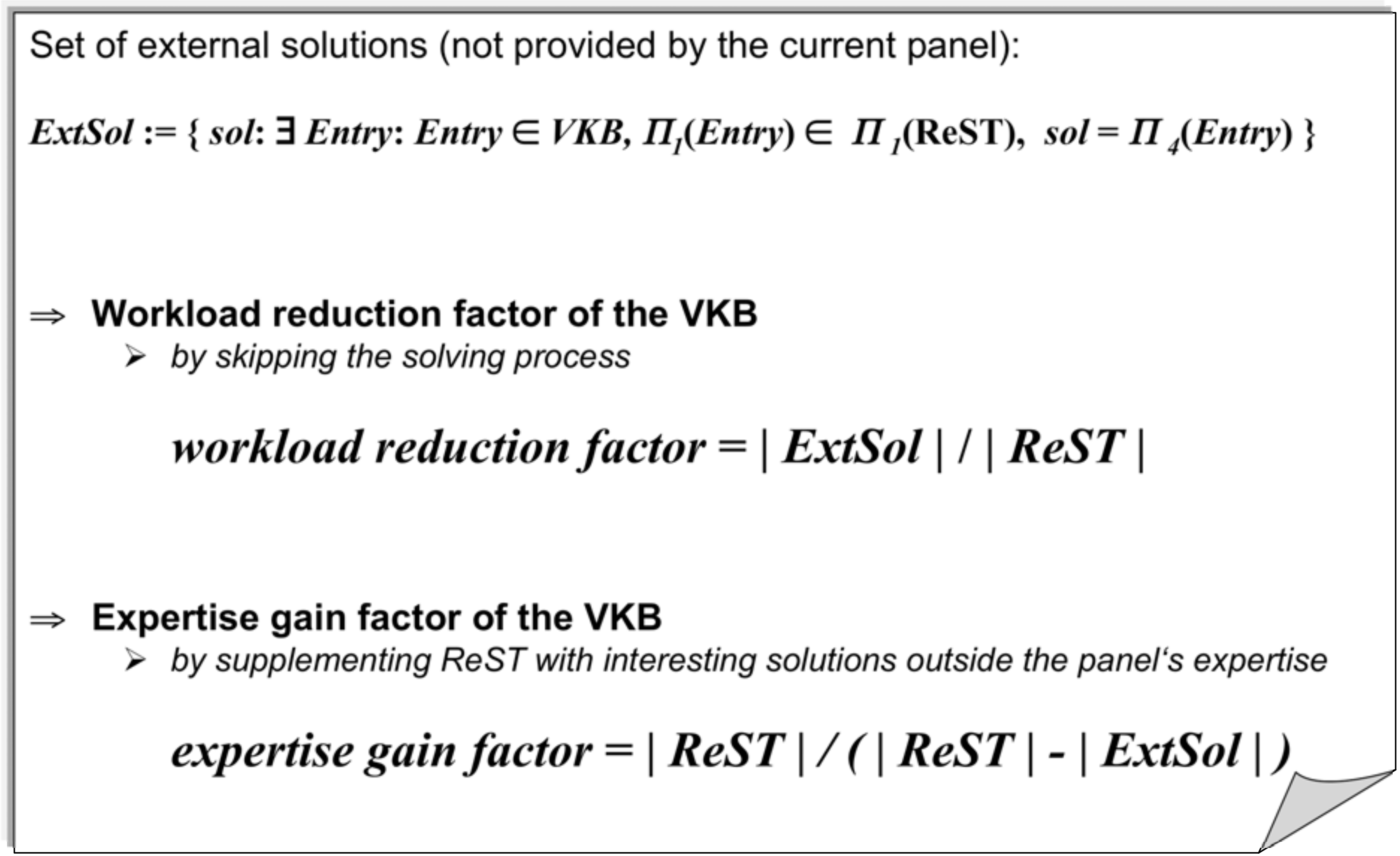


$t_j \in \Pi_1(\textbf{VKB}) \wedge$

$sol_{Kj}^{opt} =$ *external solution*

?

Quantifying the supplement of VKB to the human expertise



Set of external solutions (not provided by the current panel):

$$\mathit{ExtSol} := \{\, \mathit{sol}\colon \exists\, \mathit{Entry}\colon \mathit{Entry} \in \mathit{VKB},\ \Pi_1(\mathit{Entry}) \in \Pi_1(\mathrm{ReST}),\ \mathit{sol} = \Pi_4(\mathit{Entry}) \,\}$$

⇒ **Workload reduction factor of the VKB**

- *by skipping the solving process*

$$\textit{workload reduction factor} = |\,\mathit{ExtSol}\,| \,/\, |\,\mathit{ReST}\,|$$

⇒ **Expertise gain factor of the VKB**

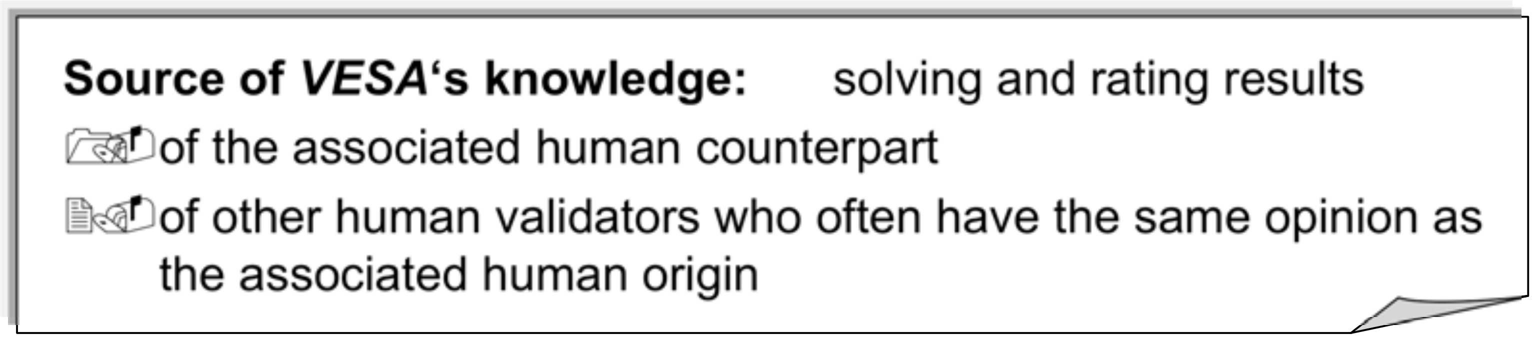
- *by supplementing ReST with interesting solutions outside the panel‘s expertise*

$$\textit{expertise gain factor} = |\,\mathit{ReST}\,| \,/\, (\,|\,\mathit{ReST}\,| - |\,\mathit{ExtSol}\,|\,)$$

5 Incorporating Validation Expert Software Agents (VESA) as Models of Individual Experiences

Objectives

- Forming a model of each validator‘s individual knowledge and behavior
- Successive refinement of this model by consecutive validation sessions



Source of *VESA*‘s knowledge: solving and rating results

- of the associated human counterpart
- of other human validators who often have the same opinion as the associated human origin

VESAs

- *are formed just in the moment of their need and „forgotten“ after their usage*
- *model just the required aspect of their human origin based on historical information of former sessions (i.e. not the current session)*
- *are requested in case its human counterpart is not available*
- *may be requested even if the human origin is present to validate the VESA concept itself by comparing the behavior of VESA with the real one of the human source.*

VESA models the solving behavior of an expert e_i for a test case t_j as follows

Step # 1

In case e_i solved (*with a solution different from „unknown“*) t_j in a former session, his/her solution with the latest time stamp τ will be provided by ***VESA***.

Step # 2

✓ All validators e', who ever delivered a solution to t_j form a set $Solver_i^0$, which is an initial dynamic agent for e_i : $Solver_i^0 := \{e' : [t_j, E_{Kj}, ...] \in VKB \wedge e' \in E_{Kj}\}$

✓ Select the most similar expert e_{sim} with the largest set of cases that have been solved by both e_i and e_{sim} with the same solution in the same session. e_{sim} forms a refined dynamic agent $Solver_i^1$ for e_i :

$$Solver_i^1 := e_{sim} : (e_{sim} \in Solver_i^0) \wedge (|\{[t_j, E_{Kj}, \_, sol_{Kj}^{opt}, \_, \_, \tau, \_] : e_i \in E_{Kj}, e_{sim} \in E_{Kj}\}| \rightarrow \max!)$$

✓ Provide the latest solution of the expert e_{sim} to t_j , i.e. the solution with the latest time stamp τ by ***VESA***.

Step # 3

If there is no such most similar expert, provide the solution $sol := unknown$ by ***VESA***.

An example of a *VESA* ‘s solving behavior compared to the human counterpart

EK^3

external knowledge (entries of the VKB) available in the 3rd session

e_2

human expert #2

$t_1, t_2, \ldots$

test case inputs

$o_1, o_2, \ldots$

solutions (outputs)

$VESA_2$

the VESA-model of expert #2

EK_3	solution of		EK_3	solution of	
	$VESA_2$	e_2		$VESA_2$	e_2
t_{29}	o_8	o_8	t_{36}	o_9	o_9
t_{30}	o_9	o_9	t_{37}	o_9	o_9
t_{31}	o_2	o_2	t_{38}	o_9	o_9
t_{32}	o_8	o_3	t_{39}	o_9	o_9
t_{33}	o_8	o_8	t_{40}	o_{23}	o_{23}
t_{34}	o_2	o_2	t_{41}	o_{19}	o_{22}
t_{35}	o_8	o_8	t_{42}	o_{23}	o_{23}

VESA models the rating behavior of an expert e_i for a test case t_j as follows

Step # 1

In case e_i rated t_j in a former session, adopt the rating with the latest time stamp τ_s and provide the same rating r and the same certainty c by ***VESA***.

Step # 2

- ✓ All validators e', who ever delivered a rating to t_j form a set $Rater_i^0$, which is an initial dynamic agent for e_i : $Rater_i^0 := \{e' : [t_j, \_, E_{Ij}, \ldots] \in VKB \wedge e' \in E_{Ij}\}$

- ✓ Select the most similar expert e_{sim} with the largest set of cases that have been rated by both e_i and e_{sim} with the same rating in the same session. e_{sim} forms a refined dynamic agent $Rater_i^1$ for e_i :

$$Rater_i^1 := e_{sim} : (e_{sim} \in Rater_i^0) \wedge (|\{[t_j, \_, E_{Ij}, sol_{Kj}^{opt}, r_{IjK}, \_, \tau, \_] : e_i \in E_{Ij}, e_{sim} \in E_{Ij}\}| \to \max!)$$

- ✓ Provide the latest rating r of the expert e_{sim} along with its certainty c, i.e. the ones with the latest time stamp τ , to the present test case t_j by ***VESA***.

Step # 3

If there is no such most similar expert, provide the rating $r := norating$ along with a certainty $c := 0$ by ***VESA***.

An example of a *VESA* ‘s rating behavior compared to the human counterpart

EK^3
external knowledge (entries of the VKB) available in the 3rd session

e_2
human expert #2

$t_1, t_2, \ldots$
test case inputs

$o_1, o_2, \ldots$
solutions (outputs)

$VESA_2$
the VESA-model of expert #2

EK_3	solution	rating of		EK_3	solution	rating of	
		$VESA_2$	e_2			$VESA_2$	e_2
t_1	o_4	0	0	t_{29}	o_3	0	0
t_1	o_6	0	0	t_{29}	o_4	0	1
t_1	o_{21}	0	0	t_{29}	o_8	1	1
t_1	o_{18}	1	1	t_{29}	o_{16}	0	0
t_2	o_2	0	0	t_{30}	o_2	0	0
t_2	o_7	0	0	t_{30}	o_4	0	1
t_2	o_{20}	0	1	t_{30}	o_9	1	1
t_3	o_2	0	0	t_{30}	o_{16}	0	0
t_3	o_3	0	0	t_{31}	o_2	1	0
t_3	o_8	0	0	t_{31}	o_4	0	1
t_3	o_{20}	1	0	t_{31}	o_8	0	1
t_4	o_{23}	0	0	t_{31}	o_{16}	0	0

6 A Prototype Test

How to find human experts who are able and willing to cooperate for free ?

By choosing an “application” with a certain “entertainment factor”:

Selection of an appropriate wine for a given dinner

6.1 The Knowledge Base

Input space: $I := [\, s_1, s_2, s_3 \,]$:

- $s_1 \in \{$ *pork, beef, veal, fowl,…, fish,…,goat cheese,…, fruit dessert, ice cream* $\}$
- $s_2 \in \{$ *non(raw), steamed, boiled, grillesd, fried,* … $\}$
- $s_3 \in \{$ *Asian, Western* $\}$

Output space: $O := \{\, o_1, o_2, \ldots, o_{24} \,\}$ with

- $o_1 =$ *Red wine, fruity, low tannin, less compound*
- $o_2 =$ *Red wine, young, rich of tannin*
- …

Rule base: $R := \{\, r_1, r_2, \ldots, r_{45} \,\}$ with

- $r_1 : o_1 \leftarrow (\, s_1 = \textit{fowl} \,)$
- $r_2 : o_1 \leftarrow (\, s_1 = \textit{veal} \,)$
- $r_3 : o_2 \leftarrow (\, s_1 = \textit{pork} \,) \wedge (\, s_2 = \textit{grilled} \,)$
- …

6.2 The Test Cases

... have been generated with a technology as introduced in former papers.

The resulting "Reasonable Set of Test Cases" (***ReST***) is:

t_{1}	pork	boiled	Asian	t_{22}	fish	steamed	Western
t_{2}	pork	grilled	any	t_{23}	fish	boiled	Asian
t_{3}	pork	fried	any	t_{24}	fish	grilled	any
t_{4}	pork	stewed	any	t_{25}	fish	fried	any
t_{5}	beef	boiled	Asian	t_{26}	fish	stewed	Asian
t_{6}	beef	grilled	any	t_{27}	fish	deep fried	Asian
t_{7}	beef	fried	any	t_{28}	hard cheese	non	Western
t_{8}	beef	stewed	any	t_{29}	hard cheese	casserole	Western
t_{9}	veal	boiled	any	t_{30}	hard cheese	deep fried	Western
t_{10}	veal	grilled	any	t_{31}	soft cheese	non	Western
t_{11}	veal	fried	any	t_{32}	soft cheese	casserole	Western
t_{12}	veal	stewed	any	t_{33}	soft cheese	deep fried	Western
t_{13}	venison	boiled	any	t_{34}	goat cheese	non	Western
t_{14}	venison	grilled	any	t_{35}	goat cheese	casserole	Western
t_{15}	venison	fried	any	t_{36}	goat cheese	deep fried	Western
t_{16}	venison	stewed	any	t_{37}	blue mold cheese	non	Western
t_{17}	fowl	boiled	any	t_{38}	blue mold cheese	casserole	Western
t_{18}	fowl	grilled	any	t_{39}	blue mold cheese	deep fried	Western
t_{19}	fowl	fried	any	t_{40}	fruit dessert	non	any
t_{20}	fowl	stewed	any	t_{41}	aromatic dessert	non	any
t_{21}	fish	non	Asian	t_{42}	ice cream	non	any

6.3 Application Conditions

The experimentation took place with

- three human experts e_1 , e_2 , e_3
- a test case set ***ReST*** = { t_1 , t_2 , …, t_{42} }
- session schedule:

session number	experts			VESAs			examined test case inputs out of $\Pi_1(ReST)$
	e_1	e_2	e_3	$VESA_1$	$VESA_2$	$VESA_3$	
1	+	+	+	-	-	-	$\Pi_1(ReST^1) := \{ t_1, \ldots, t_{28} \}$
2	⊕	+	+	+	-	-	$\Pi_1(ReST^2) := \{ t_{15}, \ldots, t_{42} \}$
3	+	⊕	+	-	+	-	$\Pi_1(ReST^3) := \{ t_1, \ldots, t_{14,} t_{29}, \ldots, t_{42} \}$
4	+	+	⊕	-	-	+	$\Pi_1(ReST^4) := \{ t_i : t_i \ mod \ 3 \neq 0 \}$

+ takes part in the session - does not take part in the session
⊕ takes part in the session only for being compared with its VESA

<u>Notational Conventions</u>

- *$\boldsymbol{VKB^i}$ denotes the **VKB** as developed after the **i** -th session*
- *$\boldsymbol{VESA_k^i}$ denotes the behavior of the **VESA** which models the behavior of expert $\boldsymbol{e_k}$ after the **i** -th session*
- *$\boldsymbol{ReST^i}$ denotes the test case set used in the **i** -th session*
- *$\boldsymbol{EK^i}$ denotes the available “external knowledge” of the **VKB** in the **i** -th session: $\boldsymbol{EK^i := \Pi_1(VKB^i) \cap ReST^i}$*

6.4 Desired Outcome of the Experiment

The experiment should provide answers to the following questions

1. *Does the VKB contribute to the validation sessions at an increasing rate with an increasing number of validation sessions?*
   - ***How many external solutions (outside the expertise of the current expert panel) are introduced into the rating process by the VKB?***
2. *Does the VKB contribute valid knowledge (best rated solutions) in an increasing rate with an increasing number of validation sessions?*
   - ***How many of the introduced solutions win the rating contest against the solutions of the current expert panel?***
3. *Does the VKB increasingly gain the human expertise as number of validation sessions increases?*
   - ***How many new best rated solutions are introduced into the VKB after a validation session?***
4. *Do the VESAs models of their human source improve with in increasing number of validation sessions?*
   - ***Do the VESAs provide the same solutions and ratings as their human counterpart?***

To quantify these measures, we computed after each session (session # ***i***)

- *the number* $\mathbf{a}_i$ *of cases from* ***VKB*** $^{i\text{-}1}$, *which were the subject of the rating session and relate it to* **|** ***EK***^{i} **|** *:* $A_i := a_i \; / \, | \, EK^i \, |$
- *the number* $\mathbf{b}_i$ *of cases from* ***VKB*** $^{i\text{-}1}$, *which provided the optimal (best rated) solution and relate it to* **|** ***EK***^{i} **|** *:* $B_i := b_i \; / \, | \, EK^i \, |$
- *the number* $\mathbf{c}_i$ *of cases from* ***VKB*** $^{i\text{-}1}$, *for which a new solution has been introduced into* ***VKB*** *and relate it to* **|** ***EK***^{i} **|** *:* $C_i := c_i \; / \, | \, EK^i \, |$
- *the number* $\mathbf{d}_i$ *of solutions and ratings, which are identical responses of* e_{i-1} *and* ***VESA*** $_{i-1}$ *and relate it to the number of required solutions and ratings:* $D_i := d_i \; /$ # ***responses***

Thus, desired answers can be formalized

1. *Does the VKB contribute to the validation sessions at an increasing rate with an increasing number of validation sessions:* $A_4 > A_3 > A_2$ *?*
2. *Does the VKB contribute valid knowledge (best rated solutions) in an increasing rate with an increasing number of validation sessions:* $B_4 > B_3 > B_2$ *?*
3. *Does the VKB increasingly gain the human expertise as number of validation sessions increases:* $C_2 > C_3 > C_4$ *?*
4. *Do the VESAs model of their human source improve with in increasing number of validation sessions:* $D_4 > D_3 > D_2$ *?*

7 Test Results

Does the VKB contribute to the validation sessions at an increasing rate with an increasing number of validation sessions: $A_4 > A_3 > A_2$?

- # of new external solutions from VKB:
    - 1 (of 14 possible in EK) in session 2
    - 2 (of 28) in session 3
    - 24 (!) (of 28) in session 4 $0.85 >> 0.071 \geq 0.071$
- *Obviously, the VKB needs to gain some "initial experience" before it contributes a remarkable number of new solutions.*
- *The desired effect became remarkable in the 4th session.*

2. *Does the VKB contribute valid knowledge (best rated solutions) in an increasing rate with an increasing number of validation sessions:* $B_4 > B_3 > B_2$?

- # of new external solutions, which won the rating session:
    - 0 (out of 14) in session 2
    - 0 (out of 28) in session 3
    - 2 (out of 28) in session 4: $0.071 \geq 0 \geq 0$
- *However, it is remarkable that 2 solutions which were not provided by the panel got very best marks by the same panel.*
- *This is what we want the VKB to do: Contributing better knowledge than the current human experts. The „collective experience“ of former panels reveals to be better than the current panel.*

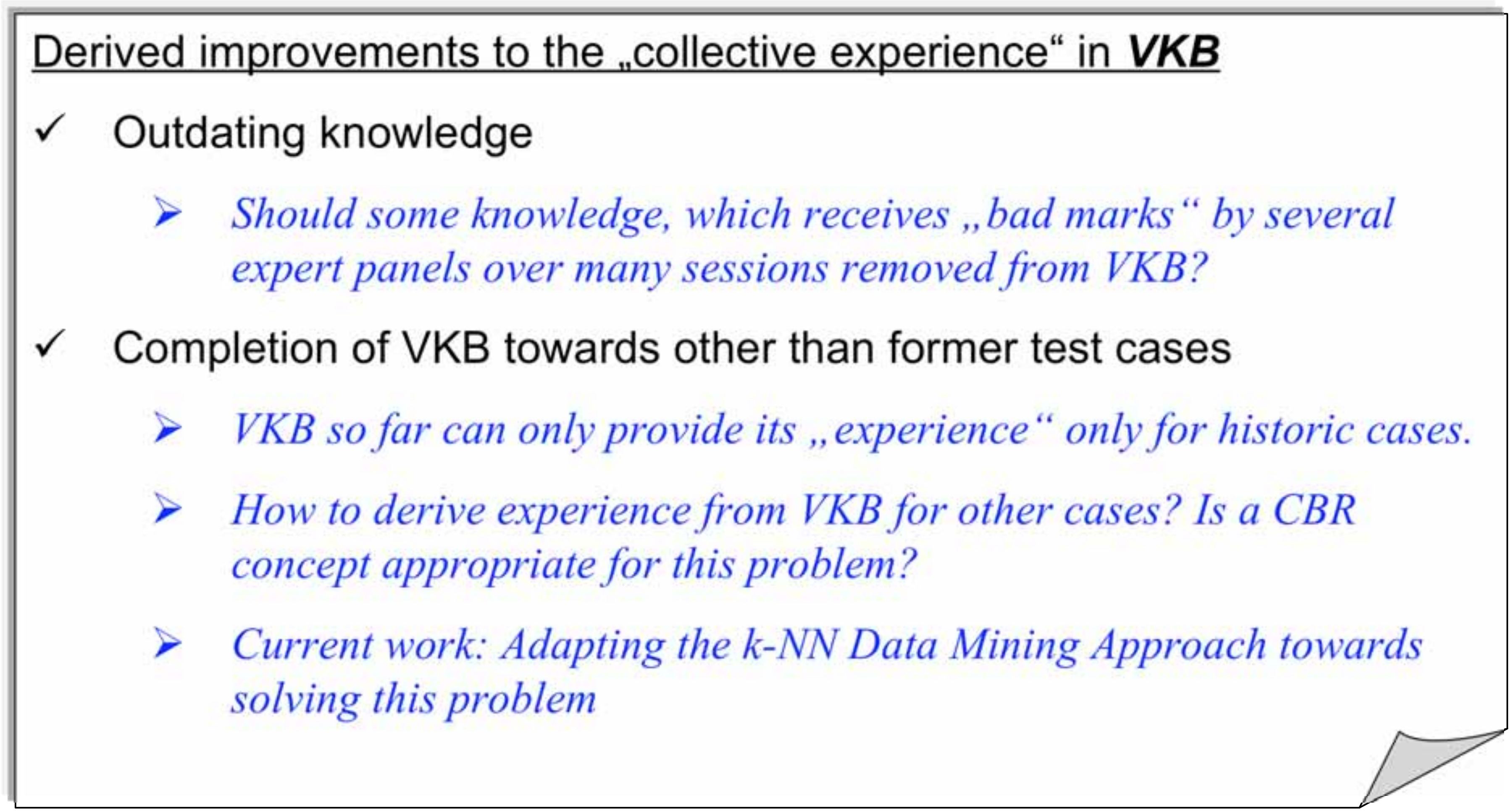
3. *Does the VKB increasingly gain the human expertise as number of validation sessions increases:* $C_2 > C_3 > C_4$?

- # of cases introduced into VKB:
  - 7 (of 14) after session 2
  - 16 (of 28) after session 3
  - 17 (of 28) after session 4: $0.5 \leq 0.57 \leq 0.61$
- *Here, our expectation was not met!*
- *The reason is probably, that the domain knowledge itself as well as its reflection in human minds changed from session to session.*
- *Most interesting problem domains are not static by nature; individual peoples' opinions are not static by nature.*

4. *Do the **VESA**s model of their human source improve with in increasing number of validation sessions:* $D_4 > D_3 > D_2$?

- # of identical responses by the expert and his/her VESA
  - 27 (of 63) in session 2
  - 78 (of 126) in session 3
  - 90 (of 150) in session 4: $0.6 \approx 0.62 > 0.43$
- *Again, we explain this as the result of changing minds by the experts.*
- *A crucial problem is*
  - *the interpretation of a verbal case description and*
  - *some latent dependence from other circumstances than the case input itself (the mood, e.g.).*

Lessons Learnt



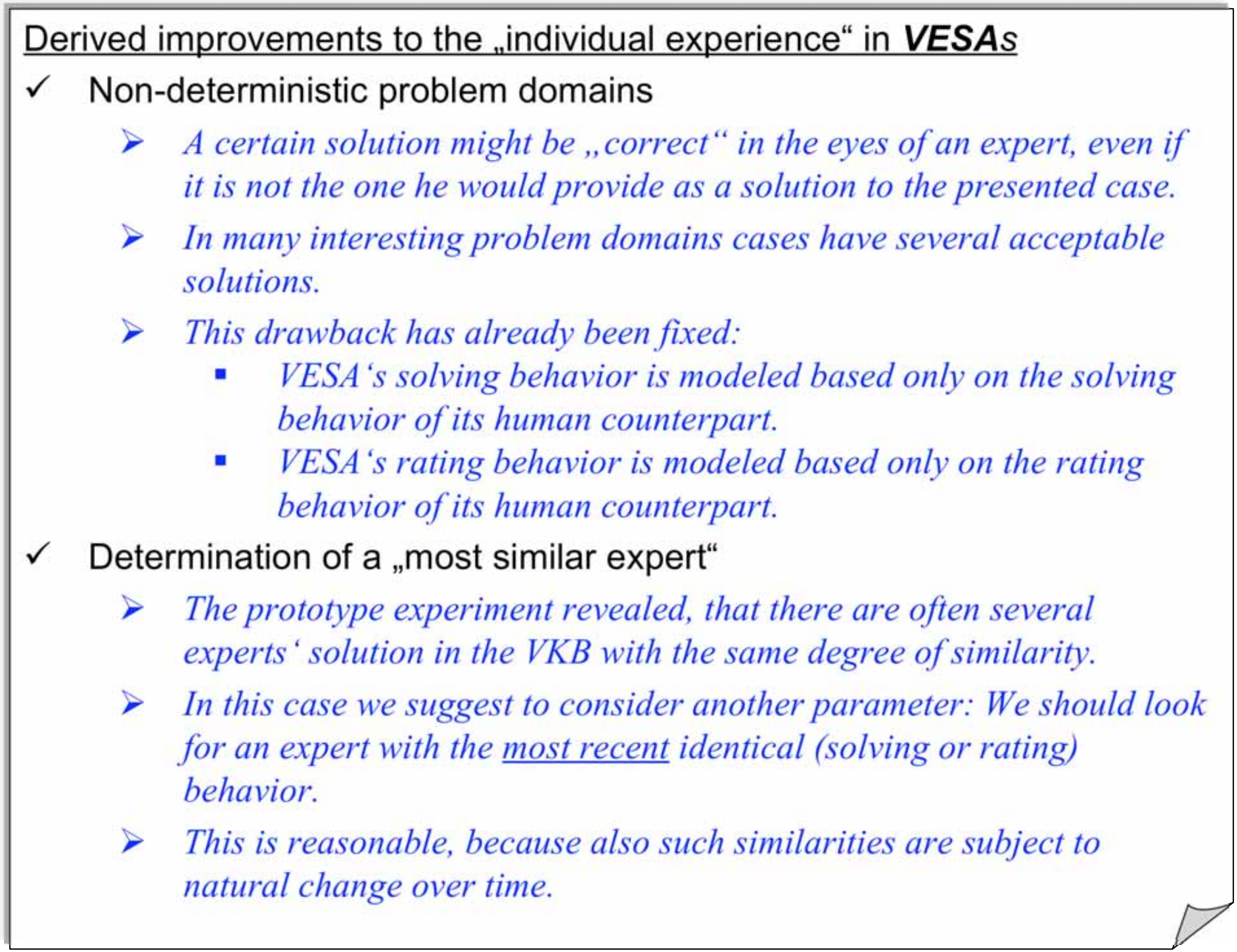
Derived improvements to the „collective experience“ in ***VKB***

✓ Outdating knowledge

- *Should some knowledge, which receives „bad marks“ by several expert panels over many sessions removed from VKB?*

✓ Completion of VKB towards other than former test cases

- *VKB so far can only provide its „experience“ only for historic cases.*
- *How to derive experience from VKB for other cases? Is a CBR concept appropriate for this problem?*
- *Current work: Adapting the k-NN Data Mining Approach towards solving this problem*



Derived improvements to the „individual experience“ in ***VESA*s**

- ✓ Non-deterministic problem domains
  - ➢ *A certain solution might be „correct“ in the eyes of an expert, even if it is not the one he would provide as a solution to the presented case.*
  - ➢ *In many interesting problem domains cases have several acceptable solutions.*
  - ➢ *This drawback has already been fixed:*
    - ▪ *VESA's solving behavior is modeled based only on the solving behavior of its human counterpart.*
    - ▪ *VESA's rating behavior is modeled based only on the rating behavior of its human counterpart.*
- ✓ Determination of a „most similar expert“
  - ➢ *The prototype experiment revealed, that there are often several experts' solution in the VKB with the same degree of similarity.*
  - ➢ *In this case we suggest to consider another parameter: We should look for an expert with the <u>most recent</u> identical (solving or rating) behavior.*
  - ➢ *This is reasonable, because also such similarities are subject to natural change over time.*

Derived improvements to the „individual experience“ in ***VESA**s* *(cont‘d)*

✓ Permanent validation of the ***VESA**s*

- ➢ *The concept will be refined by adding some permanent „self-validation“ of each VESA by*
    - ➢ *submitting VESA‘s solution to the rating process of its human counterpart and*
    - ➢ *comparing VESA‘s rating with the rating of its human counterpart.*
- ➢ *Thus, some statement about each VESA‘s quality can be derived:*
    - ☞ *The number of VESA‘s solutions, which are rated by its human counterpart as „correct“ and*
    - ☞ *the number of VESA‘s ratings which are identical with those of its human counterpart*

  are measures about the performance of the human behavior model.

✓ Completion of ***VESA**s* towards other than former test cases

- ➢ *In case there is no „most similar expert“ who ever considered (solved or rated) a current case, a concept of determining a „most likely response“ of the modeled expert needs to be developed.*

8 Summary and Conclusion

1. Ensuring validity of AI systems requests methods beyond conventional software engineering techniques. The only source of domain knowledge is often human expertise.

2. Human expertise is often uncertain, undependable, contradictory, unstable, it changes over time and is quite expensive.

3. The concept of ***VKB*** is the key to use this resource more efficiently towards valid systems. The VKB approach includes all aspects of „collective historical experience“ that have been provided by previous expert panels.

4. While ***VKB*** aims at modeling the human experts‘ collective and most accepted (best rated) knowledge, the ***VESA*** concept aims at modeling the individual human experts.

5. Experiments revealed that the ***VKB*** and ***VESA*** approach needs to be refined with respect to

   ➢ their completion towards other than (previous) test cases

      👓 Under construction: Adapting the k-NN data mining approach

   and ***VESA*** needed to be developed further with respect to

   ✓ the nature of the non-deterministic problem domains (done!)

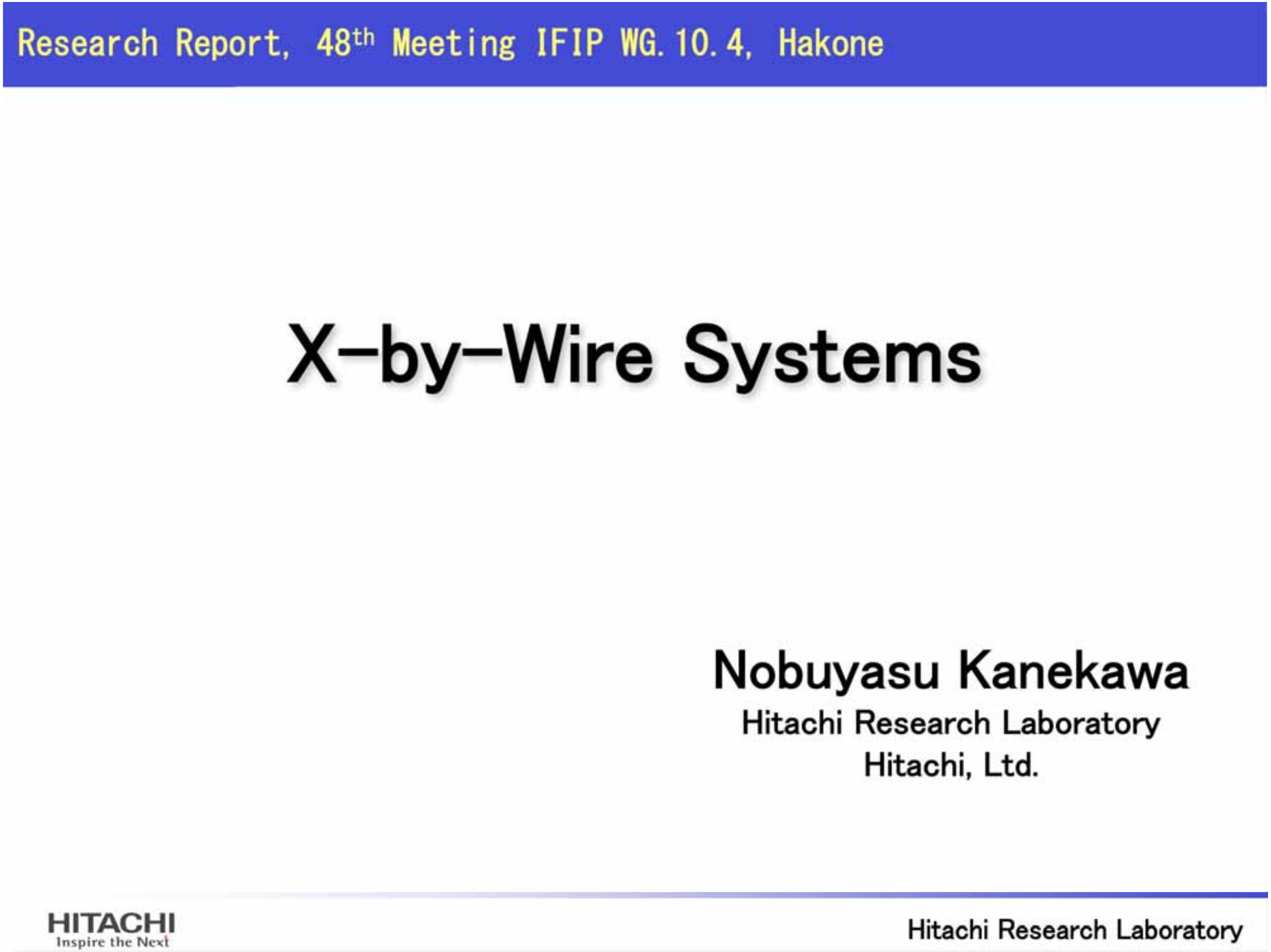
      ☝ *Solving cases based on a previous rating is not appropriate*

   ➢ their permanent validation

Session 2

Moderator

Takashi Nanya, University of Tokyo, Japan



Research Report, 48th Meeting IFIP WG.10.4, Hakone

X-by-Wire Systems

Nobuyasu Kanekawa
Hitachi Research Laboratory
Hitachi, Ltd.

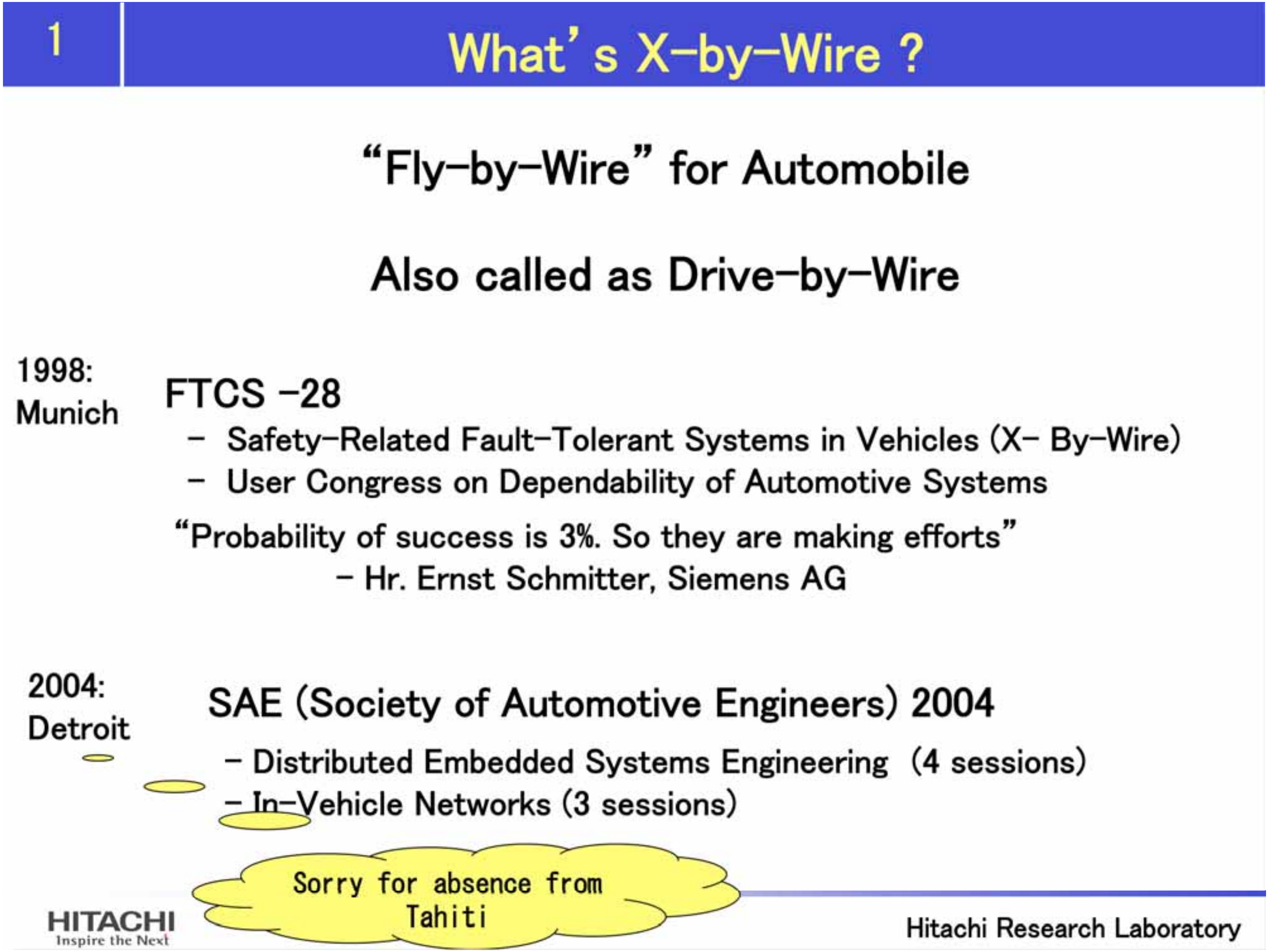
Hitachi Research Laboratory

1. What's X-by-Wire ?

2. Our Approach

1. What's X-by-Wire ?

2. Our Approach



1

What's X-by-Wire ?

“Fly-by-Wire” for Automobile

Also called as Drive-by-Wire

1998: Munich

FTCS -28

- Safety-Related Fault-Tolerant Systems in Vehicles (X- By-Wire)
- User Congress on Dependability of Automotive Systems

“Probability of success is 3%. So they are making efforts”

– Hr. Ernst Schmitter, Siemens AG

2004: Detroit

SAE (Society of Automotive Engineers) 2004

- Distributed Embedded Systems Engineering (4 sessions)
- In-Vehicle Networks (3 sessions)

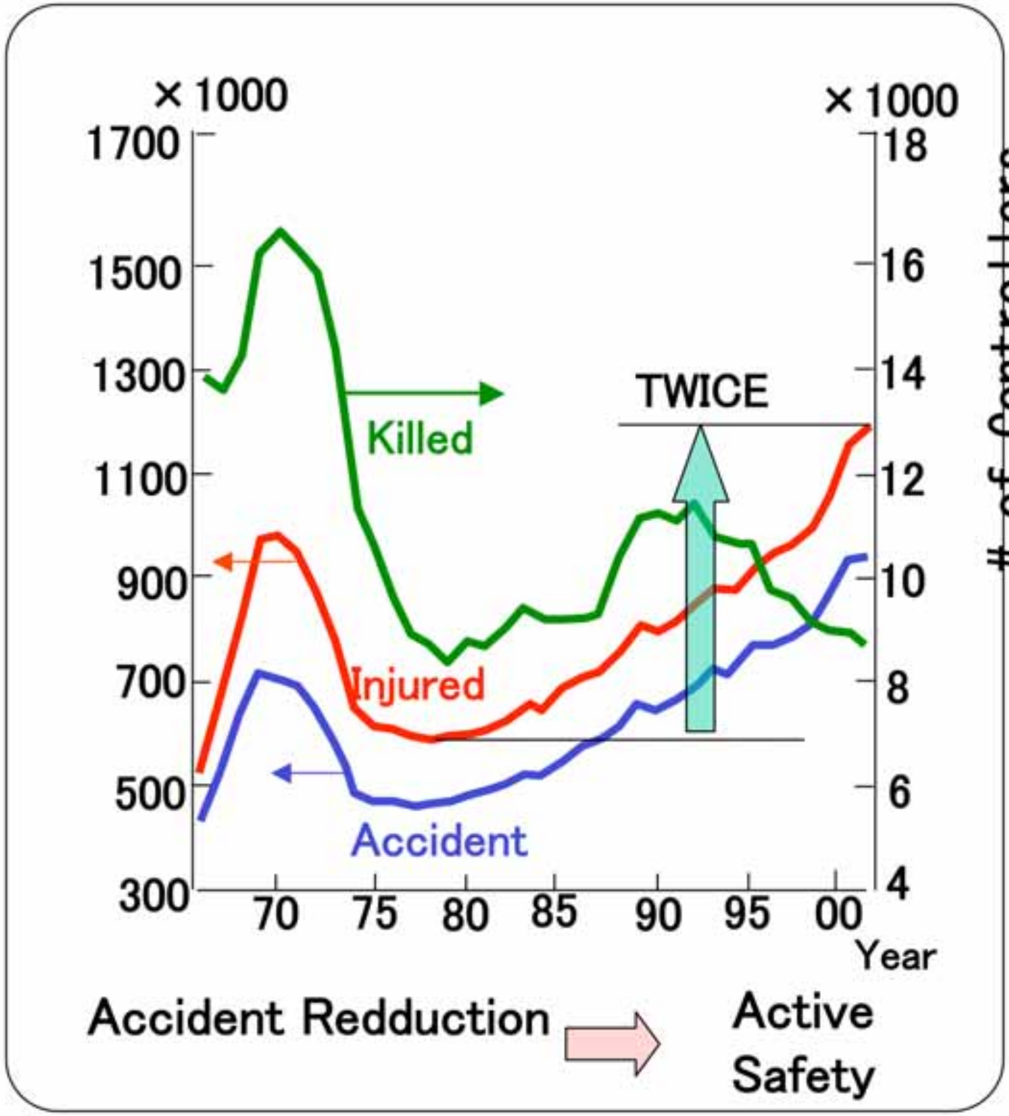
Sorry for absence from Tahiti

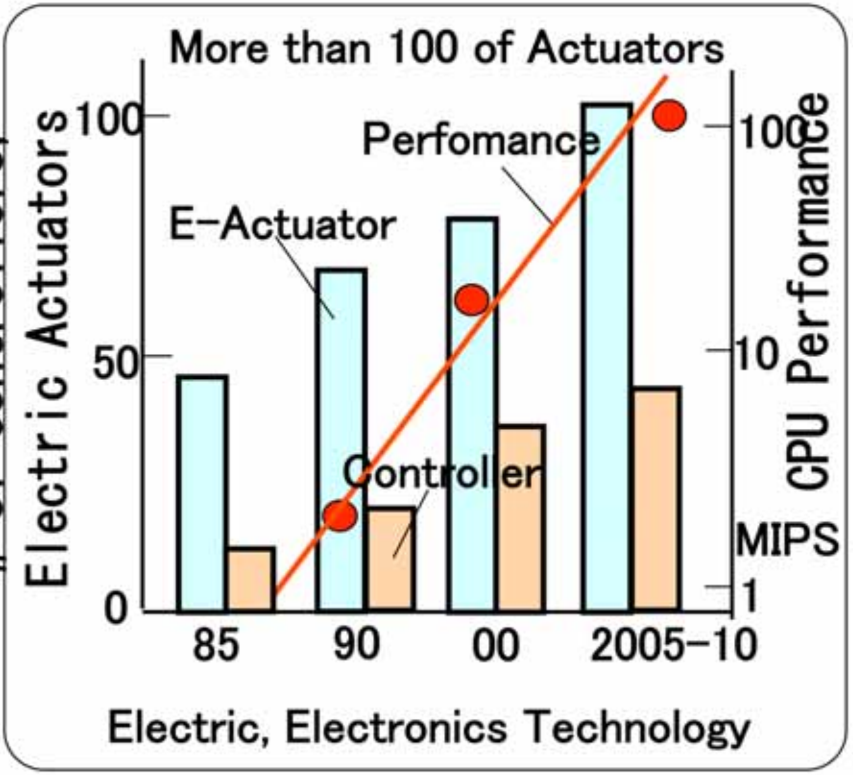
HITACHI Inspire the Next

Hitachi Research Laboratory

2

Necessity for Active Safety





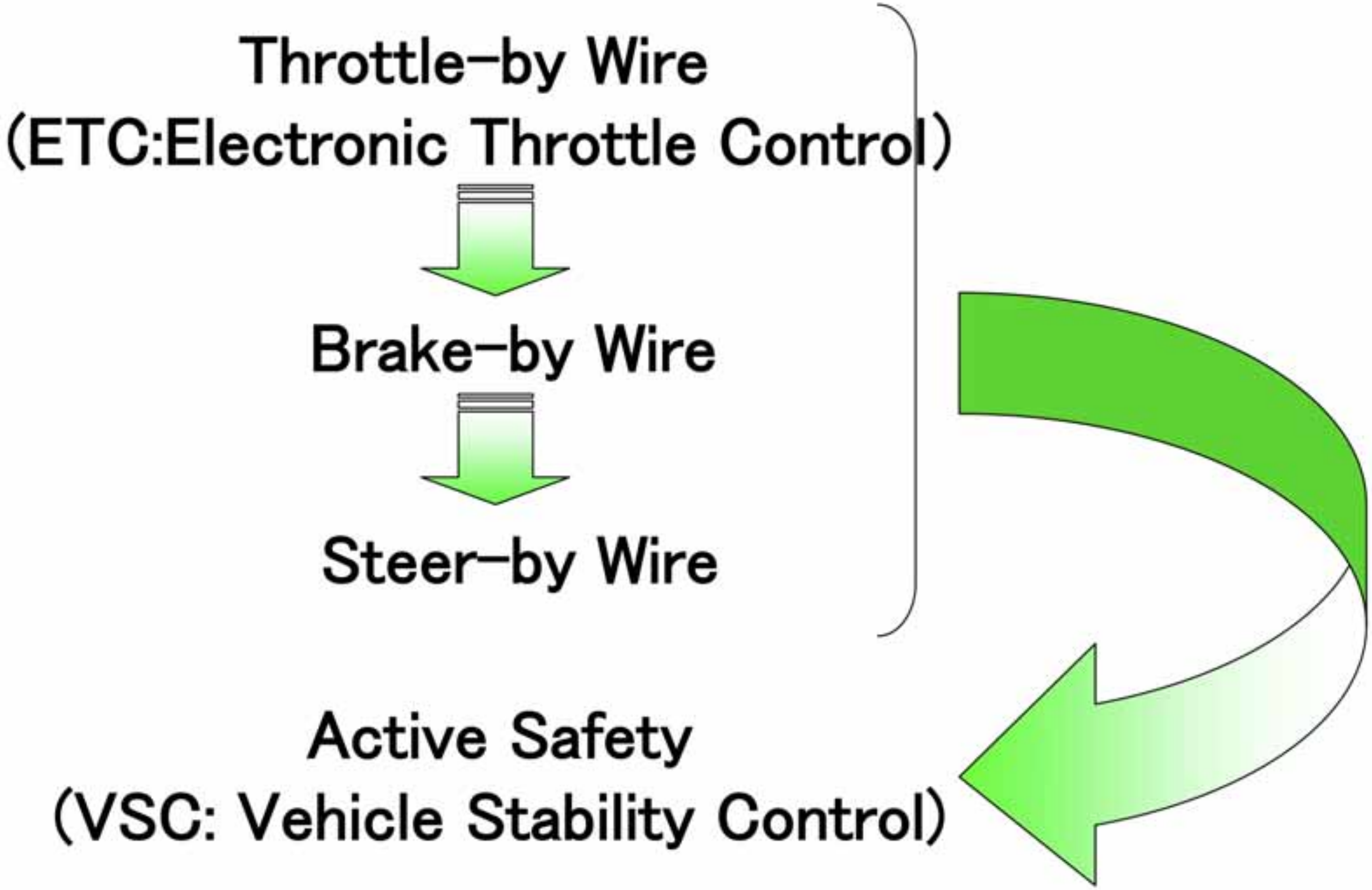
Active Safety by Electric /Electronic Technology

HITACHI
Inspire the Next

Hitachi Research Laboratory

3

X-by-Wire Road Map



Throttle-by Wire
(ETC:Electronic Throttle Control)

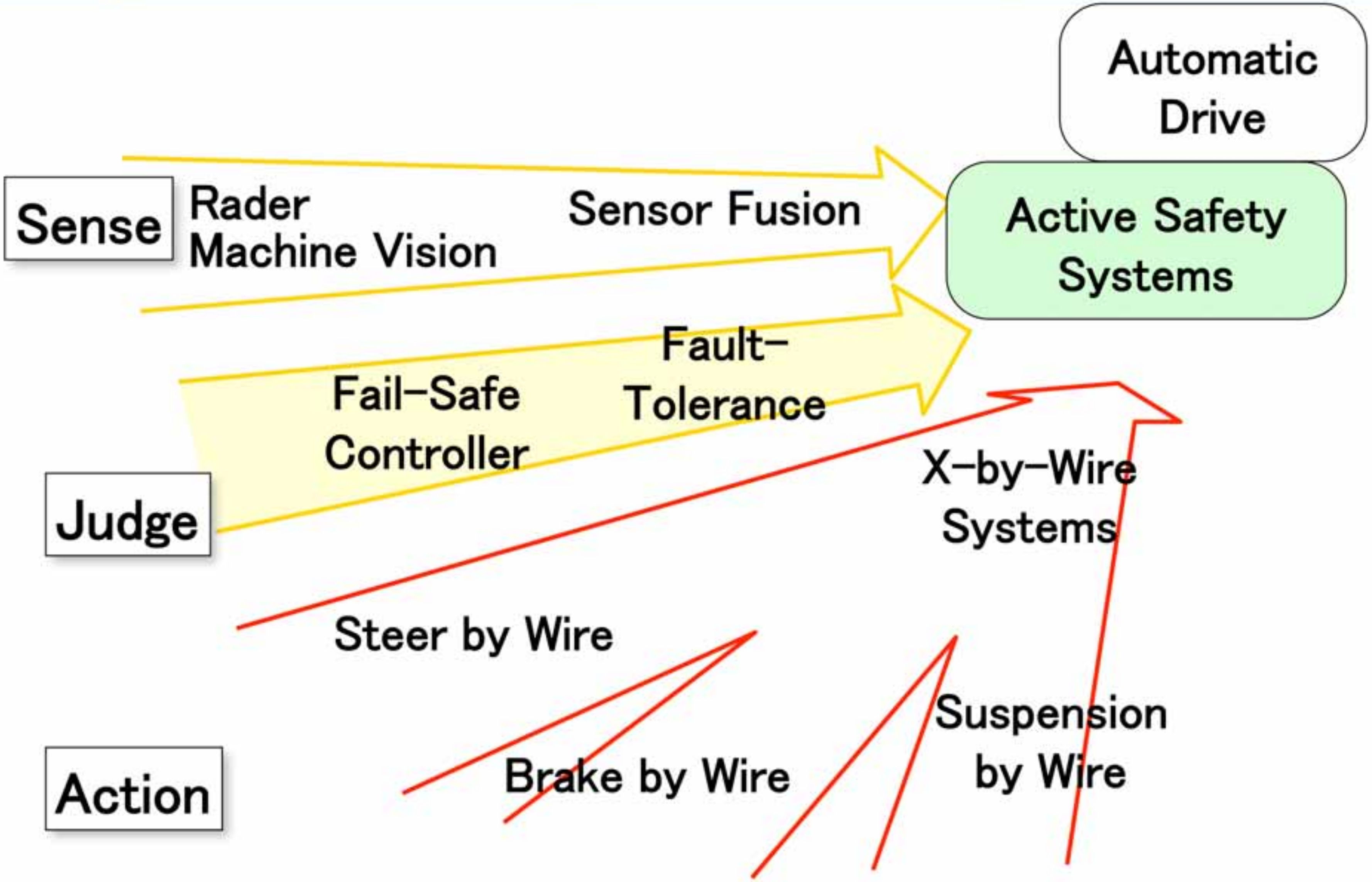
Brake-by Wire

Steer-by Wire

Active Safety
(VSC: Vehicle Stability Control)

4

X-By-Wire Road Map



Automatic Drive

Active Safety Systems

Sense

Rader
Machine Vision

Sensor Fusion

Judge

Fail-Safe Controller

Fault-Tolerance

Action

Steer by Wire

Brake by Wire

Suspension by Wire

X-by-Wire Systems

Hitachi Research Laboratory

5

Concept Cars

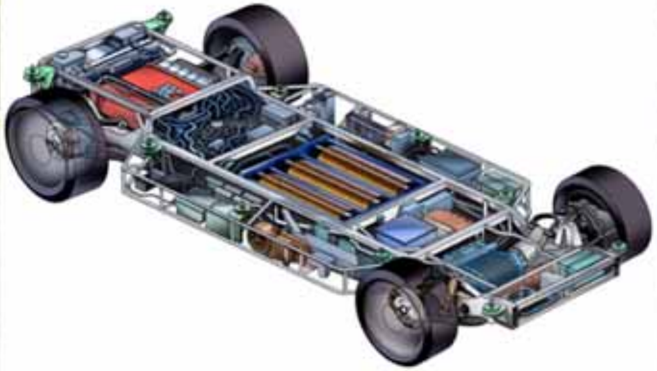
Daimler Chrysler <R129> 1997

X-by-Wire Operated with a Side Stick

Side Stick

GM <Hy-Wire> 2002 : Fuel Cell Vehicle

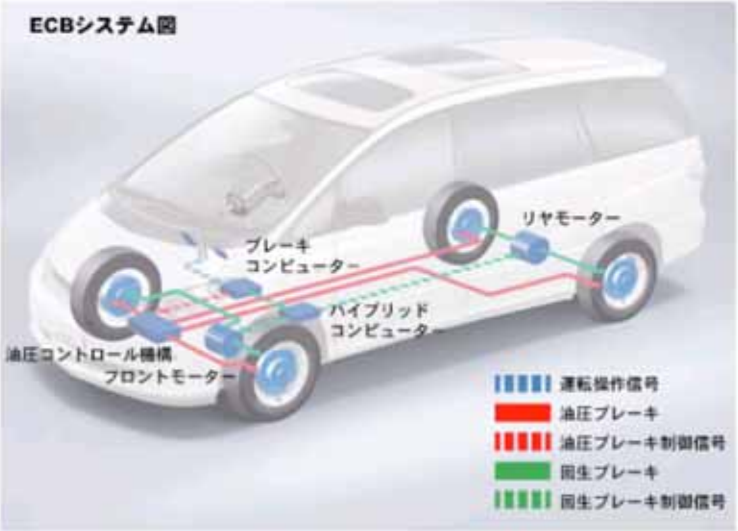
Power-train Platform with 11” Thickness. Layout-Free Cabin



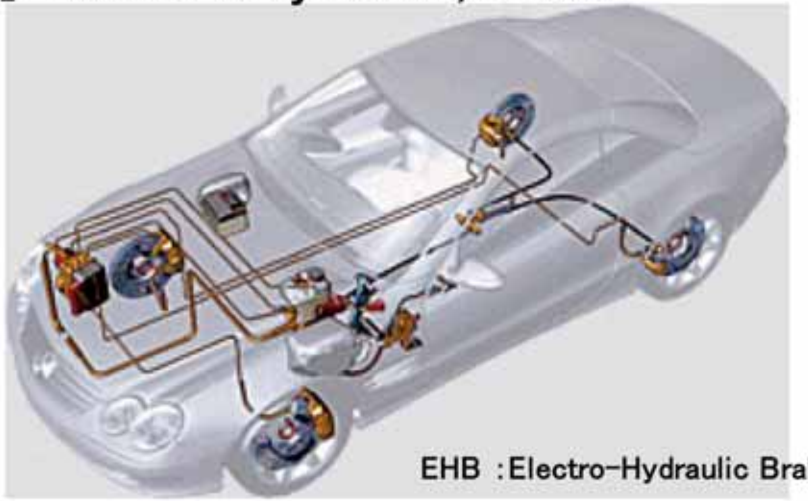
6 X-by-Wire Real Cars

Brake-by-Wire

【EHB】 Toyota／Estima-Hybrid



【EHB】 Daimlar Chrysler/SL, E-class

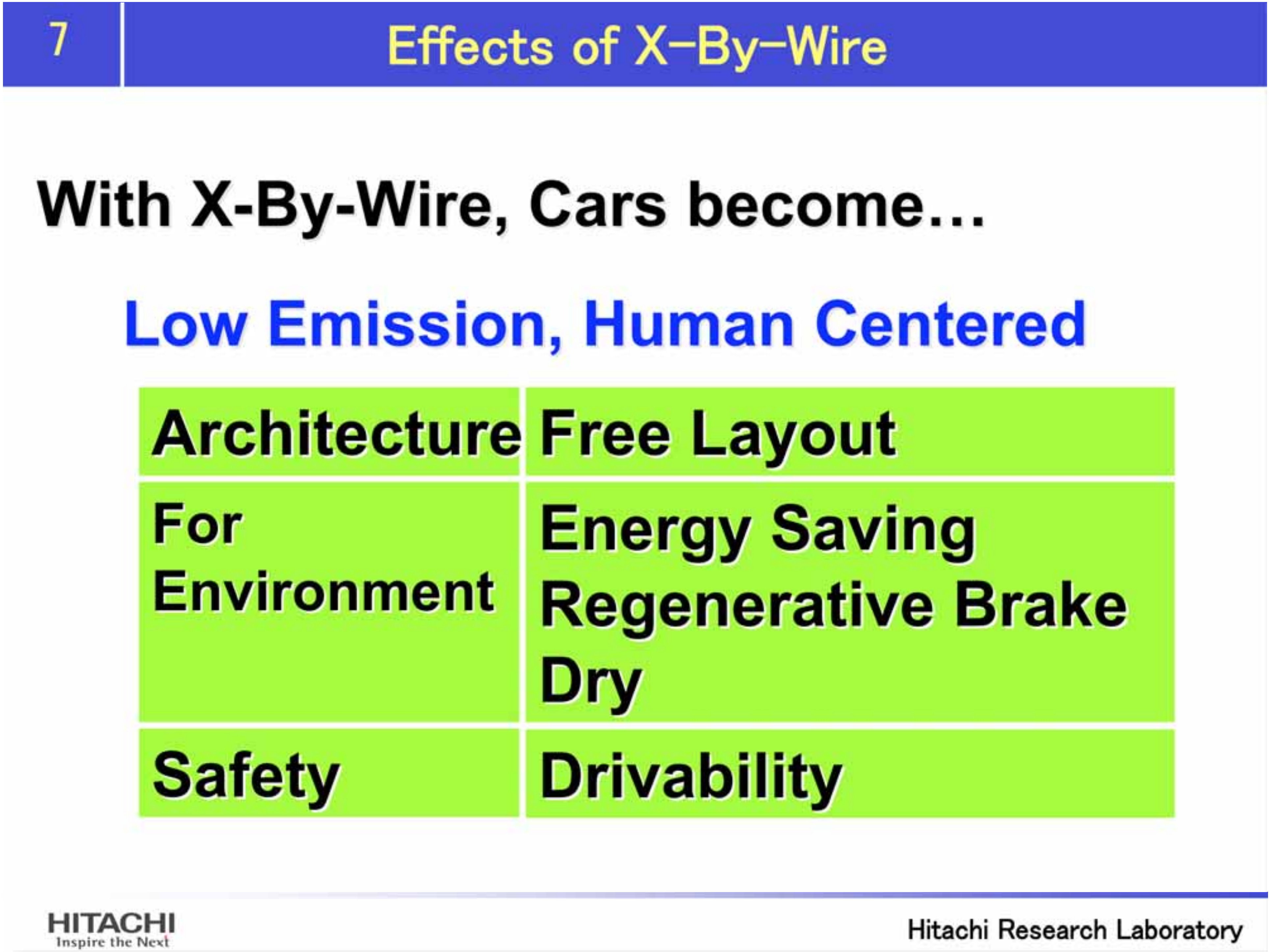


EHB :Electro-Hydraulic Brake

Steer-by-Wire

【Variable Gear Ratio】 Honda/S8000

【 Variable Gear Ratio 】 Toyota/Land Cruiser



7

Effects of X-By-Wire

With X-By-Wire, Cars become…

Low Emission, Human Centered

Architecture	Free Layout
For Environment	Energy Saving Regenerative Brake Dry
Safety	Drivability

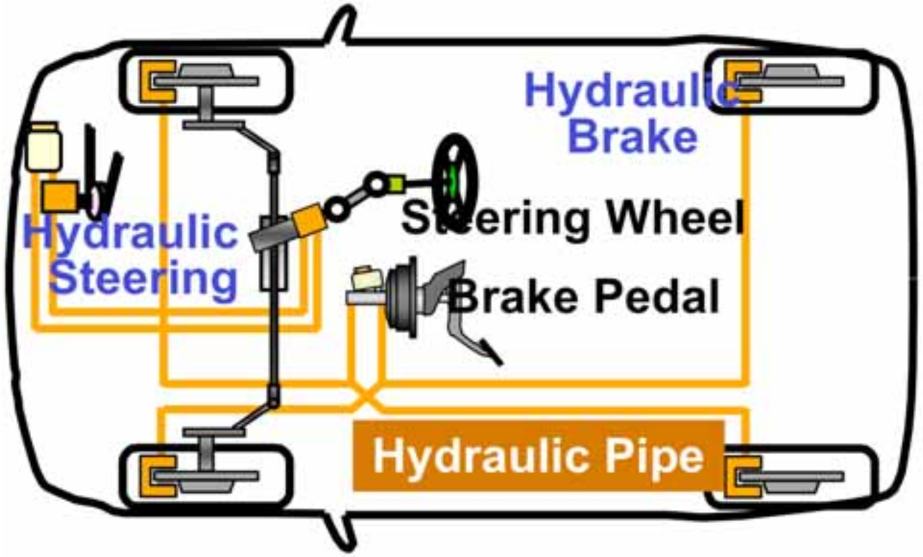
HITACHI
Inspire the Next

Hitachi Research Laboratory

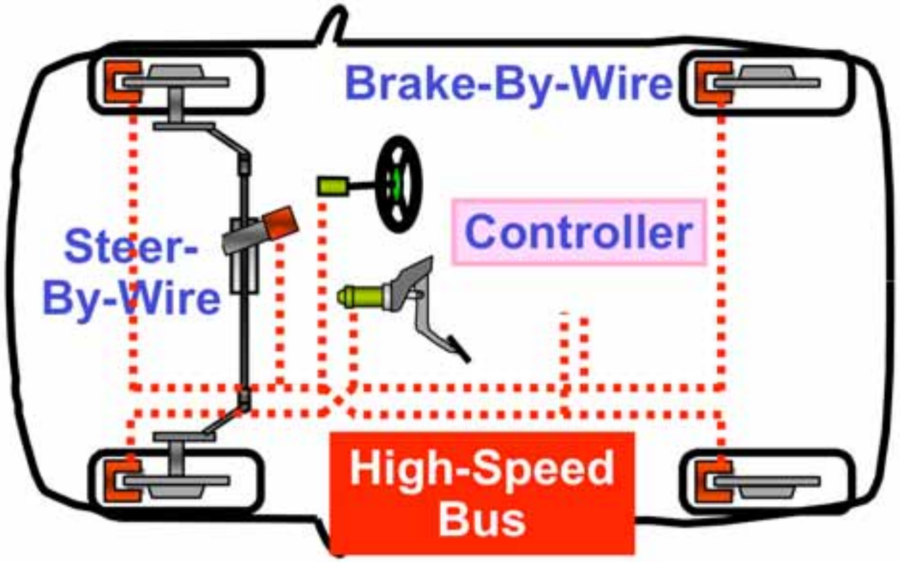
8

X-By-WireSystem

Conventional



X-By-Wire System



Mechanically Connected among Actuators and Mechanics

Integrated Control Connecting Components via High-Speed Bus

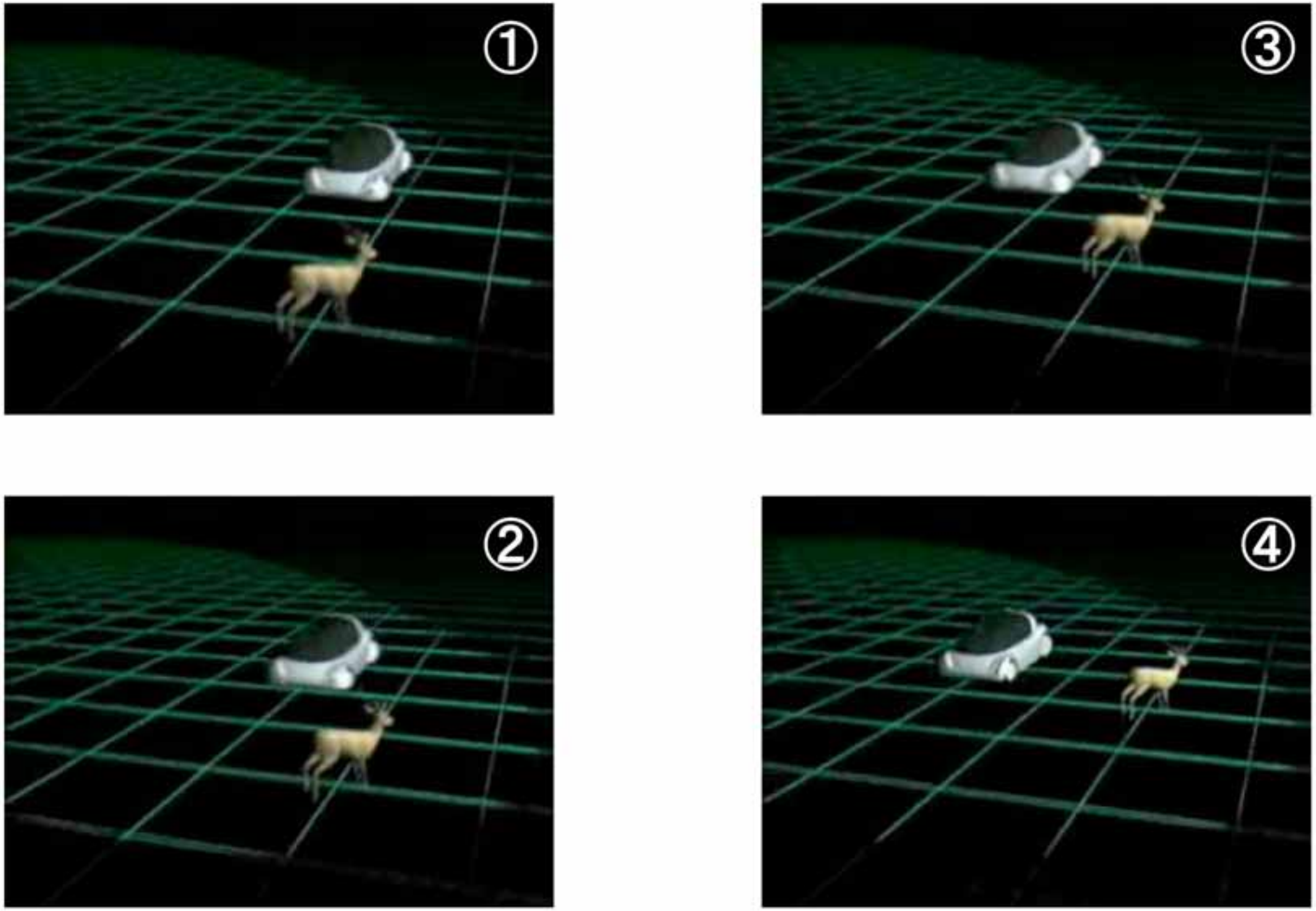
HITACHI
Inspire the Next

Hitachi Research Laboratory

9

Hitachi’s Concept Car

From Promotion Video at Tokyo Motor Show, 2003



HITACHI
Inspire the Next

Hitachi Research Laboratory

10 Inexpensive Dependability

“Aero-space is no longer high-tech.：

Reliability can be improved with cost.

X-By-Wire is the high-tech.， which realizes

dependability with low-cost.”

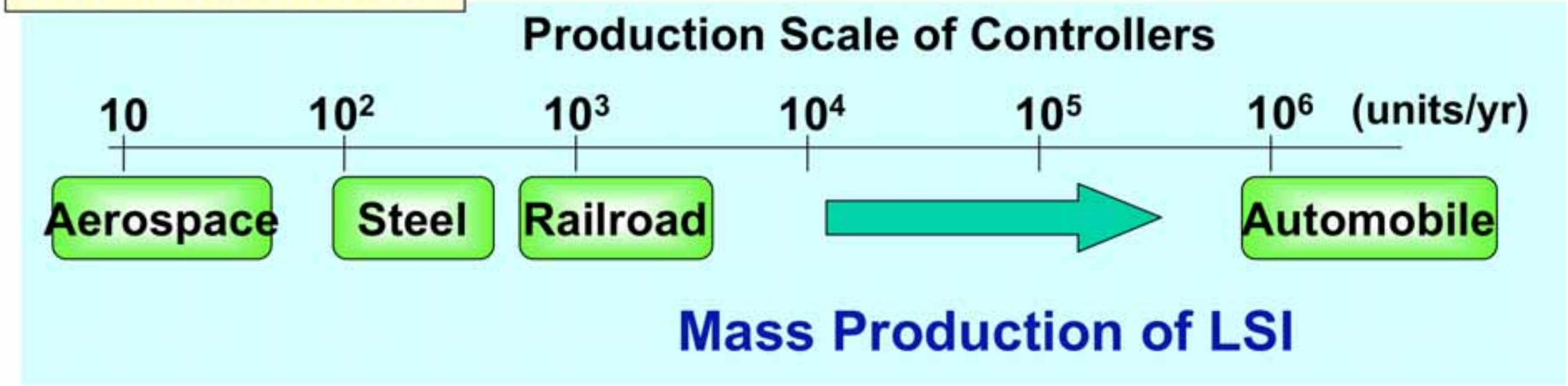
- Prof. M. Broy, Technical University of Munich (FTCS-28)

1. What's X-by-Wire ?

2. Our Approach

11 Low-Cost Dependable Technology

Cost Reduction



Production Scale of Controllers

10 | 10^2 | 10^3 | 10^4 | 10^5 | 10^6 (units/yr)

Aerospace | Steel | Railroad | Automobile

Mass Production of LSI

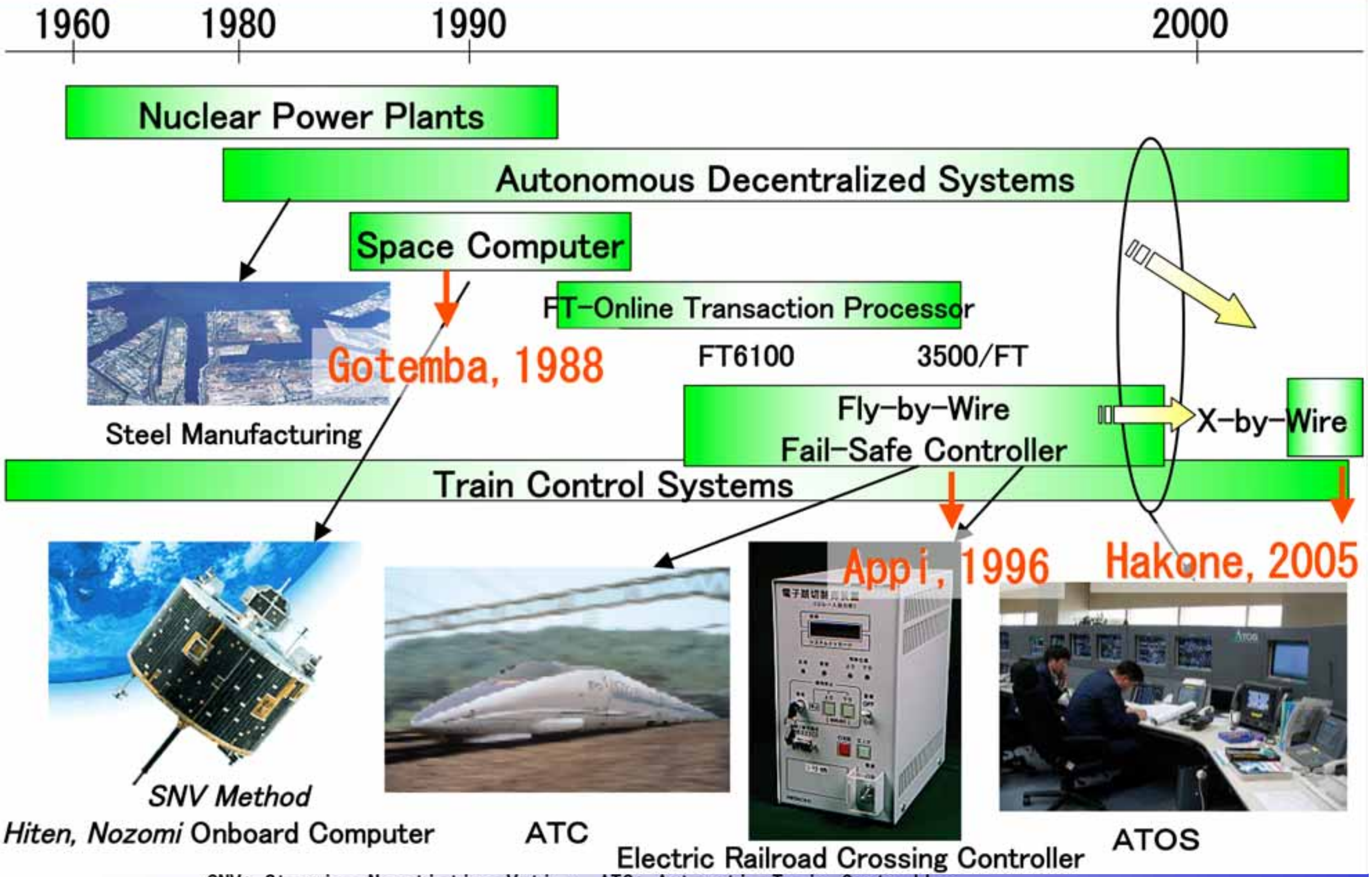
Low-Cost Dependability with LSI Technology

- ✓ Redundant CPUs in One Chip
- ✓ Self-Checking / Failsafe Technology
- ✓ Optimal Clock Diversity

and Autonomous Decentralized Concept

Hitachi Research Laboratory

12 Our Expertise in Dependability



1960 1980 1990 2000

Nuclear Power Plants

Autonomous Decentralized Systems

Space Computer

FT-Online Transaction Processor

Gotemba, 1988

FT6100 3500/FT

Steel Manufacturing

Fly-by-Wire

Fail-Safe Controller

X-by-Wire

Train Control Systems

App i, 1996

Hakone, 2005

SNV Method

Hiten, Nozomi Onboard Computer

ATC

Electric Railroad Crossing Controller

ATOS

SNV: Stepwise Negotiating Voting, ATC: Automatic Train Controller,
FTC: Fault-Tolerant Computer
ATOS: Autonomous Decentralized Transport Control System

HITACHI
Inspire the Next

Hitachi Research Laboratory

13 Making Controllers Dependable : Dual CPUs

Dual CPU Controller

Compares outputs of two CPUs.

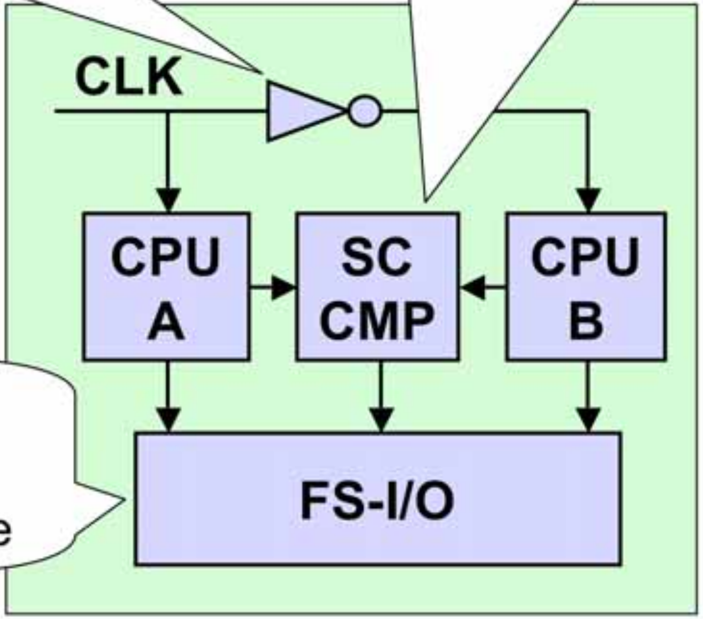
Takes fail-safe mode operation, if there is a difference.

Optimal Clock Diversity

- ➢ Operates CPUs out-of-phase
- ➢ Improves noise immunity

Self-Checking Comparator

- ➢ Self-checks its own operation
- ➢ Compares output of two CPUs



Fail-Safe I/O

- ➢ Executes fail-safe mode operation on CPU failure

Applications

- ･Digital ATC
- ･Crossing Controller

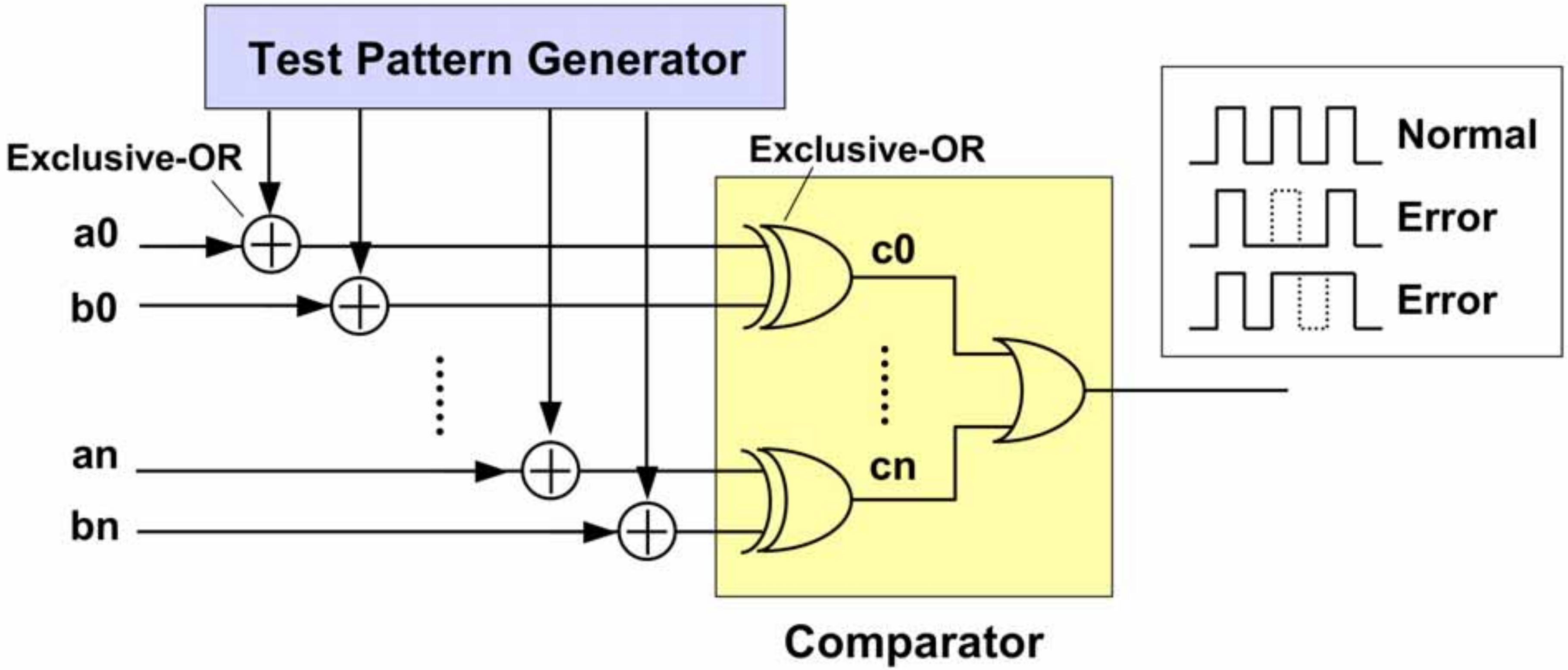
14

Self-Checking Comparator

Comparison of two outputs

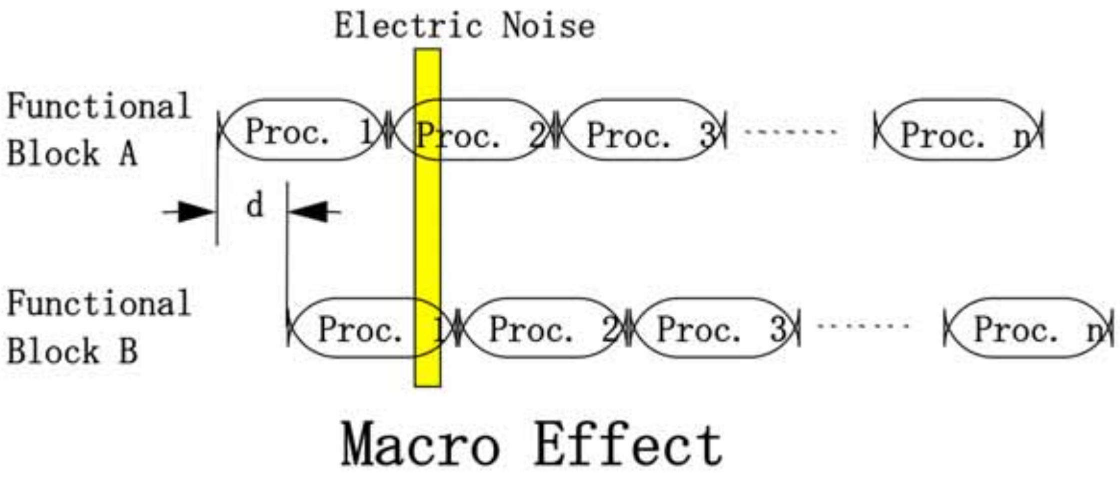
CPU A	a0	a1	a2				an
CPU B	b0	b1	b2				bn

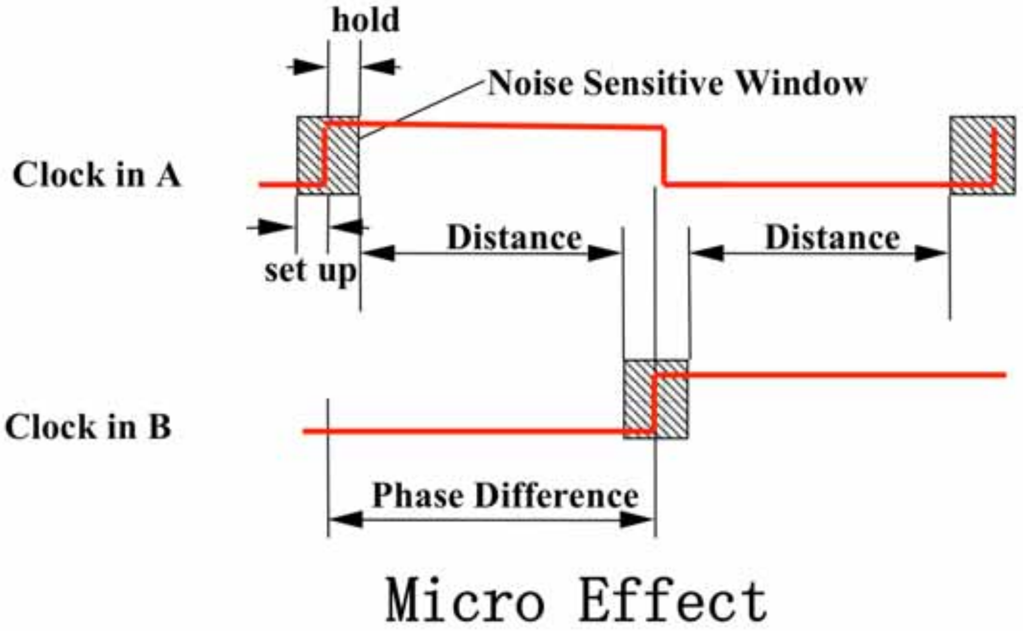
<u>Test Pattern Generator</u> inputs cyclic error signals intentionally



15

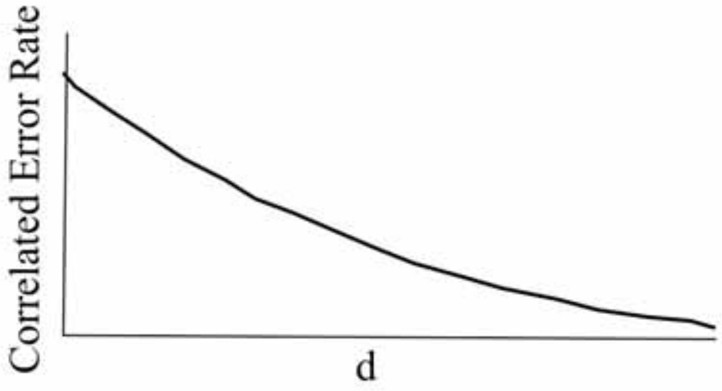
Effects of Time Diversity





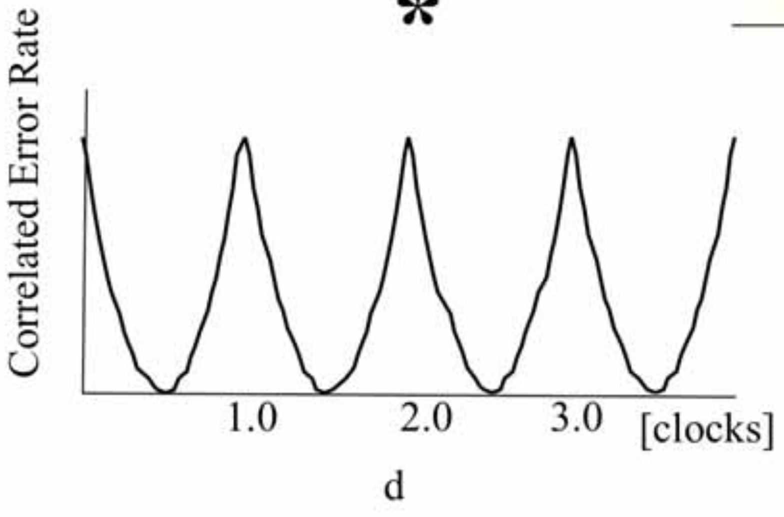
16

Effects of Time Diversity

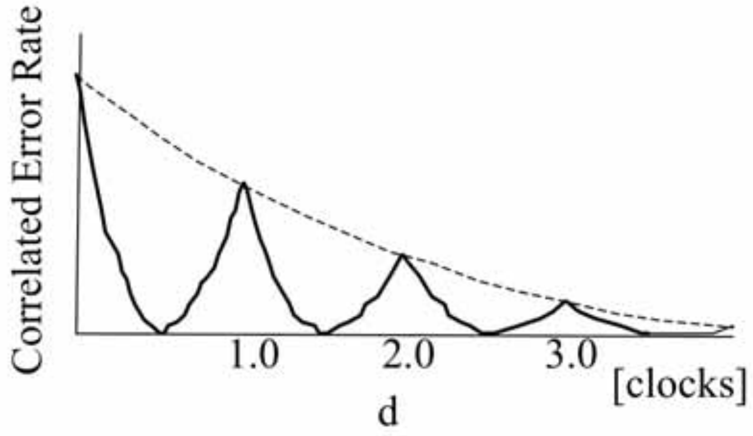


(a) Macro Effect

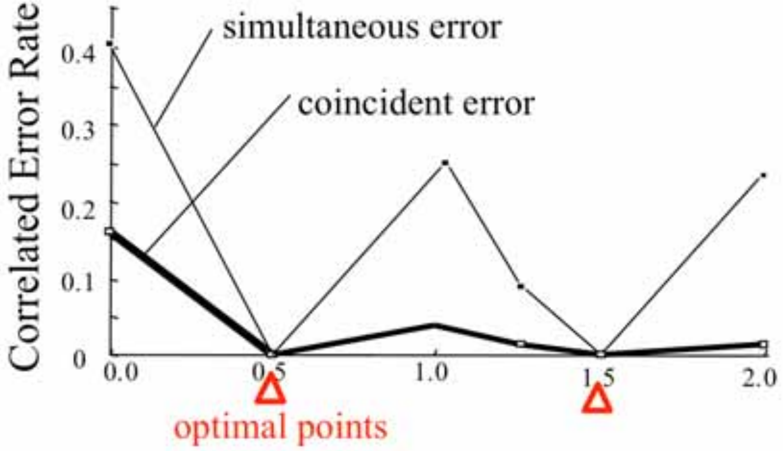
*



(b) Micro Effect



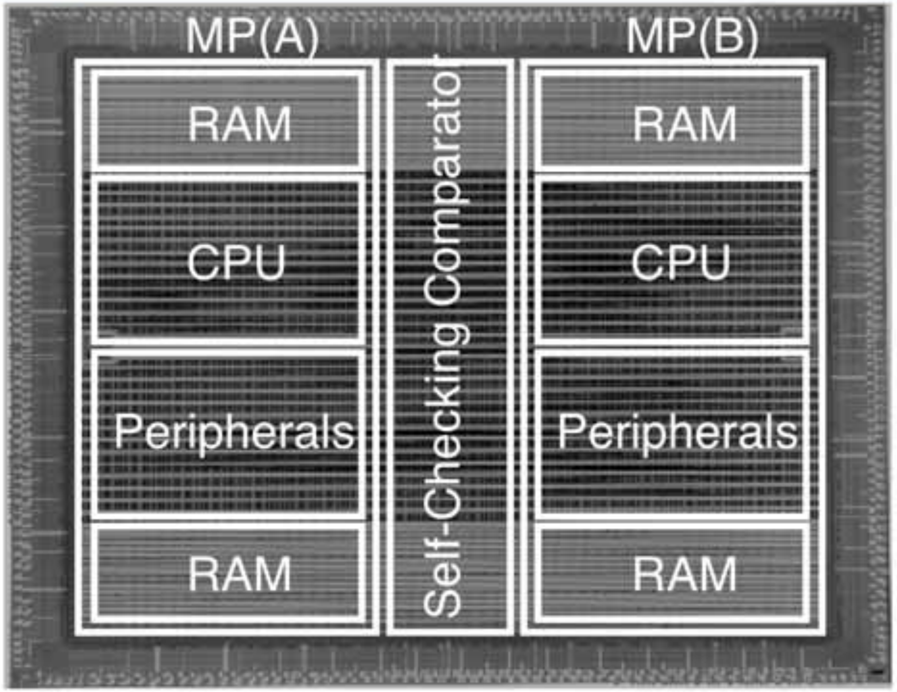
(c) Overall Effect

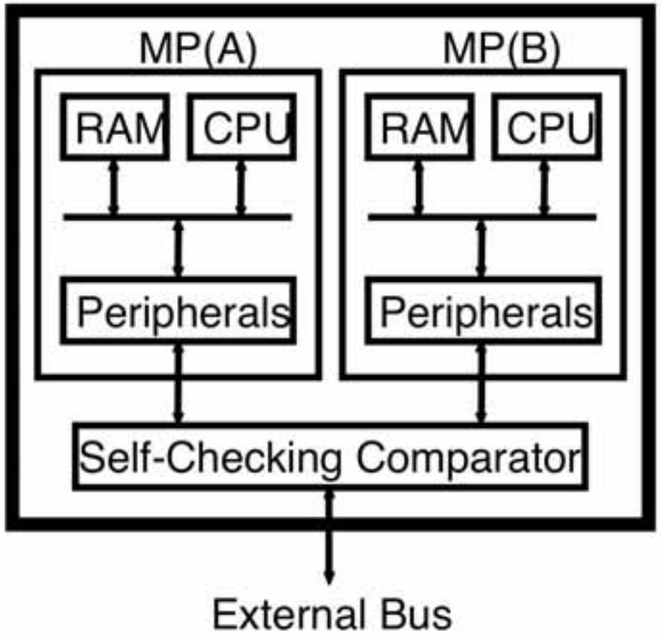


Experimental Results

17

Intra-Chip Redundancy CPU（FUJINE）

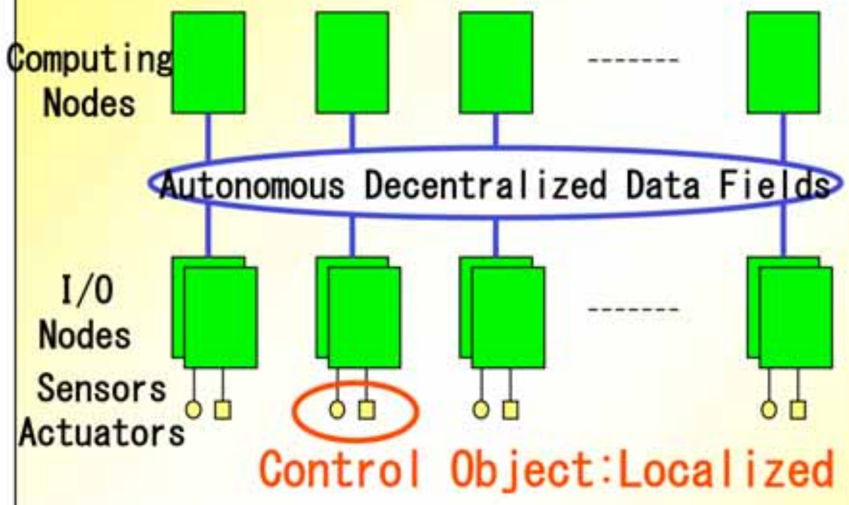




Process	0.35 µm 5 Metal CMOS
Hard Macros	PLL x 2, RAM(40KB)
Random Logic	740k gates
Chip Size	14.75 mm□
Operating Frequency	60 MHz
Power Dissipation	2.6W @ 60MHz
Package	479pin BGA

18 Autonomous Decentralized Systems

For Most of Applications



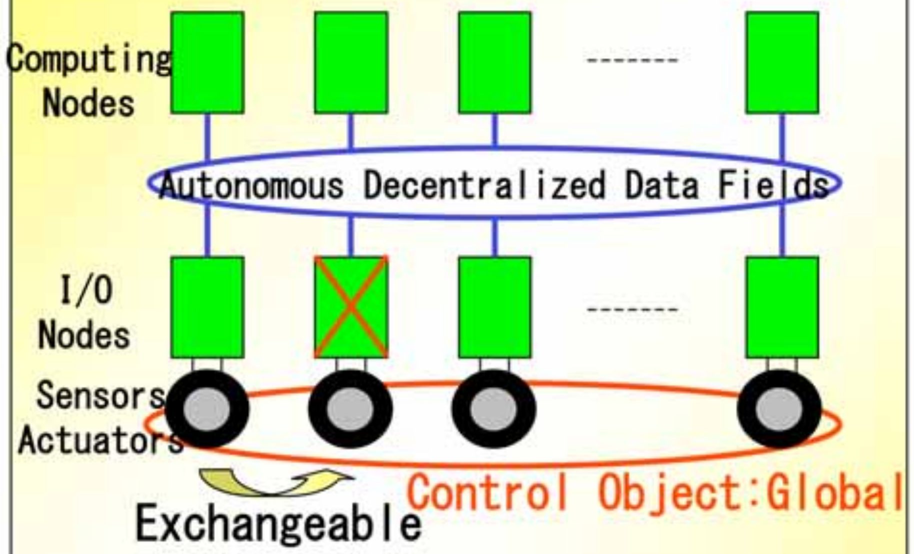
Merits;

✓ Online System-Expansion

✓ Computing Node Dependability

(I/O nodes should be redundant)

For Automotive/Aerospace

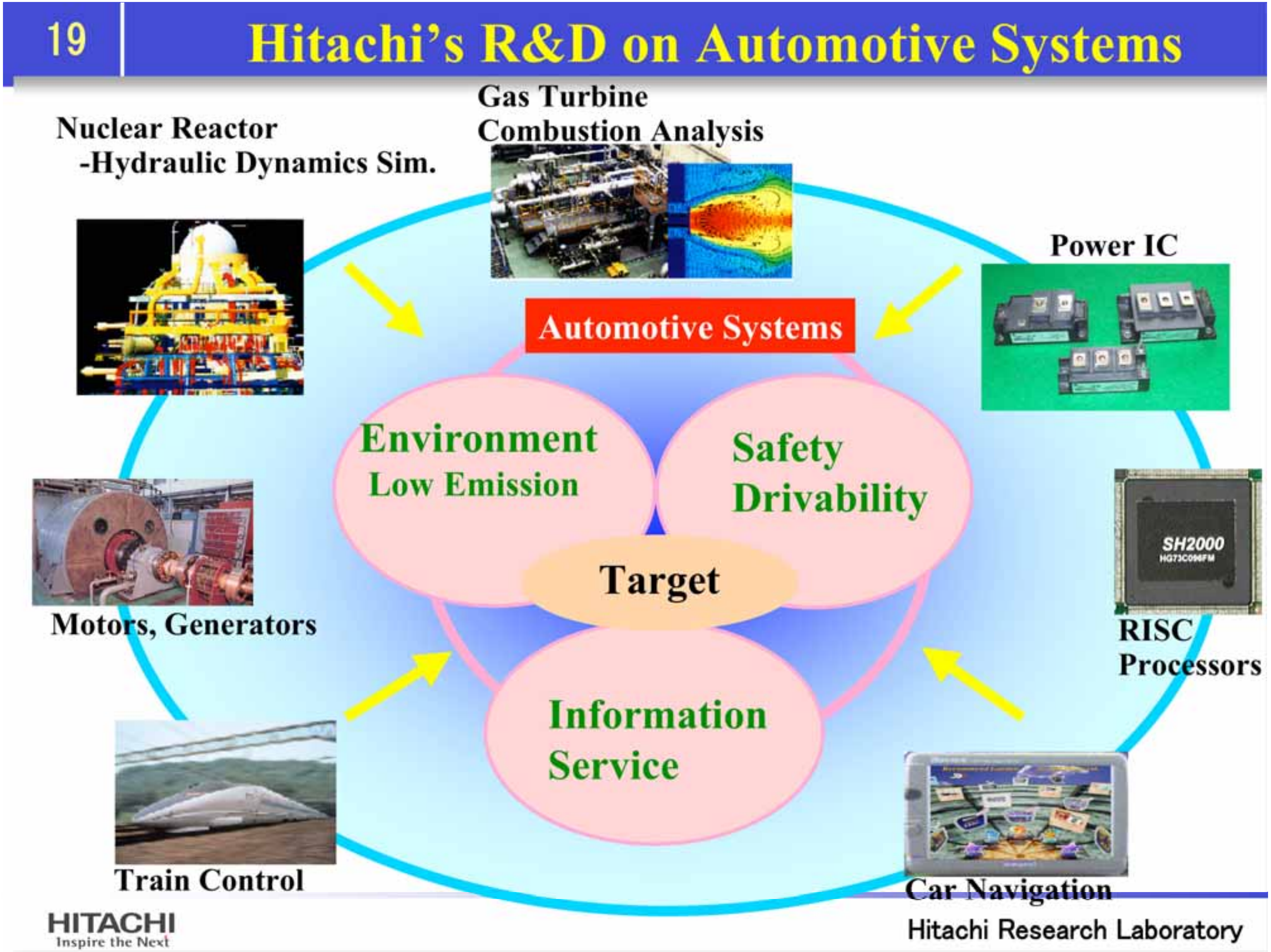


Merits;

✓ Online System-Expansion

✓ Computing Node Dependability

✓ I/O Node Dependability w/o Redundancy



19

Hitachi's R&D on Automotive Systems

Nuclear Reactor
-Hydraulic Dynamics Sim.

Gas Turbine
Combustion Analysis

Power IC

Automotive Systems

Environment
Low Emission

Safety
Drivability

Target

Information
Service

Motors, Generators

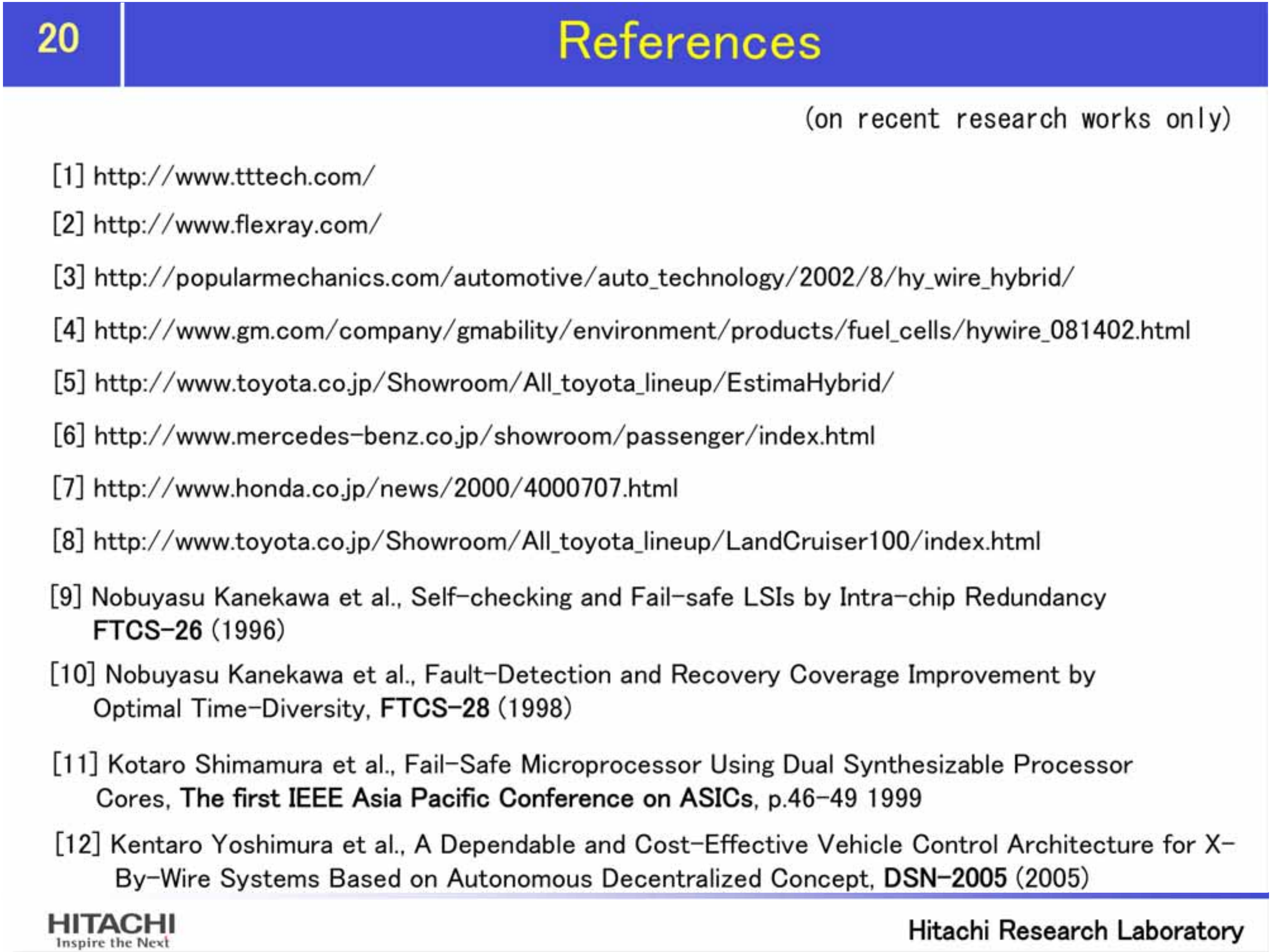
RISC
Processors

Train Control

Car Navigation

HITACHI
Inspire the Next

Hitachi Research Laboratory



20

References

(on recent research works only)

[1] http://www.tttech.com/

[2] http://www.flexray.com/

[3] http://popularmechanics.com/automotive/auto_technology/2002/8/hy_wire_hybrid/

[4] http://www.gm.com/company/gmability/environment/products/fuel_cells/hywire_081402.html

[5] http://www.toyota.co.jp/Showroom/All_toyota_lineup/EstimaHybrid/

[6] http://www.mercedes-benz.co.jp/showroom/passenger/index.html

[7] http://www.honda.co.jp/news/2000/4000707.html

[8] http://www.toyota.co.jp/Showroom/All_toyota_lineup/LandCruiser100/index.html

[9] Nobuyasu Kanekawa et al., Self-checking and Fail-safe LSIs by Intra-chip Redundancy **FTCS-26** (1996)

[10] Nobuyasu Kanekawa et al., Fault-Detection and Recovery Coverage Improvement by Optimal Time-Diversity, **FTCS-28** (1998)

[11] Kotaro Shimamura et al., Fail-Safe Microprocessor Using Dual Synthesizable Processor Cores, **The first IEEE Asia Pacific Conference on ASICs**, p.46-49 1999

[12] Kentaro Yoshimura et al., A Dependable and Cost-Effective Vehicle Control Architecture for X-By-Wire Systems Based on Autonomous Decentralized Concept, **DSN-2005** (2005)

HITACHI
Inspire the Next

Hitachi Research Laboratory

HITACHI

Inspire the Next

Reflection oriented Dependable Planning Concept (RDPC) and its Application to the learning in Education and in Intelligent Agent

Setsuo Tsuruta,
Shinichi Dohi,
Shogo Nakamura
Tokyo Denki University

Background

- Real world, such as learning, is complex, and perfect planning is difficult.
- Problems in its execution and potential capabilities are often found during execution.
- Thus, dependable planning can be defined as such contributes to attain as much as possible.
- It is inherently fault-tolerant, for plans, constrains, or even goals are changed during execution.
- They often seems opportunistically changed but reflection including profound consideration is more important

Need

- Dependable planning needs periodical repetition of plan generation / modification, plan execution, and reflective evaluation.
- Efficient or, at least, serious execution and its evaluation are necessary to attain as much as possible.
- Rather than being opportunistic, planning by reflection such as sufficient consideration about goal and execution results is needed to discover or acquire capability and to cope with encountered problems.
- Computer support for dependable planning is needed, since complex constraints exist in practical planning.
- Orientation to obtain the knowledge for using it seems necessary, since dependable planning is complicate.

Reflection oriented Dependable Planning Concept (RDPC)

- Flexible structure/condition for Planning
  - Flexible conditions such as strict and desirable levels of constraints for planning
- Repetitive planning through stepwise Reflection on evaluation results of the efficient execution
  - Stepwise Reflection and plan modification based on efficient execution & cost/performance evaluation
- Systemization and Orientation
  - Support System: Plan check & simulation tool
  - Orientation/ training to use the system/tool

Concept of Applying RDPC to Education in our school: SIE

- Flexible structure/condition for Planning
  - –No academic year but only Semester, no compulsory subject
  - –Prerequisite and recommended constraints for planning
- Repetitive planning through stepwise Reflection on evaluation results of the efficient execution
  - –Short Class Period (50 minutes class) for Efficient Execution
  - –Quick feedback of Class evaluation for Efficient Execution
  - –Credit System (tuition fee per subject) for Cost Evaluation
  - –GPA (Grade Point Average) for Performance Evaluation
- Systemization and Orientation
  - –Dynamic Syllabus tool (Systemization)
  - –Curriculum Planning class for Training (Orientation)

Quick feedback of class evaluation

(to improve quality of education for efficient learning plan execution)

- Using the Web, Students can comment and request to the class
  - Students are willing to attend classes
  - Students can see that other students also do not understand key items, by looking at the Web.
- Quick feedback on the class
  - Grasp students’ understanding level and feedback

Credit System

(Reminding students of the course price for efficient learning plan execution)

- **One unit = \15,700**
  - **Around 60k\ tuition fee per subject (4 units course)**
- **Effects**
  - **Few students drop (give up) courses**

GPA (Grade Point Average: for efficient learning plan execution)

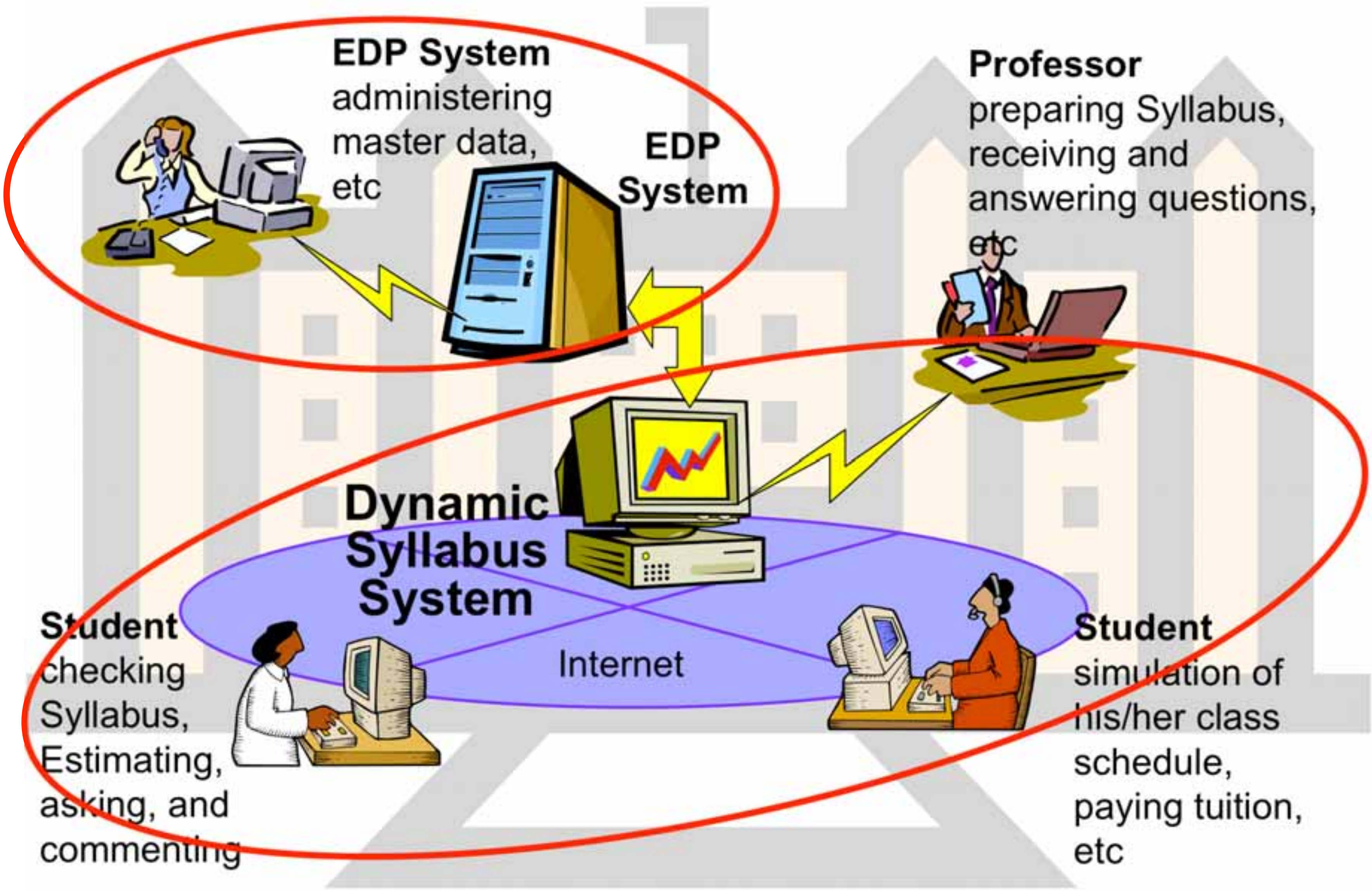
$$GPA = \frac{\sum_{i=1}^{n} u_i g_i}{\sum_{i=1}^{n} r_i}$$

Units of each course（*u*）
Grade point acquired for each course（*g*）

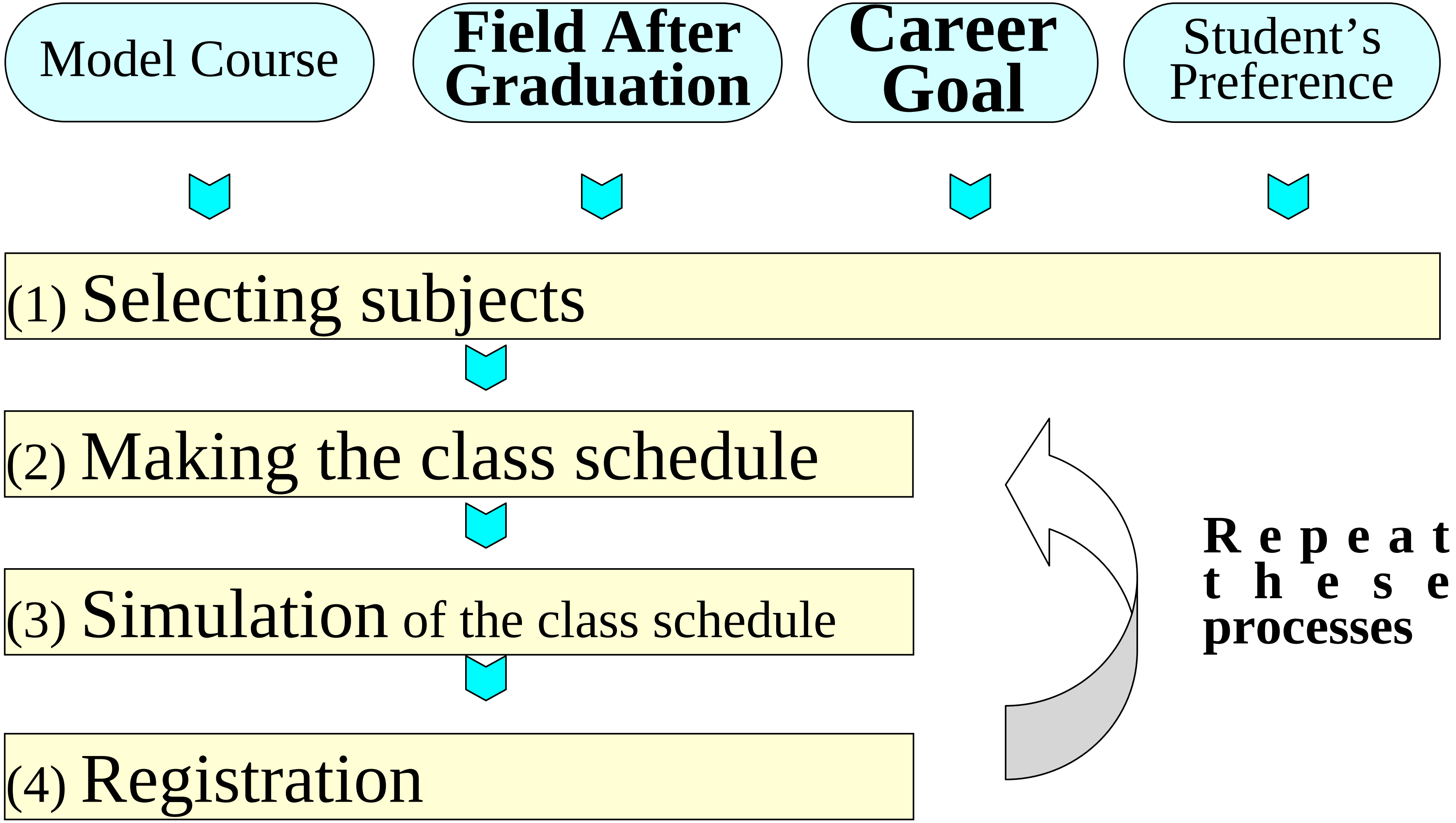
Units of registration（*r*）

Rank(Score)	Grade Point
S(90 –)	4
A(80 – 89)	4
B(70 – 79)	3
C(60 – 69)	2
D(40 – 59)	0
E(– 39)	0

Conceptual Architecture of Dynamic Syllabus (DS) tool for RDPC



Process of planning Student's own curriculum (class schedule)



Model Course

Field After Graduation

Career Goal

Student's Preference

(1) Selecting subjects

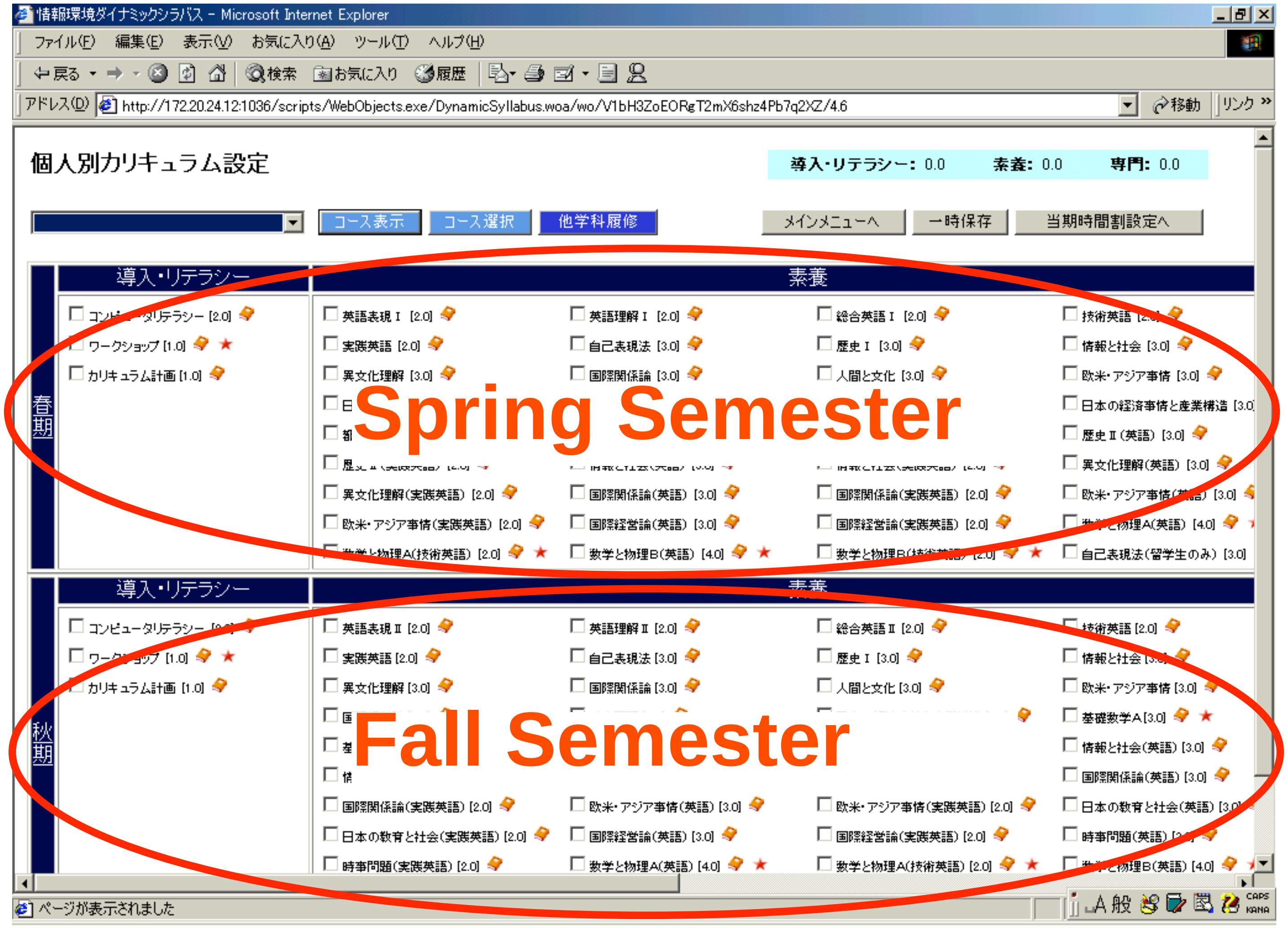
(2) Making the class schedule

(3) Simulation of the class schedule

(4) Registration

Repeat these processes

Curriculum set-up window


情報環境ダイナミックシラバス - Microsoft Internet Explorer
ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)
戻る 検索 お気に入り 履歴
アドレス(D)
移動 リンク
コース表示
コース選択
他学科履修
メインメニューへ
一時保存
当期時間割設定へ
春期
秋期
ページが表示されました

http://172.20.24.12:1036/scripts/WebObjects.exe/DynamicSyllabus.woa/wo/V1bH3ZoEORgT2mX6shz4Pb7q2XZ/4.6

個人別カリキュラム設定

導入・リテラシー: 0.0 素養: 0.0 専門: 0.0

Spring Semester

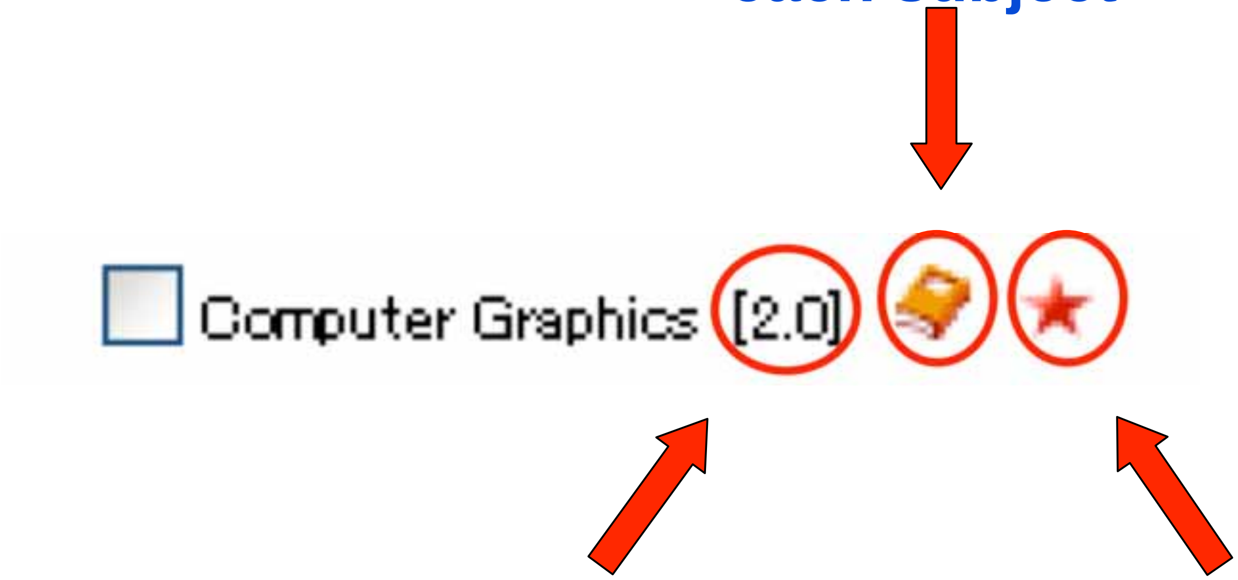
導入・リテラシー	素養			
☐ コンピュータリテラシー [2.0]	☐ 英語表現Ⅰ [2.0]	☐ 英語理解Ⅰ [2.0]	☐ 総合英語Ⅰ [2.0]	☐ 技術英語 [2.0]
☐ ワークショップ [1.0] ★	☐ 実践英語 [2.0]	☐ 自己表現法 [3.0]	☐ 歴史Ⅰ [3.0]	☐ 情報と社会 [3.0]
☐ カリキュラム計画 [1.0]	☐ 異文化理解 [3.0]	☐ 国際関係論 [3.0]	☐ 人間と文化 [3.0]	☐ 欧米・アジア事情 [3.0]
	☐ 日[illegible]	[illegible]	[illegible]	☐ 日本の経済事情と産業構造 [3.0]
	☐ [illegible]	[illegible]	[illegible]	☐ 歴史Ⅱ(英語) [3.0]
	☐ 歴史Ⅱ(実践英語) [2.0]	☐ 情報と社会(英語) [3.0]	☐ 情報と社会(実践英語) [2.0]	☐ 異文化理解(英語) [3.0]
	☐ 異文化理解(実践英語) [2.0]	☐ 国際関係論(英語) [3.0]	☐ 国際関係論(実践英語) [2.0]	☐ 欧米・アジア事情(英語) [3.0]
	☐ 欧米・アジア事情(実践英語) [2.0]	☐ 国際経営論(英語) [3.0]	☐ 国際経営論(実践英語) [2.0]	☐ 数学と物理A(英語) [4.0] ★
	☐ 数学と物理A(技術英語) [2.0] ★	☐ 数学と物理B(英語) [4.0] ★	☐ 数学と物理B(技術英語) [2.0] ★	☐ 自己表現法(留学生のみ) [3.0]

Fall Semester

導入・リテラシー	素養			
☐ コンピュータリテラシー [2.0]	☐ 英語表現Ⅱ [2.0]	☐ 英語理解Ⅱ [2.0]	☐ 総合英語Ⅱ [2.0]	☐ 技術英語 [2.0]
☐ ワークショップ [1.0] ★	☐ 実践英語 [2.0]	☐ 自己表現法 [3.0]	☐ 歴史Ⅰ [3.0]	☐ 情報と社会 [3.0]
☐ カリキュラム計画 [1.0]	☐ 異文化理解 [3.0]	☐ 国際関係論 [3.0]	☐ 人間と文化 [3.0]	☐ 欧米・アジア事情 [3.0]
	☐ 国[illegible]	[illegible]	[illegible]	☐ 基礎数学A [3.0] ★
	☐ 基[illegible]	[illegible]	[illegible]	☐ 情報と社会(英語) [3.0]
	☐ 情[illegible]	[illegible]	[illegible]	☐ 国際関係論(英語) [3.0]
	☐ 国際関係論(実践英語) [2.0]	☐ 欧米・アジア事情(英語) [3.0]	☐ 欧米・アジア事情(実践英語) [2.0]	☐ 日本の教育と社会(英語) [3.0]
	☐ 日本の教育と社会(実践英語) [2.0]	☐ 国際経営論(英語) [3.0]	☐ 国際経営論(実践英語) [2.0]	☐ 時事問題(英語) [illegible]
	☐ 時事問題(実践英語) [2.0]	☐ 数学と物理A(英語) [4.0] ★	☐ 数学と物理A(技術英語) [2.0] ★	☐ 数学と物理B(英語) [4.0] ★

Information of each subject

Syllabus of each subject

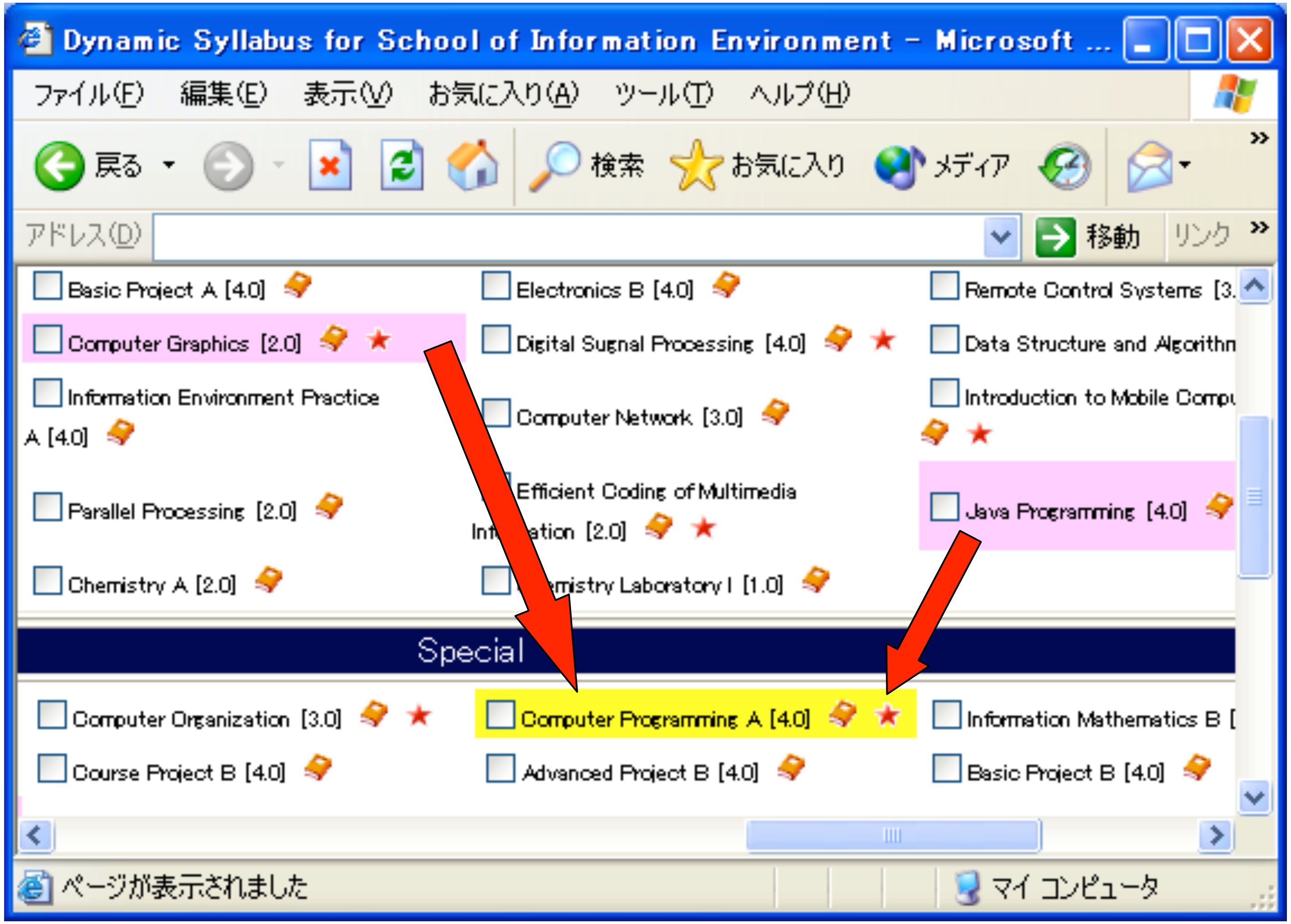


□ Computer Graphics [2.0]

The number of credits

Prerequisite condition

Prerequisite Condition


Dynamic Syllabus for School of Information Environment - Microsoft ...
ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)
戻る
検索
お気に入り
メディア
アドレス(D)
移動
リンク
ページが表示されました
マイ コンピュータ

□ Basic Project A [4.0]

□ Computer Graphics [2.0] ★

□ Information Environment Practice A [4.0]

□ Parallel Processing [2.0]

□ Chemistry A [2.0]

□ Electronics B [4.0]

□ Digital Sugnal Processing [4.0] ★

□ Computer Network [3.0]

□ Efficient Coding of Multimedia Inf ation [2.0] ★

□ mistry Laboratory I [1.0]

□ Remote Control Systems [3.

□ Data Structure and Algorithn

□ Introduction to Mobile Compu ★

□ Java Programming [4.0]

Special

□ Computer Organization [3.0] ★

□ Course Project B [4.0]

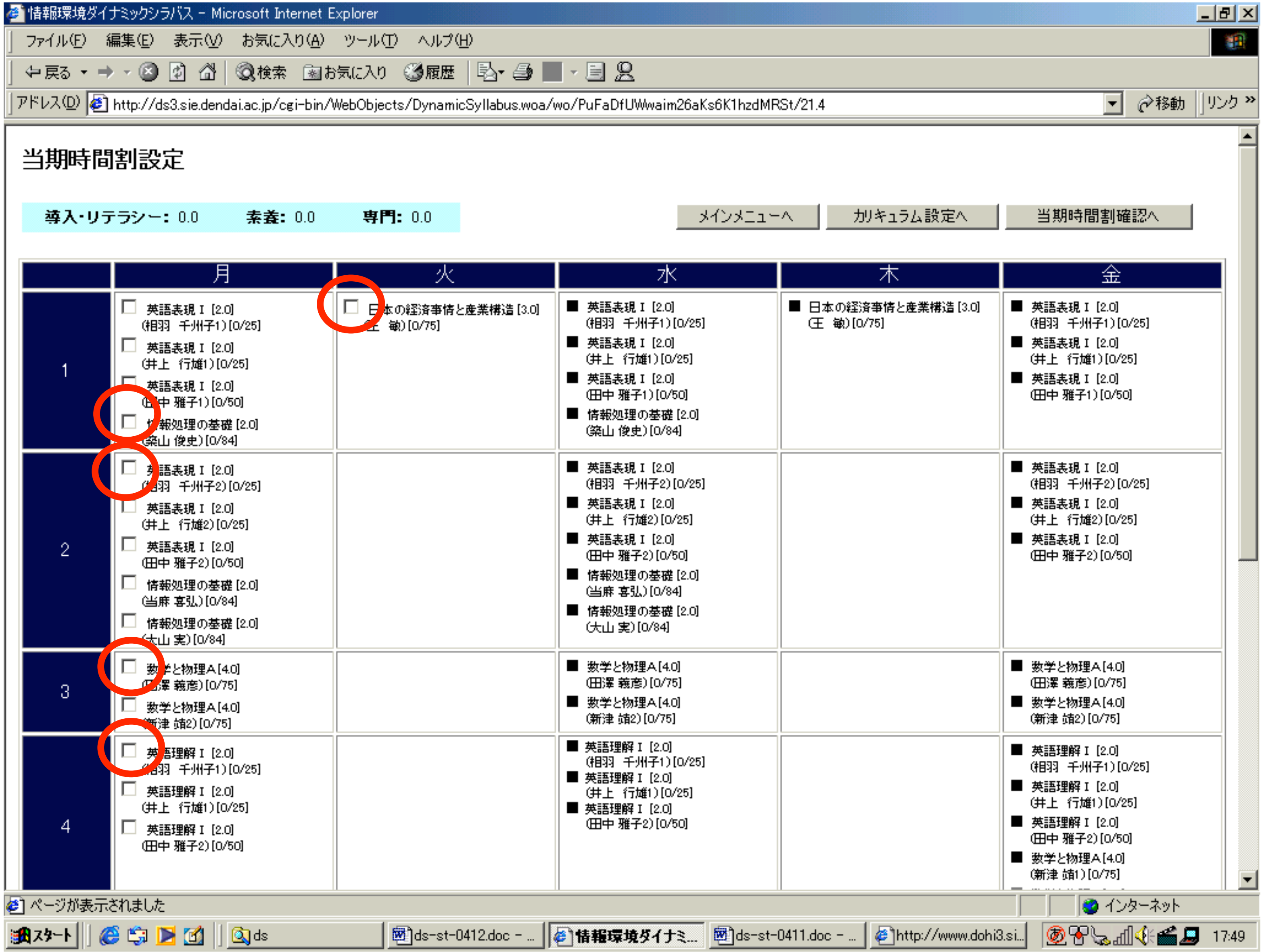
□ Computer Programming A [4.0] ★

□ Advanced Project B [4.0]

□ Information Mathematics B [

□ Basic Project B [4.0]

Prototype of Class Schedule


情報環境ダイナミックシラバス - Microsoft Internet Explorer
ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)
戻る 検索 お気に入り 履歴
アドレス(D) http://ds3.sie.dendai.ac.jp/cgi-bin/WebObjects/DynamicSyllabus.woa/wo/PuFaDfUWwaim26aKs6K1hzdMRSt/21.4
移動 リンク
ページが表示されました
インターネット
スタート ds ds-st-0412.doc - ... 情報環境ダイナミ... ds-st-0411.doc - ... http://www.dohi3.si... 17:49

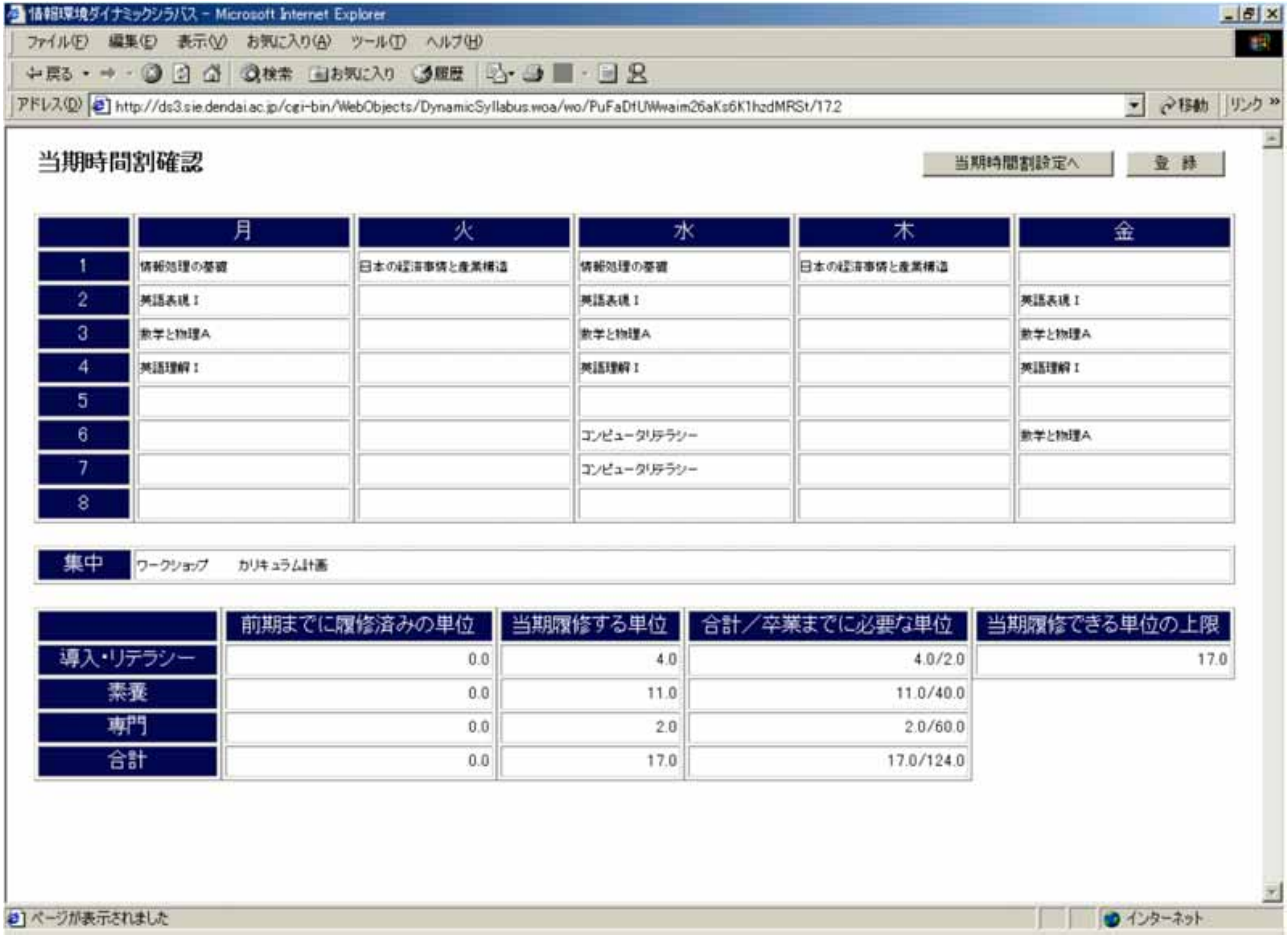
当期時間割設定

導入・リテラシー: 0.0 **素養:** 0.0 **専門:** 0.0

メインメニューへ
カリキュラム設定へ
当期時間割確認へ

	月	火	水	木	金
1	□ 英語表現Ⅰ [2.0] (相羽　千州子1) [0/25] □ 英語表現Ⅰ [2.0] (井上　行雄1) [0/25] □ 英語表現Ⅰ [2.0] (田中 雅子1) [0/50] □ 情報処理の基礎 [2.0] (築山 俊史) [0/84]	□ 日本の経済事情と産業構造 [3.0] (王　敏) [0/75]	■ 英語表現Ⅰ [2.0] (相羽　千州子1) [0/25] ■ 英語表現Ⅰ [2.0] (井上　行雄1) [0/25] ■ 英語表現Ⅰ [2.0] (田中 雅子1) [0/50] ■ 情報処理の基礎 [2.0] (築山 俊史) [0/84]	■ 日本の経済事情と産業構造 [3.0] (王　敏) [0/75]	■ 英語表現Ⅰ [2.0] (相羽　千州子1) [0/25] ■ 英語表現Ⅰ [2.0] (井上　行雄1) [0/25] ■ 英語表現Ⅰ [2.0] (田中 雅子1) [0/50]
2	□ 英語表現Ⅰ [2.0] (相羽　千州子2) [0/25] □ 英語表現Ⅰ [2.0] (井上　行雄2) [0/25] □ 英語表現Ⅰ [2.0] (田中 雅子2) [0/50] □ 情報処理の基礎 [2.0] (当麻 喜弘) [0/84] □ 情報処理の基礎 [2.0] (大山 実) [0/84]		■ 英語表現Ⅰ [2.0] (相羽　千州子2) [0/25] ■ 英語表現Ⅰ [2.0] (井上　行雄2) [0/25] ■ 英語表現Ⅰ [2.0] (田中 雅子2) [0/50] ■ 情報処理の基礎 [2.0] (当麻 喜弘) [0/84] ■ 情報処理の基礎 [2.0] (大山 実) [0/84]		■ 英語表現Ⅰ [2.0] (相羽　千州子2) [0/25] ■ 英語表現Ⅰ [2.0] (井上　行雄2) [0/25] ■ 英語表現Ⅰ [2.0] (田中 雅子2) [0/50]
3	□ 数学と物理A [4.0] (田澤 義彦) [0/75] □ 数学と物理A [4.0] (新津 靖2) [0/75]		■ 数学と物理A [4.0] (田澤 義彦) [0/75] ■ 数学と物理A [4.0] (新津 靖2) [0/75]		■ 数学と物理A [4.0] (田澤 義彦) [0/75] ■ 数学と物理A [4.0] (新津 靖2) [0/75]
4	□ 英語理解Ⅰ [2.0] (相羽　千州子1) [0/25] □ 英語理解Ⅰ [2.0] (井上　行雄1) [0/25] □ 英語理解Ⅰ [2.0] (田中 雅子2) [0/50]		■ 英語理解Ⅰ [2.0] (相羽　千州子1) [0/25] ■ 英語理解Ⅰ [2.0] (井上　行雄1) [0/25] ■ 英語理解Ⅰ [2.0] (田中 雅子2) [0/50]		■ 英語理解Ⅰ [2.0] (相羽　千州子1) [0/25] ■ 英語理解Ⅰ [2.0] (井上　行雄1) [0/25] ■ 英語理解Ⅰ [2.0] (田中 雅子2) [0/50] ■ 数学と物理A [4.0] (新津 靖1) [0/75]

Completion of Class Schedule


情報環境ダイナミックシラバス - Microsoft Internet Explorer
ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)
戻る 検索 お気に入り 履歴
アドレス(D)
移動
リンク
当期時間割設定へ
登 録
ページが表示されました
インターネット

http://ds3.sie.dendai.ac.jp/cgi-bin/WebObjects/DynamicSyllabus.woa/wo/PuFaDfUWwaim26aKs6K1hzdMRSt/17.2

当期時間割確認

	月	火	水	木	金
1	情報処理の基礎	日本の経済事情と産業構造	情報処理の基礎	日本の経済事情と産業構造	
2	英語表現 I		英語表現 I		英語表現 I
3	数学と物理A		数学と物理A		数学と物理A
4	英語理解 I		英語理解 I		英語理解 I
5					
6			コンピュータリテラシー		数学と物理A
7			コンピュータリテラシー		
8					

集中	ワークショップ　　カリキュラム計画

	前期までに履修済みの単位	当期履修する単位	合計／卒業までに必要な単位	当期履修できる単位の上限
導入・リテラシー	0.0	4.0	4.0/2.0	17.0
素養	0.0	11.0	11.0/40.0	
専門	0.0	2.0	2.0/60.0	
合計	0.0	17.0	17.0/124.0	

Effects of RDPC on dependable planning in Practical Education

- Effects of stepwise Reflection on evaluation results of the efficient execution
  - Effects of GPA and Credit System (50% dropped)
  - Effects of Short Class Period (30 % vs. 5 % failed in Exam.)
- Effects of Systemization and Orientation
  - 93.6 % created 4 years curriculum plan using DS tool/system
  - 80 % felt Training (Orientation) in Curriculum Planning Class is useful to learn how to use DS tool.

Applicability to intelligent agents (1)

- Using simulation system such as DS of RDPC, software agents can also create their learning plans, as follows.
  - Software agents should be trained how to use RDPC system (training).
  - Learning goals such as academic goals should be given.
  - Applications of learning subjects have to be derived from system functions which should be assisted by intelligent system assistants.
  - Evaluation method including grading points such as GPA and evaluation timing such a semester are also necessary.
  - All subjects should have the same evaluation timing to synchronously modify the learning plan, since they relate each other by prerequisite conditions and so on.
  - Each evaluation, the learning plans can be modified at each step, reflecting capabilities or various difficulties encountered while learning.
- Thus, they can learn efficiently and dependably towards their learning goal or attain as much as possible.

Applicability to intelligent agents (2)

- The more agents become intelligent, the more they become alike human. And they become sometimes too lazy to search, achieve, or satisfy a reasonable but hard goal or high level need.
- Application to education teaches the following.

“Through introducing severe but reasonable evaluation system e.g. GPA and a credit system, machine agents also are expected to be controlled as human lest they should be lazy or give up when they try to achieve or satisfy a difficult goal, sub-goal or need for learning.”

Applicability to intelligent agents (3)

- As to the learning goal, the DS tool for RDPC is restricted to the academic goal or the school age.
- Meanwhile, software agents should learn as long as they live or they are needed as intelligent system assistants.
- This is a kind of life learning in case of human.
- However, as to the application for intelligent software agents, it is also reasonable to have a restriction that the new learning subjects (new intelligent functions) do not appear in the same version/revision of software agents.
- Thus, the structure or order for learning subjects is fixed concerning such as prerequisite/desired conditions

Applicability to intelligent agents (4)

- If system functions increase in case of version-up, software agents will be given a new academic goal or a new system concept.
- Deriving a set of applicant learning subjects from added or modified functions of the new version, RDPC is possibly used as in an academic education.

Introduction of Story-board

- Though partly fixed or implicitly incorporated, concrete learning or didactic knowledge used in the practical education is in DS.
- Really, such knowledge is used in the practical education to exploit RDPC in the education (Curriculum Planning) of our school SIE.
- Story-board has no such concrete knowledge in its framework.
  - However it can represent concrete knowledge .
  - It is more general knowledge representation framework.
- Thus, in order to build a dependable planning system for intelligent agents' learning, Story-board is useful if the practically used learning (meta) knowledge such as those in DS of RDPC or more especially for software agents' learning is incorporated.

Conclusion

1. RDPC helped by its DS tool and training of its usage was proved effective to its application in education, namely in dependable curriculum planning by each students of our school SIE.
2. Furthermore, RDPC was found applicable to the learning of intelligent machine agents, especially through incorporating its conceptual or meta knowledge to general representation tool such as a story board

Experimental software risk assessment

Henrique Madeira
University of Coimbra, DEI-CISUC
Coimbra, Portugal

Universidade de Coimbra

Component-based software development

- **Vision:** development of systems using pre-fabricated components. Reuse custom components or buy software components available from software manufactures (Commercial-Off-The-Shelf: COTS).
- **Potential advantages:**
  - Reduce development effort since the components are already developed, tested, and matured by execution in different contexts
  - Improve system quality
  - Achieve of shorter time-to-market
  - Improve management of increased complexity of software
- **Trend** → **use general-purpose COTS components and develop domain specific components**.

Some potential problems

- COTS
  - In general, functionality descrition is not fully provided.
  - No guarantee of adequate testing.
  - COTS must be assessed in relation to their intended use.
  - The source code is normally not available (makes it impossible white box verification & validation of COTS).
- Reuse of custom components in a different context may expose components faults.

Using COTS (or reusing custom components) represent a risk!
How to assess (and reduce) that risk?

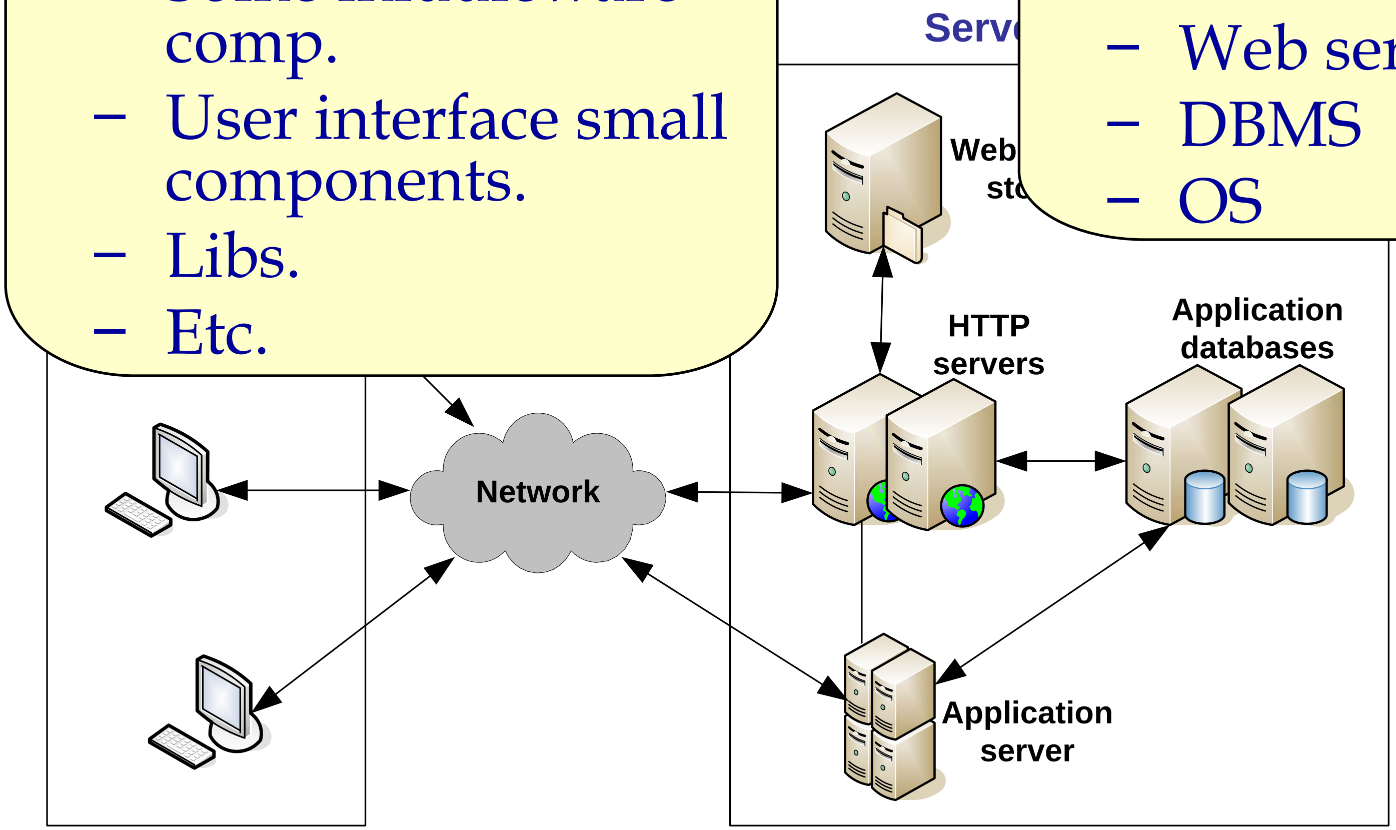
A real example: COTS in very large scale systems

Fine grain COTS:

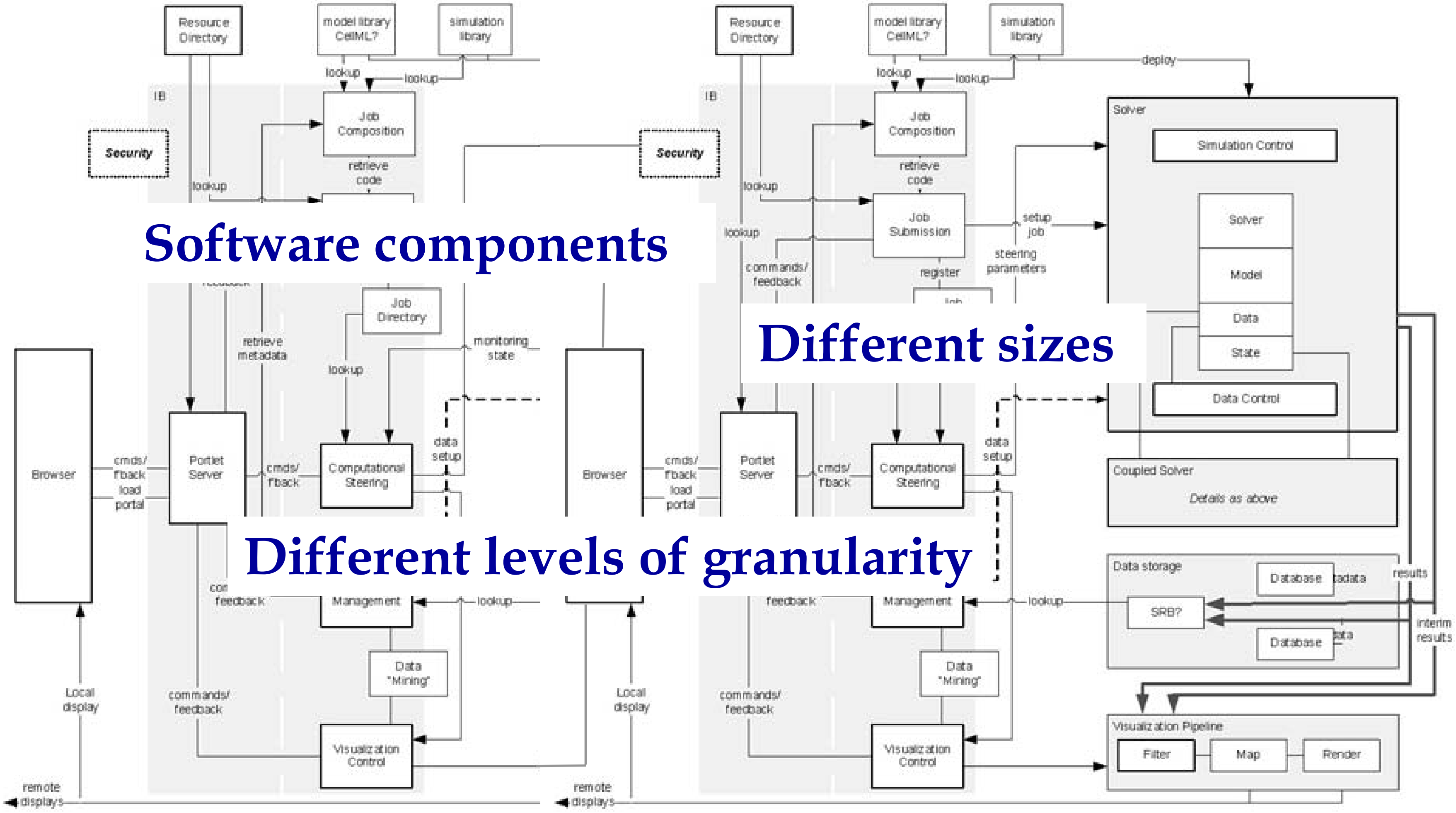
- Some middleware comp.
- User interface small components.
- Libs.
- Etc.

Coarse grain COTS:

- Middleware comp.
- Web servers
- DBMS
- OS



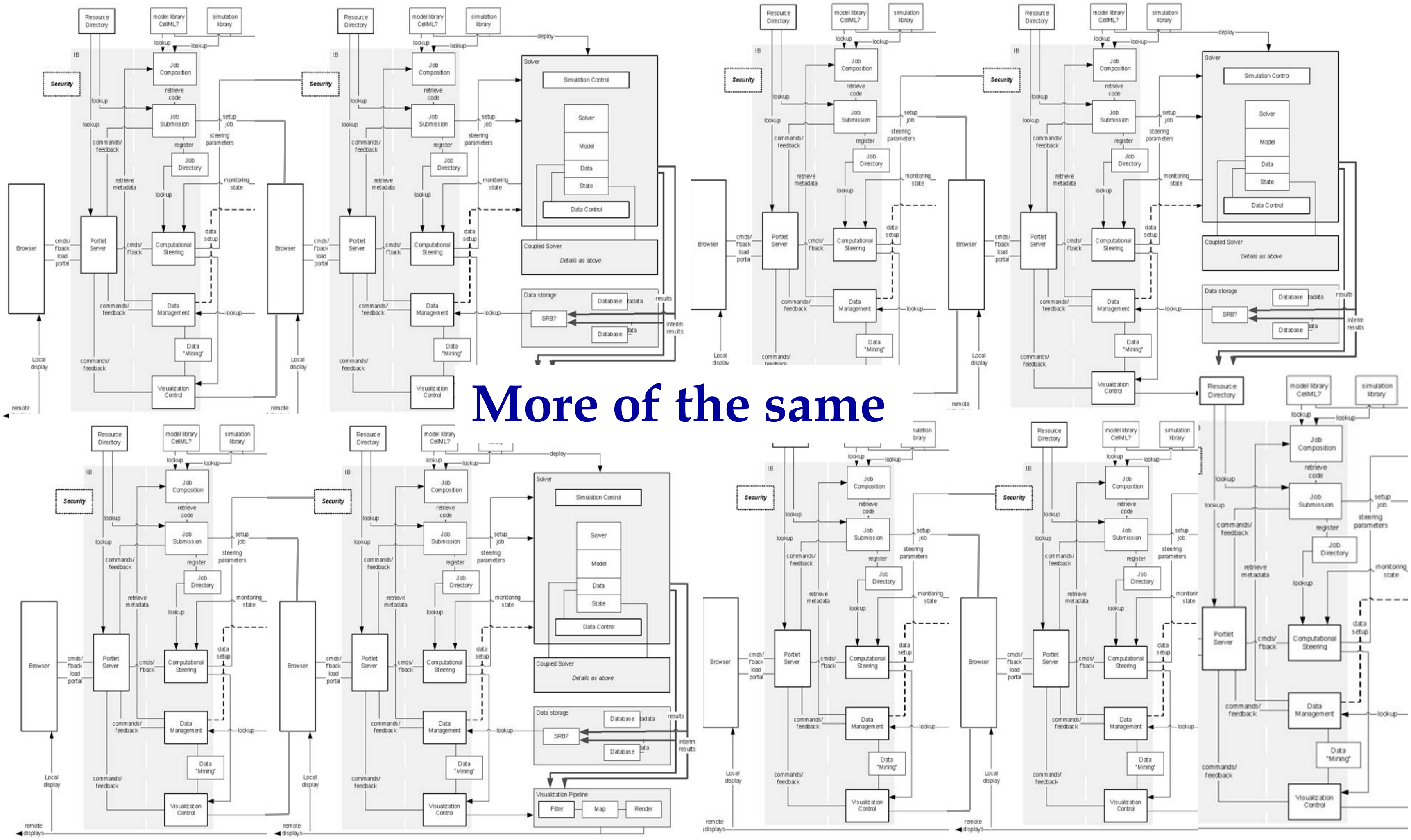
Case-study 1: *I-don't-care-about software architecture diagram*


Software components

Different sizes

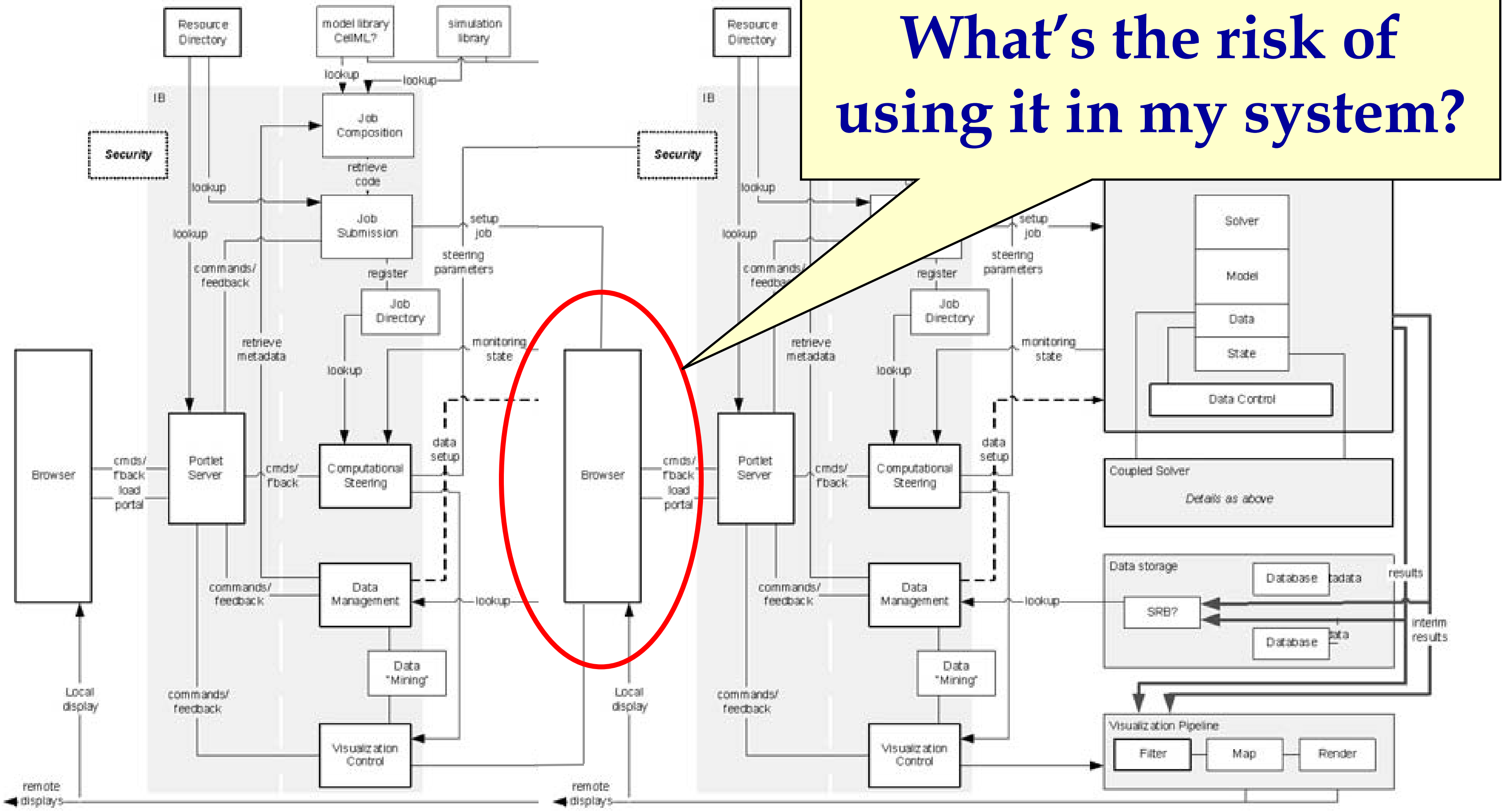
Different levels of granularity

Case-study 2: *I-really-don't-care-about* software architecture diagram



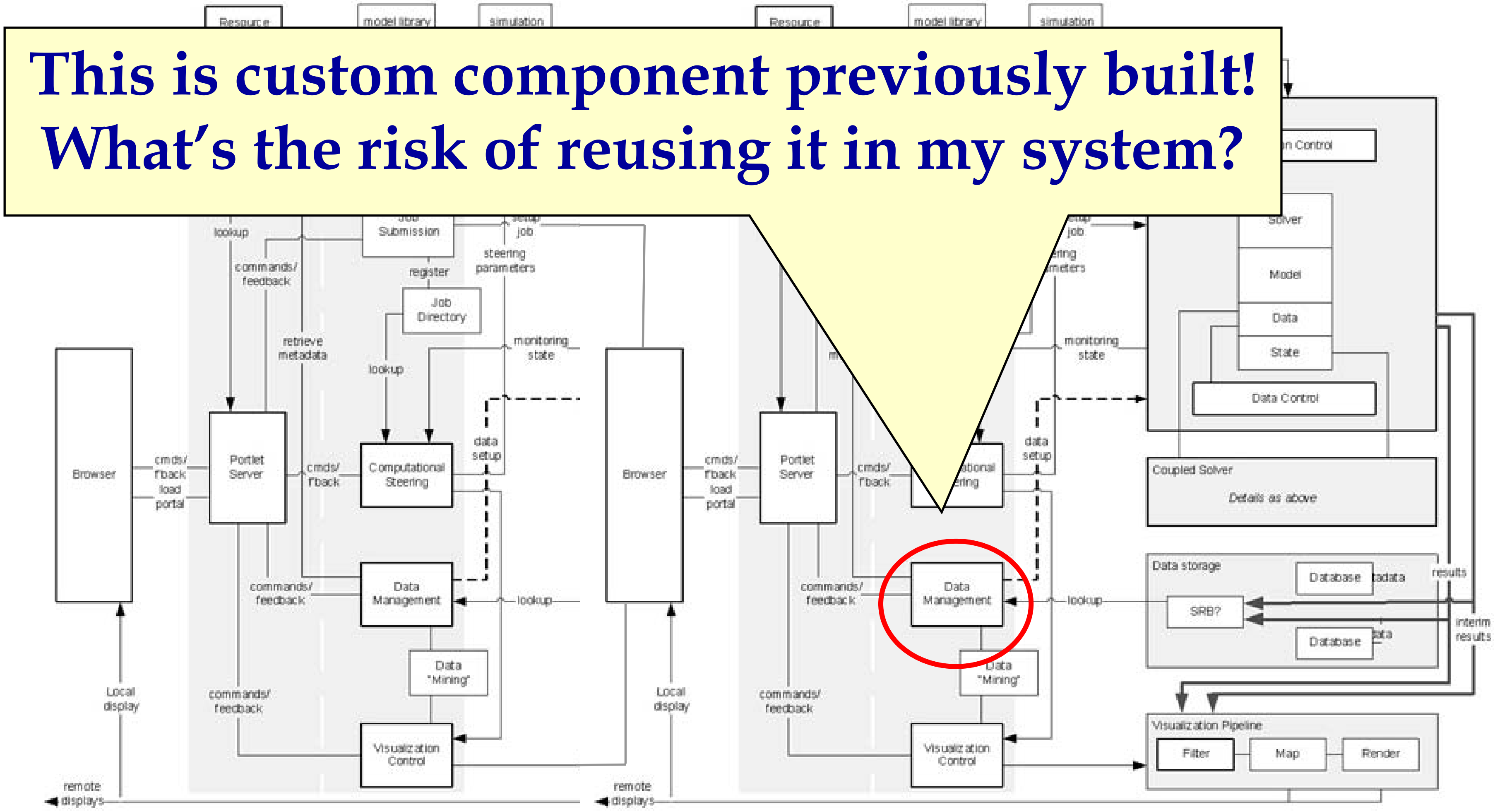
More of the same

Question 1



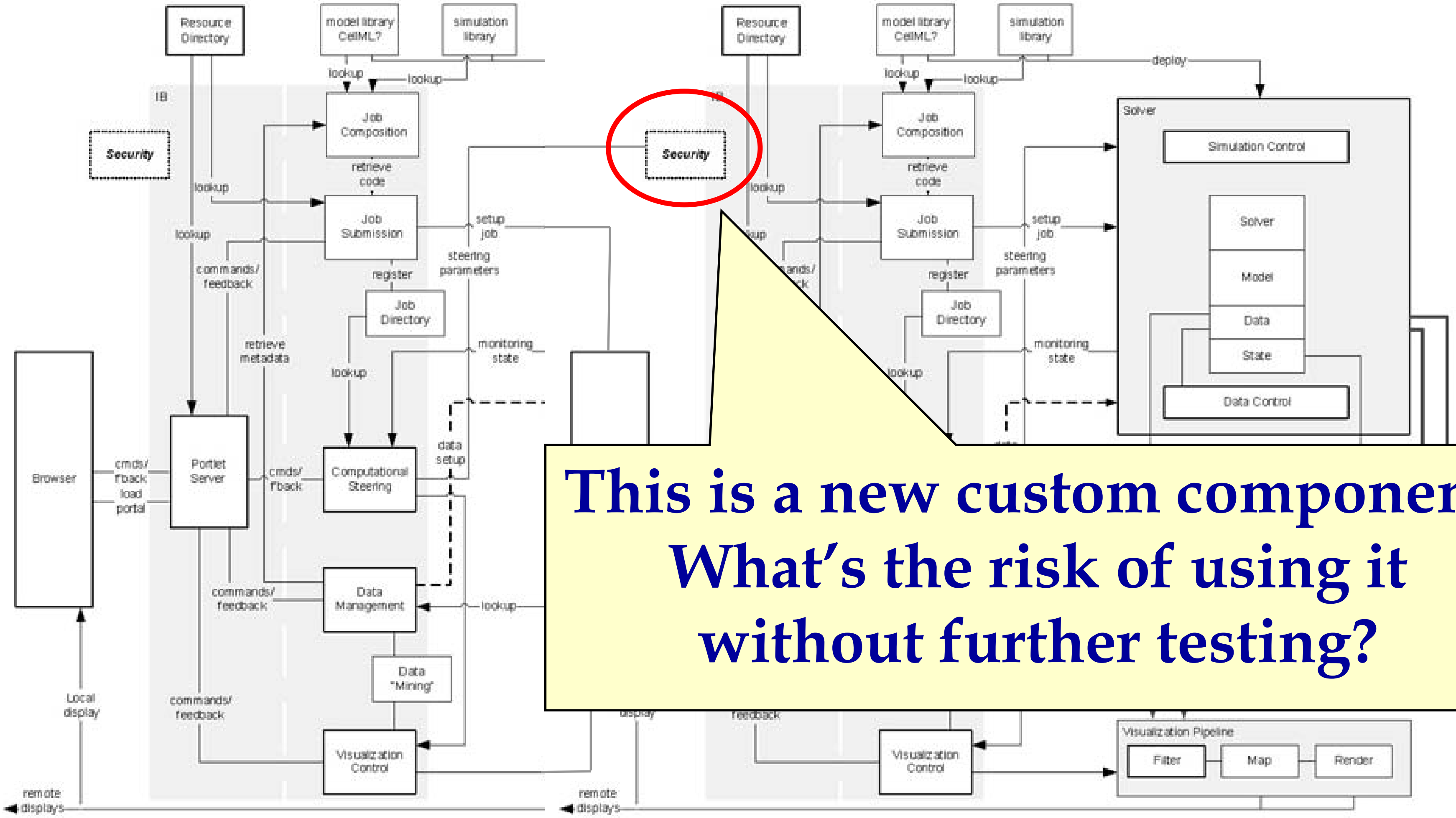
This is a COTS! What's the risk of using it in my system?

Question 2



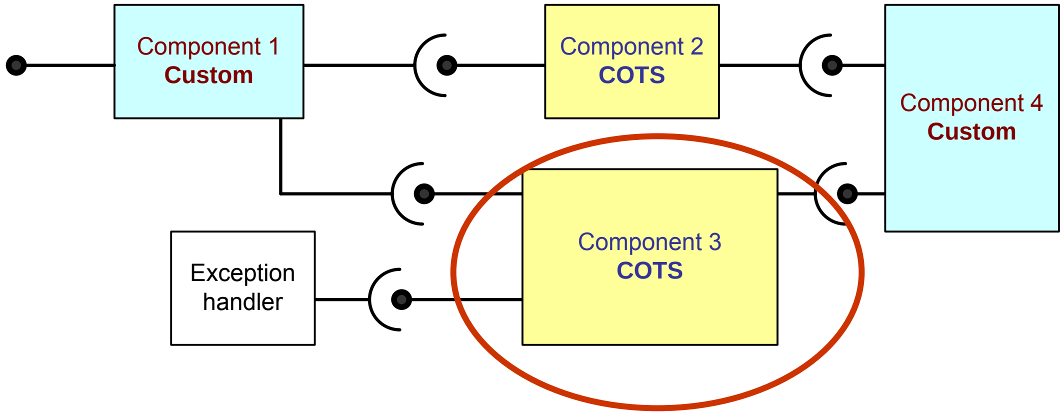
This is custom component previously built!
What's the risk of reusing it in my system?

Question 3



This is a new custom component! What's the risk of using it without further testing?

Experimental risk assessment



Example of question:

What's the risk of using Component 3 in my system?

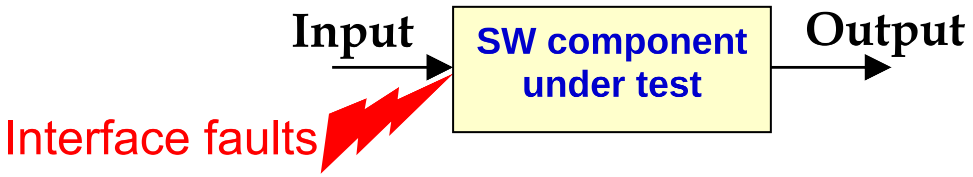
Risk = prob. of bug * prob. of bug activation * impact of bug activation

Software complexity metrics

Injection of software faults

Two possible injection points

1. Injection of interface faults in software components (classical robustness testing: Ballista, Mafalda, …)



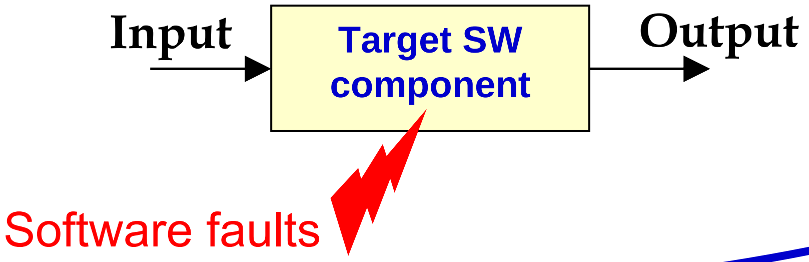
Input

SW component under test

Output

Interface faults

2. Injection of **realistic** software faults inside software components (new approach)



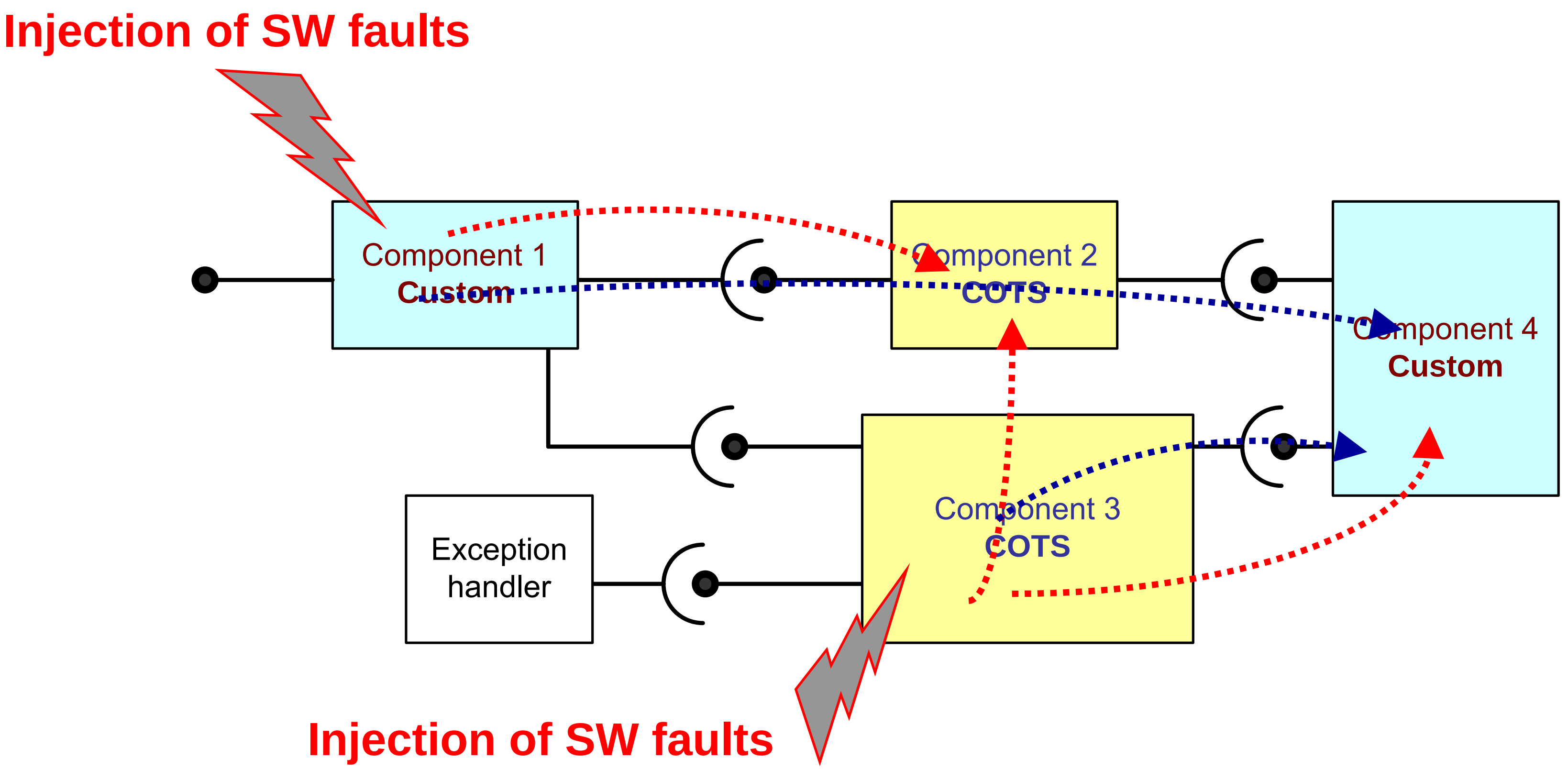
Input

Target SW component

Output

Software faults

Why injection or real software faults?

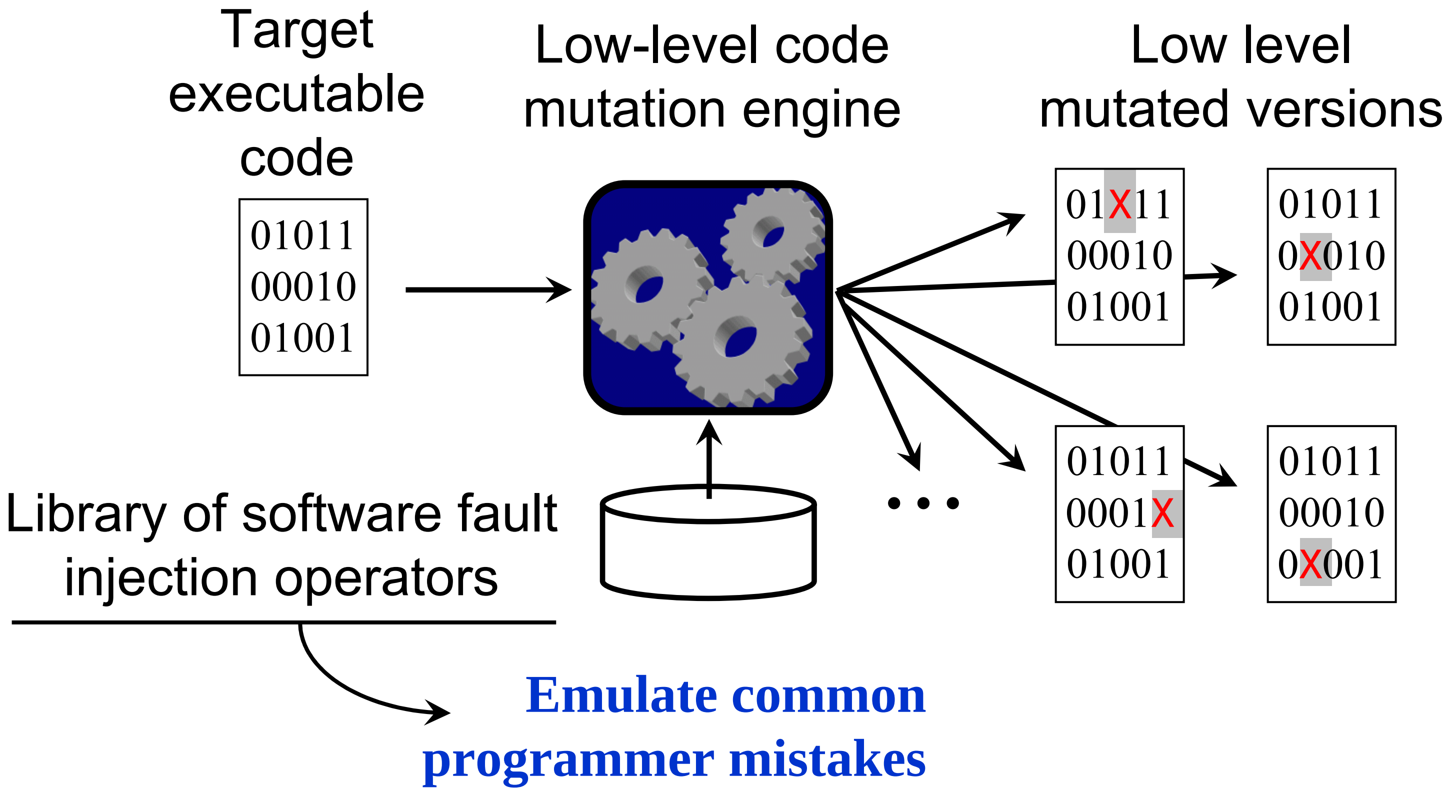


- Error propagation through non conventional channels is a reality.
- Faults injected inside components are more representative.

How to inject software faults?

- **Use G-SWFIT (ISSRE 2002, DSN 2003, DSN 2004)**
  - Injects the **top N** most common software faults.
  - This top N is based on field data (our study + ODC data from IBM) and corresponds to ~65% of the bugs found in field data.
  - Injects faults in executable code.
  - Largely independent on the programming language, compiler, etc that have generated the executable code.
- **G-SWFIT is now a reasonably mature technique.**

G-SWFIT
Generic software fault injection technique



Target executable code

Low-level code mutation engine

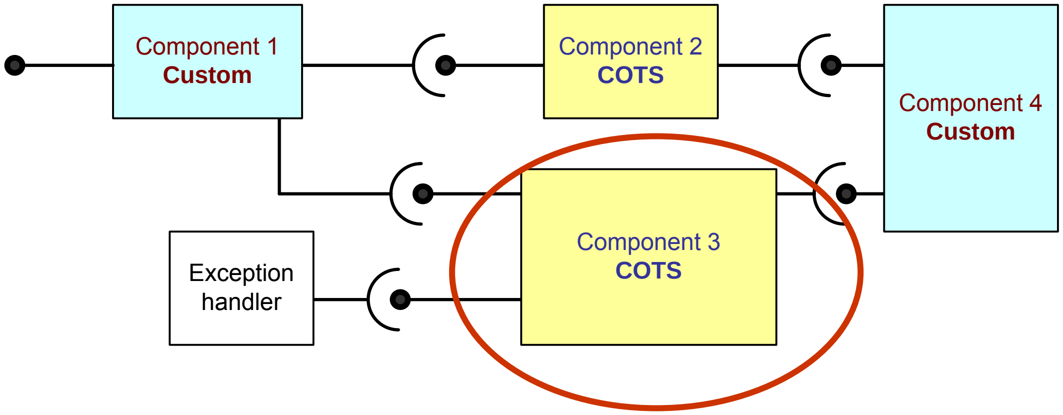
Low level mutated versions

Library of software fault injection operators

Emulate common programmer mistakes

The technique can be applied to binary files prior to execution or to in-memory running processes

Experimental risk assessment (again)



Example of question:

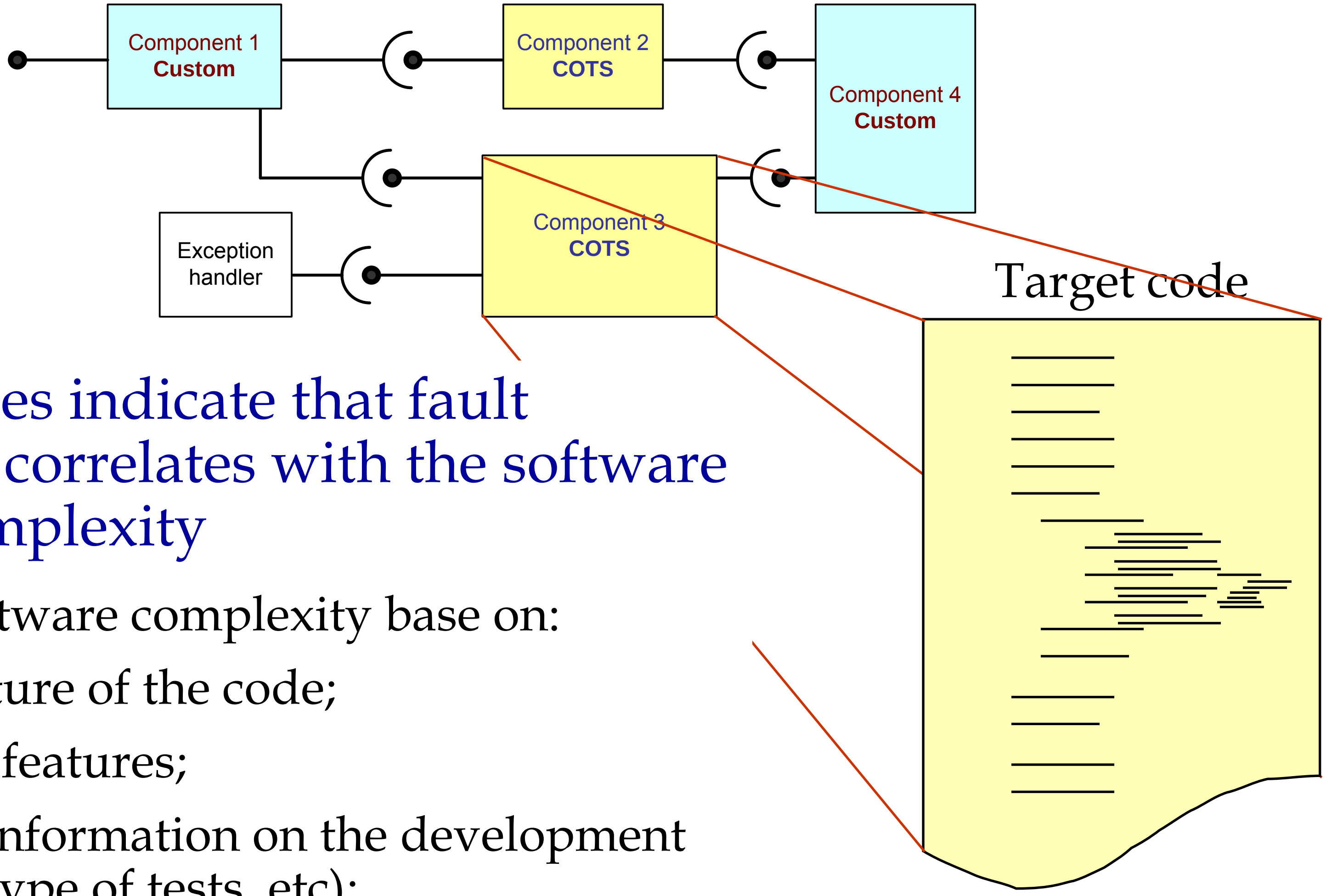
What's the risk of using Component 3 in my system?

Risk = prob. of bug * prob. of bug activation * impact of bug activation

Software complexity metrics

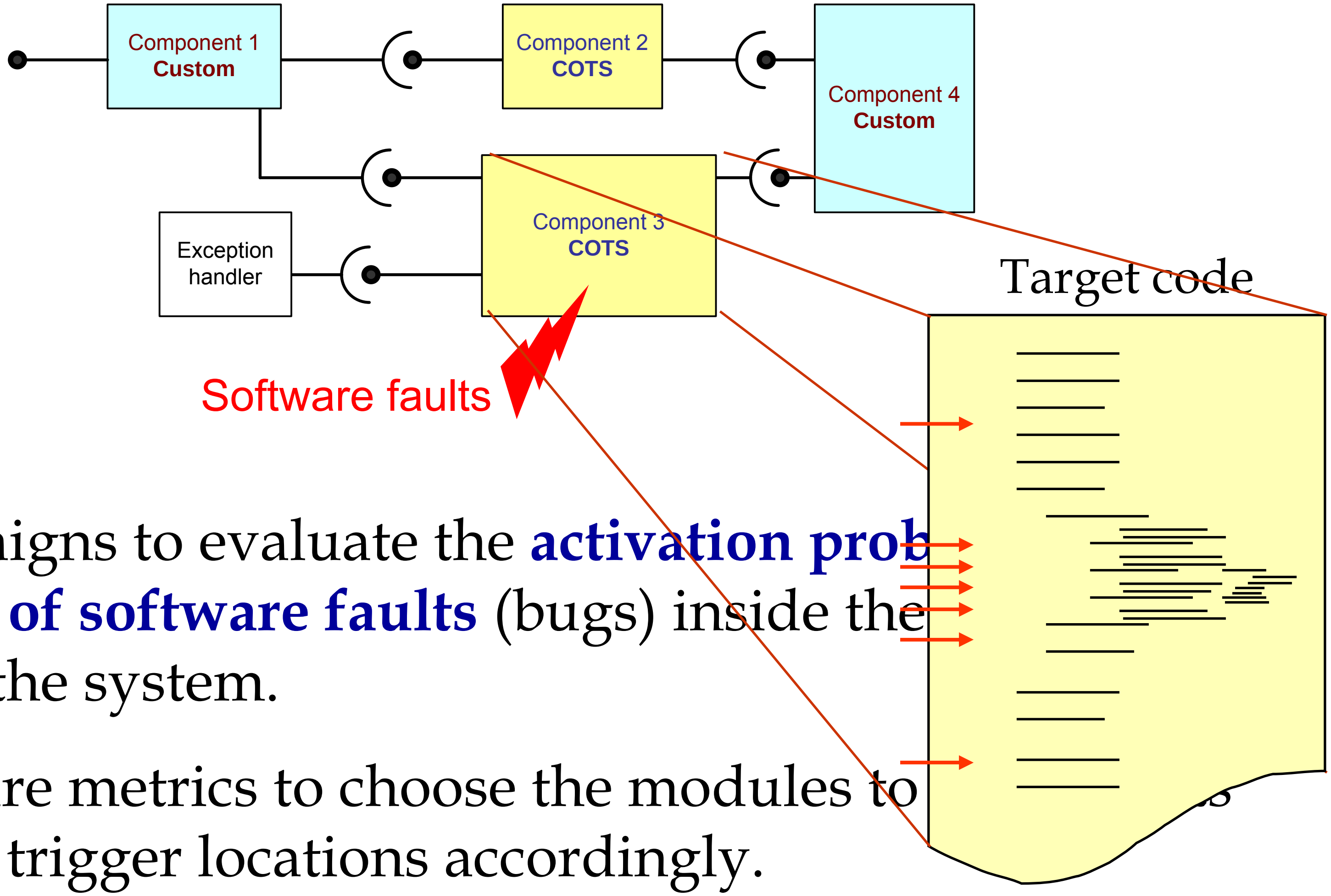
Injection of software faults

Estimation of the probability of residual bugs



- Many studies indicate that fault probability correlates with the software module complexity
- Metrics of software complexity base on:
    - Static feature of the code;
    - Dynamic features;
    - Possible information on the development process (type of tests, etc);
    - ...

Estimation of bug activation probability and bug impact

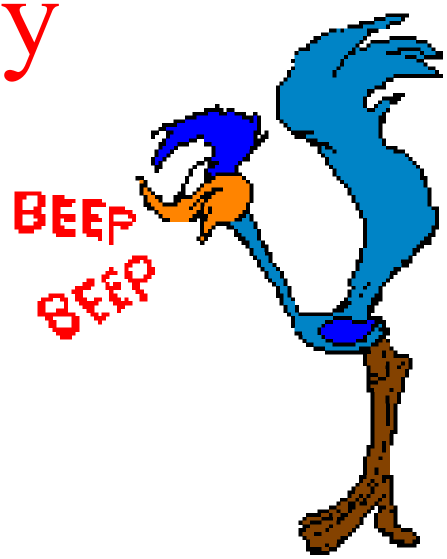


- Test campaigns to evaluate the **activation prob** the **impact of software faults** (bugs) inside the the rest of the system.
- Use software metrics to choose the modules to and define trigger locations accordingly.

Conclusions and current work on experimental risk assessment

- Experimental software risk assessment seems to be viable.
- Risk is a multi-dimensional measure. Many software risks can be assessed, depending on the property I'm interested in.
- Current work:
    - Improve the G-SWFIT technique:
        - Improving current tool.
        - Expansion of the mutation operator library
        - Construction of a field-usable tool for software fault emulation in Java environments
    - Study of software metrics and available tools.
    - Define a methodology for experimental software risk assessment.
    - Real case-studies to demonstrate the methodology.

Real Time Cryptography

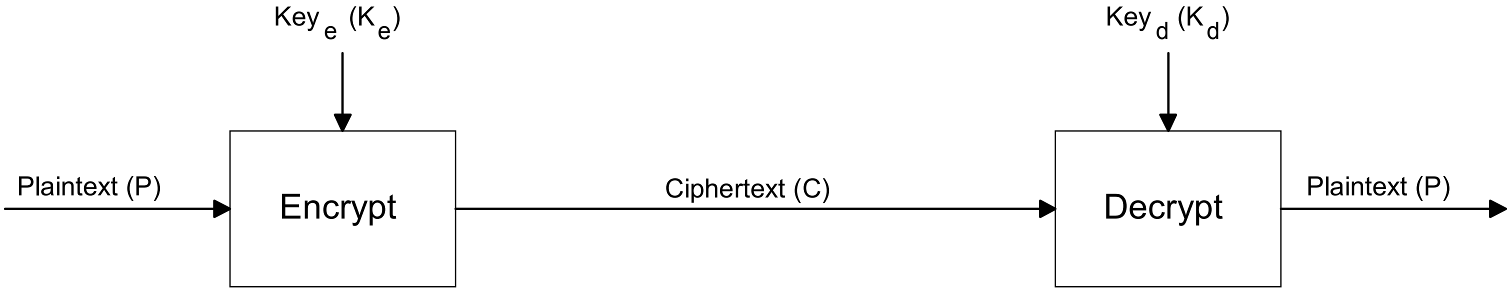


- **The application area**
  - Cryptography optimized for embedded, real-time, control systems
- **The development**
  - A new algorithm, called BeepBeep, overcomes the problems with using existing cryptography for real time systems
- **Contact:** Kevin Driscoll Kevin.Driscoll@Honeywell.com
  612-951-7263 (phone) 612-951-7438 (FAX)

Grid Security

- “need at least 1 Gbps encryption”
- BeepBeep can do that ...
  … in *software* on a 1 GHz Pentium
- No encryption hardware
  - Cheaper
  - More flexible and easier to manage
  - Allows ad hoc grids (i.e., SETI@home model)
- If no physical security at nodes
  - Must assume some nodes will be compromised
    - Portion of data and algorithm will exposed
    - Node’s keys will be exposed
      - Need unique key(s) for each node
      - Need crypto algorithm with good key agility

Encryption Basics



Key$_e$ (K_e)

Key$_d$ (K_d)

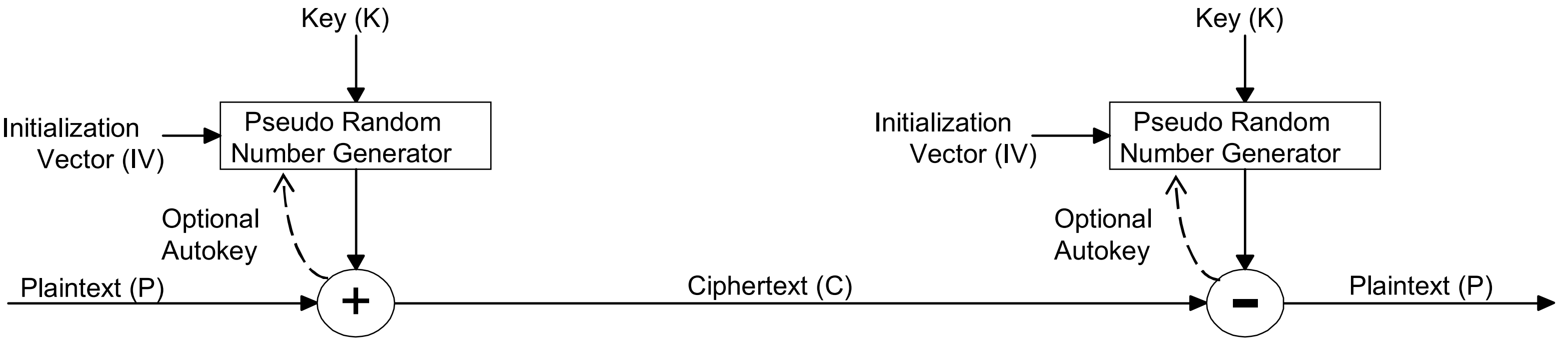
Plaintext (P)

Encrypt

Ciphertext (C)

Decrypt

Plaintext (P)



Key (K)

Key (K)

Initialization Vector (IV)

Pseudo Random Number Generator

Initialization Vector (IV)

Pseudo Random Number Generator

Optional Autokey

Optional Autokey

Plaintext (P)

+

Ciphertext (C)

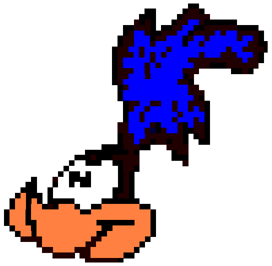
-

Plaintext (P)

Problems with Using Existing Software Cryptography for Embedded Real-Time Systems

- **Is relatively slow, particularly on start-up**
  - Messages (and sessions) are small, less text to amortize start-up cost
  - Latency (lag) is more important than throughput
  - Only worst case timing counts, average is unimportant
    - One missed deadline is not helped by finishing early at all other times
  - Systems typically use repeating execution time slots of fixed size
  - Central control changes key for each message (high key agility)
- **Uses too much data memory (cache thrashing)**
  - Real time systems are multitasking with many context switches / sec
  - Must assume cache is flushed, S-box accesses are mostly misses
- **Consumes additional communication bandwidth**
  - Ciphertext must be no bigger than plaintext
- **Uses separate secrecy and integrity algorithms (or modes)**
  - Makes execution even slower
  - Prevents “lump in the cable” retrofits
- **Most real-time cryptography will be retrofits, which exacerbates the above problems**

Benefits (vs AES, on Pentium)



- **About 2 times faster for very large messages**
- **About 40 times faster for small messages**
- **About half the memory size**
- **25 to 200 times faster than 3DES**
- **Includes integrity with secrecy (increases the above ratios)**
  - Allows "lump in the cable" (or "dongle") implementations (with possible sub-bit-time latency)
- **Several thousand times faster and smaller than public key**
- **1:1 byte replacement (to fit existing message sizes)**
  - Can eliminate need for the addition of an explicit IV
  - Can incorporate existing CRC or checksum into integrity
- **Optimized for CPUs typically found in embedded, real time, control, and communication systems**

Achieving Speed

- **Use an efficient stream cipher**
- **State stays in CPU registers (no RAM used)**
- **Ignore or circumvent conventional wisdom fears**
  - Feedback shift registers are slow in software
    - Invention to improve speed by almost 100 times
  - Multiply is slow
    - Becoming faster (from 42 clocks to 1//4 clocks)
    - Invention to use multiply in a powerful new way
  - Conditional jumps are slow on pipelined CPUs
    - Use multiplexor logic instead of conditional jump
      - Instead of: `if C then Z = A else Z = B`
      - Use this: `Z = ((A xor B) and C) xor B`
    - Use unrolled loop to eliminate other jumps
- **Speed on Pentium is better than 1 bit per clock**
  - Actual speed is 1.19 vs theoretical 1.83 bits per clock

Simple and Small

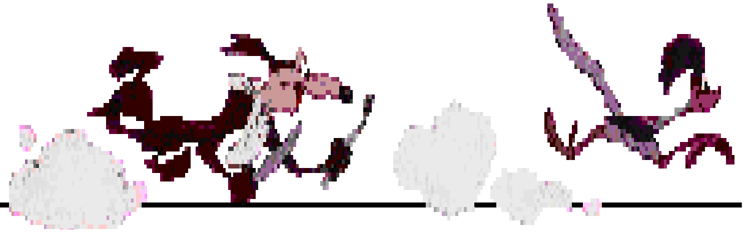
- BeepBeep’s executable code
  - One page of C code
(half of which is declarations and comments)
  - Pentium* without explicit IV 419 bytes
  - Pentium* with explicit IV 484 bytes
  - Pentium* main loop 185 bytes
- BeepBeep’s data memory
  - Pentium* MMX (data stays in registers) 0 bytes

* with MMX registers

Minimizing Message Size

- **No block padding (BeepBeep isn't a block cipher)**
- **Minimize or eliminate Initialization Vector (IV)**
  - Use existing data for IV (e.g. unencrypted header fields)
  - Use explicit or implicit message IDs (e.g. time / sequence) (most real time systems use such IDs)
  - Use Block IV Mode to eliminate IVs (see next slide)
  - Eliminate some IVs by chaining messages together
    - Can be used with reliable message delivery
    - Crypto-state is carried over between messages
- **Use existing CRC or other check data for integrity** (may need to add bits if existing check bits aren't enough)

Security



Basis: 127 bit Linear Feedback Shift Register (LFSR)

Benefits: Good statistics, period that won’t repeat

Old Attack: Berlekamp-Massey

Old Fixes: clock control, nonlinear filter, nonlinear combination of multiple LFSRs

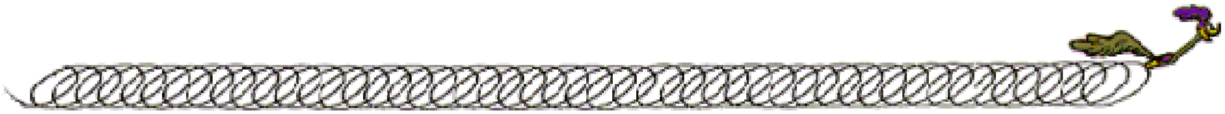
Newer Attacks: embedding, probabilistic correlation, linear consistency, best affine approximation, …

Fixes: use both clock control and nonlinear filter with state, two-stage combiner, and 5 different algebras

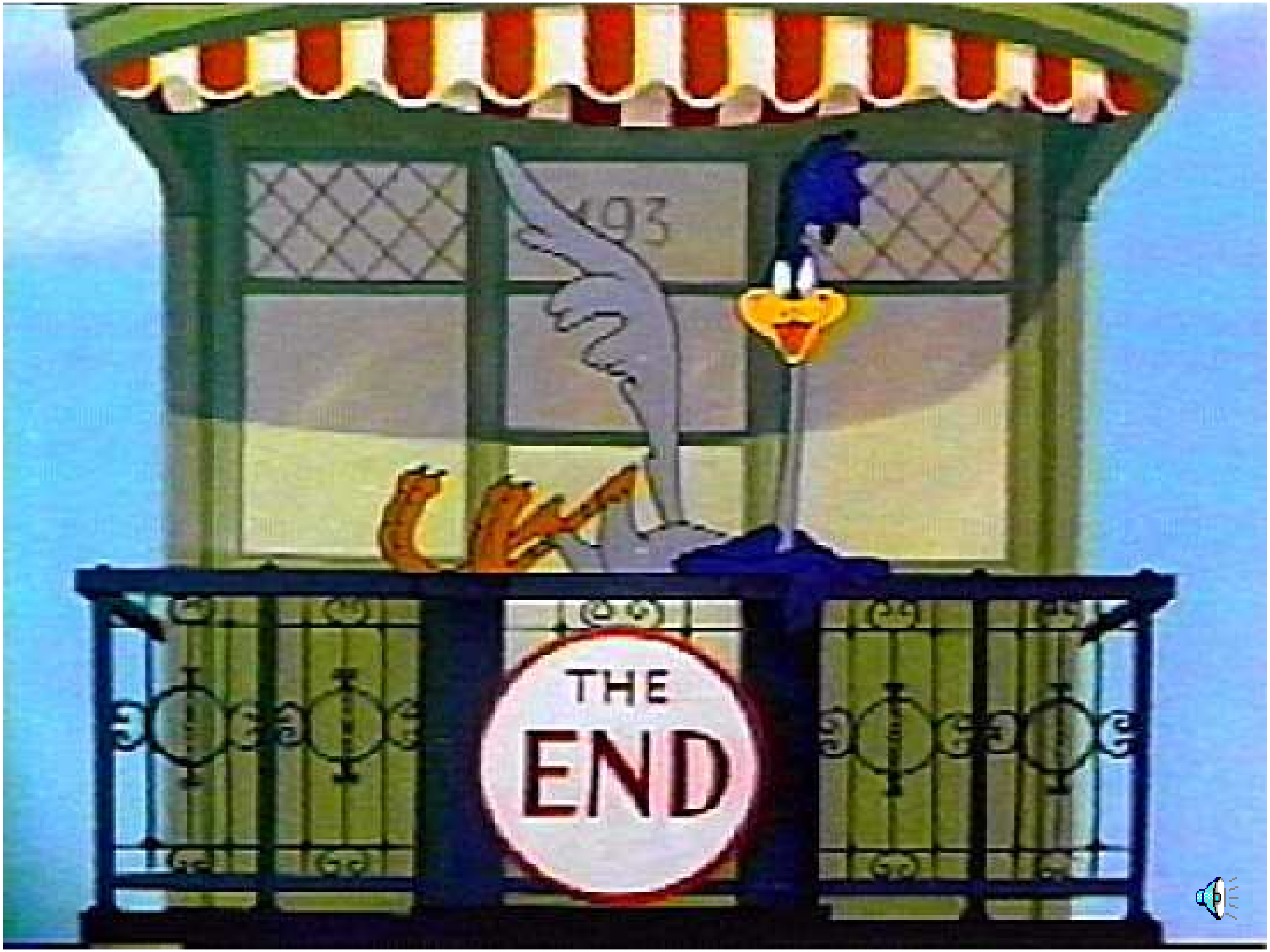
Security (continued)

- **Non-linear filter**
  - 32-bit ones complement addition and 32-bit multiply provide 128th order non-linearity
- **Integrity provided by two mechanisms**
  - Two-stage combiner's non-associative operations
    - Ciphertext $=$ (Plaintext + Key$_1$) XOR Key$_2$
    - Ciphertext $\neq$ Plaintext + (Key$_1$ XOR Key$_2$)
    - Ciphertext $\neq$ Plaintext + (Key$_x$) [for any possible key]
      - Can't directly recover running key, even with known plaintext
  - Plaintext-based autokey feeds back into the non-linear filter's state and into the clock control's state
    - Propagates any text changes through to the end of message
    - Real-time control system messages usually end with check data that can be used to detect tampering

Postulated Uses



- **Aviation**
  - Encryption of ACARS radio traffic
  - Constraints: bandwidth, real-time multitasking
    - Also memory and execution time for retro-fit applications
  - This application has been cleared for export by BXA
- **Home automation and security** (see Oct 2001 IEEE Computer)
  - Secrecy, integrity, authentication, and key management
  - Between central site and residences
  - Constraints: memory, bandwidth, small (8/16 bit) CPU
    - No other algorithm could meet memory constraints when all four security services were included
      - Limit: 1638 bytes Flash ROM, 50 bytes RAM
      - Used: 1628 28
  - This application has been cleared for export by BXA
- **Commercial buildings and industrial controls**


493
THE
END

48th Meeting of IFIP Working Group 10.4
Hakone, Japan - 5th July

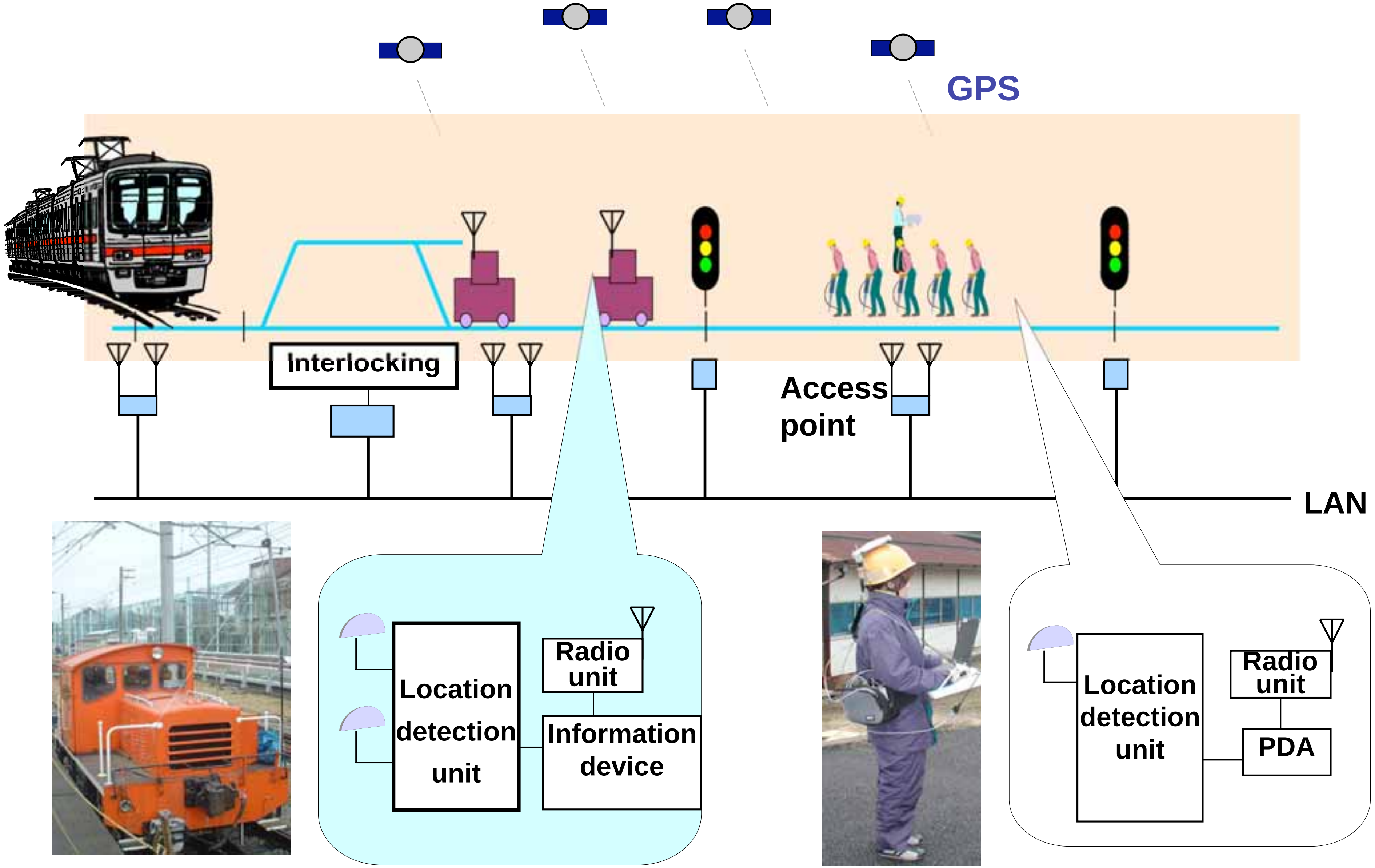
Research Report

A Railway Maintenance Staff Protection System

Yuji HIRAO

Railway Technical Research Institute

Maintenance Staff Protection System (safety-related system)



GPS

Interlocking

Access point

LAN

Location detection unit

Radio unit

Information device

Location detection unit

Radio unit

PDA

Technological interests

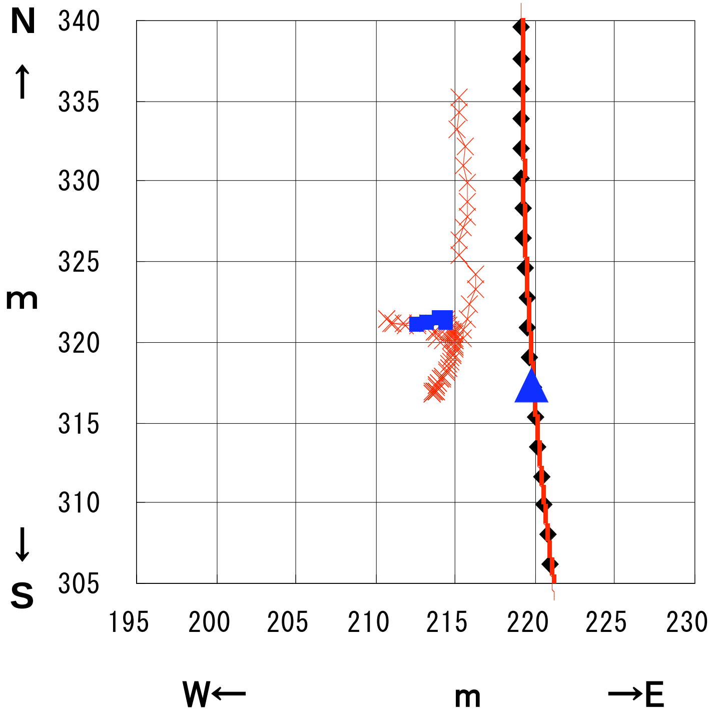
1. Enhancement of positioning accuracy
2. System safety

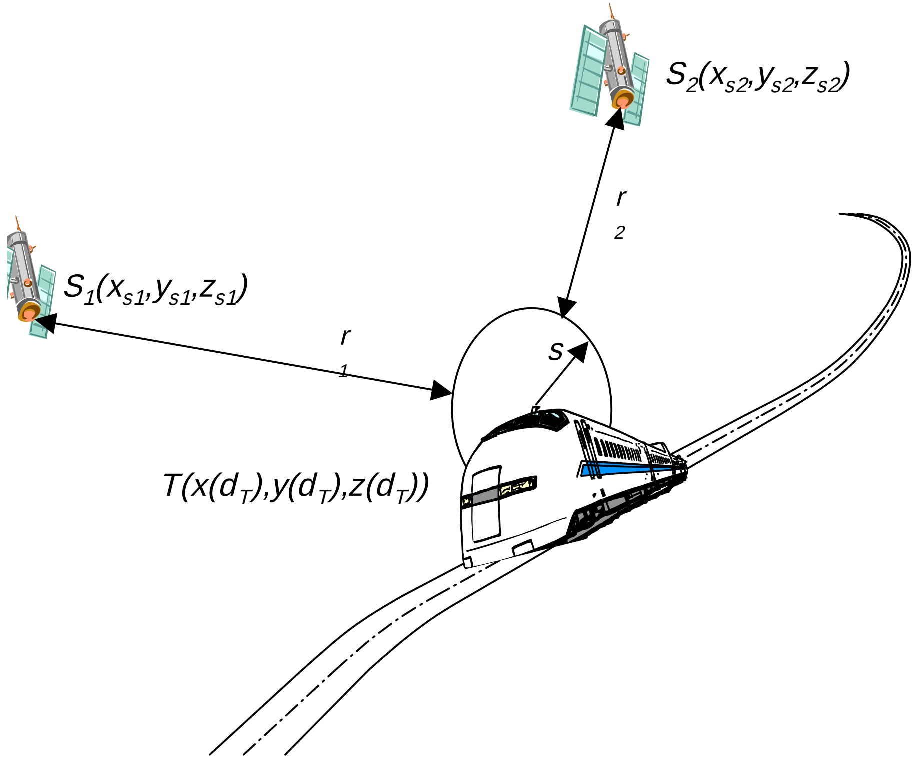
   Data transmission system

      Nomadic computing, ubiquitous ….

Positioning by GPS

— Enhancement of positioning accuracy —

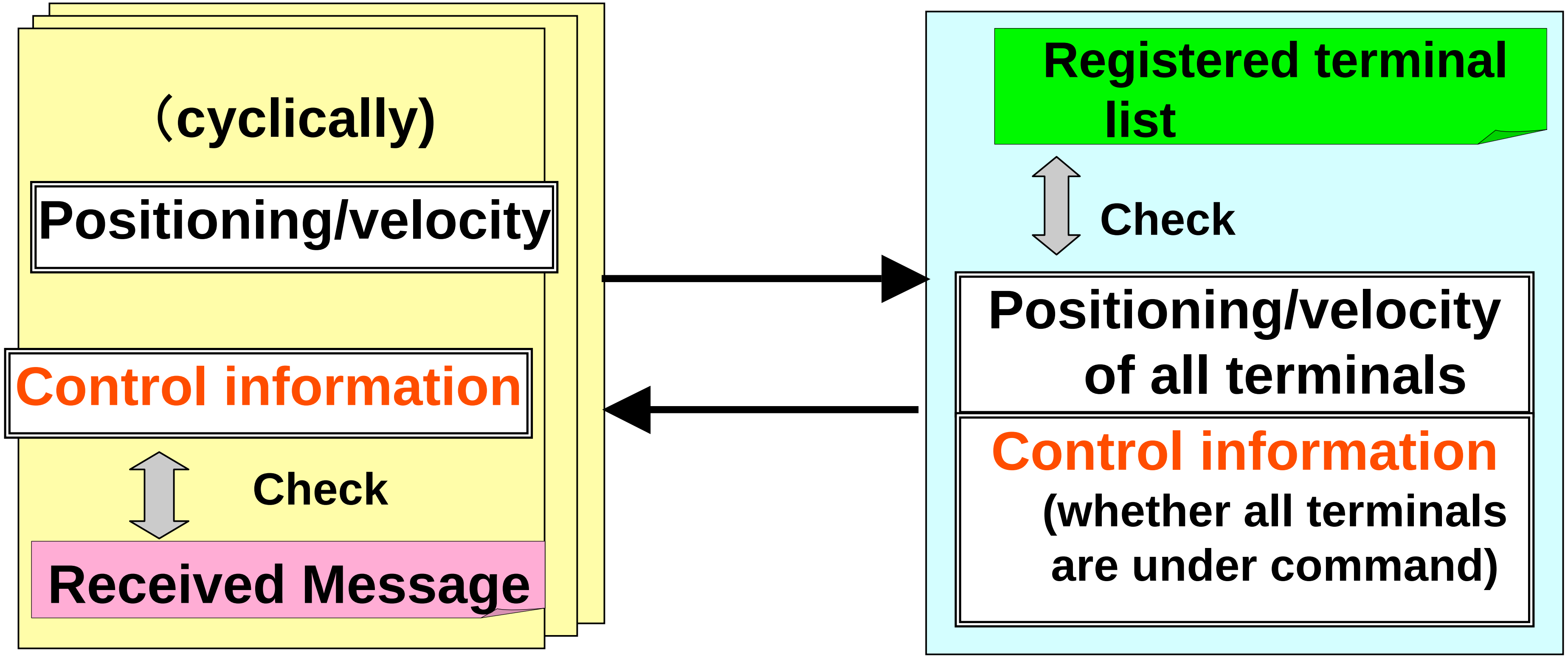



$S_2(x_{s2},y_{s2},z_{s2})$
r_2
$S_1(x_{s1},y_{s1},z_{s1})$
r_1
s
$T(x(d_T),y(d_T),z(d_T))$

System Safety

Wayside terminals

Central management



（cyclically)

Positioning/velocity

Control information

Check

Received Message

Registered terminal list

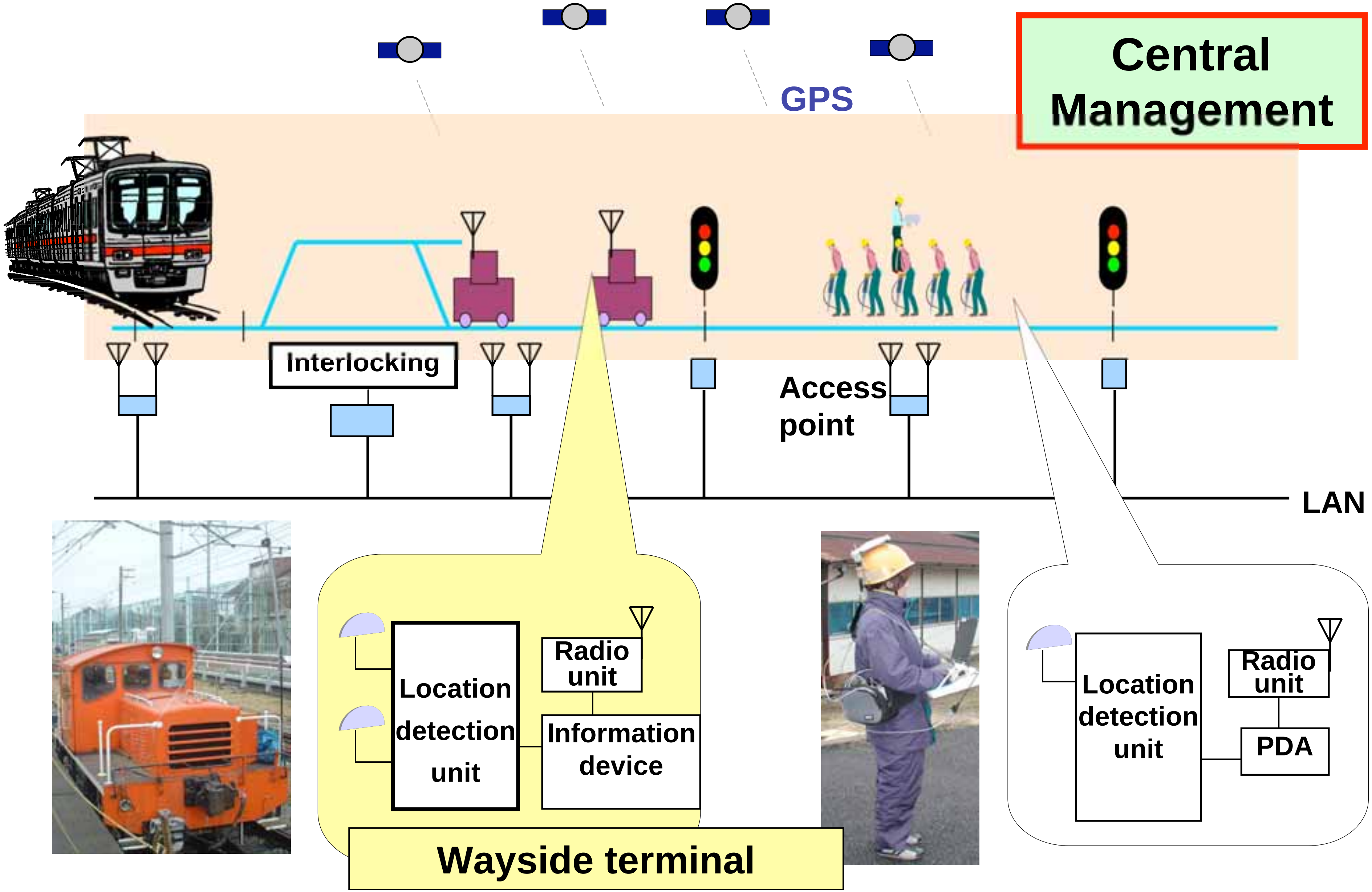
Check

Positioning/velocity of all terminals

Control information (whether all terminals are under command)

Fail-safe

Maintenance Staff Protection System (safety-related system)



GPS

Central Management

Interlocking

Access point

LAN

Location detection unit

Radio unit

Information device

Wayside terminal

Location detection unit

Radio unit

PDA

Sunset on Mount Fuji

Compiled by LAAS-CNRS