

# IEEE NETWORK<sup>®</sup>

January/February 2013, Vol. 27, No. 1

THE MAGAZINE OF GLOBAL INTERNETWORKING

[www.comsoc.org](http://www.comsoc.org)

## Cyber Security of Networked Critical Infrastructures

A Publication of the IEEE Communications Society  
in cooperation with the  
IEEE Computer Society and the  
Internet Society



IEEE



IEEE COMMUNICATIONS SOCIETY

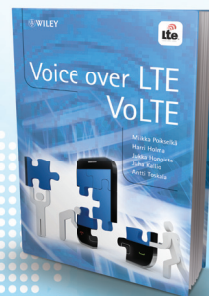
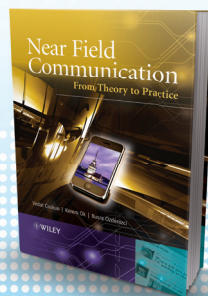
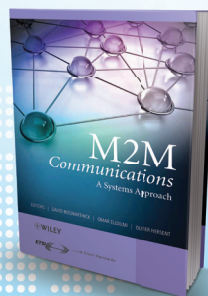
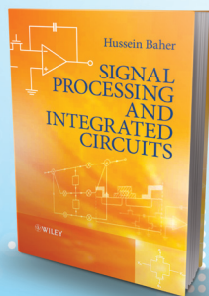
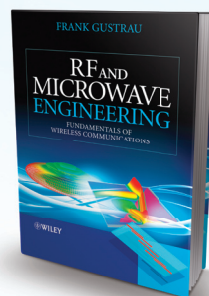
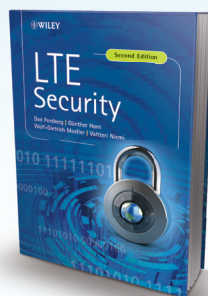
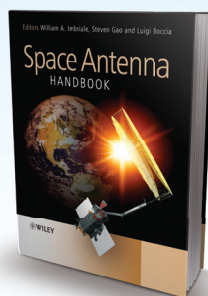
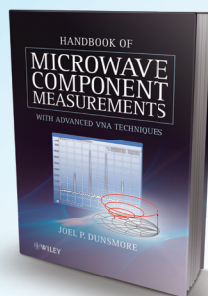
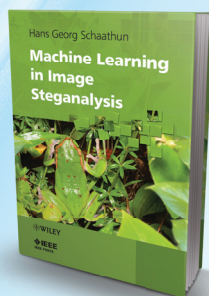


IEEE COMPUTER SOCIETY

# COMMUNICATIONS TECHNOLOGY

20% Discount On These New Books

Visit [www.wiley.com](http://www.wiley.com) and quote code VB781\* when you order



Connect with us:

Sign up for email alerts. Visit:

[www.wiley.com/commstech](http://www.wiley.com/commstech)



twitter

[www.twitter.com/WBComms](http://www.twitter.com/WBComms)



Wiley  
Communications  
Technology

\*Discount valid until 16/12/2012

### Special Issue

## Cyber Security of Networked Critical Infrastructures

### 3 Guest Editorial

Saed Abu-Nimeh, Ernest Foo, Igor Nai Fovino,  
Manimaran Govindarasu, and Thomas Morris

### 5 Authentication and Authorization Mechanisms for Substation Automation in Smart Grid Network

Binod Vaidya, Dimitrios Makrakis, and Hussein T. Mouftah

### 12 A Layered Encryption Mechanism for Networked Critical Infrastructures

Huayang Cao, Peidong Zhu, Xicheng Lu, and Andrei Gurtov

### 19 In Quest of Benchmarking Security Risks to Cyber-Physical Systems

Saurabh Amin, Galina A. Schwartz, and Alefiya Hussain

### 25 Measuring the Global Domain Name System

E. Casalicchio, M. Caselli, and A. Coletta

### Accepted from Open Call

### 32 Participatory Privacy: Enabling Privacy in Participatory Sensing

Emiliano De Cristofaro and Claudio Soriente

### 37 Converged Access of IMS and Web Services: A Virtual Client Model

Salekul Islam and Jean-Charles Grégoire

### 45 On the Analysis of Hierarchical Autonomic Control of Multiparty Services

Nuno Coutinho and Susana Sargento

### 52 Vertical and Horizontal Circuit/Packet Integration Techniques for the Future Optical Internet

Chaitanya S. K. Vadrevu, Massimo Tornatore, Chin P. Guok,  
Inder Monga, and Biswanath Mukherjee

### 59 A Tale of the Tails: Power-Laws in Internet Measurements

Aniket Mahanti, Niklas Carlsson, Anirban Mahanti,  
Martin Arlitt, and Carey Williamson

---

### Editor's Note 2

---

IEEE NETWORK ISSN 0890-8044 is published bimonthly by the Institute of Electrical and Electronics Engineers, Inc. Headquarters address: IEEE, 3 Park Avenue, 17th Floor, New York, NY 10016-5997, USA; tel: +1-212-705-8900; e-mail: [ieee.network@ieee.org](mailto:ieee.network@ieee.org). Responsibility for the contents rests upon authors of signed articles and not the IEEE or its members. Unless otherwise specified, the IEEE neither endorses nor sanctions any positions or actions espoused in *IEEE Network*.

ANNUAL SUBSCRIPTION: \$40 in addition to IEEE Communications Society or any other IEEE Society member dues. Non-member prices: \$250. Single copy price \$50.

EDITORIAL CORRESPONDENCE: Address to: Xuemin (Sherman) Shen, Editor-in-Chief, *IEEE Network*, IEEE Communications Society, 3 Park Avenue, 17th Floor, New York, NY 10016-5997, USA; e-mail: [xshen@bbcr.uwaterloo.ca](mailto:xshen@bbcr.uwaterloo.ca)

COPYRIGHT AND REPRINT PERMISSIONS: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limits of U.S. Copyright law for private use of patrons: those articles that carry a code on the bottom of the first page provided the per copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, USA. For other copying, reprint, or republication permission, write to Director, Publishing Services, at IEEE Headquarters. All rights reserved. Copyright © 2013 by the Institute of Electrical and Electronics Engineers, Inc.

POSTMASTER: Send address changes to *IEEE Network*, IEEE, 445 Hoes Lane, Piscataway, NJ 08855-1331, USA. Printed in USA. Periodical-class postage paid at New York, NY and at additional mailing offices. Bulk rate postage paid at Easton, PA permit #7. Canadian GST Reg# 40030962. Return undeliverable Canadian addresses to: Frontier, P.O. Box 1051, 1031 Helena Street, Fort Erie, ON L2A 6C7.

SUBSCRIPTIONS: Orders, address changes should be sent to IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08855-1331, USA. Tel. +1-732-981-0060.

ADVERTISING: Advertising is accepted at the discretion of the publisher. Address correspondence to *IEEE Network*, 3 Park Avenue, 17th Floor, New York, NY 10016-5997, USA.

---

Director of Magazines

Sergio Benedetto, Politecnico di Torino, Italy

Editor-in-Chief

Sherman Shen, University of Waterloo, Canada

Senior Technical Editors

Tom Chen, Swansea University, UK

Peter O'Reilly, Northeastern Univ., USA

Technical Editors

Jiannong Cao, Poly. Univ., HK

Jiming Chen, Zhejiang Univ., China

Han-Chieh Chao, National Ilan University, Taiwan

Michael Fang, Univ. of Florida, USA

Erol Gelenbe, Imperial College London, UK

Roch Glitho, Concordia Univ. Canada

Minho Jo, Korea Univ., Korea

Admela Jukan, Technische Univ. Carolo-Wilhelmina

zu Braunschweig, Germany

Nei Kato, Tohoku Univ., Japan

Xiaodong Lin, OUIT, Canada

Ying-Dar Lin, National Chiao Tung Univ., Taiwan

Ioanis Nikolaidis, Univ. of Alberta, Canada

Romano Fantacci, Univ. of Florence, Italy

Sudipta Sengupta, Microsoft Research, USA

Ness Shroff, OSU, USA

Ivan Stojmenovic, Univ. Ottawa, Canada

Joe Touch, USC/ISI, USA

Anwar Walid, Bell Labs Research,

Alcatel-Lucent, USA

Guoliang Xue, Arizona State Univ., USA

Murtaza Zafer, IBM T. J. Watson Research

Center, USA

Feature Editors

"New Books and Multimedia"

Yu Cheng, IIT, USA

IEEE Production Staff

Joseph Milizzo, Assistant Publisher

Eric Levine, Associate Publisher

Susan Lange, Online Production Manager

Jennifer Porcello, Production Specialist

Catherine Kemelmacher, Associate Editor

2013 IEEE Communications Society Officers

Vijay K. Bhargava, *President*

Sergio Benedetto, *Past President-Elect*

Leonard Cimini, *VP-Technical Activities*

Abbas Jamalipour, *VP-Conferences*

Nelson Fonseca, *VP-Member Relations*

Vincent Chan, *VP-Publications*

Alex Gelman, *VP-Standards Activities*

Stan Moyer, *Treasurer*

John M. Howell, *Secretary*

Board of Governors

The officers above plus Members-at-Large:

Class of 2013

Gerhard Fettweis, Stefano Galli

Robert Shapiro, Moe Win

Class of 2014

Merrily Hartman, Angel Lozano

John S. Thompson, Chengshan Xiao

Class of 2015

Nirwan Ansari, Stefano Bregni

Hans-Martin Foisel, David G. Michelson

2013 IEEE Officers

Peter W. Staecker, *President*

J. Roberto B. de Marca, *President-Elect*

Marko Delimar, *Secretary*

John T. Barr, *Treasurer*

Gordon W. Day, *Past-President*

E. James Prendergast, *Executive Director*

Doug Zuckerman, *Director, Division III*

## EDITOR'S NOTE



Xuemin Shen

For the *IEEE Network* 2013 January/February issue, in addition to the Special Issue on Cyber Security of Networked Critical Infrastructures, we also include five accepted open call articles as follows.

Participatory sensing is an emerging computing paradigm that enables the distributed collection of data by self-selected participants. The first article, "Participatory Privacy: Enabling Privacy in Participatory Sensing" by Emiliano De Cristofaro and Claudio Soriente, focuses on privacy issues in participatory sensing and proposes a simple privacy-enhanced infrastructure. It also provides a set of clear definitions geared to protecting the privacy of both data producers and consumers.

The second article, "Converged Access of IMS and Web Services: A Virtual Client Model" by Salekul Islam and Jean-Charles Grégoire, presents a virtual client-based converged access architecture for IP multimedia subsystem (IMS) and web services. It also describes how to implement an IMS-web hybrid service called Movie-on-Demand (MoD) by deploying a simple IMS client and web server in a surrogate, and using an open source implementation of a full IMS environment, from client to application server.

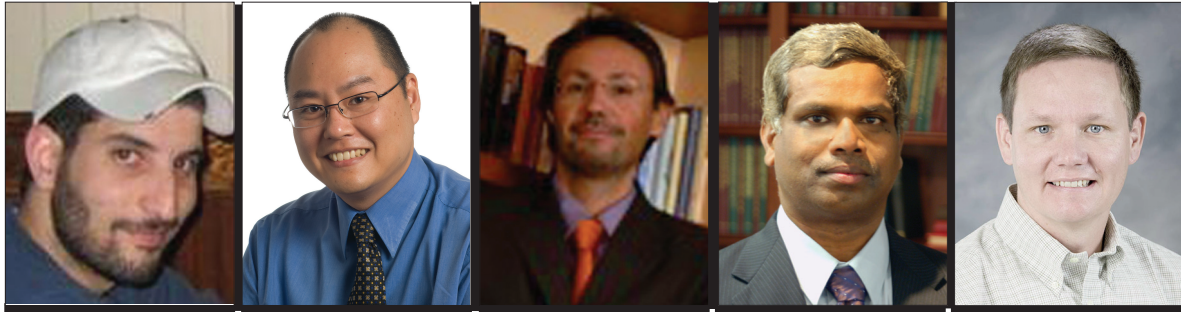
The third article, "On the Analysis of Hierarchical Autonomic Control of Multiparty Services" by Nuno Coutinho and Susana Sargento, describes a context-driven framework for multiparty content delivery and discusses the rewards of employing the Abstract Multiparty Transport concept, which provides autonomic control of personalized group-based services to users through a hierarchical strategy.

The fourth article, "Vertical and Horizontal Circuit/Package Integration Techniques for the Future Optical Internet" by Chaitanya S. K. Vadrevu *et al.*, proposes two approaches to dynamically migrate capacity between the circuit and packet sections of hybrid circuit/package networks, called vertical stacking and horizontal partitioning. In vertical stacking, the backup capacity of wavelength circuits can be dynamically exchanged between packet and wavelength services while ensuring survivability. In horizontal partitioning, the excess capacity on links in the packet section can be loaned to circuit services.

Power-laws are ubiquitous in the Internet and its applications. The fifth article, "A Tale of the Tails: Power-Laws in Internet Measurements" by Aniket Mahanti *et al.*, presents a review of power-laws with emphasis on observations from Internet measurements. It introduces power-laws and describes two commonly observed power-law distributions, the Pareto and Zipf distributions.

In closing, we would like to thank all the authors who have contributed the open call articles. We would also like to acknowledge the contributions of Associate Editors and reviewers who have participated in the review process and provided helpful suggestions to the authors on improving the content and presentation of their articles. We hope you enjoy reading the Special Issue on Cyber Security of Networked Critical Infrastructures and the open call articles in this collection. Happy New Year!

# Cyber Security of Networked Critical Infrastructures



Saed Abu-Nimeh

Ernest Foo

Igor Nai Fovino

Manimaran  
Govindarasu

Thomas Morris

A new era of cyber warfare has appeared on the horizon with the discovery and detection of Stuxnet. Allegedly planned, designed, and created by the United States and Israel, Stuxnet is considered the first known cyber weapon to attack an adversary state. Stuxnet's discovery put a lot of attention on the outdated and obsolete security of critical infrastructure. It became very apparent that electronic devices that are used to control and operate critical infrastructure like programmable logic controllers (PLCs) or supervisory control and data acquisition (SCADA) systems lack very basic security and protection measures. Part of that is due to the fact that when these devices were designed, the idea of exposing them to the Internet was not in mind. However, now with this exposure, these devices and systems are considered easy prey to adversaries.

In this Special Issue, we focus on the advances of research in the area of the security of networked critical infrastructure. SCADA systems and the Internet and its components are a few examples of networked critical infrastructure. The objectives of this Special Issue are twofold. First, we aim to present the latest advances in the security of networked critical infrastructure research. Second, we hope that this Special Issue will encourage and stimulate more activities in the field.

This Special Issue brings together some of the latest research studies in the field of networked critical infrastructure security. It comprises four articles that discuss a wide range of problems.

In the first article, Vaidya *et al.* present a lightweight multi-factor authentication and attribute-based authorization approach to protect SCADA systems. Their approach relies on public key certificates and a zero-knowledge protocol-based server-aided verification mechanism. An access control mechanism using an attribute certificate are used to authenticate remote users.

In the second article, Cao *et al.* present a layered encryption approach to protect critical infrastructure networks. The mechanism utilizes hash chain technology to protect

the data in such networks. Since critical infrastructure comprises resource-constrained devices, the proposed approach is lightweight and can be used either independently or in combination with other protocols.

In the third article, Amin *et al.* present a framework for assessing security risks in cyber-physical systems. To benchmark risks in cyber-physical systems, the framework focuses on reliability and security risks. A model-based approach to estimate these risks is proposed. Finally, they suggest that risk assessment of cyber-physical devices should consider both technology-based defenses and institutional structures to be more effective.

In the fourth article, Casalicchio *et al.* present an approach to measure the health of the Domain Name System (DNS). The Measuring the Naming System (MeNSa) framework is described. Metrics on DNS are aggregated to verify the level of service and the presence or absence of threats.

In summary, the four articles in this Special Issue represent some of the latest novel advances in the research of networked critical infrastructure security. We hope that the discussions and findings in these articles will be of value and stimulate new ideas to future research.

## Biographies

SAEED ABU-NIMEH ([sabunimeh@paypal.com](mailto:sabunimeh@paypal.com)) is a distinguished scientist at PayPal Inc. He received his Ph.D. in computer science from Southern Methodist University. His research interests include web security and machine learning.

ERNEST FOO ([e.foo@qut.edu.au](mailto:e.foo@qut.edu.au)) is an active researcher in the School of Electrical Engineering and Computer Science at Queensland University of Technology. His research is in the area of information and network security. He has broad interests, having published in the area of formal analysis of privacy and identity management protocols as well as proposing secure reputation systems for wireless sensor networks. Recently, he has been conducting research in the area of secure SCADA and critical infrastructure protection. In the past, he has worked extensively in the field of electronic commerce protocols investigating secure protocols for electronic tendering and electronic contracting in the Australian construction industry. These protocols are of particular importance in ensuring secure electronic procurement in both the government and private sectors. He has over 40 publications in internationally refereed conferences and journals.

IGOR NAI FOVINO (igor.nai@gmail.com) is head of the Research Division of the Global Cyber Security Center. He has deep knowledge in the fields of ICT Security of industrial critical infrastructure, energy and smart grids, risk assessment, IDS, and cryptography. He is an author of more than 60 scientific papers published in international journals, books, and conference proceedings; moreover, he serves as a reviewer for several international journals in the ICT security field. In May 2010 he received the IEEE HSI 2010 best paper award in the area of SCADA systems. He is also an expert in European policies (mainly in the CIIP field). Since 2012 he has been a member of the European Commission Experts Working Group on the security of ICS and smart grids. During his career he has worked as a contractual researcher at the University of Milano in the field of privacy preserving datamining and computer security, and as a contractual professor of operating systems at the University of Insubria. From 2005 to 2011 he served as scientific officer at the Joint Research Centre of the European Commission, providing scientific support to EU policies related to the EPCLIP program. Since 2007 he has been a member of the IFIP Working Group on Critical Infrastructure Protection.

MANIMARAN GOVINDARASU [SM] (gmani@iastate.edu) is a professor in the Department of Electrical and Computer Engineering at Iowa State University. His research expertise is in the areas of cyber-physical systems security of smart grid, cyber security, and real-time systems. He has co-authored over 125 peer-reviewed publications in these areas. He recently has developed a cyber security testbed for smart grid at Iowa State University and conducts vulnerability assessment and attack-defense evaluations, and develops robust countermeasures. He is a co-author of the text *Resource Management in Real-Time Systems and Networks* (MIT Press, 2001). He has given tutorials at respected conferences (e.g., IEEE INFOCOM 2004, IEEE ComSoc Tutorials Now™, and IEEE ISGT 2012, and delivered industry short courses

on the subject of cyber security. He has served in leadership roles in many IEEE conferences, symposia, and workshops. He serves on the editorial board of *IEEE Transactions on Smart Grid* and served as a guest editor for several publications including *IEEE Network* (January 2003) and an *IEEE Power & Energy Special Issue* (January 2012), and serves as the Founding Chair of the Cyber Security Task Force at IEEE PES Society PSACE-CAMS Subcommittee and Vice-Chair of the CAMS Subcommittee. He is an ABET Program Evaluator.

THOMAS MORRIS (morris@ece.msstate.edu) received his Ph.D. in computer engineering at Southern Methodist University, Dallas, Texas, with a research emphasis on cyber security. He joined the Department of Electrical and Computer Engineering at Mississippi State University (MSU) in 2008 as an assistant professor. He currently serves as director of the MSU Critical Infrastructure Protection Center (CIPC) and is a member of the MSU Center for Computer Security Research (CCSR). His primary research interests include cyber security for industrial control systems and electric utilities, and power system protective relaying. His recent research outcomes include vulnerability and exploit taxonomies, intrusion detection systems, virtual testbeds, and a relay setting automation program used by a top 20 investor owned utility. He has authored 33 peer reviewed research conference and journal articles in these areas. His research projects are funded by the Department of Homeland Security, Oak Ridge National Laboratories, NASA, the U.S. Army Corps of Engineers Engineering Research Development Center (ERDC), Pacific Gas and Electric Corporation, and Entergy Corporation. Prior to joining MSU, he worked at Texas Instruments (TI) for 17 years in multiple roles including circuit design and verification engineer, applications engineer, team leader, and program manager.

---

# Authentication and Authorization Mechanisms for Substation Automation in Smart Grid Network

**Binod Vaidya, Dimitrios Makrakis, and Hussein T. Mouftah, University of Ottawa**

---

## Abstract

Supervisory control and data acquisition systems are used extensively to control and monitor critical infrastructure including power, gas, oil, and water. To integrate intelligent electronic devices in smart grid infrastructure, the utilities are deploying substation automation systems (SASs) and extensive communication networks, but there is growing concern about SCADA security including substation security. Although there are several solutions utilized to prevent security threats in SCADA networks, existing SCADA networks still have severe shortcomings. In this article, we propose a lightweight and efficient security solution for SASs that provides multilevel multi-factor authentication and attribute-based authorization by deploying public key certificates, and zero-knowledge protocol-based server-aided verification and access control mechanisms using attribute certificates. It can be seen that the proposed approach is efficient and robust.

---

**S**upervisory control and data acquisition (SCADA) systems are real-time process control systems that monitor and control local or remote devices. They are extensively used in critical infrastructure including power, gas, oil, and water. A large number of modern intelligent electronic devices (IEDs) are installed in substation automation systems (SASs) that provide powerful tools to collect, monitor, and analyze data. In a smart grid, these devices provide valuable information that can be used to improve reliability and reduce operating costs.

Traditionally, SCADA systems were considered secure as they utilized dedicated communication lines and proprietary protocols. However, modern SCADA systems are being implemented using industry standard Transmission Control Protocol/Internet Protocol (TCP/IP) networks, different communication technologies, and SCADA protocols. In order to integrate IEDs in smart grid infrastructure, utilities are deploying SCADA systems as well as extensive communication networks including wireless access networks and IP networks in modern electric power systems. But there is growing concern regarding SCADA security, including substation security [1]. Although there are several solutions utilized to prevent security threats in SCADA networks, existing SCADA networks still have severe shortcomings. These issues vary from devices configured with default passwords to unobserved access via dialup and corporate information technology (IT) networks of the utilities.

Providing secure access to substation devices from remotely located sites is much more challenging than just allowing SCADA control center to access substation equipment. Hence, there are still several concerns on remote maintenance access at the SAS.

To guarantee safe, secure, and reliable operation of the electric power network, the North American Electric Reliability

Council (NERC) has imposed several cyber security measures. These measures currently target mostly transmission and generation in North America. NERC critical infrastructure protection (CIP) standards (CIP-002–CIP-009) provide a cyber security framework for identification and protection of critical cyber assets to maintain secure and reliable operation of electric grid systems.

SCADA provides automation solutions using several standards such as International Electrotechnical Commission's (IEC) 60870-5, IEC 61850, IEC 62351 [2], Distributed Network Protocol (DNP3) [3], and Modbus.

In this article, we propose a lightweight and efficient substation level security solution that provides multilevel multi-factor authentication and attribute-based authorization. We have deployed public key certificate and zero-knowledge proof protocol-based server-aided verification mechanism in which IEDs can authenticate any remote users with the help of a substation controller (SSC) as well as an access control mechanism using an attribute certificate.

The rest of this article is organized as follows. We highlight standards and describe the SAS. We discuss security requirements and threats, and then highlight SAS challenges. We present system architecture. We describe the proposed security solution, and discuss system analysis. Finally, we conclude the article.

## SCADA Standards

Modern SASs utilize open standards such as IEC 61850, IEC 60870-5, IEC 61850, IEC 62351, and DNP3, which are as follows.

### IEC 61850

IEC 61850 is a standard for design of substation automation and part of the IEC TC57 reference architecture for electric

power systems. It is used for protective relaying, substation automation, distribution automation, power quality, distributed energy resources, substation to control center, and other power industry operational functions.

### *IEC 60870-5*

IEC 60870-5 is a standard defined for systems used in SCADA in power system automation applications. It provides a communication profile for sending basic tele-control messages between two systems, which uses permanent directly connected data circuits between them.

### *IEC 62351*

To address security of protocols used in the electric power industry, IEC has developed IEC 62351 standards for handling security of TC57 protocols including IEC 61850, and IEC 60870-5 and its derivatives (i.e., DNP3). IEC 62351 defines numerous mechanisms to protect exchange of information in automation applications [2]. The major goal of this standardization is to provide end-to-end security in power automation systems. Some standards are as follows:

- IEC 62351-3 identifies how to ensure secure TCP/IP-based protocols using transport layer security (TLS).
- IEC 62351-5 defines security for IEC 60870-5 and its derivatives, providing different solutions for serial and networked versions. It specifies how to incorporate user and device authentication, and data integrity.
- IEC 62351-8 deals in role-based access control (RBAC) for power system management. It covers the access control of users and automated agents to data objects in power systems by means of RBAC.

The National Institute of Standards and Technology (NIST) has recommended this family of standards for the smart grid.

### *Inter-Control Center Communications Protocol*

ICCP is meant for providing data exchange over wide area networks (WANs) between utility control centers, utilities, regional control centers, and non-utility generators. ICCP is an international standard, IEC TASE.2.

### *Distributed Network Protocol*

DNP3 is a protocol that defines communications between master stations, remote terminal units (RTUs), and IEDs in SCADA. IEEE has also adopted DNP3 as IEEE 1815-2010 [3], which will be upgraded to IEEE 1815-2012.

Initially, DNP3 was designed without any security features. DNP3 is extended to DNP3 Secure Authentication (SA) [4], which was designed to meet requirements of IEC 62351-5. DNP3-SA employs techniques including symmetric cryptography and hashed message authentication codes (HMACs). Implementation presumes that both master station and outstation share a common secret key, called an update key, which is used to generate a session key.

The recently released DNP3-SA5 reinforces overall security for data information gathering, exchange, and use in SCADA systems.

### *Substation Automation System*

SAS refers to a system that uses data from IEDs, and controls and manages automation capabilities within substations and control commands from remote users to control power system devices. Field devices (i.e., IEDs, RTUs, PLCs) used in SASs are the components that perform actual sensing and actuation functions throughout the smart grid. Typically these devices have limited processing capabilities, limited memory, and

often non-standard software platforms. Thus, the possibility of susceptibilities is increased; also, difficulties during the assessment process are created.

Substations in both the transmission and distribution domains have exceptional security requirements due to their geographic location. Due to the critical nature of transmitted data and profound use of wireless communication, a specific concern arises for communication links. All communication paths between the control center and substation, along with all inter-substation communications, necessitate comprehensive analysis.

### *Security Requirements and Threats*

Although the significance of specific threats can diverge greatly depending on the assets that need to be secured, some critical threats addressed in SCADA networks are as follows: bypass controls, spoofing attack, man-in-the-middle (MiTM) attack, modification attack, replay attack, insider attack, denial of Service (DoS) attack, and compromised user.

Key requirements that must be covered by a secure SCADA system are:

- Integrity — preventing unauthorized modification or theft of information
- Authentication and authorization — evading forgery/spoofing and unauthorized usage
- Availability — preventing DoS attack and ensuring authorized access to information
- Confidentiality — avoiding disclosure of information to unauthorized persons or systems
- Non-repudiation/accountability — preventing denial of an action that took place or claim of an action that did not take place

### *Challenges in Substation Automation System*

SAS often allows remote access, since maintenance personnel may need to access a substation facility for various purposes. They have to remotely access and manage data from IEDs/RTUs to optimize operation, maintenance, and asset management. Leveraging remote access capabilities in SASs can provide opportunities for data integration solutions including fault location, data archiving, and power quality monitoring. Encroaching on these benefits are new cyber security regulations that do not easily align with control system remote diagnostics.

In existent SCADA networks, the majority of IED implementation utilizes a security model based on permission levels protected by different passwords. In other words, IEDs normally possess passwords to control access to different levels of functionality such as reading data, modifying settings, and so on.

To simplify commissioning and field maintenance, SAS deployments generally use passwords that are shared and consistent throughout the entire deployment. In most SCADA networks, the utilities often use the same password for all field devices, or the same password is used by all mobile personnel. Due to complexity in changing passwords within devices, they generally do not change. The key problem with such a naive security approach is that it is impossible to identify a misbehaving individual [5]. The only way to disable access for a malicious individual is to change passwords in all devices, which is time consuming. Moreover, accountability cannot be achieved, which means permissions cannot be granted or revoked on an individual basis, in turn making it impossible to meet the requirements of existing security standards.



Most IEDs/RTUs use role-based authentication. Security policies are maintained in RBACs through granting of rights to roles rather than individuals. This model may have flaw since it cannot distinguish individual user if password is compromised.

Since existing IEDs do not have strong authentication and authorization capabilities, SCADA, including SAs, becomes vulnerable and faces various security threats. Thus, one of the key challenges is to achieve security enhanced remote access having full compliance with NERC CIP standards.

Most SCADA networks use integrated authentication solution with enterprise security infrastructure such as Active Directory through the use of RADIUS to authenticate users. An enterprise-level remote access server authenticates a user before establishing a communication link to IEDs in the substation. Remote access is entirely dependent on a central authentication server. Thus, to perform local maintenance, a field technician needs to be authenticated by the enterprise, which requires an active communications link. But if the communications link is broken, remote access to the substation may not be possible [6].

Utilities have to deploy remote access mechanisms for accessing IEDs/RTUs in SAs using various communication modes such as dialup modems, wireless access, and IP networks, which are strongly considered in NERC CIP.

Issues addressed are how to authenticate and authorize users to IEDs in substations in such a way that access is specific to a user, authentication information is specific to each user, and control of authentication and authorization can be centrally managed across all IEDs in the substation and across all substations belonging to the utility, and updated reasonably promptly to ensure that only intended users can authenticate to intended devices and perform authorized functions.

## System Architecture

Figure 1 shows the comprehensive substation automation architecture with various access configurations. It can be seen that IEDs in SAS can be reached through either local access or remote access.

Substations can be accessed remotely for troubleshooting and maintenance operations by utilizing various communication infrastructures including a public telecom network using dial-up services, public IP network using a router, a private WAN network, or even wireless access networks using 802.11 access points.

Users have to be uniquely identified and authenticated for local and remote accesses. Unique identification of personnel should be considered for detailed accountability of activity. Thus, prime requirements for remote access are controlling access to IEDs and ensuring accountability. In this regard, access to critical cyber assets should be restricted to authorized personnel only;

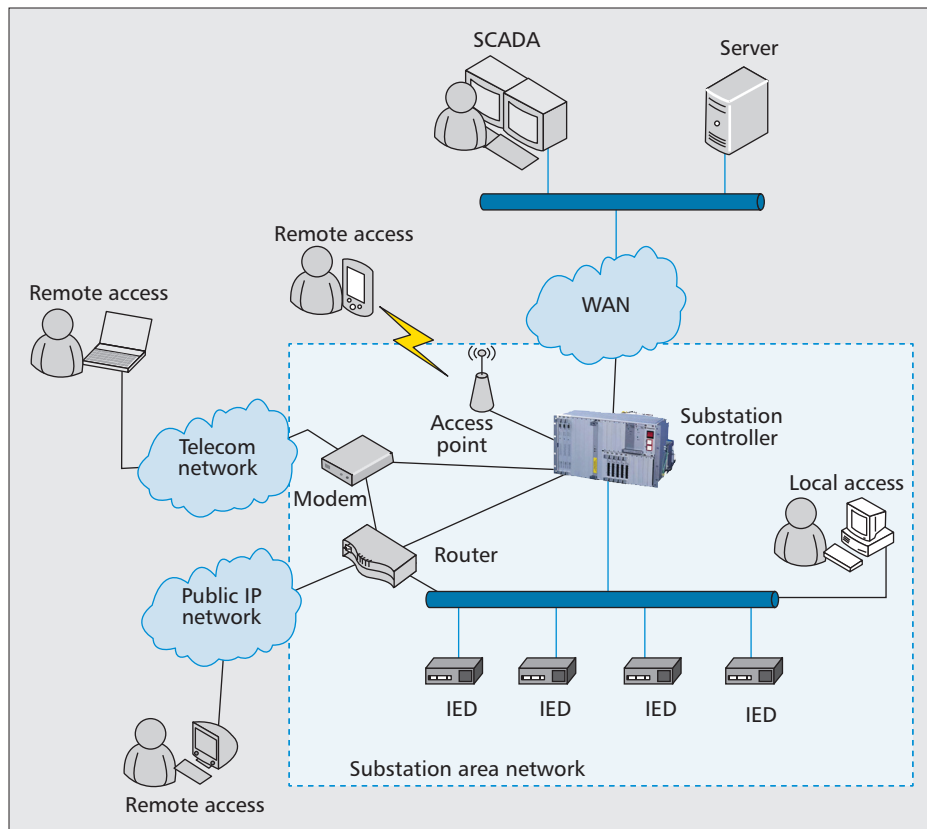


Figure 1. Comprehensive substation automation architecture with various access configurations.

detailed logs must be maintained for tracking events; and user access must be disabled immediately, whenever authority of personnel is terminated.

A solution with optimal authentication and access control should be implemented at the substation level. For that purpose, a substation controller or gateway or data concentrator can be used as a server to protect substation devices including IEDs. Such a device shall perform authentication and limit IED access to authorized users only.

In the proposed mechanism, a substation controller (SSC) is used as the central point of access control to substation equipment such as IEDs. While deploying substation-level authentication and access control, a substation controller should be able to perform the functions of authenticating users, assigning attribute certificates to users, and keeping access logs.

One of the requirements of a substation security solution is that it should not only rely on a remote central authentication server to provide user authentication and authorization. Even if a communications link is not active, remote users could still be able to securely access substation equipment using other means of communications such as a public IP network, wireless access, and dialup.

## Proposed Substation Security Approach

In this section, we propose a comprehensive substation-level security mechanism for SCADA in smart grid infrastructure that provides robust and lightweight authentication and authorization. Besides smart grid, other SCADA systems (i.e., gas, water, and pipeline), distributed control systems (DCSs) (i.e. oil and gas refineries), and other critical infrastructure systems can deploy the proposed approach because they have several common requirements.

Typical characteristics of SCADA networks make it challenging to adapt cryptographic protocols such as public-key crypto-systems into these systems. For instance, restrictions include limited computational and storage capabilities of field devices, low-rate data transmission on SCADA networks, and the necessity for real-time responses from devices across the network. Although IEC 62351 explicitly specifies RSA as a solution to protect time-critical messages in SAs, elliptic curve cryptography (ECC) has attracted increasing attention in SCADA networks since it has advantages over RSA in terms of required key lengths and processing times. ECC necessitates not only less power consumption and computation, but also reduced amounts of data transmitted and stored, so these factors are essential requirements for SCADA systems. In our proposed approaches, ECC is considered as asymmetric cryptography.

It has the following features: ECC-based public-key cryptography (PKC); zero-knowledge proof protocol with server-aided verification (SAV); multi-factor multilevel authentication (e.g., verifications of user password embedded implicit certificate at SSC as well as device password embedded response at IED); and attribute certificate (AC) [7] for authorization services.

The proposed authentication mechanism includes the following phases: initialization and registration, authentication, and authorization. The authentication phase can be either authentication scheme A or authentication scheme B depending on remote access services.

Along with authentication and authorization techniques, SCADA systems need to be integrated with intrusion detection systems (IDSs) to detect cyber attacks, and provide real-time or near-real-time warning of attempts to access system resources in an unauthorized manner as well as record information needed to legally prosecute the attacker.

### Initialization and Registration

Initially, a trusted authority (TA) has a pair of secret ( $x_{TA}$ ) and public ( $X_{TA}$ ) keys, and publishes  $X_{TA}$ . Assume that a user (UA) is the prover, whereas IED/RTU in the substation is a verifier and substation controller (SSC) serves as a gateway.

In order to access an IED in the substation, every UA has to obtain the password of that particular IED from the TA in the control center as a shared secret.

During the registration phase, using a user-selected password, every UA has to access the TA in order to securely obtain a token containing an implicit certificate ( $\sigma_A$ ) and signature parameter with which UA can compute secret ( $x_A$ ) and public ( $X_A$ ) keys.

### Authentication Schemes

There are two authentication mechanisms: authentication scheme A and authentication scheme B.

Authentication scheme A is used if UA has to access an IED through an SSC/gateway. Figure 2a shows message flows for authentication scheme A.

As shown in Fig. 2a, upon receiving UA's request containing witness ( $W$ ) and  $\sigma_A$ , SSC can verify implicit certificate verification equation (ICVE) [ $X_A = X_{TA}.h(\sigma_A) + R_A$ , where  $R_A$  is a public parameter of UA in order to validate both  $\sigma_A$  and  $X_A$ . If it holds, the SSC will forward  $\{W, X_A\}$  to the respective IED; otherwise, it will abort. With a challenge ( $c$ ) from the IED, UA computes a response ( $y$ ) (containing its hashed password  $[\sigma_A, x_A]$ , device password  $[pwd_D]$  of the respective IED as a shared secret) and send it to the SSC. Ultimately, the IED will verify response verification equation (RVE) [ $W = h(Y + h(c + pwd_D).X_A)$ ]. If it does hold, the IED accepts the response; otherwise, it discards it.

Authentication scheme B is used if UA has to access an IED directly through an IP network or a dialup modem. Figure 2b shows message flows for authentication scheme B. Upon receiving UA's request having  $W$ , the IED sends  $c$  to UA. UA will then send  $\{y, \sigma_A\}$  to the IED, which will just forward it to the SSC. SSC will verify ICVE. If it holds, the SSC will send  $\{Y, X_A\}$  to the IED; otherwise, it will abort. Next, the IED will verify RVE. If it does hold, the IED accepts the response; otherwise, it will discard it.

### Authorization

While implicit certificates are virtuous for authenticating users, there still remains to define authorization approach to specify what user can do. Attribute certificates (ACs), which are similar to PKCs, are used to store user-defined attributes. ACs are used for authorization services in many distributed environments.

Depending on need, an attribute certificate framework may have different authorization models in a privilege management infrastructure (PMI) environment. A delegation model consisting of four components is considered: source of authority (SOA), attribute authority (AA), privilege holder (PH), and privilege verifier (PV). Figure 3 shows the delegation model for PMI used in the proposed approach.

The AA is the entity that signs ACs, whereas the SOA is the root of trust of the PMI, which can delegate its power of authorization to subordinate AAs. The PH is the entity that holds a particular privilege and asserts its privileges for a particular context of use. The PV trusts the SOA as the authority for a given set of privileges for resource.

ACs are designed to be short-lived and have user-specific attributes about a given subject to facilitate flexible and scalable PMI. An AC may point to a public-key certificate that can be used to authenticate the identity of an AC holder. However, authorization information also needs to be bound to identity. The AC provides this binding; it is simply a digitally signed identity and a set of attributes. It contains serial number, issuer, holder, validity period, attribute info, and digital signature of the AA.

Using ACs, resource-constrained devices do not need to maintain access control lists that can potentially be large or always be connected to a network to access a central server.

Whenever UA requires access to an IED at the substation, both user and device can be authenticated through one of the above-mentioned authentication schemes. Once UA is authenticated, SSC then provides UA's AC, which describes UA's permissions. SSC defines user attributes for the individual UA. After computing a signature parameter that includes the hash value of the message containing user attributes, the SSC will send the AC along with an EC-based digital signature to UA.

When UA needs to access an IED, she will send a digitally signed message, an implicit certificate, and an AC. In order to accomplish the validation process, the first IED verifies the implicit certificate with the help of the SSC; then it will verify signatures on the message and AC by using the public keys of the sender and SSC. Based on information gathered from the AC, the IED will respond to UA.

If a user needs to have authorization permissions revoked, the SSC will issue an attribute certificate revocation list.

AC may be used with various security services, including access control, data origin authentication, and non-repudiation.

### Analysis of the Proposed Mechanism

We provide a security analysis of the proposed approach and efficiency analyze both authentication schemes. We discuss existing solutions and compare our approach with some existing solutions.

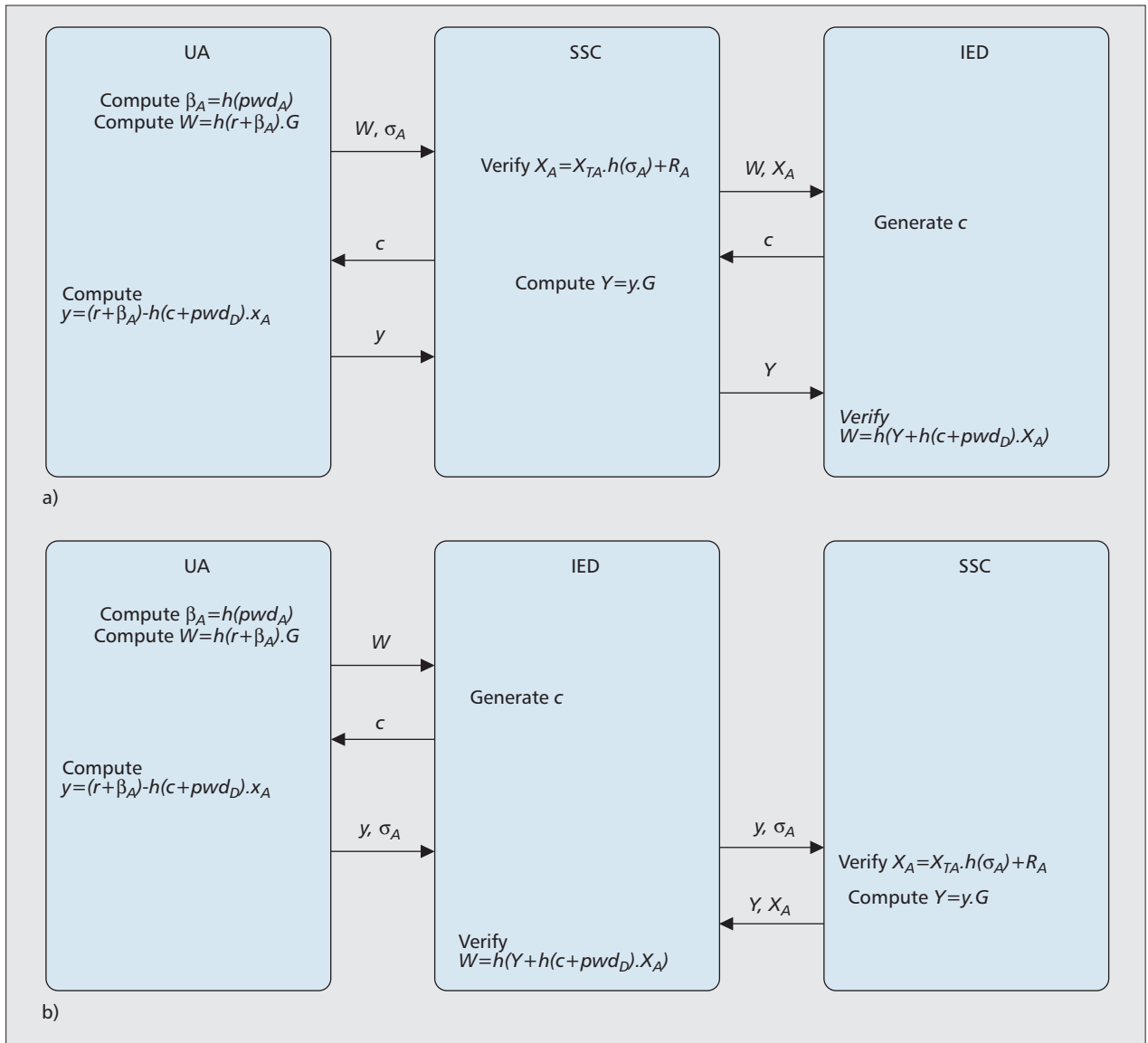


Figure 2. Message flow of authentication schemes: a) for authentication scheme A; b) for authentication scheme B.

### Security Analysis

We have provided a security analysis of the proposed security mechanisms, which can mitigate various attacks. Mathematical assumptions of proposed schemes are an EC discrete logarithm problem (ECDLP):

- Impersonation attack: If an adversary tries to impersonate a legal user, s/he needs to know the secret key, passwords, and commitment to compute valid response. Deriving the secret key and commitment is not feasible due to the intractability of ECDLP. Similarly, deriving user and device passwords is not possible due to hash functions being used.
- MiTM attack: If attacker wants to perform MiTM attacks, s/he needs to capture and modify communication flows between the UA and the SSC or IED. However, the proposed protocol can resist MiTM attacks since the adversary cannot know the commitment and secret key of UA as well as user and device passwords.
- Replay attack: If an attacker performs a replay attack, s/he will repeat a valid transmission maliciously. However, our protocol can resist a replay attack since commitment and challenge are chosen randomly during the protocol run.
- Insider attack: Using multilevel and multi-factor authentication

and attribute-based authorization, this approach can mitigate insider attacks to some extent. However, authentication and authorization are just preventative measures. Having auditing, monitoring, and logging mechanisms in the proposed system can provide detective and responsive measures for mitigating insider attacks.

- DoS attack: The proposed approach can mitigate DoS attack. Since this approach uses multi-layer authentication at the substation, SSC will allow only validated user to access IEDs/RTUs.
- Non-repudiation: Non-repudiation can be achieved, as UA has to use  $\sigma_A$  and AC to get authentication and authorization. Only the owner of a private key can send a target message/command. Since this message will be processed only if the signature is valid, the command will be executed only when its legitimate origin can be validated.

### Efficiency Analysis

We provide a performance evaluation of proposed authentication schemes. Table 1 shows an evaluation of computational costs for both schemes A and B.

For convenience, some notations are defined to denote time complexity; for example,  $t_H$  is the time complexity for

executing a hash function;  $t_M$  that for executing modular multiplication;  $t_A$  that for executing a modular addition;  $t_{ECM}$  that for executing multiplication of number and EC point; and  $t_{ECA}$  that for executing the addition of two EC points.

Although the overall time complexity of both schemes A and B is the same, they are quite different. In scheme A, verification by an SSC is accomplished before challenge and response generation, while verification by an IED is accomplished at the end of the authentication phase. In scheme B, verifications by both SSC and IED are done at the end of the authentication phase.

While considering computational cost in the authentication phase, which includes implicit certificate verification and response verification, both schemes require  $6t_H + 4t_A + 1t_M + 4t_{ECM} + 2t_{ECA}$ . However, response verification requires only  $2t_H + 1t_A + 1t_{ECM} + 1t_{ECA}$ , so computational cost at IEDs/RTUs is significantly small.

The proposed mechanisms require three message flows for communication between user and substation. It can be seen that communication cost for the proposed mechanisms is relatively low, since only a compact implicit certificate is used.

### Discussion

Several researchers have put effort into finding cyber security solutions for SCADA networks.

Khurana *et al.* [8] depicted key design principles and engineering practices that could help ensure the correctness and effectiveness of standards for authentication in smart grid protocols including DNP3.

Majdalawieh *et al.* [9] proposed DNPsec to address security in DNP3, which incorporates integrity, authentication, and non-repudiation mechanisms.

Fries *et al.* [2] discussed improvements to IEC 62351. On establishing a TLS connection, multiple users or applications

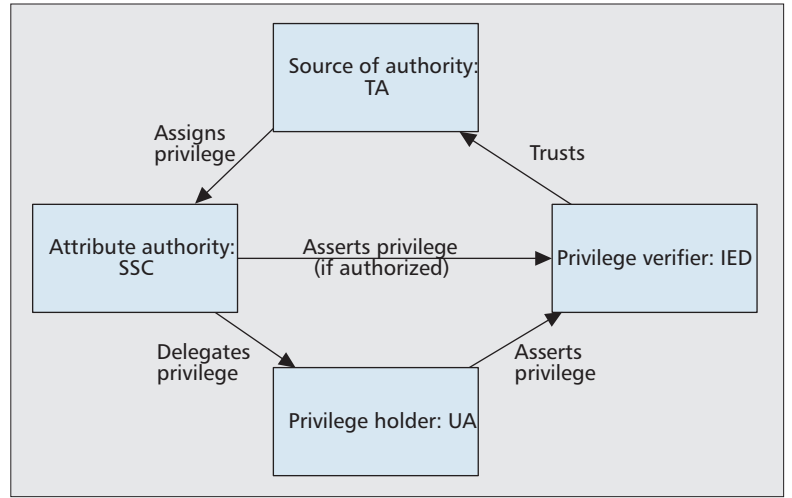


Figure 3. Delegation model for PMI.

can securely connect from the control center to field devices via a substation controller.

Vaidya *et al.* [10] proposed a substation-level authentication protocol for SCADA network. However, it does not consider authorization in a SCADA network.

Security mechanisms for DNP3 (e.g., DNP3-SA [4] and DNPsec [9]) are mainly used for secure connectivity between master station and outstation. Use of DNP3-SA and DNPsec can address various attacks including MiTM attacks. However, they do not provide any explicit authentication for the user. Furthermore, they do not work well when the link to the master station is malfunctioning.

Legacy security protocols such as IPSec can be used to create a secure virtual private network (VPN) tunnel between a utility control center and substation networks. However, the drawback is that authorization issues are not addressed.

With mounting interest in emerging smart grid technology, industrial control systems (ICSs) for electric utilities have

	UA	SSC	IED
<b>Authentication scheme A</b>			
Witness generation	$2t_H, 1t_A, 1t_{ECM}$		
Verification by SSC		$1t_H, 1t_{ECM}, 1t_{ECA}$	
Challenge and response generation	$1t_H, 2t_A, 1t_M$		
Verification by IED		$1t_{ECM}$	$2t_H, 1t_A, 1t_{ECM}, 1t_{ECA}$
Total	$3t_H, 3t_A, 1t_{ECM}, 1t_M$	$1t_H, 2t_{ECM}, 1t_{ECA}$	$2t_H, 1t_A, 1t_{ECM}, 1t_{ECA}$
<b>Authentication scheme B</b>			
Witness generation	$2t_H, 1t_A, 1t_{ECM}$		
Challenge and response generation	$1t_H, 2t_A, 1t_M$		
Verification by SSC and IED		$1t_H, 2t_{ECM}, 1t_{ECA}$	$2t_H, 1t_A, 1t_{ECM}, 1t_{ECA}$
Total	$3t_H, 3t_A, 1t_{ECM}, 1t_M$	$1t_H, 2t_{ECM}, 1t_{ECA}$	$2t_H, 1t_A, 1t_{ECM}, 1t_{ECA}$

Table 1. Evaluation of computational costs.

focused on migration to IP-based networks for control systems. For instance, ICS-specific communication protocols are being reformed with TCP/IP protocols or tunneled over IP networks. Due to growing interconnection with other enterprise networks or even the Internet means more exposure of automation systems to innumerable attacks. IP-based networks can be severely challenging to manage, since they increase the possibility of cyber security vulnerabilities and incidents. Inadequate authentication and access control in a substation may lead to unauthorized access that could provide the ability to corrupt a device. Proper and timely handling of vulnerabilities is a critical factor to reduce risk exposure to cyber security threats in SCADA networks.

With interconnection between SASs, corporate IT networks, and other third party IP networks, robust security at the substation level is fundamental to security of the SCADA network itself. Also, use of legacy security infrastructure such as firewalls, intrusion prevention/detection systems (IPS/IDS), or VPN technology is desired.

The proposed mechanism overcomes the above-stated limitations since remote access to resource-constrained devices (i.e., IEDs) in the substation can be achieved from any remote location by using a zero-knowledge protocol with the SAV technique. It also provides explicit authentication by using a user's password and token. With authentication and authorization credentials, the proposed approach provides high assurance capabilities allowing remote users to access IEDs/RTUs in the substation for authorized services, even when communication infrastructure with the control center is disrupted.

## Conclusion

In this article, we have proposed a lightweight and efficient substation-level security solution that provides multilevel multi-factor authentication and attribute-based authorization. We have devised a public-key certificate and zero-knowledge protocol-based SAV mechanism in which IEDs can authenticate any remote users with the help of an SSC as well as an access control mechanism using an AC. It can be seen that both authentication schemes are efficient and robust. We have shown that the proposed mechanism can provide better security than existing mechanisms.

## Acknowledgments

This work was supported by the Government of Ontario under the ORF-RE WISENSE project and the Natural Sciences and Engineering Research Council (NSERC) of Canada under NSERC Discovery Grant 2011-16.

## References

- [1] G. N. Ericsson, "Cyber Security and Power System Communication—Essential Parts of A Smart Grid Infrastructure," *IEEE Trans. Power Delivery*, vol. 25, issue 3, July 2010, pp. 1501–07.
- [2] S. Fries *et al.*, "Security for the Smart Grid – Enhancing IEC 62351 to

- Improve Security in Energy Automation Control," *Int'l. J. Advances in Security*, vol 3, no. 3–4, 2010, pp. 169–83.
- [3] IEEE 1815-2010 Standard for Electric Power Systems Communications – Distributed Network Protocol (DNP3), IEEE, July 2010.
- [4] G. Gilchrist, "Secure Authentication for DNP3," *Proc. IEEE Power and Energy Society General Meeting*, 2008.
- [5] Guidelines for Smart Grid Cyber Security, NISTIR 7628, vol. 1–3, NIST, U.S. Dept. of Commerce, Aug. 2010.
- [6] Z. Lu *et al.*, "Review and Evaluation of Security Threats on the Communication Networks in the Smart Grid," *Proc. IEEE MILCOM 2010*, Oct.–Nov. 2010, pp. 1830–35.
- [7] S. Farrell and R. Housley, "An Internet Attribute Certificate Profile for Authorization," RFC 5755, Jan. 2010.
- [8] H. Khurana *et al.*, "Design Principles for Power Grid Cyber-Infrastructure Authentication Protocols," *Proc. 43rd Hawaii Int'l. Conf. System Sciences*, Jan. 2010.
- [9] M. Majdalawieh, F. Parisi-Presicce, and D. Wijesekera, "DNPSec: Distributed Network Protocol Version 3 (DNP3) Security Framework," *Advances in Computer, Information, and Systems Sciences, and Engineering*, Springer, 2006, pp. 227–34.
- [10] B. Vaidya, D. Makrakis, and H. T. Mouftah "Provisioning Substation-Level Authentication in the Smart Grid Networks," *Proc. IEEE MILCOM 2011*, Nov. 2011, pp. 1189–94.

## Biographies

BINOD VAIDYA (bvaidya@site.uottawa.ca) is a postdoctoral fellow in the School of Electrical Engineering and Computer Science, University of Ottawa, Canada, since April 2010. Prior to joining the University of Ottawa, he worked as a post-doctoral researcher at Chosun University, South Korea (2007–2008), a research associate at Gwangju Institute of Science and Technology, South Korea (2008–2009), and a researcher at Instituto de Telecomunicações, Portugal (2009–2010). He also worked as a lecturer at the Institute of Engineering, Tribhuvan University, Nepal, for more than 15 years. He has authored or co-authored over 70 papers in international journals, books, conferences, and symposia. He has served as a guest editor for several reputed journals as well as an editorial board member for several international journals.

DIMITRIOS MAKRAKIS (dimitris@site.uottawa.ca) is a professor in the School of Electrical Engineering and Computer Science, University of Ottawa, and the director of the Broadband Wireless and Internetworking Research Laboratory. Prior to joining the University of Ottawa, he was an assistant and later associate professor in the Department of and Computer Engineering, University of Western Ontario, and the director of the Advanced Communications Engineering Centre (a research facility established in partnership with Bay Networks, presently Nortel, and Bell Canada). Before starting his academic carrier, he worked for the Canadian Government and Canadian industry. He has received the Premiers Research Excellence Award (PREA). He has authored or co-authored over 200 papers in international journals, conferences, and books.

HUSSEIN T. MOUFTAH (mouftah@site.uottawa.ca) joined the School of Electrical Engineering and Computer Science, University of Ottawa in September 2002 as a Canada Research Chair Professor. He worked as a full professor and associate head of the Electrical and Communications Engineering Department at Queen's University (1979–2002). He has three years of industrial experience mainly at BNR of Ottawa, now Nortel Networks (1977–1979). He served as Editor-in-Chief of *IEEE Communications Magazine* and IEEE ComSoc Director of Magazines, Chair of the Awards Committee, and Director of Education. He was a distinguished lecturer of IEEE ComSoc (2000–2007). He is the author or coauthor of six books, 40 book chapters, and more than 1000 technical papers and 10 patents. He is a Fellow of the Canadian Academy of Engineering, Engineering Institute of Canada and Royal Society of Canada: The Academy of Science.

---

# A Layered Encryption Mechanism for Networked Critical Infrastructures

Huayang Cao, Peidong Zhu, and Xicheng Lu, National University of Defense Technology, China  
Andrei Gurtov, Helsinki Institute for Information Technology, Finland

---

## Abstract

Networked critical infrastructures improve our lives, but they are attractive targets for adversaries. In such infrastructures, to secure sensitive data is vital, as the information system is a foundation of today's critical infrastructures, and data security is a main concern in such systems. Cryptography is an approach for data security, but this method should be altered according to various features of infrastructure networks. Since complex and distributed critical infrastructures usually spread over large geographic areas, different parts of those infrastructures have different levels of perimeter defense. Devices in weakly protected zones are more likely to be captured than those in well protected zones. If an adversary captures devices, s/he can bypass cyber security measures and obtain secret information directly. Such a threat requires a layered security mechanism that can prevent adversaries from invading the whole infrastructure network from these weak zones. In this article, we propose a layered encryption mechanism based on hash chain technology for protecting sensitive data. Besides showing the layered defense, the mechanism is also lightweight and has convenient key management. It can be used independently or as a supplement to existing security measures. We evaluate performance of the proposed mechanism over different kinds of devices.

---

**A** modern human society heavily relies on critical infrastructures, including energy, water, information and telecommunications, banking and finance, and emergency services, just to name a few. Attacks on or misuse of these infrastructures can lead to economic and human losses. Today, critical infrastructures have been establishing widely distributed networks, which make them more effective but also more prone to cyber attacks, as adversaries can take advantage of accessible points at various locations across the whole network.

Since people prefer to use existing network technologies in critical infrastructures, these infrastructures are susceptible to the set of security threats that are encountered by conventional computer networks as well. For example, in 2011, multiple cyber worms were detected in industrial networks [1, references therein]. These worms tried to steal information about industrial control systems, and were believed to be the precursor of another destructive attack. Generally, different critical infrastructures care about diverse security aspects, but to secure sensitive data is a common concern for many networked infrastructures. Leakage of or tampering with sensitive data may cause privacy exposure (electricity consumption information collected by an advanced metering infrastructure [AMI]), economic loss (commercial information in a financial system), catastrophic attacks (industrial process information in plant), and so on. A helpful strategy to enhance data security is to apply cryptography technologies. This approach prevents data from being eavesdropped, tampered with, or forged.

In this article, we focus on the need for data security in common complex and distributed critical infrastructures, and build a layered defense mechanism in the data plane based on

cryptography technologies. To establish such a mechanism, however, we need to consider the features and special requirements in these networked infrastructures.

The remainder of the article is organized as follows. We describe the features and special requirements in complex and distributed critical infrastructures. We provide our considered threat model. We explain our cryptography mechanism, which is based on hash chain technology, for securing data in a layered manner and evaluate its performance. We present related work in mitigation of threats in critical infrastructures. Finally, the article is concluded.

## *Complex and Distributed Critical Infrastructures*

This section describes features of critical infrastructures. Since critical infrastructures vary from industrial base to telecommunications to finance, this article does not dig into a specific architecture. Instead, we consider their common features, which are shown in Fig. 1. While real critical infrastructure networks are far more complex and larger, this sample illustrates some intrinsic properties.

Networked critical infrastructure is normally divided into multiple hierarchical zones. An infrastructure has one or several top zones, which could be infrastructure data centers, top control areas or other critical parts. These zones are well protected with strong perimeters. On the other side, affording the same perimeter defenses for every network zone is impossible due to the large scale and complexity, which gives potential attackers easy access to network components. Thus, different

network zones inherently have different security levels. We can label those zones with similar security defense measures with the same security level, as they are vulnerable to similar threats. The threat that devices with low security levels can be captured is the main point we consider in this article, which is different from computer networks where the security measures usually consider cyber attacks only.

The second feature we consider is their complex communication links. Some components are connected with hub nodes, while others are connected in a meshed network. Additionally, these links can be built over various networking technologies, and should be assumed to be unsafe, that is, cyber attacks such as eavesdropping and man-in-the-middle attacks could happen.

In today's critical infrastructures, various devices are adopted. They have different capabilities and are interconnected with lots of open or private communication protocols. It is infeasible to apply a security protocol to all kinds of devices. Instead, a lightweight security mechanism is more suitable.

We assume that devices in modern critical infrastructure networks are capable of cryptography operations. This assumption may seem unreal in today's critical infrastructures, as many legacy devices are implemented over decades. However, since most legacy devices in smart grids will be replaced in the coming decades [2], and some small devices (e.g., smart meters) have cryptography function, we can also expect updates in other critical infrastructures.

It should also be noted that data transmissions in networked critical infrastructures are dominated by machine-to-machine (M2M) communications. Such networks are expected to be available for 24 hours a day, and are not supposed to be interfered with by humans. Moreover, in the M2M communication environment, real-time communications are often required. Such features require a lightweight security mechanism that is more stable than heavyweight ones, with fewer chances of errors, better support for real-time communications, and greater ease of deployment.

There may be other features in networked infrastructures beyond the aforementioned ones. Digging into more features may be helpful for enhancing security further. However, it is out of this article's scope and will be our future work.

### Considered Threat Model

To launch an attack, an adversary should exploit entry points first. By utilizing these entries, the adversary could deliver specific cyber attacks on critical infrastructures.

**1) Entry points:** Generally, perimeter defense is required to protect devices and information within critical infrastructures. But there are always weak points due to the scale and complexity:

**Weakly protected zones:** A modern infrastructure network can be widely distributed, involving numbers of subnetwork

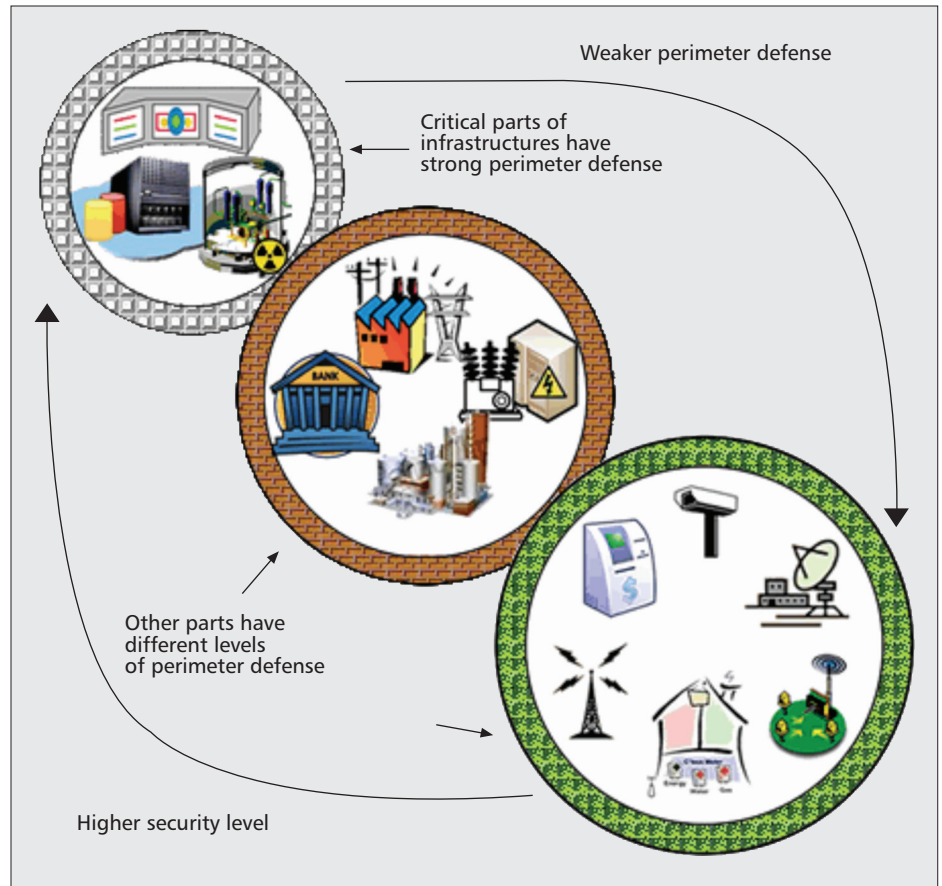


Figure 1. Illustration of critical infrastructures.

zones. As mentioned before, these zones have different security levels. Some of them are more prone to attacks. For example, meters in a smart grid reach every house and building, telecommunication and finance terminals appear in every corner in modern cities, and various monitoring devices are often deployed in remote areas. An adversary can easily access devices in these areas.

**Communication links:** In a network, links are often assumed to be insecure. Transferring data over unsecured links is the main purpose for inventing cryptography technologies. Given the scale, complexity, and diversity of critical infrastructures, it is difficult to protect communication links, which serve as entry points in infrastructure networks.

**Insider threat:** An insider agent is another kind of entry point. When an insider agent prepares to launch an attack, it means the attacker has successfully compromised the perimeter defense and invaded the corresponding security level.

**2) Attack actions:** Once an adversary gains entry points, s/he could perform corresponding actions on data.

**Compromised device attack:** Since a compromised device can pass authentication of critical infrastructures, the adversary may utilize such a device to establish communications with others for arbitrary purposes.

**Eavesdropping attack:** An adversary could obtain sensitive data by monitoring network traffic. Besides violating data confidentiality, such an attack could be used for gathering information to perpetrate further crimes. Even encrypted communications could be susceptible because an adversary could obtain encryption keys from weakly protected devices.

**Man-in-the-middle attack:** If an adversary obtains encryption keys from captured devices, s/he could penetrate other communications and manipulate data. Malicious modification on data can cause various loss events.

The threats mentioned above are not exhaustive, but they are the most typical threats on data security in networked infrastructures. It is concluded that due to the physical threats, even encrypted communications can't guarantee the data security, as encryption keys may be leaked. This threat enables an adversary with the ability to penetrate in the whole network from weakly protected zones, which is so called stepping stone attack. Assigning independent keys to each pair of devices or zones is also inapplicable due to the scale and complexity.

However, this layered feature also provides us with the possibility of applying a layered encryption mechanism which can perform layered defense and thus prevent adversaries from launching a stepping stone attack.

## Layered Encryption Mechanism

In this section, with Hash Chain-based encryption, we propose a layered defense mechanism in networked critical infrastructures.

### Hash Chain

The basic idea behind Hash Chain is applying a hash function, say  $H$ , on a pre-selected value successively until the number of results has reached the desired amount. This result-chain is called the hash chain.

$$h_0 = r$$

$$h_n = H(h_{n-1}) = H(H(H(\dots H(r) \dots))) \text{ } n \text{ times}$$

Hash Chain was first proposed by Lamport as an authentication measure [3]. In this article, we don't use it to perform authentication, but as an encryption key generation method. Among all properties that Hash Chain provides, the following appealing ones enable it suitable to our layered encryption mechanism.

- It is easy to find  $h_i$  if  $h_{i-1}$  is given
- It is computationally hard to find  $h_{i-1}$  if only  $h_i$  is given

The above properties arise from the preimage resistance which is a fundamental requirement on hash function.

### Design of Encryption Mechanism

As we mentioned before, the potential threat in physical space is the main feature of critical infrastructure networks compared to traditional computer networks. Considering the threat in physical space, we should adopt the following one-way policy in cyber space as supplementary:

- A zone is able to get all data from zones with lower or the same security level

*If adversaries could invade in a zone, they can also invade in other zones with the same or lower security levels. So this policy doesn't further harm security of the infrastructure.*

- A zone can only get partial data from zones with higher security level

*According to principle of layered defense, adversaries who invade in a system can only receive the data that are originally destined to the compromised devices, but can't get any other data that are originally transferred between higher security level zones, e.g., by eavesdropping.*

To meet this policy for data security, networked critical infrastructures need an encrypted communication mechanism which has layered defense ability. Our proposal utilizes symmetric cryptography and one-way encryption keys to meet such requirement, that is, higher security level zones can get lower level keys, but not vice-versa. One-way keys are obtained based on the preimage resistance property of Hash Chain. The associated issues of the mechanism include Key Distribu-

tion, Pair-key Negotiation, Key Update and Security Level Update. We also discuss issues about incremental deployment and authentication techniques at the end of this section.

### Key Distribution

According to layered defense principle, critical infrastructure owners should divide their network zones into several security levels with different levels of perimeter defense. The owners can refer to IATF report [4] or other standards for such operation. In our mechanism, the high security levels are assigned to zones with strong perimeter defenses, and lower levels correspond to weaker perimeter defenses. However, it is not necessary to define exact values for security levels, but just ensure the sequence of them. It is also not required to add any facilities for bonding the same-security-level zones together, but just assign the same *security level* and *key* to devices in these zones. Finally, we get an infrastructure network with multiple security levels, and several zones for each level.

We adopt a key predistribution scheme, and introduce a key management device that can be a high-performance computer located in a top security zone. The distribution process can be organized into three steps. First, the key management device generates a series of unique *secret identities* for all devices. Then it generates a hash chain. Elements in this chain are assigned to devices as keys. Finally, we put all necessary information into devices' memory as follows:

- **Key:** Used for encrypted communications. It is selected from the generated hash chain, based on a device's security level. A higher security level corresponds with an earlier generated element in the hash chain. Devices with the same security level will get the same keys.
- **Key version:** During the lifetime of an infrastructure, the encryption key may be updated many times. The initial version value is zero, and is modified after each update.
- **Security level:** Indicates the security level of a device. This field also helps to negotiate key pairs between zones. It is collected from infrastructure owners.
- **Security level version:** Similar to key version, the security level of devices may change over time.
- **Secret identity:** Generated by the key management device and unique among all devices. It is used for update processes and does not appear in any communications, so it is impossible for a third party to get this information.

### Pair-Key Negotiation

After key predistribution, encrypted communications can be enabled by first negotiating about key pairs between devices. In our mechanism, key pair negotiation is quite easy to implement. The negotiation is mainly based on the security level fields in devices' memory. Each device exchanges its own security level with its neighbors and decides the key pair between them as follows:

- If the neighbor's security level is higher than or equal to its own, it will use its key as the pair key to this neighbor.
- If the neighbor's security level is lower than its own, it will iteratively apply a hash function on the key according to the deviation value and get the result as its pair key to this neighbor.

After this exchange, a device stores neighbors' security levels and pair keys for secure communications. During exchange, the security level is transferred in plain text, so an adversary can also get this information. However, based on hash function, devices decide the key pairs locally, and do not transfer them back to neighbors, so the adversary cannot obtain secret information. Figure 2 shows that devices with the same security level share the same key, and a cross-level key can be obtained with the above rules.



## Key Update

The threat of device capture may cause key leakage. Although the layered defense mechanism helps high security level zones to resist this threat, the adversary can decrypt data in the same or lower security level zones. To mitigate this threat, keys should be updated once the security incident is sensed. Key update can also happen at the occurrence of network change.

At the beginning of update, key management device calculates a new keys-chain, and distributes them in the form of  $\{flag, vk, C(NKs)\}$  where  $flag$  is set to indicate a key update,  $vk$  means the key version in current update,  $NKs$  means the newly generated keys, and  $C(NKs)$  indicates encrypting each key with a corresponding device's secret identity and joining all results into one. We use the secret identity as the encryption key to ensure that only valid devices can get a newly generated key. We can remove any device from the infrastructure network by just removing its secret identity in the next update. It should be noted that the generated  $C(NKs)$  may be very large in size and may be fragmented when transferring, so the transmission protocol should be able to perform fragmentation and restructuring. However, this is not a major consideration in this article.

The key management device can send  $\{flag, vk, C(NKs)\}$  to any devices in an infrastructure network. After receiving, the receiver will check  $flag$  and  $vk$ . If  $vk$  is not greater than its key version, it will neglect this update; otherwise, it will try to decrypt  $C(NKs)$ . If successful, it will update its key to the new one as well as update the key version to  $vk$ , remove its own key from  $C(NKs)$ , and forward this modified key update to other neighbors. After this, the device updates its pair keys, which can be done locally, as the security levels of its neighbors have been stored. On the other hand, if it cannot extract a key from  $C(NKs)$ , this update is discarded.

In contrast, when installing a new device in an existing network, we can fill secret identity and security level as we did during the establishment, and select a proper element from the current key chain as well as the key/security level version in the key management device and fill them in this device.

## Security Level Update

With the evolving of networked critical infrastructures, the physical environment and security policy may change over time, which requires update of the security level of devices. Given the changed security level, the key of devices should also be updated accordingly.

This update process is similar to that of key update. The key management device gets a new security level for each device from infrastructure owners, generates new keys, and sends an update as  $\{flag, vl, C(NLs, NKs)\}$ , where  $flag$  is set to indicate a security level update,  $vl$  is the security level version, and  $C(NLs, NKs)$  are the encrypted and concatenated security levels and keys for each device.

Upon receiving, the receiver also checks  $flag$  and  $vl$ , and then extracts and updates the security level as well as the key. The security level version is updated to the new one, and the key version is increased by one on both the key management device and normal devices sides. The update also triggers another key pair negotiation between devices.

It can be seen from the above description that no explicit authentication is employed in our proposal. Instead, authentication is achieved by the key predistribution and secret identity scheme, as only those devices having their secret identities adopted in the update process can update keys successfully and work normally afterward. Such an authentication policy is

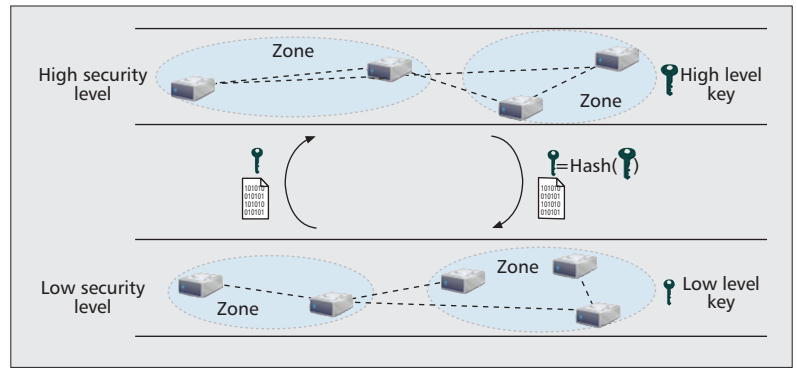


Figure 2. Security levels and key sharing.

efficient compared to public-key style authentication. However, a pair of devices may need authentication for each other in some scenarios. For this requirement, two devices can set up a shared key by using a Diffie–Hellman process, and the basic authentication feature of our mechanism ensures the validity of both devices. After that, the HMAC method can be utilized to ensure the authenticity, as well as integrity, for data security.

Another issue that may arise is applying the mechanism over legacy devices. As described before, legacy devices may not be capable of cryptography operation. Nowadays, there are many cheap encryption chips on the market that can be adopted to implement our mechanism. On the other hand, infrastructure owners can also set up gateways for legacy devices. A gateway device performs as a security proxy. It communicates with updated devices as well as other gateways in a cryptography-enabled way, and with legacy devices in the old ways. By introducing gateway devices, the infrastructure network can benefit from encrypted communications and the layered defense feature. However, communications within a zone of legacy devices can still suffer from threats. We recommend that the legacy devices should be updated according to the security requirement.

## Performance Evaluation

We assume that there are  $n$  devices with  $k$  security levels in an infrastructure network, and each device has  $m$  communication neighbors on average. According to the above description, the computational and storage cost of each device can be estimated in Table 1.

It can be seen that in most scenarios, the computational cost for normal devices is low. In some rare cases, normal devices have to perform a lot of decryption, because we only employ a basic update scheme for these processes, and such cost can be dramatically reduced by introducing a kind of fast lookup technology. Our main focus for normal devices is on the most frequent scenario (i.e., encrypted communications). We also test the performance of key generation, as it is the most time-consuming operation for a key management device.

## Performance of Encrypted Communication

One of the main negative effects of using cryptography is the extra delay during communications. We use Advanced Encryption Standard (AES) in the test, with 128 bits key and 1024 bits data to be encrypted. The algorithm implementation is based on the Cryptlib [5] toolkit. Figure 3 shows our evaluation on three different kinds of devices. A Nokia Internet tablet 770 has an ARM CPU running at 220 MHz and 64 Mbytes RAM. An N810 comes with 400 MHz CPU and 128 Mbytes RAM. A ThinkPad laptop has Intel CPU with two 2.67 GHz cores and 768 Mbytes RAM. The round-trip time (RTT) is determined by *ping* test in a wireless LAN. Both

Computational cost	Freq.	Key management device	Normal device
Key distribution	Once	$O(n)$ for generating random value $O(k)$ for generating hash value	None
Pair-key negotiation	Not often	None	$O(m)$ for comparison
Key update	Rarely	$O(n)$ for encryption $O(k)$ for generating hash value	$O(n-\log_m n)$ for decryption
Security level update	Rarely	$O(n)$ for encryption $O(k)$ for generating hash value	$O(m)$ for comparison $O(n-\log_m n)$ for decryption
Encrypted communication	Always	None	$O(1)$ for encryption/decryption
Storage cost	N/A	$O(n)$ for secret identities and corresponding security levels $O(k)$ for keys $O(1)$ for key/security level version	$O(m)$ for neighboring devices' security levels and pair-keys

Table 1. Computational and storage cost of each device.

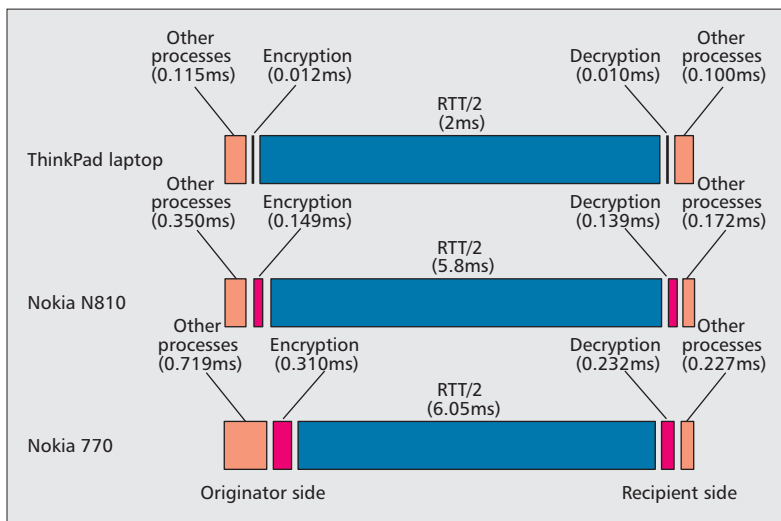


Figure 3. Performance test of encrypted communications.

	Public key	Kerberos/Needham-Schroeder	Layered encryption
Confidentiality	Yes	Yes	Yes
Integrity	Yes	Yes	Yes
Authentication	Yes	Yes	Yes
Third-party device	Required	Required	Required
Computational cost	High	Middle	Low
Storage cost	Acceptable	Acceptable	Acceptable
Operation complexity	Middle	High	Low
Protection grain	Fine-grained	Fine-grained	Coarse-grained layered defense

Table 2. Comparison among three kinds of data security measures.

RTT and cryptography performance values are arithmetic mean values of 1000 samples.

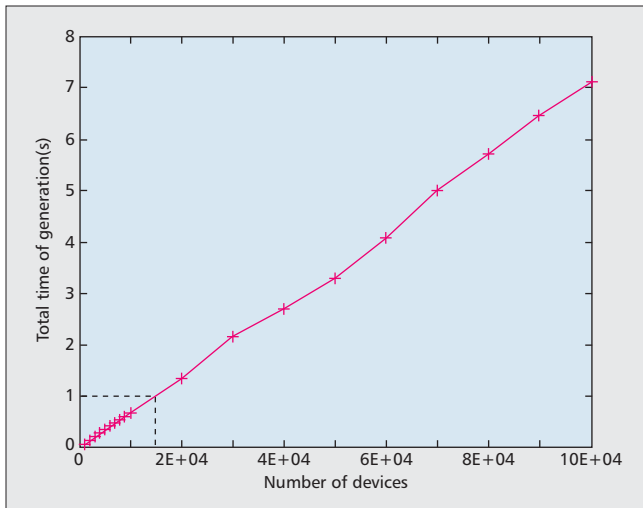
The other processes in Fig. 3 indicate data preparation and other normal operations. It can be seen that the process time varies for different devices, but transmissions always consume more than 80 percent of the whole time, and the cryptography operation even costs less time than data preparation. In a real infrastructure network, the RTT can be much longer, which makes the cryptography delay negligible for the whole process.

It is also worth noting that there is only one packet transmission during the key pair decision in our mechanism, which ensures a very low latency in communications, as RTT consumes more time than symmetric cryptography.

#### Performance of Key Generation

During the update process, the key management device must generate layered encryption keys and encrypt them with the corresponding device's secret identity. We use the ThinkPad laptop as the key management device, which will generate a series of 128-bit keys with MD5 hash function and encrypt each of them with another corresponding 128-bit secret identity and AES algorithm. Since we use the hash chain as encryption keys but not as an authentication measure, there is no need to worry about the flaws of MD5 in collision resistance [6]. Figure 4 demonstrates the time of generating a new version of keys for a target infrastructure. Here we arrange all devices into a pyramid with 10 security levels, but the hierarchy does not affect the total time in reality, as generation of every key involves the same amount of hash and encryption operation, and the total time only depends on the number of devices. It can be seen that when the number of devices is less than 10,000, all keys can be generated in 1 s. And when the number reaches 100,000, the generation can be done in about 7 s. Although we do not get the exact number of devices in a real infrastructure, we believe such performance is quite acceptable.

It should be noted that our tests in this section are based on common hardware and a C lan-



**Figure 4.** Key generation time according to the number of devices.

guage library. By special design and using an assembly language written algorithm, dedicated hardware such as encryption chips can achieve even higher performance on symmetric cryptography.

As for data security in a network environment, there are two other alternatives, public key and Kerberos/Needham-Schroeder style schemes. The latter can also ensure data security with symmetric cryptography. We do not present their details here, but have a brief comparison of our mechanism and the two alternatives in Table 2.

We can conclude from Table 2 that public key and Kerberos/Needham-Schroeder schemes can also guarantee data security. However, they are off-the-shelf security measures designed for common networks, and fail to consider features of infrastructure networks, such as diversity of devices, dominant M2M communications, and layers. Their main drawbacks are expensive computational cost and complicated interaction operations, which makes them impossible to be adopted in many infrastructure networks. In contrast, the layered encryption mechanism in this article is more suitable for a networked infrastructure environment.

### Related Research Work

The cyber security of critical infrastructures has attracted much attention recently. Since legacy systems lack sufficient security mechanisms in general, many efforts have been undertaken to provide additional security for traditional communication protocols [7, 8]. However, the deployment and key management, as well as authentication still provide difficulties in this area. Alzaid *et al.* propose a secure key management scheme for PCS/SCADA systems [9]. This work and its relevant series of works fully take into account the threat of device capture, and make efforts to enhance key management. The main scenario they consider is wireless sensor networks in which all devices suffer from the same level of physical threat. Thus, a complicated scheme is required to provide fine-grained key management. When we protect the critical infrastructure as a whole, the different levels of perimeter defense should be considered in the design of a security mechanism.

Another thread of research is providing filtering in critical infrastructures. Firewalls, as an example, are mainly used as a barrier to prevent unauthorized or unwanted communications. Stateful filtering in firewalls is generally believed to be more reliable. A formal language for describing system state is introduced in [10]. In this work, attacks on an industrial sys-

tem can be detected by monitoring the state evolution of the system. Also, Fadlullah *et al.* model the malicious attack event as a Gaussian process, and propose an early warning system to detect the abnormal traffic in a smart grid network [11]. The Dates project is an integrated monitoring system for oil, gas, and electric systems [12]. This kind of work is compatible with our proposal as they work on different layers. Also, the cooperation of them may simplify the overall security policy for infrastructure networks.

### Conclusion

In this article, we have proposed a mechanism for protecting data security in complex and distributed critical infrastructure networks. The mechanism uses cryptography techniques to protect sensitive data based on novel usage of hash chain technology. Based on the layered feature of critical infrastructures, our mechanism introduces a hash chain to achieve the following two main goals: layered defense, which prevents infrastructure networks from stepping stone attacks, and convenient key management, which enables cross-layer communications as simple as in-layer communications. Moreover, it is a lightweight solution as it only performs in the data plane. This layered encryption mechanism can be used independently or as a secure key management scheme for existing communication protocols. The performance evaluation demonstrated the effectiveness of the proposed mechanism.

It is worth noting that the insider threat is still a nut in today's infrastructure networks. In our mechanism, the insider threat means the attacker can successfully invade a security level. Although the layered defense measure still works, such a threat can have a great impact on low security level zones. In other words, infrastructure owners should pay attention to the insider threat when setting up perimeter defenses, such as adopting the least privilege principle on access control and installing anti-temper hardware in high security level zones. Besides enhancing security technology, education/training and workforce enhancement should also be emphasized.

After the work in this article, we believe that applying other special cryptography technologies (e.g., identity or attribute based cryptography) in critical infrastructure networks is worthy of deep investigation. Since a single technology can hardly meet all security requirements in a system, other technologies, such as firewalls, whitelists, and WSN security measures, can also be combined with our proposal to further enhance the cyber security of networked critical infrastructures.

### Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grant 61170285 and 61202486, Hunan Provincial Innovation Foundation for Postgraduate under Grant CX2010B030, and NUDT Innovation Foundation for Postgraduate under Grant B100605.

### References

- [1] E. J. Byres, SCADA Security 2012 Crystal Ball, Tofino Security, 2012; <http://www.tofinosecurity.com/blog/scada-security-2012-crystal-ball>.
- [2] G. W. Arnold, "Challenges and Opportunities in Smart Grid: A Position Article," *Proc. IEEE*, vol. 99, no. 6, 2011, pp. 922-27.
- [3] L. Lamport, "Password Authentication with Insecure Communication," *Commun. ACM*, vol. 24, no. 11, 1981, pp. 770-72.
- [4] Information Assurance Technical Framework (IATF) Release 3.0, National Security Agency, 2000; <http://www.dtic.mil/dtic/tr/fulltext/u2/a393328.pdf>.
- [5] Cryptlib Encryption Toolkit; <http://www.cs.auckland.ac.nz/~pgut001/cryptlib/>.
- [6] X. Wang and H. Yu, "How to Break MD5 and Other Hash Functions," *Proc. 24th Annual Int'l. Conf. Theory and Applications of Cryptographic Techniques*, 2005, pp. 19-35.
- [7] M. Majdalawieh, F. Parisi-Presicce, and D. Wijesekera, "DNPSec: Dis-

- 
- tributed Network Protocol Version 3 (DNP3) Security Framework, *Advances in Computer, Information, and Systems Sciences, and Engineering*, 2006, pp. 227–34.
- [8]. I. Nai Fovino *et al.*, “Design and Implementation of a Secure Modbus Protocol, Critical Infrastructure Protection III,” vol. 311, 2009, pp. 83–96.
- [9]. H. Alzaid *et al.*, “Mitigating Sandwich Attacks against a Secure Key Management Scheme in Wireless Sensor Networks for PCS/SCADA,” *Proc. 24th Int’l. Conf. Advanced Information Networking and Applications*, 2010, pp. 859–65.
- [10]. A. Carcano *et al.*, “A Multidimensional Critical State Analysis for Detecting Intrusions in SCADA Systems,” *IEEE Trans. Industrial Informatics*, vol. 7, no. 2, 2011, pp. 179–86.
- [11] Z. Md. Fadlullah *et al.*, “An Early Warning System against Malicious Activities for Smart Grid Communications,” *IEEE Network*, vol. 25, no. 5, 2011, pp. 50–55.
- [12]. Detection and Analysis of Threats to the Energy Sector (DATES), SRI International; <http://www.csl.sri.com/projects/dates/>.

### Biographies

HUAYANG CAO ([huayang.cao@gmail.com](mailto:huayang.cao@gmail.com)) received his M.S. degree in computer science and technology from National University of Defense Technology (NUDT), Changsha, China, in 2009, where he is currently pursuing a Ph.D. degree in computer science and technology. He visited HIIT, Finland as a visiting Ph.D. student during December 2011–November

2012. His research is focused on Internet routing security, and security for cyber-physical systems and social networks.

PEIDONG ZHU [SM] ([zpd136@gmail.com](mailto:zpd136@gmail.com)) is a professor with the School of Computer Science of NUDT. He received his Ph.D. degree in computer science from NUDT in 1999. During 2009, he was a visiting professor at St. Francis Xavier University, Canada. His research interests include large-scale cyber-physical-social systems, network routing and security, and architecture of the Internet.

XICHENG LU ([xclu@nudt.edu.cn](mailto:xclu@nudt.edu.cn)) received his B.Sc. degree in computer science from Harbin Military Engineering Institute, China, in 1970. He was a visiting scholar at the University of Massachusetts between 1982 and 1984. He is currently a professor in the School of Computer Science, NUDT. His research interests include distributed computing, computer networks, and parallel computing. He has served as a member of editorial boards of several journals and has co-chaired many professional conferences. He is an academician of the Chinese Academy of Engineering.

ANDREI GURTOV ([gurtov@hiit.fi](mailto:gurtov@hiit.fi)) received M.Sc. (2000) and Ph.D. (2004) degrees in computer science from the University of Helsinki, Finland. He is a professor at the University of Oulu and an adjunct professor at Aalto University. His research interests include network security, multipath and multicast communication, and sensor and peer-to-peer networks.

---

# In Quest of Benchmarking Security Risks to Cyber-Physical Systems

**Saurabh Amin, Massachusetts Institute of Technology**  
**Galina A. Schwartz, University of California at Berkeley**  
**Alefiya Hussain, University of Southern California**

---

## Abstract

We present a generic yet practical framework for assessing security risks to cyber-physical systems (CPSs). Our framework can be used to benchmark security risks when information is less than perfect, and interdependencies of physical and computational components may result in correlated failures. Such environments are prone to externalities, and can cause huge societal losses. We focus on the risks that arise from interdependent reliability failures (faults) and security failures (attacks). We advocate that a sound assessment of these risks requires explicit modeling of the effects of both technology-based defenses and institutions necessary for supporting them. Thus, we consider technology-based security defenses grounded in information security tools and fault-tolerant control in conjunction with institutional structures. Our game-theoretic approach to estimating security risks facilitates more effective defenses, especially against correlated failures.

---

**S**urvivability of critical infrastructures in the presence of security attacks and random faults is of national importance. These infrastructures are spatially distributed across large physical areas, and consist of heterogeneous cyber-physical components interconnected by communication networks with complex peering and hierarchies. Networked control systems (NCSs) and supervisory control and data acquisition (SCADA) systems are widely used to monitor, control, and remotely manage infrastructures over private or shared communication networks. Such cyber-physical systems (CPSs) permit synergistic interactions between physical dynamics and computational processes. Wide deployment of information and communication technologies (ICT) in CPSs results in higher reliability and lower operational costs relative to the traditional proprietary and closed systems. However, as recent incidents indicate, today's CPSs face new security threats driven by their exposure to ICT insecurities.

### Security Threats

To develop a classification of security threats to CPSs, we first outline how the operator(s) of modern CPSs typically approach the monitoring, control, and management of infrastructures. As shown in Fig. 1, they use a layered architecture consisting of *regulatory control* (layer 1), *supervisory control* (layer 2), and a *management level* (layer 3). This architecture enables robust composition of multilevel controllers, and permits CPS operators to use *defenses* to limit the effects of failures caused by *faults* and/or *attacks*.

The regulatory control layer directly interacts with the underlying physical infrastructure dynamics through a network of sensors and actuators. These field devices are connected to programmable logic controllers (PLCs) or remote terminal units (RTUs), and implement detection and regulation mechanisms that are primarily reactive in nature. These mechanisms

can also respond to localized failures of field devices and communication links. The regulatory controllers (or PLCs) interact with the supervisory controllers via a control network.

At the supervisory control layer, model-based diagnostic tools are combined with optimal control-based tools to ensure on-time response to distributed failures. The supervisory workstations are used for data logging, diagnostic functions such as fault diagnosis, and supervisory control computations such as set-point control and controller reconfigurations.

Lastly, the management (topmost) layer focuses on strategies that maximize the operator's profit while minimizing its losses due to security and reliability failures. The CPS operator and other authorized remote users can access information about the CPS processes and send specifications to the controllers at lower layers via the Internet or a corporate network.

Security threats to hierarchically managed CPSs arise from four channels. First, CPSs inherit vulnerabilities from embedded commercial off-the-shelf ICT devices, and are subject to correlated software bugs and hardware malfunctions. Second, the proprietary protocols and closed networks are being replaced with standard open Internet protocols and shared networks. Malicious attackers capable of exploiting protocol and network insecurities can target CPS operations. Third, numerous parties generate, use, and modify CPS data. This poses new challenges in access control and authorization among the strategic players such as the operators, SCADA and ICT vendors, and end users of the system. Fourth, CPSs employ a large number of remote field devices that can be accessed via short-range communications. Thus, CPSs are vulnerable to adversarial manipulation, both remote and local.

Adversaries can exploit the aforementioned threat channels via denial-of-service (DoS) and deception attacks, which result in losses of availability and integrity of sensor-control data,

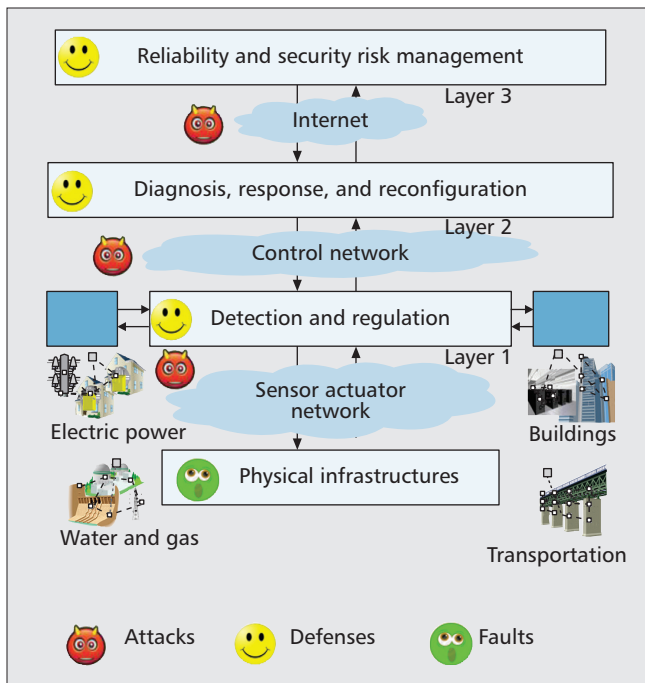


Figure 1. A layered architecture for management of CPS.

respectively. In Table 1, we present examples of security attacks on the regulatory and supervisory control layers. Attacks at the management level are similar to attacks on computer networks. We refer the reader to [1, 2] for specific discussions on security attacks to smart grid infrastructures.

### Classification of Correlated Failures

The danger of correlated failures becomes especially profound in CPSs due to the tight coupling of typically continuous physical dynamics and discrete dynamics of embedded computing processes. Correlated failures originate from one or more of the following events:

- *Simultaneous attacks*: Targeted cyber attacks (e.g., failures due to Stuxnet); non-targeted cyber attacks (e.g., failures due to Slammer worm, distributed DoS attacks [3], congestion in shared networks); coordinated physical attacks (e.g., failures caused by terrorists)
- *Simultaneous faults*: Common-mode failures (e.g., failure of multiple ICT components in an identical manner [4], programming errors); random failures (e.g., natural events such as earthquakes and tropical cyclones, and operator errors such as an incorrect firmware upgrade)
- *Cascading failures*: Failure of a fraction of nodes (components) in one CPS subnetwork can lead to progressive escalation of failures in other subnetworks (e.g., power network blackouts affecting communication networks, and vice versa) [5].

The above classification is neither fully disjoint nor exhaustive. Still, we envision that it will be useful for CPS risk assessment. We term correlated failures caused by simultaneous attacks as security failures and simultaneous faults as reliability failures. Due to the tight cyber-physical interactions, it is extremely difficult (and often prohibitively time-consuming) to isolate the cause of any specific failure using the diagnostic information, which, in general, is imperfect and incomplete. Thus, reliability and security failures in CPSs are inherently intertwined. We believe that the quest to find a mutually exclusive and jointly exhaustive partition of failure events must be abandoned. Instead, the research emphasis should shift to the analysis of *interdependent reliability and security failures*, and risk assessment.

## Information and CPS Risks

### The Interplay of Technological Defenses and Institutions

There are two types of technological means to reduce CPS risks: ICT security tools and control-theoretic tools. The *ICT security tools* include authentication and access control mechanisms, network intrusion detection systems, patch management, and security certification. In practice, the effectiveness of these security tools is limited by CPS reliability and cost considerations. For example, the frequency of security patch updates is limited by the real-time constraints on the availability of CPS data; common criteria certification is limited by the resources for CPS security and so on. The *control-theoretic tools* include model-based attack/fault detection and isolation, robust control strategies that maintain closed-loop stability and performance guarantees under a class of DoS/deception attacks, and reconfigurable (switching) control strategies to limit the effect of correlated failures. Recently, several organizations (e.g., NIST, NERC, DHS) have proposed security standards and recommendations that combine the ICT-specific security defenses with control theoretic tools.

While technology-based defenses for CPS are the main channel to improve their survivability against correlated failures, the mere existence of these defenses is not sufficient. It is well established that the lack of private parties' incentives for security improvements is a severe impediment to achieving socially desirable improvements of CPS security [6]. Indeed, large-scale critical infrastructures are typically managed by profit-driven private entities. Proper implementation of technological defenses and resilient operation requires compliance of relevant entities. Below we highlight the informational deficiencies that negatively affect the incentives for security.

### Informational Deficiencies

Due to the prohibitively high costs of information acquisition, it is often too costly to determine the following:

- Which hardware malfunctions and software bugs have caused a system failure
- Whether the system failure was caused by a reliability failure or security failure or both

In many cases, this information varies significantly across different entities (players), such as CPS operators, SCADA and ICT vendors, network service providers, users, and local/federal regulatory agencies (or government). Informational deficiencies arise from the conflicting interests of individual players whose choices affect the CPS risks. One may say that interdependent failures cause *externalities* that result in misaligned player incentives (i.e., the individually optimal CPS security defenses diverge from the socially optimal ones).

Moreover, in environments with incomplete and also asymmetric (and private) information, the societal costs of a correlated CPS failure typically exceed the losses of the individual players whose products and services affect CPS operations, and on whose actions the CPS risks depend. Specifically, interdependencies between security and reliability failures in CPS are likely to cause negative externalities. In such environments, the individual players tend to underinvest in security relative to a socially optimal benchmark. This requires design of institutional means to realign the individual players' incentives to make adequate investments in security. Examples of institutional means include *regulations* that require players to certify that they possess certain security capabilities, and *legal rules* which mandate that players share information about security incidents with government agencies and/or the public through established channels.

	Control layer	
	Regulatory control	Supervisory control
Deception attacks	Spoofing, replay	Set-point change
	Measurement substitution	Controller substitution
DoS attacks	Physical jamming	Network flooding
	Increase in latency	Operational disruption

**Table 1.** Cyber-attacks to CPS control layers.

Clearly, these individual players cannot completely eliminate the risk of CPS failures even in the presence of advanced technological defenses and institutional measures, which aim to reduce (or even eliminate) incentive misalignment between individual and socially optimal security choices. For example, consider a benchmark case when security defenses are optimally chosen by the social planner for a given technological and institutional environment. There still remains a residual risk driven by fundamental physical limits. Indeed, when security defenses are chosen by individual players, the risk is only higher. Thus, non-negligible (public) residual risks are characteristic for CPSs that are subjected to correlated failures.

So far, the occurrence of extreme correlated failures have been statistically rare. However, with the emergence of organized cyber-crime groups capable of conducting intrusions into NCS/SCADA systems, the risks of such rare failure events cannot be ignored. Unsurprisingly, cyber-warfare is projected to become the future of armed conflict, and managing CPS risks must be at the core of any proactive defense program.

### Benchmarking CPS Risks

Due to the aforementioned challenges, benchmarking CPS risks is a hard problem, and several questions remain unanswered [7–9]. Our goal in this article is twofold:

- We suggest a game-theoretic framework that assesses security risks by quantifying the misalignment between individually and socially optimal security investment decisions when the CPS comprises interdependent NCS.
- We advocate that better information about these risks is a prerequisite to improvement of CPS security via a combination of more sophisticated technology-based defenses and the advancement of their supporting institutions.

Improved assessment of the CPS risks will lead to several beneficial developments, such as improved risk management at both the individual and societal levels. Thus, a standardized framework should be established that can assess and compare different technological and institutional means for risk management. At the very least, better knowledge of CPS risks will permit the players to make more informed (and therefore better and cheaper) choices of security defenses, thus improving the societal welfare.

### Framework to Benchmark CPS Risks

We now present a risk assessment framework from the perspective of CPS operators. Our setup can readily be adapted to assess risks from the perspective of other players.

#### CPS with a Centralized Control System

Consider a CPS with  $m$  independent components managed by a single operator (i.e., centralized control system). For the  $i$ th component, let  $\Omega^i$  denote the set of all hardware flaws, software bugs, and vulnerability points that can be compromised

during any reliability and/or security failure event. The failure events form a collection of subsets of  $\Omega^i$ , which we denote by  $\mathcal{F}$ . Let the random variables  $X_R^i: \Omega^i \rightarrow \mathbb{R}$  and  $X_S^i: \Omega^i \rightarrow \mathbb{R}$  represent the reliability and security levels of the  $i$ -th component, respectively, with joint (cumulative) distribution function:

$$F_{X_R^i, X_S^i}(x_R^i, x_S^i) = P\{\omega \in \Omega^i \mid X_R^i(\omega) \leq x_R^i, X_S^i(\omega) \leq x_S^i\},$$

where the measure  $P$  assigns probabilities to failure events. Notice that the reliability level  $X_R^i$  and security level  $X_S^i$  are defined on the same measure space  $(\Omega^i, \mathcal{F})$ , and they are not mutually independent, that is,

$$F_{X_R, X_S}(x_R, x_S) \uparrow F_{X_R}(x_R) \cdot F_{X_S}(x_S).$$

Unfortunately, the CPS operator does not have perfect knowledge of these distributions. Reasonable estimates of  $F_{X_R}(x_R)$  may be obtained from historical failure data. However, estimating the joint distribution  $F_{X_R, X_S}(x_R, x_S)$  is difficult as attackers continue to find new ways to compromise security vulnerabilities.

In general, the random vector  $(X_R^i, X_S^i)$  is influenced by:

- Action set of the CPS operator  $\mathcal{A} = \mathcal{U} \cup \mathcal{V}$ , where  $\mathcal{U} := \{\mathcal{U}^1, \dots, \mathcal{U}^m\}$  and  $\mathcal{V} := \{\mathcal{V}^1, \dots, \mathcal{V}^m\}$  denote the set of control and security choices, respectively
- Action set of other players  $\mathcal{B}$ , such as vendors, attackers, service providers, users, and regulatory agencies
- Environment  $\mathcal{E}$ , including the technological, organizational, and institutional factors

For given reliability and security levels  $x_R^i, x_S^i$ , let the function  $L^i(x_R^i, x_S^i)$  denote the losses faced by the CPS operator when the  $i$ th component fails (e.g., the cost of service disruptions, maintenance/recovery costs, and penalties for users' suffering). Then, for CPS with  $m$  independent components, the aggregate risk can be expressed as:<sup>1</sup>

$$\mathcal{R} = \sum_{i=1}^m \mathcal{R}^i(L^i(X_R^i, X_S^i)), \quad (1)$$

where the functional  $\mathcal{R}^i$  assigns a numerical value to each random variable  $L^i$  with distribution function  $F_{L^i}$ . Henceforth, we use the expected (mean) value of loss,  $\mu(L^i) = E[L^i(X_R^i, X_S^i)]$ , as a metric of  $\mathcal{R}^i$ , but caution that it is inadequate to capture risk of extreme failure events.<sup>2</sup> From Eq. 1, we observe that the aggregate risk is also influenced by actions  $\mathcal{A}$ ,  $\mathcal{B}$ , and environment  $\mathcal{E}$ . To emphasize this dependence, we will use  $\mathcal{R}(\mathcal{A}, \mathcal{B}, \mathcal{E})$  to denote the aggregate CPS risk.

For a given environment  $\mathcal{E}$  and fixed choices  $\mathcal{B}$  of other players, the CPS operator's objective is to choose security actions  $\mathcal{V}$  and control actions  $\mathcal{U}$  to minimize the total expected cost  $\mathcal{J}(\mathcal{U}, \mathcal{V})$  of operating the system:

$$\mathcal{J}(\mathcal{U}, \mathcal{V}) = \mathcal{J}_I(\mathcal{V}) + \mathcal{J}_{II}(\mathcal{U}, \mathcal{V}), \quad (2)$$

where  $\mathcal{J}_I(\mathcal{V}) := \sum_{i=1}^m \ell^i(\mathcal{V}^i)$  denotes the operator's cost of employing security choices  $\mathcal{V}$ , and  $\mathcal{J}_{II}(\mathcal{U}, \mathcal{V})$  is the expected

<sup>1</sup> The assumption of independent components can easily be relaxed to include parallel, series, and interlinked components.

<sup>2</sup> Other commonly used choices of risk  $\mathcal{R}^i$  include the mean-variance model:  $\mu(L^i) + \lambda^i \sigma(L^i)$ , where  $\lambda^i > 0$  and  $\sigma(L^i)$  is the standard deviation of  $L^i$ ; and the value-of-risk model:  $\text{VaR}_{\alpha^i}(L^i) = \min\{z \mid F_{L^i}^i(z) \geq \alpha^i\}$ , which is the same as  $\alpha^i$ -quantile in distribution of  $L^i$ .

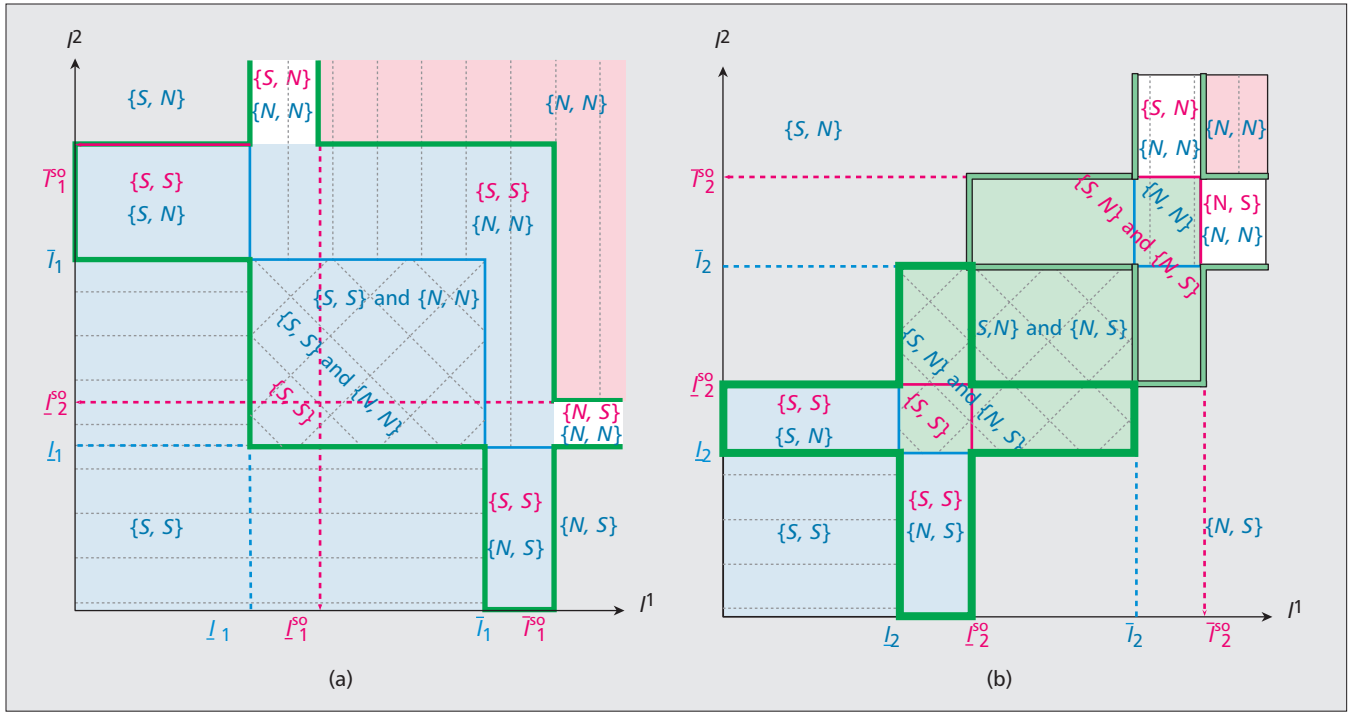


Figure 2. Individual optima (Nash equilibria) and social optima.

operational cost. From Eq. 2, when the CPS operator's security choices are  $\mathcal{V}$ , s/he chooses control actions  $\mathcal{U} = \mu^*(\mathcal{V})$  to minimize total expected cost, where  $\mu^*(\mathcal{V})$  is an optimal control policy. Let the CPS operator's minimum cost for the case when security choices are  $\mathcal{V}$  and  $\{\emptyset\}$  (i.e., no security defenses) be defined as  $\bar{\mathcal{J}}(\mathcal{V}) := \mathcal{J}(\mu^*(\mathcal{V}), \mathcal{V})$  and  $\mathcal{J}^0 := \mathcal{J}(\mu^*(\{\emptyset\}), \{\emptyset\})$ , respectively. To evaluate the effectiveness of  $\mathcal{V}$ , we use the difference of corresponding expected costs:

$$\Delta(\mathcal{V}) := \mathcal{J}^0 - \bar{\mathcal{J}}(\mathcal{V}). \quad (3)$$

Thus,  $\Delta(\mathcal{V})$  denotes the CPS operator's gain from employing security choices  $\mathcal{V}$ . It can be viewed as the reduction of operator's risk when s/he chooses  $\mathcal{V}$  over no defenses, that is,

$$\mathcal{R}(\mathcal{A}^0, \mathcal{B}, \mathcal{E}) - \mathcal{R}(\mathcal{A}(\mathcal{V}), \mathcal{B}, \mathcal{E}) = \Delta(\mathcal{V}), \quad (4)$$

where  $\mathcal{A}(\mathcal{V})$  and  $\mathcal{A}^0$  denote the action set corresponding to security choices  $\mathcal{V}$  and  $\{\emptyset\}$ , respectively. The problem of choosing optimal security choices  $\mathcal{V}^*$  can now be viewed as an optimization problem over the set of security defenses:

$$\max_{\mathcal{V}} \Delta(\mathcal{V}), \text{ subject to the constraint } \mathcal{J}(\mathcal{V}) \leq K,$$

where  $K$  is the available budget for security investments.

The residual risk after the implementation of optimal security choices  $\mathcal{V}^*$  can be obtained as  $\mathcal{R}(\mathcal{A}^0, \mathcal{B}, \mathcal{E}) - \Delta(\mathcal{V}^*)$ . Risks from failure events (those resulting from security attacks, random faults, cascading failures, etc.) can thus be estimated and compared, and the best security defenses  $\mathcal{V}$  corresponding to anticipated failure types can be selected by the CPS operator.

The above analysis assumes that the choices  $\mathcal{B}$  of other players do not change in response to the CPS operator's choices  $\mathcal{A}$ . When players are strategic, the optimal security choices must be computed as best responses to the other players' (Nash) strategies. Finally, government or regulatory agencies can also influence the environment  $\mathcal{E}$ .

### CPS with Interdependent Networked Control Systems

Let us focus on the issue of misalignment between individual and socially optimal actions in the case when a CPS comprises multiple NCSs communicating over a shared network. In contrast to the above, we now assume that each NCS is managed by a separate operator. The NCS operators choose their security levels to safeguard against network-induced risks (e.g., due to distributed DoS attacks). Each NCS is modeled by a discrete-time stochastic linear system, which is controlled over a lossy communication network:

$$\begin{aligned} x_{t+1}^i &= Ax_t^i + v_t^i B u_t^i + w_t^i \\ y_t^i &= \gamma_t^i C x_t^i + v_t^i \end{aligned} \quad t \in \mathbb{N}_0, \quad i \in M, \quad (5)$$

where  $M$  denotes the number of players,  $x_t^i \in \mathbb{R}^d$  the state,  $u_t^i \in \mathbb{R}^m$  the input,  $w_t^i \in \mathbb{R}^d$  the process noise,  $y_t^i \in \mathbb{R}^p$  the measured output, and  $v_t^i \in \mathbb{R}^p$  the measurement noise, for player  $\mathbf{P}_i$  at the  $t$ th time step. Let the standard assumptions of linear quadratic Gaussian (LQG) theory hold. The random variables  $\gamma_t^i$  (resp.  $v_t^i$ ) are i.i.d. Bernoulli with the failure probability  $\tilde{\gamma}^i$  (resp.  $\tilde{v}^i$ ), and model a lossy sensor (resp. control) channel.

We formulate the problem of security choices of the individual players as a non-cooperative two-stage game [10]. In the first stage, each  $\mathbf{P}_i$  chooses to make a security investment ( $\mathcal{S}$ ) or not ( $\mathcal{N}$ ). The set of player security choices is denoted  $\mathcal{V} := \{\mathcal{V}^1, \dots, \mathcal{V}^m\}$ , where  $\mathcal{V}^i = \mathcal{S}$  if  $\mathbf{P}_i$  invests in security and  $\mathcal{N}$  if not. Once player security choices are made, they are irreversible and observable by all the players. In the second stage, each  $\mathbf{P}_i$  chooses a control input sequence  $\mathcal{U}^i := \{u_t^i, t \in \mathbb{N}_0\}$  to maintain optimal closed-loop performance. The objective of each  $\mathbf{P}_i$  is to minimize his/her total cost:

$$\bar{\mathcal{J}}^i(\mathcal{V}, \mathcal{U}) = \bar{\mathcal{J}}_1^i(\mathcal{V}) + \bar{\mathcal{J}}_{\text{II}}^i(\mathcal{V}, \mathcal{U}), \quad i \in M, \quad (6)$$

where the first stage cost is denoted  $\bar{\mathcal{J}}_1^i(\mathcal{V}) := (1 - \mathcal{I})\ell^i$ , and  $\bar{\mathcal{J}}_{\text{II}}^i(\mathcal{V}, \mathcal{U})$  denotes second stage cost (the average LQG cost). Here  $\ell^i > 0$  is the security investment incurred by  $\mathbf{P}_i$  only if



s/he has chosen  $\mathcal{S}$ , and the indicator function  $\mathcal{I}^i = 0$  when  $\mathcal{V}^i = \mathcal{S}$ , and  $\mathcal{I}^i = 1$  otherwise.

In order to reflect security interdependencies, in our model, the failure probabilities  $\tilde{\gamma}^i$  and  $\tilde{v}^i$  depend on the  $\mathbf{P}^i$ 's own security choice  $\mathcal{V}^i$  and on the other players' security choices  $\{\mathcal{V}^j, j \neq i\}$ . Following [10], we assume

$$P[\gamma_i^j = 0 \mid \mathcal{V}] = \tilde{\gamma}^i(\mathcal{V}) := \mathcal{I}^i \bar{\gamma} + (1 - \mathcal{I}^i) \alpha(\eta^{-i}).$$

In Eq. 7, the first term reflects the probability of a *direct failure*, and the second term reflects the probability of an *indirect failure*. The interdependence term  $\alpha(\eta^{-i})$  increases as the number of players, excluding  $\mathbf{P}^i$ , who have chosen  $\mathcal{N}$  increase, where  $\eta^{-i} := \sum_{j \neq i} \mathcal{I}^j$ ; similarly for  $v_i^j$ . The social planner objective is to minimize the aggregate cost:

$$\mathcal{J}^{\text{SO}}(\mathcal{V}, \mathcal{U}) = \sum_{i=1}^m \mathcal{J}^i(\mathcal{V}, \mathcal{U}). \quad (8)$$

Consider a two-player game, where the interdependent failure probabilities are given by Eq. 8. To derive optimal player actions (security choices  $\mathcal{V}^i$ ), we distinguish the following two cases: *increasing incentives* and *decreasing incentives*. For the case of increasing incentives, if a player secures, other player's gain from securing increases, that is,  $\mathcal{J}_{\text{II}}^*(\{\mathcal{N}, \mathcal{N}\}) - \mathcal{J}_{\text{II}}^*(\{\mathcal{S}, \mathcal{N}\}) \leq \mathcal{J}_{\text{II}}^*(\{\mathcal{N}, \mathcal{S}\}) - \mathcal{J}_{\text{II}}^*(\{\mathcal{S}, \mathcal{S}\})$ , where  $\mathcal{J}_{\text{II}}^*(\cdot)$  denotes the optimal second stage cost. Similarly, for the case of decreasing incentives, a player's gain from investing in security decreases when the other player invests in security, that is,  $\mathcal{J}_{\text{II}}^*(\{\mathcal{N}, \mathcal{N}\}) - \mathcal{J}_{\text{II}}^*(\{\mathcal{S}, \mathcal{N}\}) \geq \mathcal{J}_{\text{II}}^*(\{\mathcal{N}, \mathcal{S}\}) - \mathcal{J}_{\text{II}}^*(\{\mathcal{S}, \mathcal{S}\})$ .

Figure 2a (resp. Fig. 2b) characterizes the Nash equilibria (individually optimal choices) and socially optimal choices of the game for the case of increasing (resp. decreasing) incentives, where we assume  $\ell_1^{\text{SO}} < \ell_1$  (resp.  $\ell_2 > \ell_2^{\text{SO}}$ ). For  $i \in \{1, 2\}$ , the thresholds  $\underline{\ell}_i, \bar{\ell}_i, \ell_i^{\text{SO}}$ , and  $\bar{\ell}_i^{\text{SO}}$  are given in [10].

Consider the case of increasing incentives (Fig. 2a). If  $\ell^i < \underline{\ell}_1$  (resp.  $\ell^i > \bar{\ell}_1$ ), the symmetric Nash equilibrium  $\{\mathcal{S}, \mathcal{S}\}$  (resp.  $\{\mathcal{N}, \mathcal{N}\}$ ) is unique. Thus,  $\underline{\ell}_1$  (resp.  $\bar{\ell}_1$ ) is the cutoff cost below (resp. above) which both players invest (resp. neither player invests) in security. If  $\underline{\ell}_1 \leq \ell^i \leq \bar{\ell}_1$ , both  $\{\mathcal{S}, \mathcal{S}\}$  and  $\{\mathcal{N}, \mathcal{N}\}$  are individually optimal. However, if  $\ell^1 < \underline{\ell}_1$  &  $\ell^2 > \bar{\ell}_1$  (resp.  $\ell^1 > \bar{\ell}_1$  &  $\ell^2 < \underline{\ell}_1$ ), the asymmetric strategy  $\{\mathcal{S}, \mathcal{N}\}$  (resp.  $\{\mathcal{N}, \mathcal{S}\}$ ) is an equilibrium. Now, if  $\ell^i < \ell_i^{\text{SO}}$  (resp.  $\ell^i > \bar{\ell}_i^{\text{SO}}$ ), the socially optimal choices are  $\{\mathcal{S}, \mathcal{S}\}$  (resp.  $\{\mathcal{N}, \mathcal{N}\}$ ). If  $\ell^1 \leq \underline{\ell}_1^{\text{SO}}$  &  $\ell^2 \leq \bar{\ell}_2^{\text{SO}}$  (resp.  $\ell^2 \leq \underline{\ell}_2^{\text{SO}}$  &  $\ell^1 \leq \bar{\ell}_1^{\text{SO}}$ ), socially optimal choices are  $\{\mathcal{S}, \mathcal{N}\}$  (resp.  $\{\mathcal{N}, \mathcal{S}\}$ ). Similarly, we can describe individually and socially optimal choices for the case of decreasing incentives (Fig. 2b).

For both cases, we observe that the presence of interdependent security causes a negative externality. The individual players are subject to network-induced risks and tend to under-invest in security relative to the social optimum. From our results, for a wide parameter range, regulatory impositions to incentivize higher security investments are desirable (discussed later). The effectiveness of such impositions on the respective risks faced by individual players (NCS operators) can be evaluated in a manner similar to Eqs. 3–4.

## Challenges in CPS Risk Assessment

### Technological Challenges

A significant challenge for the practical implementation of our CPS risk assessment framework is to develop data-driven, stochastic CPS models, which account for dynamics of CPS with interdependent reliability and security failures. Each of these singular/basic models should account for CPS dynamics

and focus on a specific failure scenario. The basic models can be composed into a *composite* model to represent various correlated failure scenarios, including simultaneous attacks, common-mode failures, and cascading failures. By using of quantitative techniques from statistical estimation, model-based diagnosis, stochastic simulation, and predictive control, we can automatically generate new failure scenarios from real-time sensor-control data. These techniques enable the synthesis of operational security strategies and provide estimates of residual risks in environments with highly correlated failures and less than perfect information. Thus, theoretical guarantees and computational tools are needed for the following:

- Compositions of stochastic fault and attack models
- Inference and learning of new failure scenarios
- Fast and accurate simulation of CPS dynamics
- Detection and identification of failure events
- Operational ICT and control based strategies

The DETERLab testbed [11] provides the capability to conduct experiments with a diverse set of CPS failure scenarios, where the controllable variables range from IP-level dynamics to introduction of malicious entities such as distributed DoS attacks. The cyber-physical aspects of large-scale infrastructures can be integrated together on DETERLab to provide an experimental environment for assessing CPS risks. Specifically, the DETERLab provides a programmable network emulation environment, and a suite of tools that allow a user to describe the experimentation “apparatus,” and monitor and control the experimentation “procedure.” Multiple experimentations can be executed at the same time by different users if computational resources are available.

The main challenge for CPS experimentation on the DETERLab testbed is to compose physical system dynamics (real/simulated/emulated) with communication system emulation. The experimentation “apparatus” should model the communication network, the physical network, and their dynamic interactions. The experimentation “procedure” should describe the sensing and actuation policies that are the best responses to strategic actions of other players.

### Institutional Challenges

The design of institutional means is a chicken-and-egg problem. On one hand, institutional means such as imposition of legal liability, mandatory incident disclosure, and insurance instruments improve the information about CPS risks. On the other hand, substantial knowledge of CPS risks is required for their design and successful deployment.

Given the limitations of currently available risk assessment tools, the CPS operators find it hard (and, as a result, costly) to manage their risks. This problem is especially acute for risk management via financial means, such as diversification, reallocation to other parties, and insurance. For example, insurance instruments of CPS risks management are meager: the premiums of cyber-security contracts are not conditioned on the security parameters. It would be no exaggeration to say that so far, the cyber-insurance market has failed to develop. For example, the volume of underwritten contracts is essentially unchanged in a decade, despite multiple predictions of its growth by independent researchers and industry analysts. In fact, even the existing superficial “market” is largely sustained by non-market (regulatory) forces.

Indeed, the leading reason for CPS operators to acquire insurance policies at the prevailing exuberant prices is their need to comply with federal requirements for government contractors. Citizens (i.e., federal and state taxpayers) are the final bearers of these costs. We expect that this situation will remain “as is” unless information on CPS risks drastically improves.

Another related problem is that of suboptimal provider incentives (as seen in Fig. 2). A CPS operator's estimates of his/her own risk tend to be understated (relative to societal ones), even when failure probabilities are known to him/her. In such cases, the gap between individually and socially optimal incentives could be reduced via adjustments of legal and regulatory institutions. For example, it would be socially desirable to introduce limited liability (i.e., a due care standard) for individual entities whose products and services are employed in CPSs. This would improve providers' incentives to invest in their products' security and reliability. However, due to information incompleteness, currently there is no liability regime for providers of CPS components and services, for neither security nor reliability driven failures. Indeed, any liability regime is based on knowing (the estimate[s] of) failure probabilities and the induced losses. This again requires benchmarking of CPS risks.

## Concluding Remarks

Benchmarking of CPS risks is a hard problem. It is harder than the traditional risk assessment problems for infrastructure reliability or ICT security, which so far have been considered in isolation. Estimation of CPS risks by naively aggregating risks due to reliability and security failures does not capture the externalities, and can lead to grossly suboptimal responses to CPS risks. Such misspecified CPS risks lead to biased security choices and reduce the effectiveness of security defenses.

Modern, and especially upcoming, CPSs are subjected to complex risks, of which very little is known despite the realization of their significance. In this article we are calling on our colleagues to embark on the hard task of assessing interdependent CPS risks. The effectiveness of security defenses can be increased only when our knowledge of CPS risks improves.

## Acknowledgments

We are grateful to the anonymous reviewers for their feedback, and thank Professors S. Shankar Sastry (UC Berkeley) and Joseph M. Sussman (MIT) for useful discussions.

## References

- [1] Y. Mo *et al.*, "Cyber-Physical Security of A Smart Grid Infrastructure," *Proc. IEEE*, vol. 100, no. 1, Jan. 2012, pp. 195–209.
- [2] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber-Physical System Security for the Electric Power Grid," *Proc. IEEE*, vol. 100, no. 1, Jan. 2012, pp. 210–24.

- [3] A. Hussain, J. Heidemann, and C. Papadopoulos, "A Framework for Classifying Denial of Service Attacks," *Proc. 2003 ACM Conf. Applications, Technologies, Architectures, and Protocols for Computer Communications*, 2003, pp. 99–110.
- [4] S. Amin *et al.*, "Cyber Security of Water SCADA Systems – Part II: Attack Detection Using Enhanced Hydrodynamic Models," *IEEE Trans. Control Systems Technology*, 2012.
- [5] S. Buldyrev *et al.*, "Catastrophic Cascade of Failures in Interdependent Networks," *Nature*, vol. 464, no. 7291, Apr. 2010, pp. 1025–28.
- [6] C. Hall *et al.*, "Resilience of the Internet Interconnection Ecosystem," *Proc. 10th Wksp. Economics of Information Security*, June 2011.
- [7] T. Alpcan and T. Basar, *Network Security: A Decision and Game Theoretic Approach*, Cambridge Univ. Press, 2011.
- [8] P. Grossi and H. Kunreuther, *Catastrophe Modeling: A New Approach to Managing Risk*, Springer, 2005, vol. 25.
- [9] Y. Y. Haimes, *Risk Modeling, Assessment, and Management*, 3rd ed., Wiley, 2009.
- [10] S. Amin, G. A. Schwartz, and S. S. Sastry, "On the Interdependence of Reliability and Security in Networked Control Systems," *CDC-ECE, IEEE*, 2011, pp. 4078–83.
- [11] T. Benzel, "The Science of Cyber Security Experimentation: The Deter Project," *Proc. 27th ACM Annual Computer Security Applications Conf.*, 2011, pp. 137–48.

## Biographies

SAURABH AMIN (amins@mit.edu) is an assistant professor in the Department of Civil and Environmental Engineering, Massachusetts Institute of Technology (MIT). His research focuses on the design and implementation of high-confidence network control algorithms for critical infrastructures, including transportation, water, and energy distribution systems. He received his B.Tech. in civil engineering from the Indian Institute of Technology Roorkee in 2002, M.S. in transportation engineering from the University of Texas at Austin in 2004, and Ph.D. in systems engineering from the University of California at Berkeley in 2011.

GALINA A. SCHWARTZ is a research economist in the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley. Her primary expertise is game theory and microeconomics. She has published on the subjects of network neutrality, cyber risk management and modeling of cyber-insurance markets, and security and privacy of cyber-physical systems. In her earlier research, she has applied contract theory to study the interplay between information, transaction costs, institutions and regulations. She has been on the faculty in the Ross School of Business at the University of Michigan, Ann-Arbor, and has taught in the Economics Departments at the University of California, Davis and Berkeley. She received her M.S. in mathematical physics from Moscow Institute of Engineering Physics, Russia, and Ph.D. in economics from Princeton University in 2000.

ALEFIYA HUSSAIN is a computer scientist at the University of Southern California's Information Sciences Institute (USC/ISI). Her research interests include statistical signal processing, protocol design, cyber security, and network measurement systems. She received her B.E. in computer engineering from the University of Pune, India, in 1997 and Ph.D. in computer science from University of Southern California in 2005. Prior to joining USC/ISI, she was a senior principal scientist at Sparta Inc.

---

# Measuring the Global Domain Name System

**E. Casalicchio, Univ. of Rome Tor Vergata**


**M. Caselli and A. Coletta, Global Cyber Security Center (GCSEC)**

---

## Abstract

The Internet is a worldwide distributed critical infrastructure, and it is composed of many vital components. While IP routing is the most important service, today the Domain Name System can be classified as the second most important, and has been defined as a critical infrastructure as well. DNS enables naming services used by every networked application and therefore by every networked critical infrastructure. Without DNS all services used in daily life activities (e.g., commerce, finance, industrial process control, logistics, transportation, health care) become unavailable. A big challenge is to guarantee the proper level of DNS health. Providing DNS health requires monitoring the system, analyzing its behavior, and planning and actuating corrective actions. There are several initiatives in this field, all claiming to be able to measure the DNS health from a local perspective. The reality is a bit different and many challenges are still open: no standard metric exist (only a shared list of five health indicators); no common rules to compute health indicators are agreed; no common concept of regular DNS behavior is defined. The Measuring the Naming System (MeNSa) project proposes a formal and structured methodology and a set of metrics for the evaluation of the DNS health and security levels. This article discusses the problem of measuring the DNS health level and introduces the main concepts of the MeNSa project. Finally, using a real case study, the problem of metrics aggregation is discussed.

---

ritical infrastructures are increasingly incorporating in a massive way networked components. This trend obviously allowed the services provided to be enhanced and optimized, distributed self-orchestration mechanisms to be implemented, and remote installations to be managed efficiently. On the other hand, as a consequence, the Internet infrastructure used to realize these services must be considered now as part of the critical infrastructure themselves. For example, in [1] the authors proposed an architecture to secure the Domain Name System (DNS) with the final goal of protecting critical infrastructures. A recent study [2] showed how an attack on the DNS might impact several levels of the operation of energy smart grids.

The DNS, as part of the Internet infrastructure, constitutes the backbone of the modern cyber world. In 2007, the Internet Engineering Task Force's (IETF's) DNS Extensions Working Group (DNSEXT) identified the DNS as "a critical Internet infrastructure" because it resolves billions of queries per day in support of global communications and commerce. In 2011, Steve Gibbard, at the spring DNS-OARC<sup>1</sup> workshop in San Francisco, California, stressed that the DNS is a critical infrastructure. However, the DNS being a completely distributed infrastructure and completely seamless to end users has con-

tributed to making it one of the less considered infrastructures of the Internet when speaking of cyber security. This is no longer true; Kaminsky's exploit, and attacks that in the last years have taken advantage of the weaknesses of the DNS in order to damage cyber-infrastructure, have posed serious questions about the security, safety, and stability of this system.

The community has widely discussed the problem of the security of DNS and its impact on the cyber society. In 2010 [3] the concept of DNS health came out as a way of defining when the DNS system is well functioning, taking as an example human body health. The concept of DNS health is developed around five main indicators: availability, coherency, integrity, resiliency, and speed. We believe that such a list needs to be extended with the concepts of stability and security, since a system cannot be described as "healthy" if it is not stable and can become quickly unhealthy if it is not secure. Stability is intended as the desired DNS feature to function reliably and predictably day by day (e.g., protocols and standards). Stability facilitates universal acceptance and usage. Hereafter, when we talk about DNS health we include the concepts of security, stability, and resiliency.

The concepts expressed in [3] remained at the abstract level, without suggesting how to assess such a complex property. What is indeed unclear in this definition is the way in which it would be possible, from real, measurable observation, to obtain "numbers" or, better, indices, that can be used to quantify these global properties and, at last, to summarize the "level of health" of the portion of DNS under analysis. Shortly speaking, no standard frameworks for DNS measurement

---

<sup>1</sup> DNS-OARC, the Domain Name System Operation Analysis and Research Center, is a non-profit membership organization that seeks to improve the security, stability, and understanding of the Internet's DNS infrastructure. The main DNS operators are members of DNS-OARC.

exist, and no standard metrics have been defined. Moreover, how to define a common concept of *regular* DNS behavior and how to develop a standard framework for data/information sharing are still open issues [4].

In the literature there are many studies related to DNS traffic measurement and performance metrics (e.g., [5–7]), but very few report on the measurement and quantification of security, stability, and resiliency, or, in general, DNS health. Moreover, to the best of our knowledge, a complete framework dealing with DNS health evaluation has not yet been developed.

In this article, we present our “answer” to this challenge. After a brief description of the MeNSa project [8], we show how it is possible to aggregate several different metrics related to the DNS system, after identifying a well defined measurement point of view, in order to obtain aggregate indicators of its health. Nevertheless, in the article we consider the end-user perspective, and the solution we propose can be applied to any observation point and for any set of metrics.

The article is organized as follows. We briefly describe DNS vulnerabilities and show examples of real incidents. We present the MeNSa project. We discuss the metrics aggregation problem and describe the proposed solution. We introduce a case study to show how metric aggregation can be computed and used to evaluate a real case. We then conclude the article.

## Vulnerabilities and Incidents

The DNS threats can be broadly classified into three main categories: *data corruption*, *denial of service*, and *privacy*.

**Data corruption** clusters all types of incidents related to the unauthorized modification of DNS data. These incidents can happen in every part of the DNS propagation chain and can be related to corruption of repositories (e.g., databases containing resource records, zone files, DNSSEC keys, and so on); cache consistency [9]; alteration of the authenticity of DNS responses; and protocol issues, which deal with design flaws of the DNS protocol. Examples of protocol issues are cache poisoning (i.e., the well known Kaminsky’s attack), route injections, and man-in-the-middle threats. The first security flaws in the protocol were discovered in the early 1990s: in [10, 11] the authors pointed out how it would be possible to fool name-based authentication systems by means of cache contamination attacks. To solve this problem a security extension, DNSSEC, was proposed (IETF RFCs 2065 and 2535). Recent vulnerabilities discovered by Dan Kaminsky via cache poisoning attacks finally led to the development of the latest generation specifications, IETF RFC 4033, 4034, and 4035.

Remaining in the context of data corruption, several DNS hijacking attacks have been reported since 2008, where the attacked domain name registrars were the target. For example, in 2008, a large e-bill payment site was compromised (targeting its domain name registrar) by redirecting its visitors to a crafted web address, later attempting to install malicious code on the visitors’ machines. A major case of a successful cache poisoning attack against the DNS infrastructure was reported in Brazil in 2009, against one of the major Brazilian banks: the login page redirection toward a fraudulent site caused the theft of users’ access credentials. In 2012, the \*.ke domains had a good share of data corruption attacks; for example, 103 of the Government of Kenya’s websites (.go.ke) were hacked in one night [12].

**Denial of service** attacks are aimed at impacting the DNS infrastructure composed of DNS servers and the network con-

nections. There have been two major reported distributed denial of service (DDoS) attacks on the root servers, in 2002 and 2007. The first attack covered a timeline of around one hour and targeted simultaneously all of the 13 root DNS servers, affecting overall performance and in particular degrading the availability level for some of them. In light of this attack, the Anycast protocol was implemented in several root servers, mitigating the second wave of global coordinated DDoS attacks, which occurred in 2007.

A successful Denial of Service attack on a regional name server or on a name server in a higher position in the hierarchy can completely make inaccessible distributed applications (from web site to control systems) in a entire DNS zone or geographical region.

Finally, *privacy* threats are related, for example, to snooping of DNS caches. Privacy, despite its relevance, is out of the scope of our investigation.

## The MeNSa Project

The scope of the MeNSa project [8] is to define a methodology and a set of metrics to quantify the global health level of the DNS. The DNS community agrees on the fact that while it is a common practice to individually monitor the DNS subsystems to observe if the traffic parameters deviate from the average, it is a challenge to extract knowledge on the more global DNS behavior and its “normality” [4].

The key actions we propose to face this challenge are:

- To refine and improve existing metrics for DNS health indicators
- To define a metric aggregation model to merge measured metrics into a few indicators
- To identify metric threshold levels that allow the DNS community to trigger when the behavior is normal or abnormal

While in the long term the MeNSa project would provide a solution to all the above items, in this article, we concentrate our attention on *metric aggregation*.

The most relevant concepts behind the MeNSa methodology are summarized in the following.

**The DNS reference model** defining the boundaries of the system we want to measure. Figure 1 shows the simplified architecture we consider. The *end user application* (e.g., browser, Apps, thin/fat clients) generates DNS queries, and can have advanced features such as prefetching and internal caching. Name servers work at a different level of the hierarchy, from root zones to local caches. Also of great importance are the Anycast resolvers.

**The set of metrics** to quantify the health and security level of the DNS. The metrics we propose are intended to evaluate the health of the DNS by measuring the DNS along three dimensions: vulnerabilities, security, and resiliency. Examples of metrics, clustered by threat category, are reported in Table 1. A comprehensive description can be found in [8].

**The set of measurement techniques and tools** put in place to gather information needed to compute metrics. How measurements are implemented depends on two main factors:

- What can be measured from which point
- The time horizon of data collection (e.g., seconds, hours, days, or months)

Measurement techniques and data collection issues are out of the scope of the project (and of this article).

**The concept of point of view (PoV).** A PoV is intended as the perspective of a DNS actor/component in observing, using, operating, and influencing the global DNS. Potential users of the MeNSa methodology fall into one of the following categories: *end users*, who are mostly unaware of the DNS function and operation; *service providers*, such as the Internet

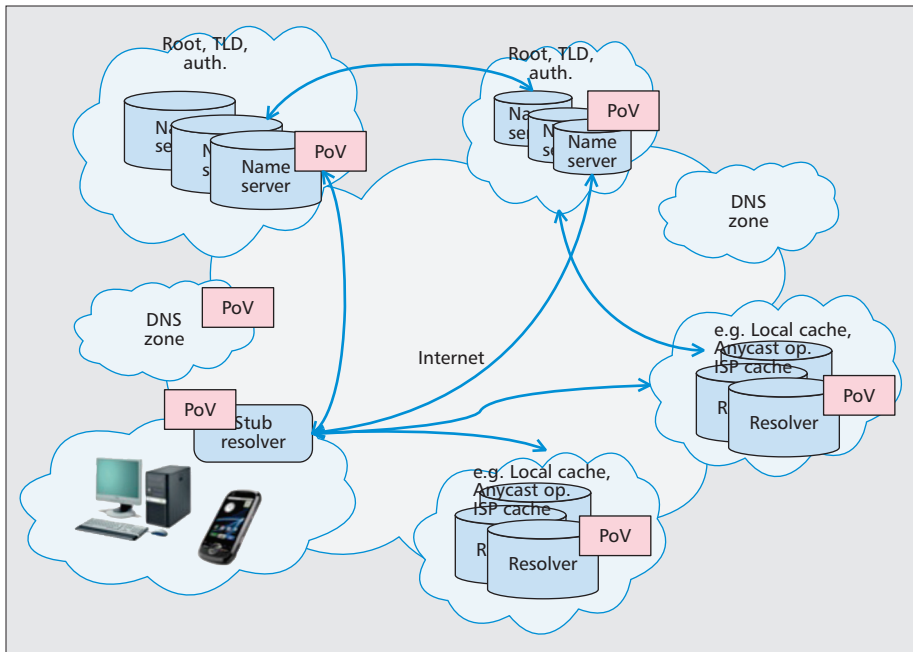


Figure 1. DNS reference architecture considered in the MeNSa project.

and application service providers; and *operators* (e.g., resolvers, name servers, registrars). The definition of different points of view is intended to categorize which components can be observed and measured by a specific DNS actor and what information is needed from other DNS actors to properly assess the level of DNS health perceived. This categorization will allow, for each PoV, defining a set of health indicators and a set of measurable metrics needed to evaluate the indicator of interest. The six points of view we defined are end-user PoV, application service provider PoV, resolver PoV, name server PoV, zone PoV, and global PoV (Fig. 1).

From each PoV it is possible to measure the perceived health level mixing two sources of information (Fig. 2): direct observations of the global DNS behavior and the Internet traffic (e.g., through active and passive measurement); and data shared by other PoVs (e.g., in the form of aggregated measures or anonymized data). The strength of this distributed approach is that measurement is kept local, and only aggregated information is shared. Local data collection is usually performed by DNS operators that daily collect tons of data to manage their infrastructure; therefore, there is no overhead. Sharing aggregated information (e.g., health indices) introduces negligible network overhead. The only additional tasks for a DNS operator are data processing to compute health and security metrics and data sharing. Both tasks imply the agreement of providers on a set of metrics and sharing rules. As reported in [4], the community is moving in this direction.

Using this approach, it should be possible to build a global picture of the health state of the DNS.

### Methodology in Action

The methodology operation is organized in three macro phases:

- *Preliminary diagnosis* that, from the chosen PoV, performs a first evaluation of the health level perceived conducting simple measurements and assessments.
- *The definition of the service level objectives (SLOs) and scenario phase*, given that the PoV is allowed to select one or more threat scenarios and the measurable and representative indices.
- *The detailed diagnosis and measurement phase* assessing the

DNS health level perceived; the achievable SLOs; the causes of SLO violation and improvement actions. The detailed diagnosis and measurement phase is organized in three stages: selection of metrics, measurement, and aggregation.

At the *aggregation* stage, all the measures collected are combined to provide aggregated indices summarizing the DNS health level perceived by the PoV, what the achievable SLOs are, and finally, what the cause of health degradation could be and possible solutions. Measures are performed using network and DNS measurement tools typically used by the community.

### Metrics Aggregation

Internet measurement theories and practices suggest that to understand the system behavior, it is better to

look at a set of metrics rather than a single indicator. Trade-offs exist between different metrics [13], and a single value can be misleading. This is true from a technical or scientific perspective, while non-technically skilled users or decision makers are typically affected by “mononumerosis,” an undue focus on a single measured value (as defined in the 1990s by Cindy Bickerstaff of the IETF IPPM metrics working group).

The concept of DNS health is multi-faceted and, as above mentioned, can hardly be captured considering the measurements on a single metric. However, non-technically skilled users do not appreciate the details of a multitude of metrics and prefer one or very few simple indicators. To overcome this problem we propose to compute a limited set of indicators that have three fundamental features:

- They provide aggregate technical level metrics.
- They provide a general understandable (by any user) system state description.
- They give a measure of the security level intended not only as prevention against unauthorized access but also for performance, stability, and resiliency.

Let us now discuss how DNS health metrics can be aggregated. Formally, a PoV is associated with:

- A set of  $M$  metrics  $\{m_1, \dots, m_M\}$ . Let  $D_i$  be the *domain* of the metric  $m_i$  (i.e., the measured values  $v_{i1}, v_{i2}, \dots$ ) of  $m_i$  belongs to  $D_i$ .
- A set of  $M$  quality mappings  $q_i : D_i \rightarrow [0,1]$ , one for each metric  $m_i$ . The mapping  $q_i$  transforms the measured value  $v_{ij}$  into a dimensionless quality value  $q_{ij} = q(v_{ij})$ , where 0 indicates the lowest quality and 1 indicates the highest one.
- A set of *aggregated indicators*. Each indicator  $I_k$  is fully defined by its *vector of weights*  $w_k = (w_{k1} \dots w_{kM})$  such that  $\sum_{i=1}^M w_{ki} = 1$ .

Different techniques can be used to aggregate metrics. These techniques do not depend on the specific PoV and should satisfy two properties. First, the aggregation process should not depend on the metrics to be aggregated and the aggregated index. Second, the timescale of the phenomena observed should not influence the aggregation, that is, the aggregation process should be capable of handling variable timescales and sampling frequencies.

For example, given a metric  $m_i$  and an observation time period  $T$ , and a chosen sampling time interval, it is possible to

Category	Measure	Metric
Repository Corruption	Data Staleness	Percentage of differing SOA serial numbers across all auth. servers numbers over a time period.
	Zone drift/Zone trash	Probability of incurring in zone drift and zone thrash status
	NS Parent/Child Data Coherence	Percentage of differences between the responses to NS queries to the parent zone with the responses to NS queries among all authoritative servers for the zone within one serial number
System Corruption	Cache Poisoning	Percentage of differences between the contents of caches vs. authoritative data
	Zone Transfer failure	Number of failed zone transfer operations
	DNS spoofing	Probability of being spoofed and probability of being spoofed over a time period
Denial of Service	Variation of DNS Request per Second	Variation of the requests number per second
	Incoming bandwidth consumption	Percentage of the available bandwidth
	Incoming traffic variation	Variation of incoming DNS traffic
Resiliency	Mean Time to Incident Discovery	Average value over a long observation period
	Operational Mean Time Between Failures	Average value over a long observation period
	Operational Availability	Percentage of the mean time an ICT system is running at the normal service level over the observation time period
Security	Attack Surface	Percentage of nodes of a target system that is susceptibility to a certain type of attack.
	Attack Deepness	Percentage of impacted nodes of a system as consequence of an attack
	Attack Escalation Speed	Attacks in a time unit variations
	Annualized Loss Expectancy	Dollars loss as consequence of incidents per years

**Table 1.** Examples of DNS health and security metrics.

identify  $S_i$  sessions  $\{s_{ij}\}$  with  $j = 1 \dots S_i$ , and to compute  $S_i$  values of the metric.

Having computed the  $v_{ij}$  values of  $m_i$ , it is possible to evaluate the quality values  $q_{ij}$  through the quality function  $q_i$ . Then the mean value  $\bar{q}_i$  and the standard deviation  $\Delta q_i$  over the  $S_i$  sessions are computed as the quality value of the metric  $m_i$  and the corresponding uncertainty level, respectively.

The aggregated indicators are computed as weighted averages using their vectors of weights.

An estimation of the uncertainty can be expressed by the squared weighted average, as is standard in error theory. Formally, the  $k$ th aggregator and the error estimations are computed as

$$I_K = \sum_{i=1}^M w_{ik} q_i \quad \Delta I_k = \sqrt{\sum_{i=1}^M w_{ik}^2 \Delta q_i^2}$$

Figure 3 shows an example of the aggregation process. In the example six metrics are considered, which are IBC, ITV, TT, CP DNSR and RRQ (see the next section for the descriptions of their meanings). Metrics are measured and transformed into quality values at step 1. At step 2 the average value and the error for each metric are computed. Finally, at

step 3, the values representing each single metric are aggregated using the weighted average. The choice of weight  $w_{ik}$  is arbitrary and depends on the relevance of metric  $m_i$  for the health index  $I_k$ . An example is provided later.

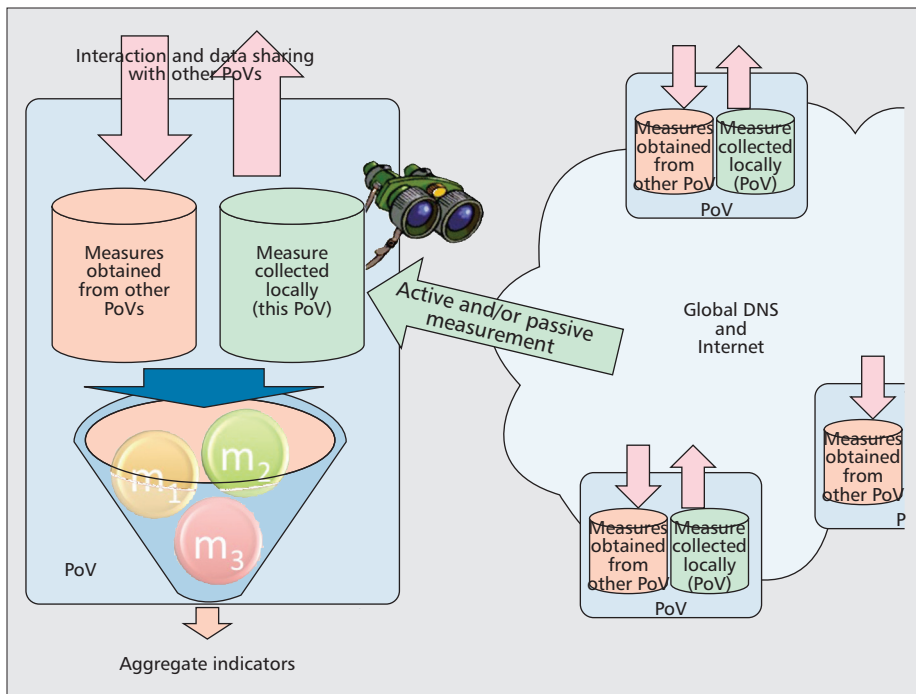
### A Case Study

As an example to clarify the use of the proposed methodology we consider the end-user PoV, which represents the perspective from which each user can evaluate the naming service. From the end-user PoV, the components involved in the resolution process are the end-user application, the stub resolver, and the network, while the only operation of interest is the DNS lookup process.

This case study and the set of experiments that follows are designed with the following objectives in mind: to show examples of aggregated indicators, to show how a set of the metrics can be computed and how these metrics can be aggregated, and to explain how the values assumed by the aggregated indexes should be interpreted.

The aggregate indicators we consider are the following:

- Total evaluation (TE) index. It gives a global assessment of the PoV aggregating all the considered measurable metrics



**Figure 2.** Concept of PoV. The PoV creates its own knowledge of the health state of the DNS mixing local and remote information. Locally collected information is shared with other PoVs.

- Protocol issues (PI) index. It estimates possible DNS protocol problems (e.g., cache poisoning)
- Denial of service (DoS) index. It evaluates how improbable a DoS is in a given scenario.
- Network (NET) index. It estimates the performance of the network components.
- Stub resolver (SR) index. It evaluates the performance of the stub resolver (i.e., the operating system libraries that implement the DNS queries).

We chose these indices because:

- They are common in PoVs.
- They suit well the metrics we have chosen and those that are used by the DNS community. Such health indices are versatile, and indeed could be bound with different metrics depending on the specific PoV and available data. This flexibility also allows us to cope with evolving security threats.
- They give a measure of the security level intended not only as prevention against unauthorized access, but also for performance, stability, and resiliency.

Moreover, these indices are pragmatic (the DNS community does not like theoretical solutions very much) and are simple enough to be understood by non-technical people.

### Measurements and Metrics

We set up a testbed where two client machines running Windows OS and Firefox 8.0 query the DNS from two different Internet service providers (ISPs). To be specific, a client was connected to the Internet through the Italian ISP Fastweb using as its access point the GCSEC laboratory (located in South East Rome, Italy); and the other client was connected to the Internet through the GARR network using as its access point the University of Rome “Tor Vergata” (UTV). DNS resolutions are demanded by Fastweb and UTV resolvers, respectively.

During the tests we collected traffic from 12 web browsing sessions each. Every session lasted from 10 min to a total of 2 h. Collected traffic is analyzed to get a measure of the metrics.

In the MeNSa project we identified a large set of metrics

useful to assess DNS health and security. As explained in the project deliverables [8], we started by considering all the most important threat scenarios for the DNS. A metric is interesting if it is capable to track system changes and deviation from normal behavior. In the following experiments we select only the metrics computable in the end-user PoV that are able to represent the system dynamic in a timespan of 2 h. These metrics are:

- Incoming bandwidth consumption (IBC), the ratio between the total amount of incoming bits during a session and the duration of the session.
- Incoming traffic variation (ITV), defined, for each session  $i$ , as the variation of IBC measured in session  $i$  in respect to the value of IBC measured in section  $i - 1$ .
- Traffic tolerance (TT), measuring the round-trip time (RTT) of an IP packet flowing between the end-user node and the ISP’s recursive resolver.
- Stub resolver cache poisoning

(CP), measuring the percentage of poisoned entries of the cache. Every entry of the cache is checked against a set of known recursive resolvers.

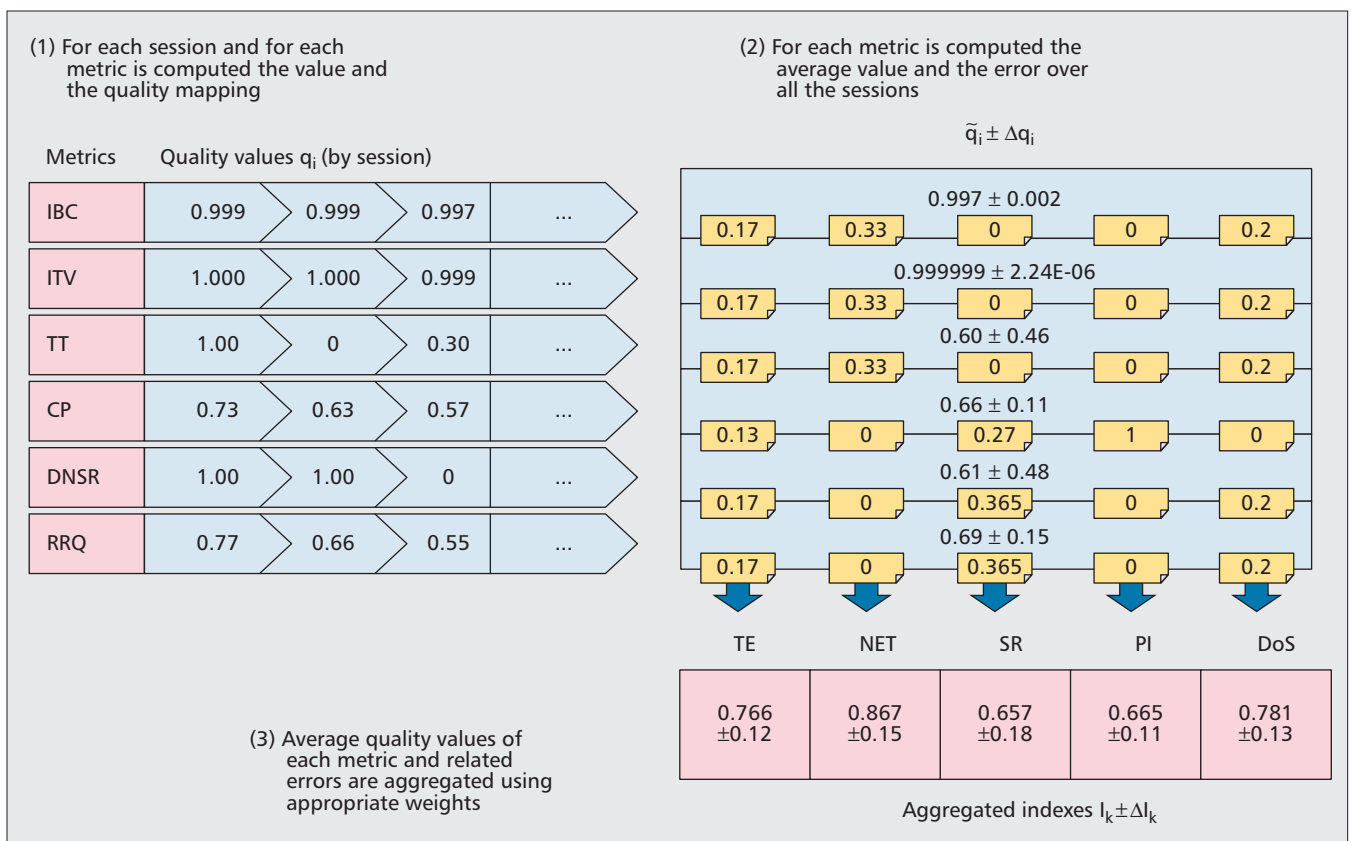
- DNS requests per seconds (DNSR), giving the total number of DNS queries in the session.
- Rate of repeated queries (RRQ) returns the number of repeated DNS queries in a session. During normal behavior, in a short time period, a name should be resolved only once because of DNS caching. If there are many DNS queries for the same name in the same session, this could be a marker of misbehavior.

IBC and ITV are measured using NetAnalyzer; TT is measured using ping. DNSR and RRQ are measured monitoring the session with WireShark and analyzing the resulting PCAP file. Finally, CP is measured by dumping the cache and parsing its content vs. authoritative DNS servers. The comparison is done immediately after the section to avoid resolver caches expiration.

### DNS Health Evaluation

Figure 3 shows an example of quality values computed for each session and the results of the metric-based aggregation explained earlier. The figure also shows the weight values. The TE index gives an overall evaluation; thus, it must aggregate all the available metrics with the same weight. The PI index in our case only refers to the cache poisoning problems, because in our test we decided to measure only this protocol issue. Thus, the corresponding vector of weights consists only of the cache poisoning metric. The DoS index aggregates all the metrics but the cache poisoning one because it focuses on network traffic. The NET index focuses only on network related metrics, equally considered. The SR index focuses only on stub resolver measures, giving more importance ( $\approx 75$  percent) to DNSR and RRQ.

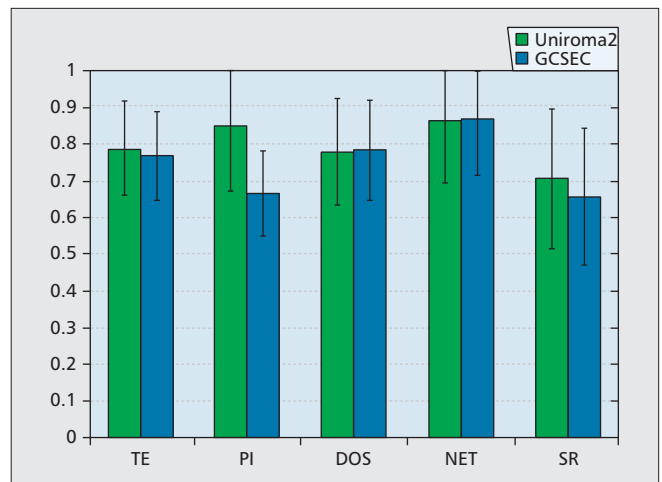
Using our testbed we set up two experiments: one reproduces normal working conditions and the other a cache incoherency scenario. The first experiment consists of two different sets of measurements that can be compared: one



**Figure 3.** An example of quality values and results of the metric-based aggregation. Steps 1–3 are executed in sequence. The weights that define the aggregation (represented in yellow boxes) are tuned in the validation phase of the methodology.

collected at GCSEC and another one at UTV (hereafter referred to as the GCSEC laboratory and Uniroma2 tests). Figure 4 shows the results of the first experiment. The TE index values computed at Uniroma2 and GCSEC are 0.79 and 0.76, respectively, showing that the overall performance is good enough in both cases. A further analysis of the other indices may be useful to increase the performance. The PI and DoS indexes give insights on the possible *issues* of the system, while the NET and SR indexes focus on the performance of the *system components*. The DoS index shows an equally good result in both settings (about 0.8). Indeed, there was no DoS issue, and no further investigation is needed. On the other hand, in the GCSEC experiment the PI index (about 0.67) is lower than the corresponding index in the Uniroma2 experiment (about 0.85). Thus, it emerges that the GCSEC access network should improve DNS security (e.g., changing its Internet access provider or changing contract or managing its own cache). The NET index shows that the network component worked properly in both tests (about 0.86). Instead, the SR index also shows good results, but the value 0.65 measured in the laboratory test suggests some possible improvement in the performance of the stub resolver used.

In the second experiment we simulated some cache poisoning in order to validate our methodology. We manually corrupted 10 percent of the DNS cache entries in the GCSEC laboratory. The TE decreases to 0.7. The NET index is still evaluated around 0.8, but the SR assessment goes down to 0.6. These results entail the presence of problems in the DNS libraries of the operating system as expected. Going further, through the measurement process we discovered that we clearly suffer from some protocol issues since  $PI = 0.38$ . The DoS indicator, however, remains above 0.75. Figure 5 contains the results of this experiment,



**Figure 4.** Comparison of the health and security level perceived by end users using the DNS from two different ISP.

where the normal behavior data were measured in the laboratory.

The results can lead to practical actions. Comparing the SR and PI indicators enables spotting the cache poisoning problem. Indeed, the result of the analysis should suggest refreshing the DNS cache. Repeating the same evaluation afterward would further validate this suggested action.

The results we obtained cannot be generalized and must be validated with a larger set of experiments. Our goal was to show that measurement is possible, and how metrics can be used and aggregated to investigate DNS health.



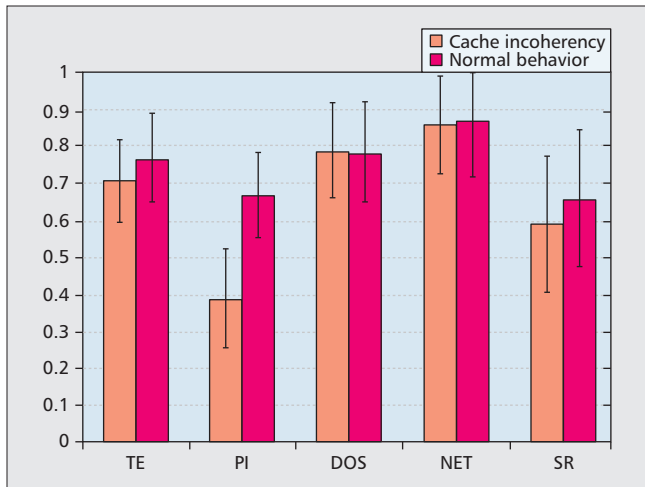


Figure 5. Comparison of the health and security level perceived by end users when a cache is poisoned. Poisoning has been artificially introduced in our collected data set.

## Concluding Remarks

The Domain Name System constitutes the hidden backbone of the Internet. Without its services almost all the applications making use of the public network would not be able to operate in an efficient manner. The massive use of information and communications technology systems in critical infrastructures puts the DNS under the lights as a new potential source of disservices.

The DNS community in the last few years has started to reflect on the need of methodologies for assessing the health of the global DNS system. In this article, after describing at a high level the MeNSa project, designed to fulfill this need, we provided the results of the first tests on field, showing how, from the end-user PoV, metrics can be aggregated and used as a tool to verify the level of service perceived and the presence or absence of threats. The aggregation method proposed is general enough to be applied to any PoV. Of course, the implementation, that is, the choice of aggregated indices, of the set of metrics and the definition of the quality mapping functions are strictly related to the PoV and the goal of the analysis.

## References

- [1] Y. Huang, D. Arsenault, and A. Sood, "SCIT-DNS: Critical Infrastructure Protection Through Secure DNS Server Dynamic Updates," *J. High Speed Networks*, vol. 15, no. 1/2006, IOS Press, pp. 5–19.
- [2] I. Nai Fovino, S. Di Blasi, and A. Rigoni, "The Role of the DNS in the Secure and Resilient Operation of CIs: The Energy System Example,"

- CRITIS, Lucerne, Switzerland, Sept. 2011.
- [3] ICANN, "Measuring the Health of the Domain Name System," Report of the 2nd Annual Symp. DNS Security, Stability, & Resiliency, Kyoto, Japan, 2010.
- [4] ICANN, GCSEC, DNS-OARC, "Final Report of the 3rd Global DNS Stability, Security and Resiliency Symposium," 2011, Rome, Italy.
- [6] R. Liston, S. Srinivasan, and E. Zegura, "Diversity in DNS Performance Measures," *Proc. 2nd ACM SIGCOMM Wksp. Internet Measurement*, 2002 ACM, New York, NY, USA, pp. 19–31.
- [5] S. Castro et al., "A Day at the Root of the Internet," *ACM SIGCOMM Comp. Commun.*, 2008, Rev. 38, 5, pp. 41–46.
- [7] B. Ager et al., "Comparing DNS Resolvers in the Wild," *Proc. 10th ACM SIGCOMM Conf. Internet Measurement*, pp. 15–21.
- [8] Global Cyber Security Center (GCSEC), "Measuring the Naming System (MeNSa) Project," <http://www.gcsec.org/activity/research/dns-security-and-stability>.
- [9] X. Chen et al., "Maintaining Strong Cache Consistency for the Domain Name System," *IEEE Trans. Knowledge and Data Engineering*, vol. 19, no. 8, Aug. 2007.
- [10] S. M. Bellovin, "Using the Domain Name System for System Break-ins," *Proc. 5th USENIX UNIX Security Symp.*, Salt Lake City, UT, June 1995.
- [11] P. Vixie, "DNS and BIND Security Issues," *Proc. 5th USENIX UNIX Security Symp.*, Salt Lake City, UT, June 1995.
- [12] J. G. Kagwe and M. Masinde, "Survey on DNS Configurations, Interdependencies, Resilience and Security for \*.ke Domains," *Proc. 2nd ACM Symp. Computing for Development*, 2012.
- [13] N. Brownlee, C. Loosley, "Fundamentals of Internet Measurement: A Tutorial," *CMG J. Computer Resource Management*, vol. 102, 2001.

## Biographies

EMILIANO CASALICCHIO (emiliano.casalicchio@uniroma2.it), Ph.D., is a researcher in the Department of Civil Engineering and Computer Science of the University of Roma "Tor Vergata." Since 1998, his research has mainly focused on large-scale distributed systems, with a specific emphasis on performance oriented design of algorithms and mechanisms for resource allocation and management. Domains of interest have been distributed web servers, grid, service oriented architectures, cloud systems, as well as critical infrastructure protection. He is the of about 70 international publications, and his research is and has been funded by the Italian Ministry of Research, CNR, ENEA, the European Community, and private companies.

MARCO CASELLI (marco.caselli@gcsec.org) received his B.S. degree from the University of Palermo and his M.S.E. degree from Sapienza University, Rome. He is currently working toward his Ph.D. degree at the University of Twente. His research interests include critical infrastructure protection as well as industrial control systems' security. He has published scientific papers on DNS security in international conferences. Before starting his Ph.D. he worked for GCSEC.

ALESSIO COLETTA (alessio.coletta@gcsec.org) graduated in computer science at Scuola Normale Superiore di Pisa and possesses a background in computer science, specifically related to ICT security from both theoretical and practical points of view. He had research experience at the University of Pisa, Italy. He worked as a scientific officer at the Joint Research Centre of the European Commission, and he currently works at GCSEC in Rome, on research activities about ICT and industrial security, security policies, malware, critical infrastructure protection, digital identity, and incident response teams.

---

# Participatory Privacy: Enabling Privacy in Participatory Sensing

**Emiliano De Cristofaro, Palo Alto Research Center (PARC)  
Claudio Soriente, ETH Zurich, Switzerland**

---

## Abstract

Participatory sensing is an emerging computing paradigm that enables the distributed collection of data by self-selected participants. It allows the increasing number of mobile phone users to share local knowledge acquired by their sensor-equipped devices (e.g., to monitor temperature, pollution level, or consumer pricing information). While research initiatives and prototypes proliferate, their real-world impact is often bounded to comprehensive user participation. If users have no incentive, or feel that their privacy might be endangered, it is likely that they will not participate. In this article, we focus on privacy protection in participatory sensing and introduce a suitable privacy-enhanced infrastructure. First, we provide a set of definitions of privacy requirements for both data producers (i.e., users providing sensed information) and consumers (i.e., applications accessing the data). Then we propose an efficient solution designed for mobile phone users, which incurs very low overhead. Finally, we discuss a number of open problems and possible research directions.

---

In the last decade, researchers have envisioned an outbreak of wireless sensor networks (WSNs) and predicted the widespread installation of sensor (e.g., in infrastructures, buildings, woods, rivers, or even the atmosphere). This has triggered a lot of interest in many different WSN topics, including identifying and addressing security issues, such as data integrity, node capture, and secure routing. On the contrary, privacy has not really been a concern in WSNs, as sensors are usually owned, operated, and queried by the same entity. (For instance, the National Department of Transportation deploys sensors and collects traffic information related to national highways.)

On the other hand, the proliferation of mobile phones, along with their pervasive connectivity, has propelled the amount of digital data produced and processed every day. This has driven researchers and IT professionals to discuss and develop a novel sensing paradigm, where sensors are not deployed in specific locations, but are carried around by people. Today, many different sensors are already deployed in our mobile phones, and soon all our gadgets (e.g., even our clothes or cars) will embed a multitude of sensors (GPS, digital imagers, accelerometers, etc.). As a result, data collected by sensor-equipped devices becomes of extreme interest to other users and applications. For instance, mobile phones may report (in real time) temperature or noise level; similarly, cars may inform on traffic conditions.

This paradigm is called participatory sensing (PS) — sometimes also referred to as *opportunistic* or *urban* sensing [3]. It combines the ubiquity of personal devices with sensing capabilities typical of WSNs. As the number of mobile phone subscriptions exceeds 5 billion, PS becomes a cutting-edge and effective distributed computing (as well as business) model. We argue that PS appreciably expands the capabilities of WSN applications by, for example, allowing effective monitor-

ing in scenarios where the setup of a WSN is either not economical or infeasible.

However, its success is strongly related to the number of users actually willing to commit personal device resources to sensing applications, and thus to associated privacy concerns. Observe that sensing devices are no longer “dull” gadgets owned by the entity querying them. They are personal devices that follow users at all times, and their reports often expose personal and sensitive information. Consider, for instance, a PS application like <http://www.gasbuddy.com/> where gas prices are monitored via user reports, and information announced by participants inevitably exposes their current and past locations, and hence their movements. If users have no incentive to contribute sensed data or feel that their privacy might be violated, they will (most likely) refuse to participate. Thus, not only traditional security but also privacy issues must be taken into account. In this article, we focus on privacy protection in PS. We define privacy in this new context, present a privacy-enhanced PS infrastructure, and elaborate on a number of desirable features that constitute challenging research problems. The proposed privacy-protecting layer can easily be adopted by available PS applications to enforce privacy and enhance user participation.

## Participatory Sensing

### What Is Participatory Sensing?

PS is an emerging paradigm that focuses on the seamless collection of information from a large number of connected, always on, always carried devices, such as mobile phones. PS leverages the wide proliferation of commodity sensor-equipped devices and the ubiquity of broadband network infrastructure to provide sensing applications where deployment of a WSN infrastructure is not economical or infeasible.

PS provides fine-grained monitoring of environmental trends without the need to set up a sensing infrastructure. Our mobile phones *are* the sensing infrastructure, and the number and variety of applications are potentially unlimited. Users can monitor gas prices (<http://www.gasbuddy.com/>), traffic information (<http://www.waze.com/>), and available parking spots (<http://spotswith.com/>), just to cite a few. We refer readers to [4] for an updated list of papers and projects related to PS.

### *What Is Not Participatory Sensing?*

PS is not a mere evolution of WSNs, where nodes are replaced by mobile phones. Sensors are now relatively powerful devices, such as mobile phones, with much greater resources than WSN nodes. Their batteries can easily be recharged and production cost constraints are not as tight. They are extremely *mobile*, as they leverage the ambulation of their carriers. Moreover, in traditional WSNs, the network operator is always assumed to manage and own the sensors. On the contrary, this assumption does not fit most PS scenarios, where mobile devices are *tasked* to participate in gathering and sharing local knowledge. Hence, a sensor (or its owner) might choose whether to participate or not. As a result, in PS applications, different entities coexist and might not trust each other.

### *Participatory Sensing Components*

A typical PS infrastructure involves (at least) the following parties:

- **Mobile nodes** are the union of a carrier (i.e., a user) with a sensor installed on a mobile phone or other portable wireless-enabled device. They provide reports and form the basis of any PS application.
- **Queriers** subscribe to information collected in a PS application (e.g., “temperature in Irvine, CA”) and obtain corresponding reports.
- **Network operators** manage the network used to collect and deliver sensor measurements (e.g., they maintain GSM and/or third/fourth generation, 3G/4G, networks).
- **Service providers** act as intermediaries between queriers and mobile nodes, in order to deliver reports of interest to queriers.

Queriers can subscribe to the appropriate service provider for one or more types of measurements. For example, assume that Alice subscribes to “available parking spots on W 16th Street, New York,” or Bob is interested in the “temperature in Central Park, New York.” In turn, mobile nodes share local knowledge — either voluntarily or in return for some profit — with one or more service providers, which make information available to queriers. For example, assume Carol’s mobile phone sends report “3 available parking spots on E 56th, New York,” while John’s device sends “74°F in Central Park, New York.”

As mobile nodes and queriers have no direct communication or mutual knowledge, service providers route reports matching specific subscriptions to their original queriers. In fact, mobile nodes ignore which queriers (if any) are interested in their reports. For example, the service provider forwards John’s temperature report to Bob; Carol’s parking report is not sent to Alice as it refers to a different location.

### *Privacy Concerns*

PS provides an effective solution to a wide range of applications; however, it prompts several security and privacy concerns that need to be carefully addressed.

On one hand, issues such as confidentiality or integrity can be mitigated using state-of-the-art techniques. For instance, all parties can be protected from external eavesdroppers using Secure Socket Layer (SSL)/Transport Layer Security (TLS).

The latter provides a secure channel between any two parties, so communications between mobile nodes and service providers or between service providers and queriers are kept confidential.

On the other hand, the need for privacy protection stems from the potential leakage of personal information to *internal adversaries*. Indeed, as the service provider collects all data (i.e., reports and queries), it might learn a considerable amount of sensitive information about both mobile nodes and queriers, and violate the privacy of their movements, interests, habits, and more. For instance, the service provider learns that both Bob and John are located in Central Park, New York. It also learns that Alice is driving on West 16th Street, looking for parking. The continuous collection of information over long periods allows the service provider to meticulously profile users.

Furthermore, as data collected through PS applications becomes available to external entities and organizations (i.e., the queriers), query interests also become sensitive and need to be hidden. For instance, service providers should not learn which interests are “hot.”

Finally, there is a tension between privacy and accountability as PS business models may require, at the very least, that reports are available only to entitled (e.g., authorized or paying) members.

However, we claim there is one main reason to protect privacy. If users feel that their privacy is endangered, they will deny sharing their reports. Specifically, it is required that the service provider performs report/query matching but learns no information about query interests. Also, data reports should not reveal to the service provider, the network operator, or unauthorized queriers any information about a mobile node’s identity, its location, the type of measurement (e.g., temperature), or the quantitative information (e.g., 74°F).

## *A Novel Privacy-Enhanced Participatory Sensing Infrastructure*

We now present our innovative solution for a Privacy-Enhanced Participatory Sensing Infrastructure (PEPSI). We describe its architecture and privacy desiderata, and overview our instantiation. Finally, we discuss efficiency costs introduced by the privacy-protecting layer.

### *PEPSI Architecture*

PEPSI protects privacy using efficient cryptographic tools. Similar to other cryptographic solutions, it introduces an additional (offline) entity, the registration authority. It sets up system parameters and manages mobile nodes or queriers registration. However, the registration authority is not involved in real-time operations (e.g., query/report matching); nor is it trusted to intervene for protecting participants’ privacy.

Figure 1 illustrates the PEPSI architecture. The registration authority can be instantiated by any entity in charge of managing participants registration (e.g., a phone manufacturer). A service provider offers PS applications (used, e.g., to report and access pollution data) and acts as an intermediary between queriers and mobile nodes. Finally, mobile nodes send measurements acquired via their sensors using the network infrastructure, and queriers are users or organizations (e.g., bikers) interested in obtaining reports (e.g., pollution levels).

PEPSI allows the service provider to perform report/query matching while guaranteeing the privacy of both mobile nodes and queriers. It aims at providing (provable) privacy by design, and starts off with defining a clear set of privacy properties.

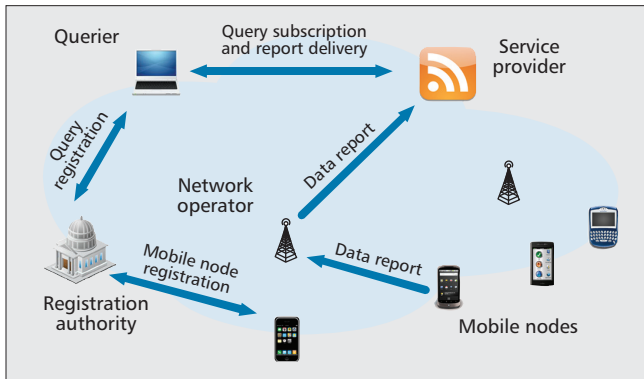


Figure 1. Privacy-enhanced participatory sensing infrastructure.

### Privacy Desiderata

The *privacy desiderata* of PS applications can be formalized as follows:

- **Soundness:** Upon subscribing to a query, queriers in possession of the appropriate authorization always obtain the desired query results.
- **Node privacy:** Neither the network operator, the service provider, nor any unauthorized querier learn any information about the type of measurement or the data reported by a mobile node. Also, mobile nodes should not learn any information about other nodes' reports. Only queriers in possession of the corresponding authorization obtain reported measurements.
- **Query privacy:** Neither the network operator, the service provider, nor any mobile node or any other querier learn any information about queriers' subscriptions.
- **Report unlinkability:** No entity can successfully link two or more reports as originating from the same mobile node. However, as we discuss below, we do not pursue report unlinkability with respect to the network operator.
- **Location privacy:** No entity can learn the current location of a mobile node (again, excluding the network operator).

In realistic scenarios, it appears unlikely — if not impossible — to guarantee report unlinkability and location privacy with respect to the network operator. In fact, PS strongly relies on the increasing use of broadband 3G/4G connectivity. In these networks, current technology does not allow providing user anonymity with respect to the network operator. Mobile nodes are identified through their International Mobile Subscriber Identity, and any technique for identifier obfuscation would lead to service disruption (e.g., the device would not receive incoming calls). Furthermore, the regular usage of cellular networks (e.g., incoming/outgoing phone calls), as well as heartbeat messages exchanged with the network infrastructure, irremediably reveal a device's location. To provide report unlinkability/location privacy with respect to other parties, we need to trust the network operator (who routes mobile nodes' reports to service providers) not to forward any information identifying the mobile nodes (the identifier, the cell from which the report was originated, etc.).

### PEPSI Construction

One of the main goals of PEPSI is to hide reports and queries from unintended parties. Thus, those cannot be transmitted *in the clear*, but must be encrypted. In this section, we discuss how to achieve, at the same time:

- Secure encryption of reports and queries
- Efficient and oblivious matching by the service provider

Due to space limitations and to ease presentation, we only provide an overview of our construction (with no technical details). We refer interested readers to [5] for a complete

description of our techniques, as well as formal cryptographic proofs.

*A Naive Solution* — Traditional confidentiality means are not suited for PS applications. Recall that in our context, mobile nodes and queriers have no mutual knowledge or common history; that is, mobile nodes provide reports oblivious of (any) potential receiver, while queriers subscribe to data reports not knowing who (if anyone) will provide measurements of interest. Hence, we cannot assume that each mobile node shares a unique pairwise secret key with each querier or that reports are encrypted under that key via a symmetric key cipher (e.g., Advanced Encryption Standard, AES). Even if we were to allow interactions between mobile nodes and queriers, we would still need the former to encrypt reports under each key shared with queriers. This would generate a number of ciphertexts quadratic in the number of measurements. Alternatively, we could use a public key encryption scheme and provide mobile nodes with the public keys of queriers. Still, scalability would be an issue as each report would be encrypted under the public key of each querier. In general, because of scalability and loose coupling between data producers and consumers, mobile nodes cannot provide measurements intended for a specific querier, and the latter cannot ask for data from a given mobile node.

Our main building block is Identity-Based Encryption (IBE) — a cryptographic primitive, based on bilinear map pairings, that enables asymmetric encryption using any string (“identity”) as a public key. In IBE, anyone can derive public keys from some unique information about the recipient's identity. Private decryption keys are generated by a third party, called the private key generator (PKG). *Our intuition is to use a tagging mechanism on top of IBE.*

*Report Encryption* — We assume that each report or subscription is identified by a set of labels, or keywords. These are used as “identities” in an IBE scheme. For example, labels “Temperature” and “Central Park, NY” can be used to derive a unique public encryption key, associated to a secret decryption key. Thus, mobile nodes can encrypt sensed data using a report's labels as the (public) encryption key. Queriers should then obtain the private decryption keys corresponding to the labels of interest. Those are obtained, upon query registration, from the registration authority, which, in practice, acts like a PKG.

*Efficient Matching using Cryptographic Tags* — After enabling encryption/decryption of reports, we need to allow the service provider to efficiently match them against queries. In fact, the application of IBE to PS settings is not trivial: with straightforward use of IBE, oblivious matching of queries and reports would be impossible. In other words, the service provider would forward *all* (encrypted) reports to all queriers; each of them will only be able to decrypt reports of interests (i.e., the ones for which they hold the decryption keys). However, given the large amount of reports produced by mobile nodes, this would incur a considerable overhead for the querier, which must try to decrypt all reports using each of her decryption keys. To address this problem, we propose an efficient tagging mechanism: mobile nodes tag each report with a cryptographic token that identifies the nature of the report only to authorized queriers, but does not leak any information about the report itself. Tags are computed using the same labels used to derive encryption keys. Similarly, queriers compute tags for the labels defining their interests (using the corresponding decryption keys) and provide them to the service provider at query subscription.

Our main contribution, in this context, is to exploit the mathematical properties of bilinear mapping: we ensure that whenever a report matches a query, corresponding tags also match. In other words, a tag computed by John using the encryption key derived from label “temperature in Central Park, New York,” is equal to the tag computed by Bob using the decryption key computed over the same label. Specifically, mobile nodes upload reports along with the respective tags, while queriers define their subscriptions uploading the tags they compute at the service provider. The latter can find matches (i.e., a tag related to a report equals the tag related to a subscription) without learning any information about underlying queries/reports.

### PEPSI Operations

Figure 2 shows how PEPSI works. The upper part of the figure depicts the offline operations where the registration authority is involved to register both mobile nodes and queriers.

*Querier Registration* — In the example, querier  $\mathcal{Q}$  (the laptop on the right side) picks “Temp” among the list of available queries and obtains the corresponding decryption key (yellow key).

*Mobile Node Registration* — Similarly, mobile node  $\mathcal{M}$  (the mobile phone on the left side) decides to report the temperature in its location and obtains the corresponding secret used for tagging (grey key).

The bottom part of Fig. 2 shows the online operations where the service provider is involved.

*Querier Subscription* —  $\mathcal{Q}$  subscribes to queries of type “Temp” in “Irvine, CA” using these keywords and the decryption key acquired offline to compute a (green) tag; the algorithm is referred to as  $\leftarrow \text{TAG}()$ . The tag leaks no information about  $\mathcal{Q}$ ’s interest and is uploaded at the service provider.

*Data Report* — Any time  $\mathcal{M}$  wants to report on temperature, it derives the public decryption key (red key) for reports of type “Temp” (via the  $\leftarrow \text{IBE}()$  algorithm) and encrypts the measurement; encrypted data is pictured as a vault.  $\mathcal{M}$  also tags the report using the secret key acquired offline and a list of keywords characterizing the report; in the example,  $\mathcal{M}$  uses keywords “Temp” and “Irvine, CA.” Our tagging mechanism leverages the properties of bilinear maps to make sure that if  $\mathcal{M}$  and  $\mathcal{Q}$  use the same keywords, they will compute the same tag, despite each of them using a different secret key ( $\mathcal{M}$  is using the grey key while  $\mathcal{Q}$  is using the yellow one). As before, the tag and the encrypted report leak no information about the nature of the report or the nominal value of the measurement. Both the tag and encrypted data are forwarded to the service provider.

*Report Delivery* — The service provider only needs to match tags sent by mobile nodes with the ones uploaded by queriers. If the tags match, the corresponding encrypted report is forwarded to the querier. In the example of Fig. 2 the green tag matches the blue one, so the encrypted report (the vault) is forwarded to  $\mathcal{Q}$ . Finally,  $\mathcal{Q}$  can decrypt the report using the decryption key and recover the temperature measurement.

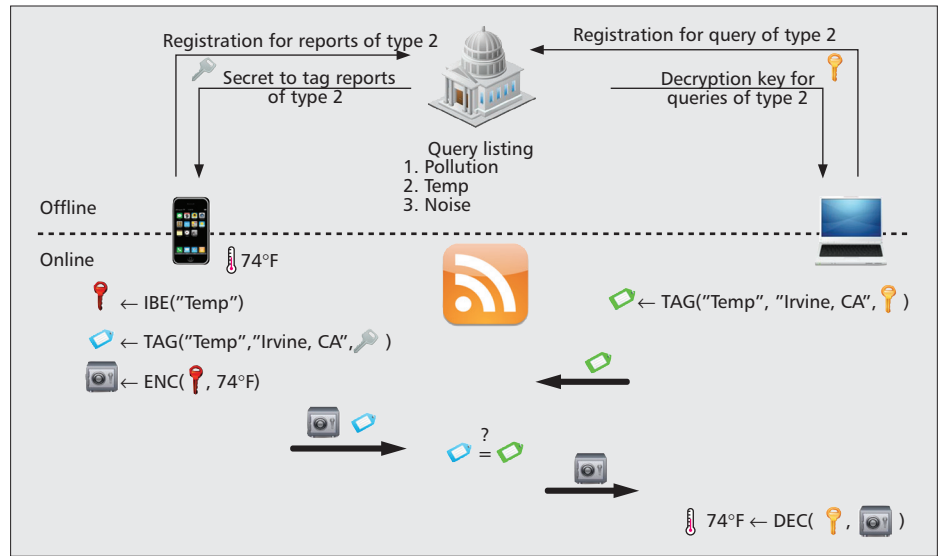


Figure 2. PEPSI operations.

### PEPSI Overhead

Resources in PS are not as constrained as in WSNs; nonetheless, overhead incurred at mobile nodes should still be minimized. To foster the adoption of our solution in current PS applications we provide an experimental evaluation of the cost of cryptographic operations used to achieve intended privacy features. We implemented protocol operations executed by mobile nodes on a Nokia N900 (equipped with a 600 MHz ARM processor and 256 Mbytes RAM). Computation overhead, for every report, is due to the computation of the tag and the encryption of the measurement. In our experiments, we experience an average time (over 100 trials) of 93.47 ms to perform these operations.

Communication overhead is merely due to the transmission of the tag, which is the output of a hash function (e.g., SHA-1); thus, it is relative small (160-bit). The encryption of the measurement generates almost no overhead, since, using state-of-the-art symmetric-key ciphers (e.g., AES), the ciphertext’s length is almost the same as plaintext’s.

Tag computation by queriers is performed only once, during query subscription. Upon reception of measurement of interests, queriers perform symmetric-key decryption, which incurs a negligible overhead.

Finally, note that the service provider incurs no communication or computational overhead; its task is limited to comparing output of hash functions (i.e., tags) and forwarding reports. From a functional point of view, the work of the service provider is no different from that in a non privacy-preserving solution. Thus, privacy protection incurs no overhead at the service provider and enjoys scalability to large-scale scenarios. We conclude that our architecture is practical enough, today, to be deployed for real-world PS applications.

### Related Work

#### Participatory Sensing Projects

In the last few years, PS initiatives have multiplied, ranging from research prototypes to deployed systems. Due to space limitations we briefly review some PS application that apparently expose participant privacy (location, habits, etc.). Each of them can easily be enhanced with our privacy-protecting layer. Interested readers can find a larger list of PS applications at [4]. Quake-Catcher [1] aims at building the world’s largest low-cost strong-motion seismic network by utilizing accelerometers embedded in any Internet-connected device.

Kim *et al.* [10] use the power of PS for meaningful places (e.g., home, office) discovery. PS has been shown to be an effective means to monitor levels of air pollution [14], noise pollution [13], and water quality [11]. PS to aid health care providers in patient monitoring has been investigated in [12].

### Privacy

Only a little attention has been paid to arising privacy issues in PS [16]. The authors of [2] study privacy in participatory sensing relying on weak assumptions: they attempted to protect *anonymity* of mobile nodes through the use of Mix Networks. (A Mix Network is a statistical-based anonymizing infrastructure that provides *k*-anonymity; i.e., an adversary cannot tell a user from a set of *k*.) However, Mix Networks are unsuitable for many PS settings. They do not attain provable privacy guarantees and assume the presence of a ubiquitous WiFi infrastructure used by mobile nodes, whereas PS applications do leverage the increasing use of broadband 3G/4G connectivity. In fact, a ubiquitous presence of open WiFi networks is not realistic today or anticipated in the near future. By contrast, our work aims at identifying a minimal set of realistic assumptions, and defining system properties and clear privacy guarantees to be achieved with provable security.

The work in [15] studies privacy-preserving data aggregation (computation of sum, average, variance, etc.). Similarly, [7] presents a solution for community statistics on time-series data, while protecting anonymity (using data perturbation in a closed community with a known empirical data distribution). Finally, [8] aims at guaranteeing integrity and authenticity of user-generated contents, by employing trusted platform modules (TPMs).

The main technical challenge in providing provable privacy in participatory sensing infrastructure stems from the simultaneous presence of several mutually untrusted (and potentially unknown) entities, including data producers, data consumers, and service providers. A similar scenario arises in the context of publish-subscribe networks [6], which face similar privacy concerns. However, state-of-the-art solutions (e.g., [9]) assume an a priori knowledge (and key exchange) between publishers and subscribers, while PS application requires loose coupling between mobile nodes and queriers. This makes it impossible to apply them to the PS scenario, where data producers and consumers may not know each other. Our solution protects their privacy while requiring no direct interaction between the two parties.

### Conclusion and Open Problems

Participatory sensing is a novel computing paradigm that bears great potential. If users are incentivized to contribute personal device resources, a number of novel applications and business models will arise. In this article we discuss the problem of protecting privacy in participatory sensing. We claim that user participation cannot be afforded without protecting the privacy of both data consumers and data producers. We also propose the architecture of a privacy-preserving participatory sensing infrastructure and introduce an efficient cryptographic solution that achieves privacy with provable security. Our solution can be adopted by current participatory sensing applications to enforce privacy and enhance user participation, with little overhead.

This work represents an initial foray into robust privacy guarantees in PS; thus, much remains to be done. Items for future work include (but are not limited to):

- Protecting query privacy with respect to the registration authority. Recall, in fact, that querier Alice needs to obtain the IBE decryption keys from the registration authority, which would then learn Alice's query interests.
- Protecting node privacy with respect to the network opera-

tor. Current technology does not allow users' locations and identities to be hidden from the network operator. Hence, it is an interesting challenge to guarantee node anonymity in broadband networks.

- Addressing collusion attacks, where multiple entities might collaborate in order to violate the privacy of mobile nodes or queriers.
- Improving the syntax of supported query types. In fact, PEPSI so far allows query/report matching based on the tags provided by both mobile nodes and queriers. However, PS applications might require more complex queries where queriers are interested in an aggregate of the reports (e.g., average or sum), or even complex query predicates (e.g., comparisons). While simple aggregate function evaluation over encrypted data is viable with available cryptographic techniques (e.g., homomorphic encryption), enabling *efficient* evaluation of complex predicates remains an open challenge.

### References

- [1] E. S. Cochran *et al.*, "The QuakeCatcher Network: Citizen Science Expanding Seismic Horizons," *Seismological Research Letters*, vol. 80, 2009, pp. 26–30.
- [2] C. Cornelius *et al.*, "AnonySense: Privacy-Aware People-Centric Sensing," *6th Int'l. Conf. Mobile Systems, Applications, and Services*, 2008, pp. 211–224.
- [3] D. Cuff, M. H. Hansen, and J. Kang, "Urban Sensing: Out of the Woods," *Commun. ACM*, vol. 51, no. 3, 2008, pp. 24–33.
- [4] E. De Cristofaro and C. Soriente, "Privacy-Preserving Participatory Sensing Infrastructure," <http://sprout.ics.uci.edu/PEPSI/index.php?page=projects.php>.
- [5] E. De Cristofaro and C. Soriente, "Privacy-Enhanced Participatory Sensing Infrastructure," <http://sprout.ics.uci.edu/PEPSI/TR-2011-01.pdf>.
- [6] P. T. Eugster *et al.*, "The Many Faces of Publish/Subscribe," *ACM Computing Surveys*, vol. 35, no. 2, 2003, pp. 114–31.
- [7] R. K. Ganti *et al.*, "PoolView: Stream Privacy for Grassroots Participatory Sensing," *6th Int'l. Conf. Embedded Networked Sensor Systems*, 2008, pp. 281–94.
- [8] P. Gilbert *et al.*, "Toward Trustworthy Mobile Sensing," *11 Wksp. Mobile Computing Systems and Applications*, 2010, pp. 31–36.
- [9] M. Ion, G. Russello, and B. Crispo, "Supporting Publication and Subscription Confidentiality in Pub/Sub Networks," *6th Int'l. ICST Conf. Security and Privacy in Communication Networks*, 2010, pp. 272–89.
- [10] D. H. Kim *et al.*, "Discovering Semantically Meaningful Places from Pervasive RF-Beacons," *11th Int'l. Conf. Ubiquitous Computing*, 2009, pp. 21–30.
- [11] S. Kuznetsov and E. Paulos, "Participatory Sensing in Public Spaces: Activating Urban Surfaces with Sensor Probes," *ACM Conf. Designing Interactive Systems*, 2010, pp. 21–30.
- [12] B. Longstaff, S. Reddy, and D. Estrin, "Improving Activity Classification for Health Applications on Mobile Devices Using Active and Semi-Supervised Learning," *4th Int'l. Conf. Pervasive Computing Technologies for Healthcare*, 2010, pp. 1–7.
- [13] N. Maisonneuve *et al.*, "NoiseTube: Measuring and Mapping Noise Pollution with Mobile Phones," *4th Int'l. ICSC Symp. Information Technologies in Environmental Engineering*, 2009, pp. 215–28.
- [14] E. Paulos, R. J. Honicky, and E. Goodman, "Sensing Atmosphere," *Sensing on Everyday Mobile Phones in Support of Participatory Research*, 2007, pp. 1–3.
- [15] J. Shi *et al.*, "PriSense: Privacy-Preserving Data Aggregation in People-Centric Urban Sensing Systems," *29th IEEE INFOCOM*, 2010, pp. 758–66.
- [16] K. Shilton, "Four Billion Little Brothers?: Privacy, Mobile Phones, and Ubiquitous Data Collection," *Commun. ACM*, vol. 52, no. 11, 2009, pp. 48–53.

### Biographies

EMILIANO DE CRISTOFARO (edc@parc.com) received his B.Sc. degree in computer science from the University of Salerno, Italy, and his Ph.D. degree in networked systems from the University of California, Irvine. Currently, he is a research scientist in the Security and Privacy group at the Palo Alto Research Center (PARC, a Xerox Company). His research interests include security, privacy, and applied cryptography. More information can be found at <http://www.emilianodc.com>.

CLAUDIO SORIENTE (claudio.soriente@inf.ethz.ch) received a Ph.D. in networked systems from the University of California, Irvine. He is currently a post-doctoral researcher at ETH Zurich, Switzerland. His research interests include privacy and distributed system security.

# Converged Access of IMS and Web Services: A Virtual Client Model

Salekul Islam, United International University, Dhaka, Bangladesh  
Jean-Charles Grégoire, Institut National de la Recherche Scientifique, Montreal, Canada

## Abstract

This article presents a virtual client-based converged access architecture for IMS and web services. A virtual client transfers most of the signaling and session maintenance loads to a remote server named surrogate, which implements the IMS client. We describe how we have implemented an IMS-web hybrid service, Movie-on-Demand (MoD), by deploying a simple IMS client and web server in a surrogate, and using an open-source implementation of a full IMS environment, from client to application server.

In spite of the continuing evolution of Internet-based services, up to the recent emergence of social networks, we still see a gap between media-oriented and data-oriented services, where media is most often embedded into a service platform in a way seldom compatible with other environments. We study here a way to achieve true converged service integration, which is close to the user and flexible, but with a limited impact on the user's computer platform. We further show how virtualization on the client side provides an interesting solution to these issues. Virtual clients, first as remote desktops but later as access to remote applications, have proven to be cost effective for corporate use (e.g., for employees in offices). They have only recently emerged in the general public as a trend toward the application as a service, running on a remote server, accessed through the Internet through a web browser. The browser essentially executes the user interface of the application (graphical user interface — GUI), and the user no longer has to worry about upgrades or software installation.

We present and study a virtual client architecture that integrates the IP multimedia subsystem (IMS) [1], a standard platform for media services, with web-based services. The virtual client keeps its simplicity by offloading signaling and session management tasks to a remote server we name a *surrogate*. Note that the term surrogate is also used in RFC 3040 to address a different type of network node. The surrogate implements the IMS client and accomplishes communication with the IMS core on behalf of the virtual client. The surrogate also deploys a web server to provide a web-based GUI to the virtual client. To illustrate the proposed converged access architecture, we implement a hybrid service of IMS and web, Movie-on-Demand (MoD), that uses an open source implementation of a full IMS environment, from client to application server (AS).

The rest of the article is organized as follows. After presenting a rationale for this work, we summarize early work on IMS clients and converged access with our critical comments. We present a virtual client-based convergence architecture; we illustrate the use of the architecture through a use case while we discuss the benefits of our model as well as a number of open issues. We conclude the article.

## A Rationale for Convergence

The universal move to IP networks as a unique telecommunications infrastructure has created an opportunity to integrate different services into more complex applications, such as universal messaging. Actually, we see that many “social” environments, even in professional settings, more and more often integrate various forms of media communications with some form of browsing or data exchange.

Such integration presents difficulties of various natures. Among those, ease of deployment and management of such network-based applications is of particular interest to us. As basic telecommunications services are typically structured in a client-server model, their integration can be done either at the client or at the server. In the first case, having the user discover and integrate these services as they emerge and evolve is a major issue. In the second case, the challenge is to identify forms of integration that will appeal to the largest number of users and provide them with a suitable interface.

Each approach has advantages and drawbacks. Downloading and upgrading software on the client's computer while keeping it well integrated with independent evolutions of the platform itself tends to be the most challenging task. This has led to an evolution of client software with the emergence of web browser-based interfaces (GUIs) to services, where most functions are executed on the server. Furthermore, it is difficult to define a unique target client platform in terms of hardware and software configurations and capabilities, so using the browser as the GUI has led to a simplification of, not to mention independence from, environments provided by specific manufacturers or platforms.

At the same time, interactive control of applications can be more difficult to achieve when the server is remote. Furthermore, interactive services tend to rely on the use of the Session Initiation Protocol (SIP) [2], which is not integrated in the web model, but rather interacts with an infrastructure such as IMS.

The IMS architecture is sketched in Fig. 1. The reader may recall that IMS is a SIP-based infrastructure created to allow integration of all multimedia services in a single, unique

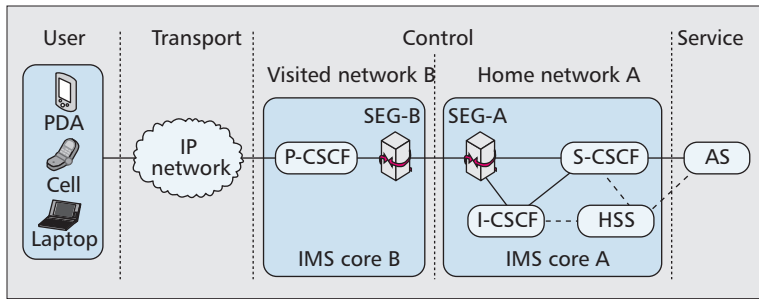


Figure 1. Architecture of next-generation networking (NGN) IMS.

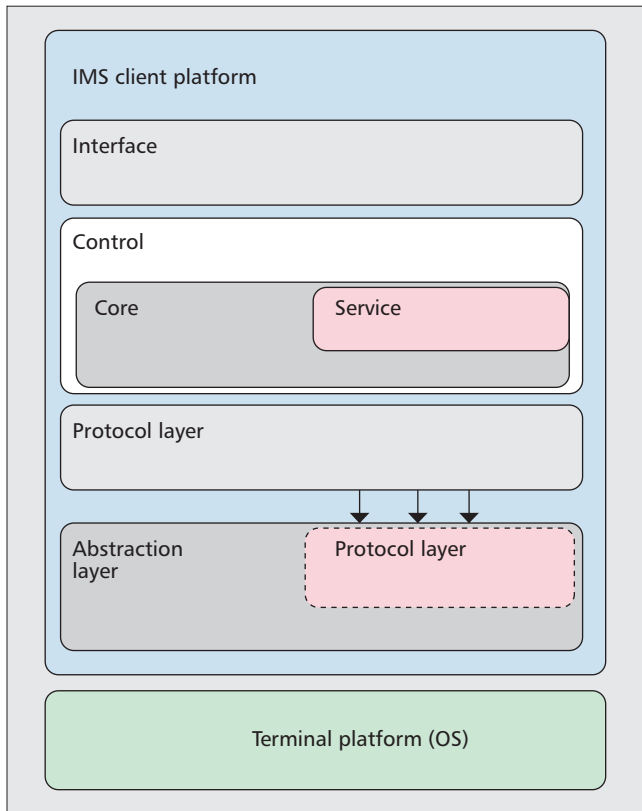


Figure 2. Architecture for an IMS client.

framework also supporting, among other features, authentication and roaming. It supports user-to-user or user-to-network services, with the support of ASes. Its core functional components for call control include different specialized call session control functions (CSCFs): the proxy CSCF (P-CSCF), the interrogating CSCF (I-CSCF), and the serving CSCF (S-CSCF). All can be considered extensions of SIP proxies. The P-CSCF is the first proxy encountered by user requests. They are forwarded to the signaling plane's central switching node, the S-CSCF, which can process requests based on the user profile. When roaming, requests go through a P-CSCF in the visiting network to the S-CSCF in the home network through a special relay located at the edge of the home network, the I-CSCF. The home subscriber server (HSS) is the master user database, and supports the IMS network entities to handle calls and service sessions. The reader should refer to [1] for further details.

Convergence therefore goes beyond deciding where the application code is executed but also, and more fundamentally, how different yet complementary infrastructures, such as IMS and the web, can be simultaneously accessed and com-

bined into new applications, in the most dynamic and flexible way. We propose that, to support the convergence of IMS and web models, we need a middle way between client-based, typically media applications, and server-based, rather data/messaging-oriented applications.

### Earlier Work

The need for convergence of web-based and SIP-based services is a strong requirement because, in essence, media communications do not exist in isolation. A communication can indeed be associated with another service (e.g., eCommerce), require browsing (content, contact, etc.), or be embedded in a more generic application (e.g., universal messaging). How to achieve such convergence is complex because of the use of two protocols based on very different premises: SIP usage is *stateful*, whereas the web, based on the Hypertext Transfer Protocol (HTTP) [3], is *stateless*.

The essence of this difference is beyond the scope of this article. Suffice it to say that SIP clients need to perform complex operations and processing, whereas HTTP clients perform only simple requests — but potentially rather complex processing of the data received; SIP is peer-oriented, whereas HTTP is client-server oriented. Integration of such different models for convergence therefore presents many challenges. Furthermore, the term *convergence* is multidimensional: it could be converged access, converged service, or even converged signaling of SIP and web domains. Different approaches have been tried.

### Client Side

On the client side, applications embed communications with both infrastructures and give a vision of uniformity to the user. We have already discussed the limits of such a model: applications are complex and inflexible, and are difficult to install, update, and customize. Most media clients remain application-based, as opposed to web-based, especially when interactive communications are involved. As a witness to this challenge, we can note that few general-purpose clients are available for IMS.

To illustrate our point further, let us consider proposed IMS client frameworks, which are the focus of many activities, and their application. Eurescom has designed an open and extensible IMS client framework in the P1656 project,<sup>1</sup> and there are other IMS client platforms, often based on the Java Community Process JSR-281 (e.g. [4]). Such an IMS client framework presents a modular, expandable, multilayered architecture, which is shown in Fig. 2.

This layering of the IMS client framework (shown in Fig. 2) results in highly modular and extensible IMS client architectures. Using the proper application programming interfaces (APIs) (e.g., JSR-281 APIs), an application can be assembled using the underlying IMS services, and the otherwise usual one-to-one mapping between an application and IMS service session can be overcome. The operators can respond to market dynamics and quickly bring out new applications. Despite these advantages, the existing models still have severe drawbacks, which are due to the implementation of the entire client framework inside the user equipment (UE). Decoupling the UE — the device used by the user — from the real IMS client is an important feature of our model.

<sup>1</sup> <http://www.eurescom.eu/public/projects/P1600-series/p1656/>



---

## Hybrid Attempts at Convergence

Moving the point of convergence away from the client implies that it will interface with a unique infrastructure. The alternative to SIP, on the web side, is to use web services, a technology created to support transactional services over the web. The Simple Object Access Protocol (SOAP) was developed by the W3 consortium to support web services, and, as such, it has acted as a focal point to study integration issues with SIP.

One approach proposes a SOAP $\leftrightarrow$ SIP gateway (GW) [5] that receives SOAP messages (over HTTP) from the user equipment (UE) and generates corresponding SIP messages for UE that has not implemented a SIP client and vice versa. It allows SIP services to be invoked for UE that has not implemented a SOAP client.

A converged service is a hybrid service, which is a mash-up of IMS and web services. The WIMS 2.0 [6] project,<sup>2</sup> named from the integration of Web 2.0 and IMS, which is based on SIP, offers hybrid SIP services by exposing the IMS capabilities through open web APIs. One example of such a hybrid is Movistar Contacta,<sup>3</sup> a Facebook application that allows a Facebook user to send SMS and invoke click-to-call with the user's Facebook friends.

Another way to achieve convergence is through a form of combined signaling that implements both SOAP and SIP clients at the UE and uses SIP and SOAP in parallel, where SIP is used for signaling in the control plane and SOAP is responsible for data transmission in the user plane [7]. A unique web service ID is agreed on between the two user terminals through the Service Description Protocol (SDP) during session establishment.

### Criticism

The existing convergence models illustrated above fail to treat both SIP-based and web services as equals. In many cases, one core protocol — either SIP or SOAP — is tunneled through the other, and hence loses some of its functionality. Hence, opportunities in the distribution of implementation and computing loads between the user end (i.e., UE), intermediate nodes, and the terminating network end are not exploited. Designing a converged control plane for SIP and SOAP messages without resolving the issue of bulk media delivery (e.g., audio/video streaming) under a SOAP foundation is only a partial solution to the convergence problem. While vying to create a converged access architecture, we must follow a balanced way and preserve the features of each infrastructure.

To conclude, the challenge we address here is to find a way to achieve service integration closer to the user, in a flexible way but with a limited impact on the user platform, which excludes the client-based solution, as discussed above. At the same time we do not want to impose undue restrictions on the features offered by either infrastructure.

## A Convergence Architecture

SIP has not yet experienced the same popularity HTTP has, and there is no equivalent of the browser to easily deploy SIP-based services. Reliance on a stateful signaling protocol appears to be a hindrance in that respect. While the protocol itself is stable, its use is constantly expanding, which may

result in changes in fields in headers and their interpretation as new uses emerge. All of this could require frequent updates to clients, which is problematic, and quite certainly so for small devices (e.g., phones and PDAs). Moving to an HTTP-controlled virtual client appears to us to be a realistic solution for these issues, with minimal cost.

Unlike earlier solutions presented in the literature, which appear to have been based on the extension of an existing software base, we propose here a “greenfield” approach, independent of any specific technology. The virtual client-based convergence architecture we propose is shown in Fig. 3. In this architecture, most of the signaling load is transferred to a server collocated with a P-CSCF, which we shall call a *surrogate*. The surrogate acts as a virtual server for the user to access, organize, provision, and monitor her SIP-based services.

### Accessing Web-Based Services

Since the UE is assumed to be equipped with a standard web browser, accessing web-based services that are not related to any IMS service is straightforward. However, accessing media content needs the support of corresponding audio/video codecs and also the software to play back the media that might even require proprietary software (e.g., flash player). However, the ongoing specification of HTML5,<sup>4</sup> the latest standard revision of the *lingua franca* of the web, supports a large number of new tags, including “audio” and “video” type tags, which have increasingly become critical elements of web content. Hence, audio and video contents could eventually be delivered directly through the web browser (not all web browsers support HTML5 at this moment) without any external, proprietary player.

### Accessing IMS Services

Accessing IMS services requires instantiating the IMS client installed within the surrogate. The surrogate presents an IMS client to the core, acting as a server side of the virtualized client for the user. However, the surrogate is transparent for the IMS operations in the network and has no impact on the IMS core architecture. The surrogate implements a web server, which receives the users' input through the GUI running on the web client inside the UE. In this model, any end-user device (e.g., mobile device, laptop, PC, IP phone) with IP connectivity could be used as UE. This is clearly an advantage over the present IMS where an end-user device can be used as UE only if the IMS client can be installed on it. A middle layer is needed between the web server and the IMS client to establish communications between them. This layer transfers the GUI's input to the IMS client applications and IMS session status (e.g., IMS registration success) to the web server. The IMS client framework inside the surrogate has a different top layer protocol stack from the generic client framework of Fig. 2. Since the user interface layer is an IMS application layer already implemented in the UE (through the web-based GUI), an IMS applications layer must be implemented to expose different IMS applications to the web server. As shown in Fig. 3, the UE communicates with the surrogate using HTTP, while the surrogate communicates with the IMS core using SIP messages. Accessing IMS services is also related to UE authentication by the IMS core and how media would be delivered to the user's platform.

*Authentication* — Each IMS subscription is associated with an IP multimedia private identity (IMPI) and one or more IP multimedia public identities (IMPUs). The identity is established through an authentication process based on an application, the IP multimedia SIM (ISIM) [8], which runs on the

---

<sup>2</sup> <http://www.wims20.org>

<sup>3</sup> <http://www.facebook.com/MContacta>

<sup>4</sup> <http://dev.w3.org/html5/spec/Overview.html>

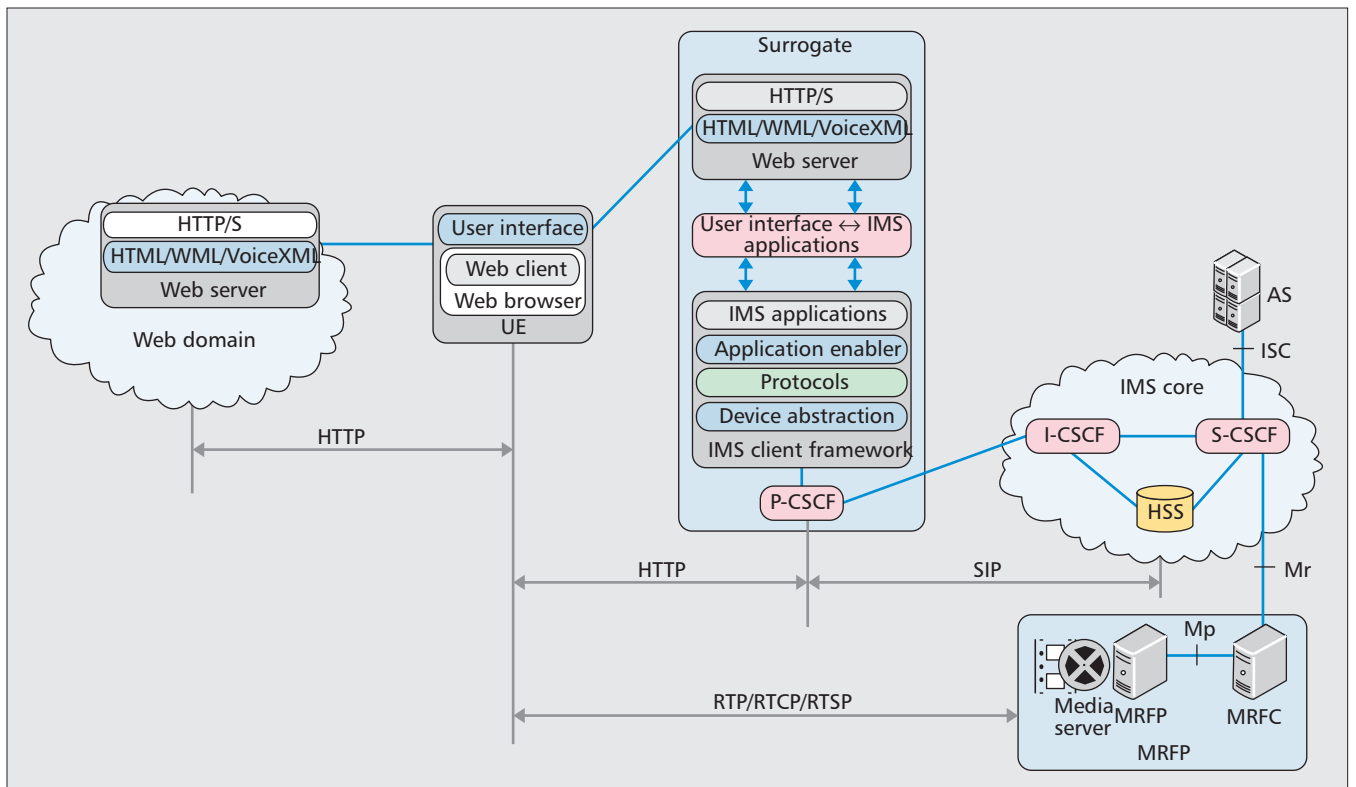


Figure 3. Virtual IMS client-based converged access architecture.

universal integrated circuit card (UICC) smartcard present on any terminal. This UICC smartcard securely stores the IMS subscriber's identity and credentials. In our proposed architecture, however, users can switch terminals and connect to the IMS core network from anywhere they want. If the user wishes to use any terminal device, it is not possible to deploy IMS information on the user's terminal device. An ISIM application depends on the UICC smartcard for hosting the application, which — by its very essence — could not be deployed inside the surrogate. Hence, an IMS soft client that instantiates a virtual ISIM application is required for our architecture. The soft client the surrogate deploys needs to perform many more tasks than a state-of-the-art IMS client (installed in a PC or laptop) does. For example, multiple users from heterogeneous devices and networks communicate with this IMS client through the web server. The IMS client must instantiate virtual ISIMs as required and maintain separate states for each user.

*Media Delivery* — Although Fig. 3 shows straight-through media delivery where media is directly delivered to the user's platform, media flow could be intercepted by the surrogate for additional processing before delivering to the user's platform. Alternative media delivery methods present specific benefits, and each service can choose the more suitable one, transparent to the user and the IMS core, since the required SIP processing is performed by the surrogate itself. If the UE does not have the appropriate audio/video codecs to decode the received media, the surrogate may intercept the media and transcode the media to another format to be understandable by the user's platform.

#### Building Web-IMS Hybrid Services

The proposed converged architecture could be used for delivering merged web-IMS hybrid services, such as the WIMS 2.0 Movistar Contacta service mentioned previously. In a hybrid service, a user may request IMS services (e.g., click-to-call)

while accessing interactive content-rich web sites. We need a convergence of web server and IMS client for building such next-generation services. The surrogate provides exactly the same platform. Next, we discuss how we have implemented a tentative hybrid service, named Movie-on-Demand (MoD), which is accessed by the end users through a web site and the implicit establishment of an IMS session.

#### Proof-of-Concept: Movie-on-Demand (MOD)

The MoD service is assumed to be a third party service, which is hosted by a service provider other than the IMS core network provider. Using the MoD service, a successfully registered user (with proper subscription to the MoD service) will be allowed to request a movie from a list of available movies. The movie will be delivered to the UE through video streaming in a video-on-demand fashion.

#### Implementation

Figure 4 shows the prototype implementation architecture for the MoD service, which has been accomplished by following the converged access architecture shown in Fig. 3. First, a specific domain, mist.org, has been created, and all communicating entities, including the virtual client, the surrogate, the IMS core, the AS, and the media server (Darwin Streaming Server) have been deployed in different machines inside this domain. The surrogate hosts the IMS client, and the UE hosts the virtual client. By virtual client we are referring to an end-user machine with Internet connectivity, able to run a standard browser and equipped with necessary hardware to receive multimedia data. Therefore, we are using UE and virtual client interchangeably in the rest of the article.

For communicating with the UE using HTTP, we have installed an Apache web server in the surrogate. The web server hosts a PHP programme which provides a GUI to the user and establishes communications between the UE and the IMS client. The middle layer that establishes communications

between the Web server (PHP program) and the IMS client is implemented using classical network socket APIs. We have implemented a simple IMS client using the C eXtended osip (eXosip) library. Our IMS client performs IMS registration to the IMS core, generates and forwards the SIP INVITE message in response to the user request for a specific movie. It receives a URL for the media server and returns it to the UE (through the web interface) to open an RTSP session.

We have used Open IMS Core,<sup>5</sup> an open source IMS core implementation developed by the Fraunhofer Institute for Open Communication Systems (FOKUS). The Open IMS Core implements all three CSCFs and a lightweight HSS, which all together are the core elements of IMS or next generation network (NGN) architecture.

In the application plane, we have deployed the UCT Advanced IPTV AS<sup>6</sup> as the application server. This program has been developed by the Communications Research Group at the University of Cape Town as a standard implementation of an IMS-based IPTV service. In the media plane, the Darwin Streaming Server (DSS),<sup>7</sup> the open version of the Apple QuickTime Streaming Server, has been installed as a media server. The movie clips that our MoD service offers to a subscribed user are being stored and streamed (on request) by this streaming server.

### Message Sequence

The MoD message sequence is shown in Fig. 5. In our implementation, we assume that all links are secured. Therefore, we do not deploy any user authentication beyond IMS registration or access control. The only access control we perform is to verify that only a registered user can request a movie clip. A trusted link between two communicating entities (e.g. IMS client and P-CSCF) could be established through *z*<sub>a</sub> (for interdomain communication) or *z*<sub>b</sub> (for intradomain communication) interface following the Network Domain Security (NDS)/IP specification [9]. It is also assumed that the IMS client is aware of the available resources (e.g., A/V codecs, audio and graphics processing power) at the UE, and hence no Session Description Protocol (SDP) [10] exchange is required between the UE and the IMS client.

First, the user accesses the web site hosted by the surrogate's web server. Next, the user sends a registration request to the IMS client to trigger IMS registration. On successful registration, the IMS client sends a Registration Success message to the user. The UE (virtual client) only triggers the IMS registration and never maintains any IMS registration related information. This issue is further elaborated on in the following section.

Next, the user selects the movie clip she wants to receive,

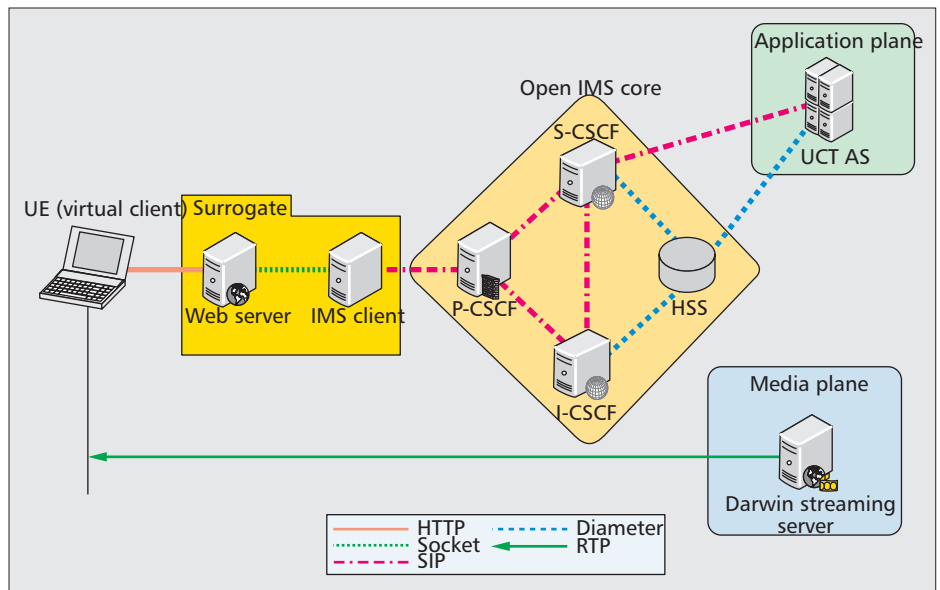


Figure 4. Implementation architecture for Movie-on-Demand service.

and a corresponding Movie Request is sent to the IMS client through the web server. The IMS client sends an INVITE message to the IMS core addressed to the URI of the movie (e.g., sip:movie1@mod.mist.org, where movie1 is the identity of the movie and mod.mist.org is the domain name of the UCT-AS) and the initial SDP offer.

The IMS core performs service control, that is, checks the registration status of the IMS client and verifies if the client is authorized to originate such an INVITE message. The HSS database must be properly provisioned with initial filter criteria so that the S-CSCF, upon successful service control, forwards the INVITE message with mod.mist.org to the UCT-AS.

The UCT-AS requests of the DSS the URL of the movie clip. Using the identity of the movie (i.e., movie1), the DSS retrieves the URL (e.g., rtsp://dss.mist.org/movie1.mp4) and returns it back to the UCT-AS. The communication between the UCTAS and the DSS is not shown in Fig. 5 since it has been implemented through proprietary methods. Next, the URL is forwarded inside a 200 OK message by the UCT-AS to the IMS core. Although it was not done in our implementation, the DSS may send a session-specific, randomly generated, unique URL (which is not advertised to the outside world) for authenticating the user during the media session (in step 23 of Fig. 5).

Upon successful authorization, the IMS core forwards the 200 OK response to the IMS client, which forwards the movie URL to the UE (through the web interface program). Finally, a media session is initiated by the VLC media player installed in the UE with the DSS, and the requested movie clip is delivered through an RTSP session.

Finally, note that although the initial SDP offer is sent by the IMS client to the AS, this SDP offer has no effect on our implementation (and hence no end-to-end resource reservation is required) due to the use of RTSP for MoD streaming. This is further discussed later.

## Discussion

### Benefits

*Beyond IMS and Web Services* — The end user benefits from converged access of all popular services from a single client. Our modular design makes provision for future extensions by focusing the addition of new applications at the surrogate. For

<sup>5</sup> <http://www.openimscore.org/>

<sup>6</sup> [http://uctimsclient.berlios.de/uctiptv\\_advanced\\_howto.html](http://uctimsclient.berlios.de/uctiptv_advanced_howto.html)

<sup>7</sup> <http://dss.macosforge.org/>

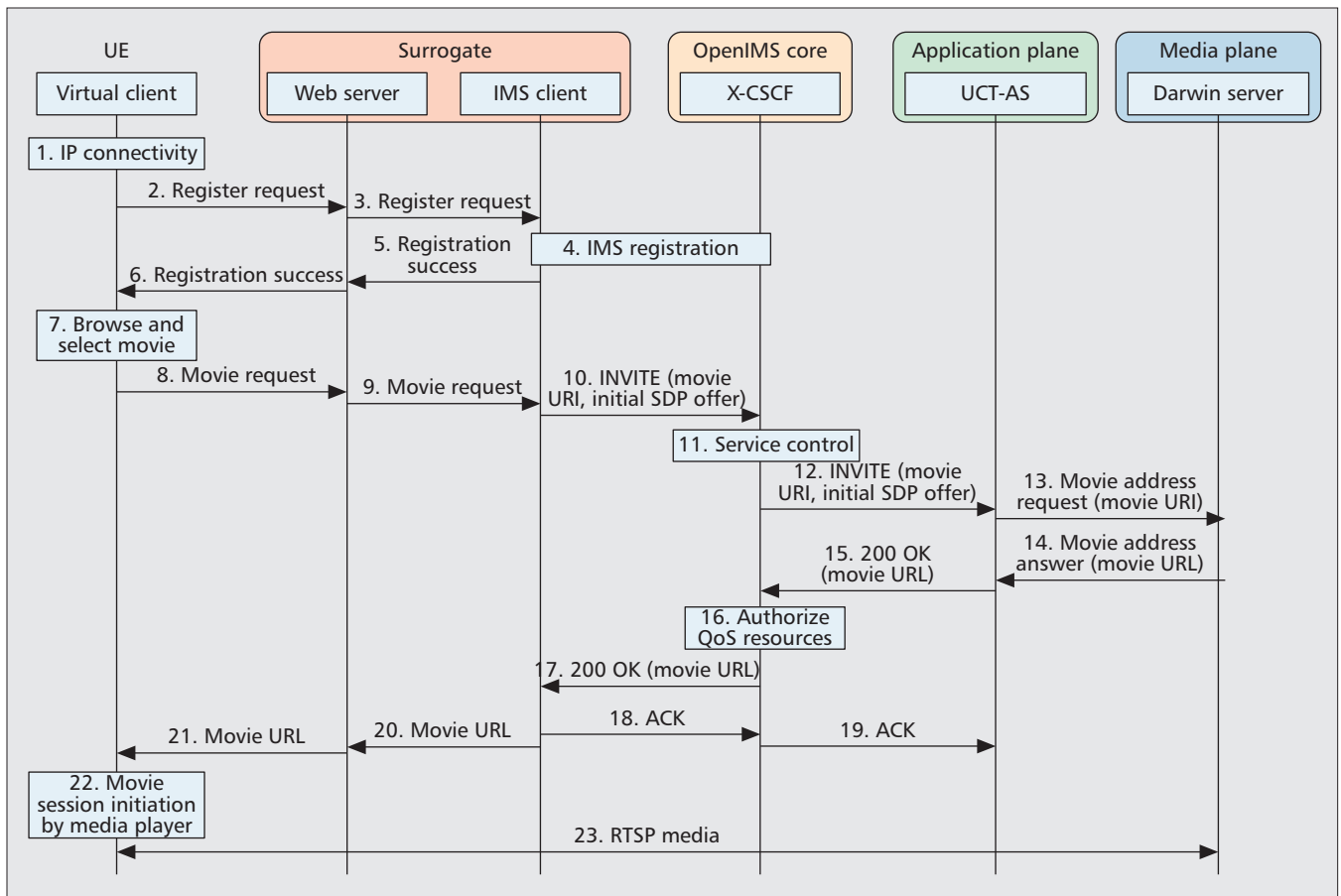


Figure 5. Message sequence for Movie-on-Demand service.

example, a peer-to-peer (P2P) client could be implemented inside the surrogate, and the received data could be downloaded directly to the user's device.

*Lightweight UE* — The UE is free from the burden of supporting IMS clients even though it enjoys all the benefits of IMS services. This also applies to future extensions of our model to other services, where most of the complexity (e.g., implementing an IMS client) is transferred to the surrogate.

*Network Access Mode Agnosticism* — IMS relies on the ISIM application for completing registration through IMS authentication and key agreement (AKA). The proposed solution works without the ISIM and independent of the network access method. The only network connectivity required is the reliable transmission of HTTP messages between the UE and the surrogate.

*Multiplexing IMS Registrations* — IMS supports multiple IMPUs attached to a single IMPI. An IMPU is assigned by the home network operator. IMS also supports an IMPU-specific service profile, which is a collection of service and user related data. Moreover, a globally routable user agent URI (GRUU) identity could be used to identify a unique combination of IMPU and UE instance that allows a SIP request to be addressed to a specific combination of IMPU and UE. These features will create an opportunity to share a single IMS registration among multiple users.

#### Limitations and Open Issues

IMS decouples the control plane from the media plane. Following that principle, our proposed model extends only the control plane and thus should be applicable for any IMS communication services including presence, voice, conferencing,

video streaming, mobile gaming, and so on. Although the IMS client is now split over the UE and the surrogate, the latter is expected to be located in close proximity to the user, most likely within the domain of the operator that is providing Internet connectivity to the user. Therefore, even for a highly interactive and delay-constrained service (e.g., mobile gaming through the IMS platform), the introduction of the surrogate should not unduly increase the latency due to the control messages. The latency of media delivery will be unaffected, while media is not manipulated by the surrogate. However, in case of the interception of media delivery (e.g., if the media is transcoded by the surrogate), a delay-constrained application may suffer from decreased quality of experience (QoE).

The proposed architecture presented in Fig. 4 outlines a high-level solution of the convergence problem. Several further concerns should be addressed for the development of a successful solution; we now discuss them.

*Implementation of the Surrogate* — The surrogate is the focal point of computing and processing in our solution. The surrogate will have to maintain hundreds of sessions, connections, and state information from different users. In grid computing, load balancing is a common technique to improve performance of remote servers, and such infrastructure would lend itself to surrogate support. The surrogates could be implemented by any party: the IMS core network provider, the Internet service provider, the administrator of a corporate network, or even an independent operator. With the advent of cloud computing, the surrogate can be deployed inside the cloud infrastructure [11]. The surrogate must implement the required services for IMS signaling, including the IMS client, the GUI for the IMS client, and the SDP negotiator. Additional services that intercept the media might also need to be implemented. Following

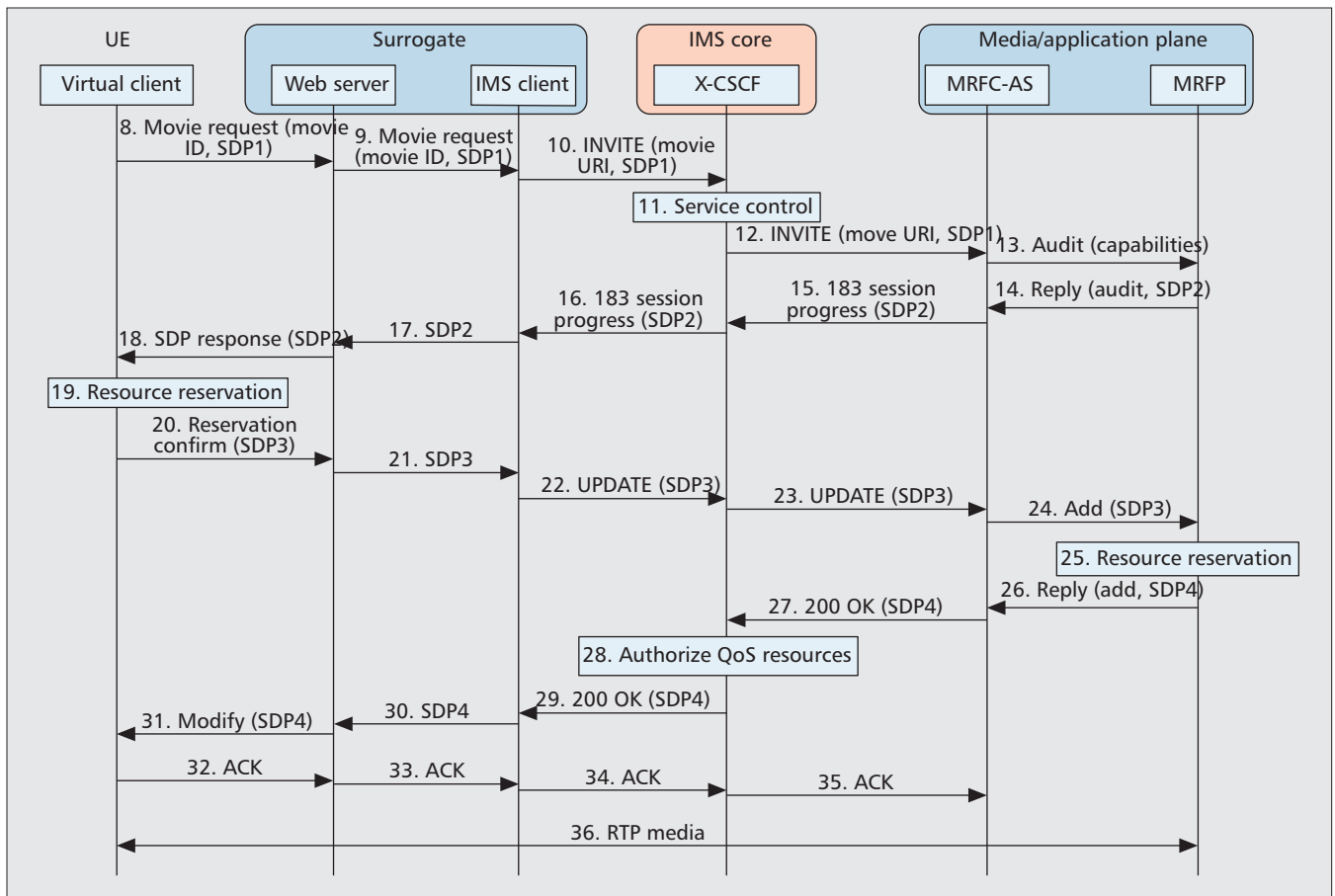


Figure 6. Message sequence for Movie-on-Demand service with the AS acting as a media controller.

the software-as-a-service (SaaS) model, if multitenant services are implemented, a single implementation could be reused for different IMS communication services.

*Authentication and Access Control* — The HTTP communication between the UE and the web server must be secured using the HTTPS (i.e., HTTP with Secure Socket Layer [SSL]/Transport Layer Security [TLS]) protocol. While the end user controls the IMS client through the web interface, the end user identity authentication and access control must be enforced by the web server before providing her access to any secured web interface. Moreover, should the link between the web server and the IMS client need to be secured (e.g., they are implemented in different domains), SSL/TLS should be deployed.

*End-to-End Media Negotiation and Resource Reservation* — Since the IMS client and the UE are separate entities, the IMS client negotiates media on behalf of the UE. If RTP media transport (instead of RTSP media) is needed, end-to-end SDP offer/response-based media negotiation and resource reservation are performed by the UE and the multimedia resource function controller AS (MRFC-AS). The AS is assumed to be collocated with the MRFC. Figure 6 shows a similar partial message sequence for our MoD service. Note that steps 1–7 would be same as in Fig. 5. The UE sends the initial SDP offer, SDP1 (either embedded as an S/MIME message or through TLS-protected HTTP [10]) to the web server, which forwards SDP1 to the IMS client. SDP1 carries the port number (which the UE has allocated) and other required information to receive the RTP media. The MRFC and the media resource function processor (MRFP) can communicate using Media Gateway Control Protocol (MEGACO)

or some other media control protocol [12]. The MRFP sends its capabilities through an SDP response (SDP2), which is forwarded to the UE. The UE and the MRFC exchange another round of SDP offer/response if resource reservation is supported, and finally reserves necessary resources. On completion of the SDP offer/response exchange, the MRFP initiates a media session with the UE, and the requested movie clip is delivered through an RTP session.

*Roaming* — In the proposed model, support for roaming depends on the implementation of surrogates. If we assume that surrogates are distributed over different networks and that a trust relationship exists among them, a user device could, while roaming, communicate with the surrogate located in the visited network, and this surrogate will in turn communicate with the P-CSCF of the visited network. The visited network's surrogate may then communicate with the home network's surrogate to pull any user information if required. If no surrogate is available in the visited network, the user device should communicate with the home network's surrogate. However, since the home network's surrogate returns back to the visited network's P-CSCF, this method would introduce delays in the exchange of signaling messages.

It is clear, however, that while keeping in the spirit of the IMS architecture, these considerations take us further away from a strict IMS model.

## Conclusion

This model is a perfect match for the emerging generation of portable devices, with readily available Internet connectivity through multiple wireless network interfaces and more limited

---

compute and memory resources than traditional home/portable computers. The use of standard interfaces to access services, rather than having a multiplicity of different components to manage, certainly matches the philosophy of use of these devices.

It also exposes new trade-offs between the breakdown of application processing, bandwidth consumption, and latency. Adding a processing step can lead to increased latency, which is detrimental for interactive (mostly voice) services. On the other hand, streaming or other services that are more delay-tolerant should not be inconvenienced. The exploration of these trade-offs is the current focus of our research.

### References

- [1] 3GPP, "Technical Specification Group Services and System Aspects; IP Multimedia Subsystem (IMS), Stage 2," TS 23.228 V11.4.0, Mar. 2012.
- [2] J. Rosenberg *et al.*, "SIP: Session Initiation Protocol," RFC 3261, June 2002.
- [3] R. Fielding *et al.*, "Hypertext Transfer Protocol – HTTP/1.1," RFC 2616, June 1999.
- [4] P. Kessler, "Ericsson IMS Client Platform," *Ericsson Review*, vol. 2, 2007, pp. 50–59.
- [5] R. Levenshteyn, and I. Fikouras, "Mobile Services Interworking for IMS and XML Web Services," *IEEE Commun. Mag.*, Sept. 2006, pp. 80–87.
- [6] D. Lozano, L. A. Galindo, and L. Garcia, "WIMS 2.0: Converging IMS and Web 2.0. Designing REST APIs for the Exposure of Session-Based IMS Capabilities," *Proc. 2nd Int'l. Conf. Next Generation Mobile Applications, Services, and Technologies*, 2008, pp. 18–24.

- [7] G. Gehlen *et al.*, "Mobile P2P Web Services Using SIP," *Mobile Information Systems*, vol. 3, 2007, pp. 165–85.
- [8] 3GPP, "Technical Specification Group Core Network and Terminals; Characteristics of the IP Multimedia Services Identity Module (ISIM) Application," TS 31.103 V10.1.0, Apr. 2011.
- [9] 3GPP, "Technical Specification Group Services and System Aspects; 3G Security; Network Domain Security; IP Network Layer Security," TS 33.210 V11.3.0, Dec. 2011.
- [10] M. Handley *et al.*, "SDP: Session Description Protocol," RFC 4566, July 2006.
- [11] S. Islam, and J.-Ch. Grégoire, "Network Edge Intelligence for the Emerging Next-Generation Internet," *Future Internet*, vol. 2, no. 4, pp. 603–623, 2010.
- [12] J.-Ch. Grégoire, and A. Jukan, "On the Support of Media Functions within the IMS," *IP Multimedia Subsystem (IMS) Handbook*, M. Ilyas and S.A. Ahson, Eds., CRC Press, 2008.

### Biographies

SALEKUL ISLAM (salekul@cse.uju.ac.bd) is an assistant professor at United International University, Bangladesh. He worked as an FQRNT postdoctoral fellow at INRS, a constituent of the Université du Québec. He has a Bachelor's degree from Bangladesh University of Engineering and Technology in computer science and engineering, and Master's and Ph.D. degrees, both in computer science, from Concordia University, Canada. His research interests are in the design, analysis, and validation of protocols for telecommunication networks and secure multicast.

JEAN-CHARLES GRÉGOIRE (gregoire@emt.inrs.ca) is an associate professor at INRS, a constituent of the Université du Québec with a focus on research and education at the Master's and Ph.D. levels. His research interests cover all aspects of telecommunication systems engineering, including protocols, distributed systems, network design and performance analysis, and, more recently, security. He also has made significant contributions in the area of formal methods.

---

# On the Analysis of Hierarchical Autonomic Control of Multiparty Services

**Nuno Coutinho and Susana Sargento, Instituto de Telecomunicações, Aveiro, Portugal  
University of Aveiro**

---

## Abstract

The increasing interest in group-based multimedia services, followed by the larger resource demands and the quest for seamless mobility support, have been propelling research on novel approaches capable of overcoming the challenges posed by future networking environments. One of those challenges is heterogeneity, which can also be leveraged in favor of more enriched services through context awareness and thus enhance user service perception. In this article we describe a context-driven framework for multiparty content delivery and discuss the rewards of employing the abstract multiparty transport concept, which provides autonomic control of personalized group-based services to users through a hierarchical strategy. Since scalability is the major concern when dealing with group-based services, we evaluated the framework and its embedded concepts regarding this feature. Herein, we describe an analytical study focused on quantifying the necessary reconfigurations in the network due to any type of context change. The outcomes of this study show that through the concept of abstract multiparty trees, we obtain considerable gains regarding link savings and consequent network control operations, thus increasing the scalability of the autonomic control architecture.

---

**H**eterogeneity is seen as a prevailing feature in future networks, bringing together the different layers and elements of a communication environment (e.g., network devices, protocols, applications, access and transport technologies, users, and the environment). This diversity, despite being a challenge when designing future network architectures, can also be seen as an opportunity to differentiate services. This added value in future communications is accomplished by employing context awareness, enabling networks to evolve toward higher levels of service personalization, taking advantage of any kind of information that may affect the service perception.

This way, IPTV, Internet video, and other multiparty services can be enriched by employing context, providing both personalized and context-aware group-based communications. In this approach, groups can be dynamically created according to the selected context, and both network and services can be adapted to these groups of users.

Due to the volatile characteristics of users, applications, and environment, context-aware architectures are inherently dynamic, and their control and management needs to be performed in an autonomous way. Some architectures in the literature already support this autonomic behavior [1], although not addressing multiparty support. This article analyzes a hierarchical architecture that employs the concept of abstract multiparty trees (AMTs) to control multiparty content distribution in an autonomous way, decentralizing the control and management functions. The AMT principle consists of allowing end-to-end multicast content transport over network segments with different transport technologies (i.e., unicast and multicast), and also providing independence between source

and listener trees (through different AMTs), and seamless resilience support. Central elements only need to control the edges of the AMTs, and all the adaptations and reconfigurations inside the AMTs are performed through these edge nodes, the overlay nodes, which implement proxy functionalities to mediate overlay connections and control the quality of service (QoS) of the multiparty trees independent of the others. The end-to-end path is therefore performed through several independent multiparty trees. This approach increases the flexibility of the distribution trees as well as hides the network dynamics and heterogeneity of the networks.

This article assesses the efficiency of the autonomous hierarchical control approach through the overlay concept, presented in [2], which shows how the intelligence introduced in the network management framework is able to use the context information available toward the provision of more personalized services while handling network resources efficiently. The evaluation study measures the reconfiguration savings by employing abstract trees compared to a more traditional multicast approach. The results show that the AMT-based approach significantly decreases the required network reconfigurations due to context changes, highlighting the scalability of the abstract trees for a large number of users.

The remainder of the article is organized as follows. We provide an overview of the related work on multicast, application layer multicast, and context-aware and autonomic architectures. We describe the abstract multiparty transport architecture and control features. An analytical model of the AMT behavior is detailed. We present a set of experiments and results obtained through simulation, and conclude the article.

---

## Related Work

The multicast transport concept is, theoretically, the most suitable for content delivery to a large number of users (group-based services). However, despite the advantages, IP multicast has seen slow commercial adoption given several deployment issues [3]. Its major drawback is the costly routing, involving much more information than unicast routing, which impairs scalability. This drawback has been fuelling the research for more scalable solutions, where strategies based on incremental deployment have gained some prominence, as is the case of application layer multicast (ALM) [4]. In ALM systems, multicast is implemented at the application layer, which does not require an upgrade of the network infrastructure. This way, ALM is a more scalable solution since routers do not need to maintain per-group state. However, the content delivery trees created by ALM are non-optimal, which leads to longer latencies and possible packet duplication on the same link.

Overlay multicast [5] is another incremental deployment strategy, where multicast functionalities can also be supported by constructing a backbone overlay of intermediate proxies, which create multicast trees among themselves. End hosts may communicate with these proxies via unicast or multicast. In order to combine the best of both strategies (overlay and native multicast), its main idea is to achieve a solution that benefits from the incremental deployment of the overlay concept and also the delivery efficiency of IP multicast.

Regarding multicast efficiency, there are several studies in the literature that provide a quantitative analysis of the network load reduction achieved by this technology. The groundwork on this topic was performed by Chuang and Sirbu [6], calculating the number of links  $L$  in a multicast delivery tree that connect the source to  $m$  distinct network locations. The simulations performed for a wide panoply of networks (real and generated) show that the efficiency gains can be reasonably described by  $L(m) \propto m^{0.8}$ . However, later studies show that only for a small or moderate number of members is the Chuang-Sirbu law a fair approximation.

Considering that nowadays an efficient content distribution depends on several factors, such as the transport technology, the information and control involved, and the possibility to enrich user experience, we support the idea that context awareness in multicast services is a key feature for next generation network (NGN) architectures. Given the importance and novelty of the concept, there has been considerable research effort on the development of context-aware mechanisms [7]. In order to enable networks with this consciousness, three main mechanisms are required: gathering, modeling, and reasoning. Considering the complexity added by these mechanisms, recent works are more focused on distributed solutions employing nature-inspired schemes to improve context dissemination and availability in dynamic scenarios [8]. Moreover, since the collected information is usually in a raw state, ontology-based solutions have gained prominence in context modeling, since the formatted information can then easily be used as input of the more evolved reasoning mechanisms. Regarding this topic, there have been great advances given the development of increasingly complex algorithms able to compute context-based decisions even with uncertain information [9].

Context-aware reasoning can be part of a broader concept, autonomic network management, which is commonly seen as the best approach to tackle the increasing complexity of present and future network management architectures. This management strategy is based on endowing the network elements with self-properties (configuring, healing, optimizing, and pro-

tecting), enabling a cooperative management strategy that is capable of autonomously reacting to context changes. A thorough study about autonomic network architectures is presented in [1]. These architectures, however, do not support multicast services. In [2, 10], we presented the concept of abstract multiparty trees for multicast-based services to develop a hierarchical autonomic control architecture, where the distributed control nodes are chosen based on context information to optimize the user's quality of experience. This architecture is described in the following section.

## Multiparty Architecture Overview

The aim of the multiparty architecture is to enable context-aware content delivery in heterogeneous networking environments, mainly concerning group-based services. Given the dynamics imposed by architectures sensitive to context changes and events (users' mobility, services requirements), it is required to develop control and management frameworks that can cope with this added complexity. The focus of this article relies on the assessment of the scalability and flexibility of the framework, considering a wide set of group members in a constantly changing environment.

The control and management framework, called the multiparty transport framework, is part of a more general architecture that includes two more frameworks: multiparty session management, and context detection and distribution [11]. The context framework acquires the context related to network elements and surroundings through distributed sensors that feed a central repository named the context broker (CB). The gathered information can then be retrieved to the other frameworks that use it in management and control procedures. Multiparty session management is responsible for creating a session's context, dynamically building groups according to members' context, codecs used, and QoS requirements.

The integration of these three frameworks enables more personalized end-to-end communication through a novel multiparty transport approach that uses context toward better content delivery. In order to efficiently enforce in the network the decisions triggered by context or group member events, the multiparty transport framework employs a hierarchical control strategy, which makes use of an overlay network so that decisions can be scaled. In the following sections we describe the multiparty transport framework, and its concepts and mechanisms.

## Multiparty Transport Framework

As previously stated, the multiparty transport framework aims to enable effective content delivery in heterogeneous networking, and also enhance the service provided by taking advantage of the context diversity of such environments. By including more information in the control processes, complexity increases, and consequently scalability issues cannot be neglected. The main goal of such a framework is to evolve the control mechanisms to react in a scalable way to any context change without impairing the service quality. For this purpose, a hierarchical autonomic control model was adopted, considering the scalability concerns that usually come along with context-sensitive architectures. The strategy employed consists of two levels of intelligence (Fig. 1): the higher level is centralized on an entity (which can be replicated for fault tolerance), named network use management (NUM), and the lower level is distributed throughout an overlay network.

The NUM enforces the selected content distribution tree in the substrate network, which is made by the second level of control that consists of a network of overlay nodes (ONs) that



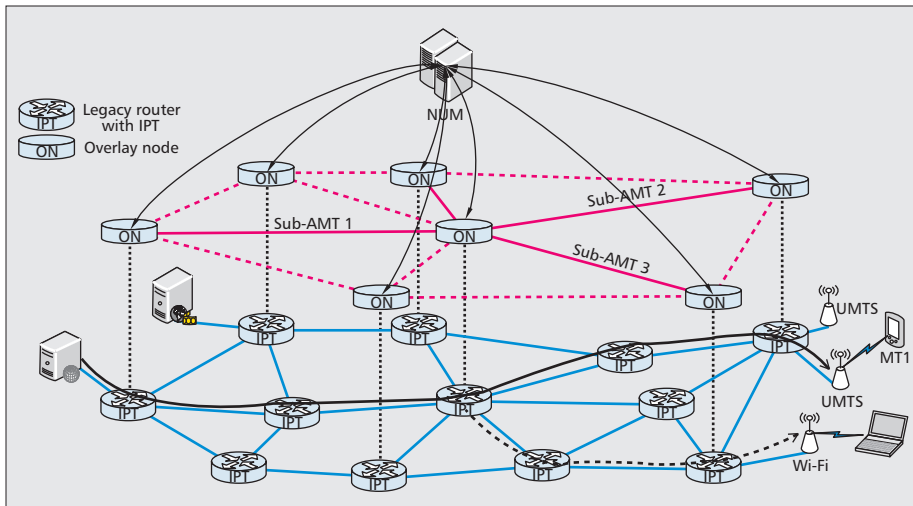


Figure 1. Hierarchical control architecture.

we call an abstract multiparty tree (AMT). The AMT principle enables end-to-end multicast content transport over network segments with different transport technologies (i.e., unicast and multicast), and provides independence between source and listener trees, which directly improves seamless resilience support. Each AMT controls the content distribution of a specific session, providing a generic and scalable transport solution for group communications. Local network segments, delimited by two ONs, define a sub-AMT (Fig. 1), which can be mapped in several physical links and nodes. Each ON performs resource and QoS control in a distributed manner through the IP Transport (IPT) components present in each network router. These abstract trees operate on top of the IP layer, allowing network dynamics and reconfigurations of the multiparty delivery tree to be hidden.

The NUM resorts to the knowledge gathered and maintained at the CB to perform context-aided decisions. Two network selection algorithms were defined, for access and core networks, that may be influenced by network resources, environment, user preferences, or terminal capabilities. These intelligent network decision procedures aim at enhancing the quality of the service perceived by each user while efficiently managing the available network resources. The access selection algorithm is performed first, attempting to offer the best combination between users and the most suitable access networks in a heterogeneous system (implicitly defining the edge ON). The optimization of this compromise is achieved by being aware of users' requirements and access technologies' characteristics [10]. The core network selection determines the feasible paths for the content delivery tree in the core network, through several sub-AMTs (and ONs) up to the access point (edge ON) chosen by the access selection algorithm. This selection considers the requirements of each service and the available resources in the network [2]. At the NUM, the network selection is always focused on group-based applications. Since it is almost impossible to meet all users' preferences simultaneously, users are grouped according to their context (location, capabilities, etc.), and then the selection procedures are applied. This feature enhances scalability, since only some sub-groups can be adapted, avoiding affecting all users of an ongoing session.

### Abstract Multiparty Transport

The concept of AMT essentially relies on a set of ONs, throughout the network, that are endowed with more resources, intelligence, and functionalities in order to better control content distribution. These nodes are used to build for each flow of a certain multiparty session a control overlay net-

work, which we denote the AMT. Thus, it is assumed that a unique session may have different associated flows (e.g., video, audio, and data). Each of these flows has a specific set of constraints and requirements that can be addressed specifically by each AMT, since they are independent of each other (disjoint control overlay network and delivery tree).

The AMTs operate on top of the IP network layer to support general transport control of the physical network delivery tree. Furthermore, a higher level of granularity is defined by dividing the end-to-end AMT in several sub-AMTs, bounded by the ONs. Each sub-AMT can be seen as

a logical link between two ONs. The abstraction level provided by the overlay network and the division of the trees enables end-to-end multicast content transport over heterogeneous network segments in terms of IP multicast capability or IPv4/v6 support (each sub-AMT is associated with a unique multicast address). This abstraction level also hides network dynamics (the reconfigurations inside each sub-AMT can be performed locally to the sub-AMT), and provides independence between source and clients (several sub-AMTs are supported independently). Applying this overlay paradigm at the transport layer, it is possible to provide a scalable transport service for group communications.

All core nodes composing a sub-AMT must implement the same transport technology. In the scope of a sub-AMT:

- An ingress ON is viewed as a session source.
- An egress ON is viewed as a leaf node.
- Core nodes simply perform IP forwarding operations,

### Abstract Multiparty Trees Control

The process of AMT selection is made by the NUM, which coordinates the edges of each sub-AMT for mediating overlay connections; in this sense, ONs may control network resources and the QoS of each sub-AMT independent of each other. A description of the NUM's features and selection procedures is given in Fig. 2. In general, there are two main events that may trigger the NUM's control procedure: the arrival of a new client interested in a non-initiated multiparty session, or an already established and ongoing session.

Whenever it is necessary to establish a new multiparty session (MS), the NUM obtains from the CB the session related context, which includes the session's source and the associated set of flows. Following the AMT approach, a corresponding control overlay network, the AMT, is defined for each flow, shaped according to the individual context information of the flow (intended clients, delay, and bandwidth constraints). Once again, it is assumed that this knowledge is available at the CB.

Hence, in order to build the AMT, the two network selection algorithms are deployed. Focusing on the control scheme described in Fig. 2, the access selection for each client of a certain flow is performed first, according to the algorithm detailed in [10]. The access selection scheme receives the set of access points within the client's range and returns the access that best meets the user preferences and network capabilities.

Subsequently, and assuming that each access point is only connected to a unique egress ON of the core network, it must be assessed if the corresponding egress ON is already part of the AMT that is being built. If this is the case, there is no

need for further operations, and the client may start receiving the content as soon as the egress ON replicates the content toward the selected access point. If this is not the case, the core network selection scheme is triggered to add the selected egress ON to the AMT. To accomplish this, *Dijkstra's* algorithm is executed taking into consideration that only the physical links with enough resources (those that meet session requirements in terms of bandwidth and/or delay) can be used to connect the new egress ON to the current AMT.

In order to enhance the scalability and flexibility of our AMT concept, every new egress ON added to an already established AMT should be connected to the closest ON within the AMT (including the content source itself). Following this principle, it is possible to aggregate as much as possible the content distribution tree, taking advantage of multicast principles and consequently avoiding packet duplication. Moreover, all ONs within the new branch toward the egress ON will become part of the corresponding AMT, defining at least a new sub-AMT. This way, possible context changes that affect at least this new sub-AMT will be handled within its scope, and will not affect the remaining AMT.

We believe that by enabling this content distribution tree modularity through the control overlay network concept, it is possible to reach a new level of flexibility. This way, the AMT strategy endows multiparty content distribution with the ability to transparently react to any networking event that may occur, such as client arrival, departure, handover, or even source mobility. A thorough analysis of these advantages is detailed in the following section.

Beyond establishing a new multiparty session, the NUM functionalities can also be triggered when it is required to update an ongoing session due to one of two events (Fig. 2): a new client arrives or changes its access point. Both events have in common the fact that a new access point is used, and the corresponding egress ON needs to be added to the matching AMT. Thus, if the selected egress ON does not yet belong to the corresponding AMT, a similar overlay network extension process to the one detailed above is performed, connecting the egress ON to the closest ON within the AMT. However, in case of client handover, it is necessary to update not only the new ONs that are found in the new AMT branch, but also to remove the ones of the previous branch if the egress ON to which it was connected has associated any other client interested in the same flow.

Moreover, a core network event may also imply network rearrangements due to link and node failures or service degradation. In this case, and given the hierarchical intelligence strategy, the ONs that delimit the impaired sub-AMT attempt to redefine it in order to overcome the failure, performing a local reconfiguration. If the ONs attempt fails, the NUM defines a new AMT from the source toward all the egress ONs with associated clients.

Note that the output of the NUM procedures is always an AMT per each flow of an MS, or a redefinition of one AMT. Although the NUM is a central element, the defined AMT is enforced in a distributed manner on the network, taking advantage of the control overlay network.

## Scalability Analysis

Scalability and flexibility are major guidelines in the development of an architecture control scheme. Since the architecture is focused on the development of a context-sensitive frame-

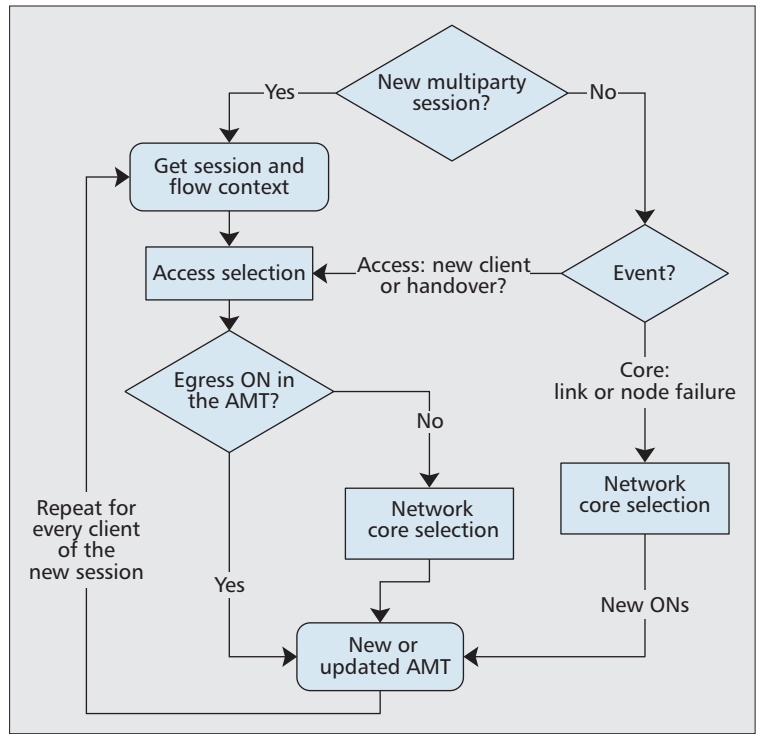


Figure 2. AMT control scheme.

work, it is critical to evaluate the ability to perform network adaptations and reconfigurations in a large-scale networking environment. In this section, we describe an analysis to quantify how well the AMT concept scales according to the number of clients and network size.

For simplicity reasons, we analyze the case of an MS that has associated only one flow. In this case, adopting the AMT strategy, a unique control overlay network is built per flow, which does not provide loss of generality in the analysis given the independence between coexisting AMTs. Thereby, the first client of the MS will trigger the creation of an AMT, which means that it will be connected to the content's source through the shortest path according to the process described in Fig. 2.

Nevertheless, the expansion of the AMT when other clients join the MS has a different behavior since, through the AMT concept, novel clients are first connected to the closest ON within the respective AMT. Thus, given that each client may only connect to a certain access point and, subsequently, to a specific egress ON ( $v_{O_e}$ ), the probability of a client connecting to an egress ON is independent and uniform:

$$P_{v_{O_e}} = \frac{1}{|V_{O_e}|}, \quad (1)$$

where  $V_{O_e}$  is the set of available egress ONs. Thus, the probability of a new client connecting to an egress ON  $v_{O_e}$  that is not already part of the control overlay network  $O_f$ , is given by

$$\overline{P_{v_{O_e}}} = \left( \frac{1}{|V_{O_e}|} \right)^i \quad (2)$$

where  $i$  is the number of clients of the MS.

Consequently, for further clients, the probability of an AMT requiring an extension is given by Eq. 2, following the scheme described in Fig. 2. Otherwise, with a probability  $1 - \overline{P_{v_{O_e}}}$ , the corresponding AMT does not need to be expanded,

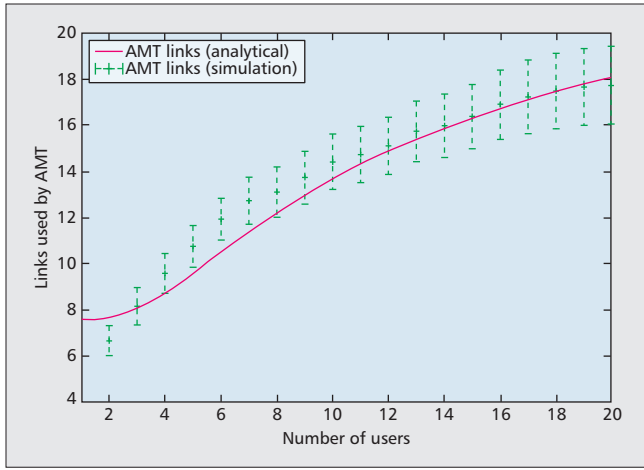


Figure 3. AMT links as a function of users (simulation vs model).

and this probability becomes higher along with the number of clients.

The estimation of the resources needed to establish this extension is of crucial importance, since it measures how well the concept of AMTs scale. Thus, considering the multiparty content delivery tree, Eq. 3 translates the content delivery tree growth (links used) into a function of the number of clients and network size:<sup>1</sup>

$$E[t(r)] = \gamma e^{\left(\frac{\theta + \nu}{N + r}\right)} + \tau \quad (3)$$

Through this expression, it is possible to determine the number of additional links used by the AMT to reach the latest client.

In the analysis, the probability density function (PDF) of the number of links in the tree that are reused when adding a new client is also obtained. Let  $X_{r_l}$  denote the random variable of reused number of edges by new clients; the PDF of reusing  $k$  links is<sup>2</sup>

$$\Pr[r_l = k] = \sigma^{-1} e^{\frac{k-\mu}{\sigma}} e^{-\frac{k-\mu}{\sigma}} \quad (4)$$

For more information on the analysis process, please refer to [12], which provides a detailed description of the methodology adopted to obtain the scaling behavior of the AMT control approach.

## Evaluation

For comparison purposes, the Protocol-Independent Multicast Source Specific Multicast (PIM-SSM) was chosen, since it can be considered a particular case (unique source) of PIM sparse mode (PIM-SM), the most employed multicast protocol at present (e.g., IPTV and triple play solutions). To build its multicast distribution tree, PIM-SSM uses the unicast routing tables to forward *Join* messages until it finds the respective source tree. Thus, considering that clients are on the same group sparsely located in the network, it is likely that each one has an associated tree whose set of nodes and links is disjoint from another receiver's source tree. In this case, a tree rearrangement triggered by any kind of event would require a complete setup of the path between the

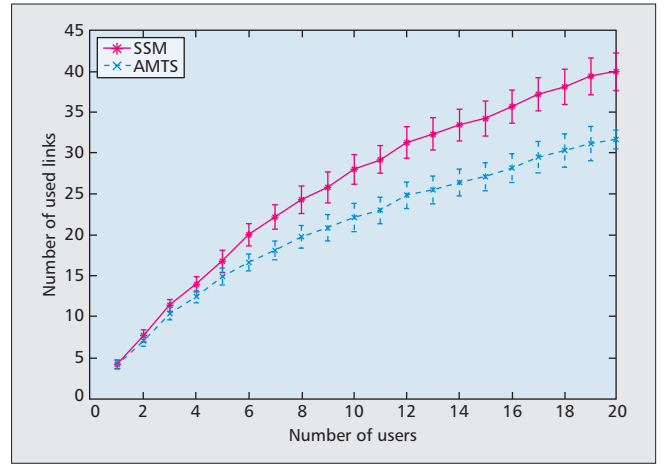


Figure 4. Comparison between PIM-SSM and AMTs strategies.

client and the source, being unaware of tree branches that could be reused.

An implementation of the behavior of both approaches was performed, focusing the analysis on the reconfigurations and link savings when dealing with clients' joins or leaves, allowing us to understand the ability of the approach to deal in a scalable way with constant changing environments. In order to resemble real core networks with router-level topologies, different network topologies were generated using the Waxman random network generator. Following this model, the nodes of the network are uniformly distributed in the plane, and edges are added according to probabilities that depend on the distances between the nodes.

The methodology followed in the implementation consists of performing a set of instructions for each node that is consecutively added to the tree, keeping track of the changes observed regarding the number of links modified (added and eliminated). This way, it is possible to compare the behavior of each approach in a large network environment and with many possible listeners. The following results were determined repeating 50 times the experience for each user, obtaining confidence level intervals of 95 percent.

## Model vs. Simulation

Figure 3 shows a comparison between the expected number of links obtained through the analytical approach (Eq. 2) for a certain scenario ( $N = 200$  nodes and  $r = 20$  clients) and the results obtained by the simulation of the same scenario. Considering the results depicted on the figure, it is shown that the analytical model fairly represents the behavior of the expansion of a multiparty delivery tree following the AMT concepts and control procedures. This fact can be attested by the majority of simulation samples whose confidence intervals include the theoretical/expected value, obtaining a more notable resemblance for a higher number of users.

## Comparison of AMTs vs PIM-SSM

This study assesses the level of traffic aggregation provided by the AMT strategy compared to PIM-SSM. Considering the implementation of PIM-SSM, Dijkstra's algorithm is used to compute the path from the content source to the listeners, since this protocol uses the unicast routing table of the routers to build its distribution tree. The AMT approach, as explained in the previous section, builds its distribution tree based on the ONs available in the network; in the simulation framework they are randomly mapped in certain nodes according to a given density. Dijkstra is also used to connect the ONs that build the abstract tree.

<sup>1</sup>  $\gamma = 101.8$ ;  $\theta = -345.9$ ;  $\nu = -10.9$ ;  $\tau = 7.6$

<sup>2</sup>  $\mu = 5.28112$ ;  $\sigma = 0.565152$

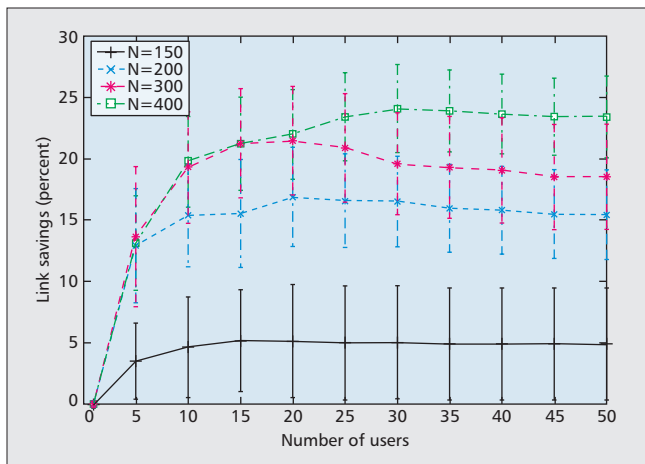


Figure 5. Gains of AMTs over SSM for different network sizes.

The results depicted in Fig. 4 show the advantages of the AMT strategy over PIM-SSM, which is even more notorious for a higher number of clients. This fact highlights the scalability of the abstract trees for a higher number of users. Although the difference between the numbers of changed links is not very high, this is related to the relatively small number of hops of each path from the source to the listener. The superposition of the confidence intervals of each curve occurs due to the randomness of each network topology created and the random distribution of ONs throughout the network nodes.

#### Impact of Network Size

Figure 5 shows the link saving gains with the network topology dimension. This gain is determined by the ratio between the links used by both approaches (AMT and PIM-SSM) in their distribution tree to reach every client. As expected, as the network topology gets larger, the gains obtained by the AMT approach are higher, since the paths are longer and the routing possibilities increase. However, for higher numbers of users, one may observe that the gains remain constant, which is due to the ever decreasing probability of a new client connects to an Egress ON that is not already part of an AMT.

#### Impact of ON Density

The results obtained for the impact of ON density (Fig. 6) show that the increase in the number of nodes with overlay control functionalities in the network increases the link savings, due to the increase on the number of sub-AMTs, enhancing network segmentation and increasing the set of branching points closer to new destinations. It is also notably similar behavior to that of the curves, which stabilize nearby 30 users. Compared to the results of the previous section, notice that this set of results concerns a network topology with 250 nodes; thus, the gains obtained would fit between the curves for 200 and 300 (being not directly comparable). Also regarding Fig. 6, there is a significant difference of gains between ON densities of 20 and 80 percent in the network (roughly 10 percent), which highlights the benefits of the AMT control strategy. Although not shown here, the number of reconfigured links increases with the percentage of ON nodes in the network, since an increase in the number of ON nodes increases the number of sub-AMTs, and consequently increases the need for their reconfiguration.

Overall, these results show that scalable architectures can be envisaged, and hierarchical control approaches may be the path for context-aware multicast autonomous control.

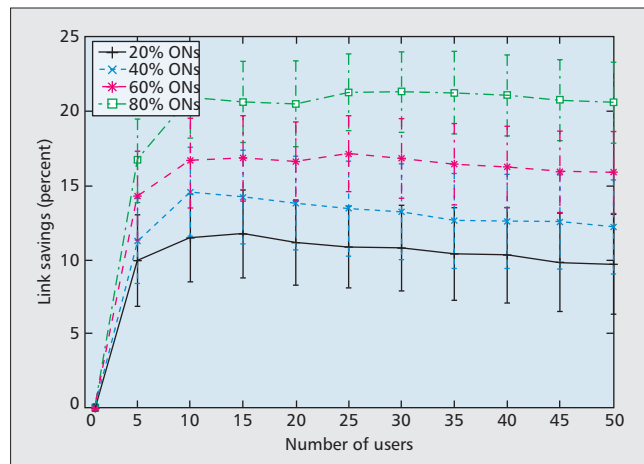


Figure 6. Gains of AMTs over SSM for different ONs densities.

## Conclusion

This article presents and studies the behavior of a hierarchical autonomous architecture sensitive to context that employs the concept of abstract multiparty trees to control content distribution in a more intelligent and scalable fashion. A brief description of the main architecture was given, followed by an explanation about the principles of AMTs and how they can benefit content delivery architectures.

To address the scalability of the hierarchical approach, an analysis was performed with respect to the probability of extension of the AMTs and the number of links that can be reused so that control messages and operations could be spared. The evaluation results showed that the hierarchical control performed through AMTs obtains considerable gains regarding link savings and consequent network control operations, thus increasing the scalability of the autonomous control architecture. We can then conclude that the distribution of control operations through some nodes in the network can significantly improve the scalability of autonomous context-aware architectures, and that a decentralized approach is the path for autonomous networking.

## References

- [1] Z. Movahedi et al., "A Survey of Autonomous Network Architectures and Evaluation Criteria," *IEEE Commun. Surveys Tutorials*, vol. PP, no. 99, 2011, pp. 1–27.
- [2] N. Coutinho et al., "Multiparty Seamless Transport," *GLOBECOM '10*, 2010, pp. 1–6.
- [3] C. Diot et al., "Deployment Issues for the IP Multicast Service and Architecture," *IEEE Network*, vol. 14, no. 1, 2000, pp. 78–88.
- [4] M. Hosseini et al., "A Survey of Application-Layer Multicast Protocols," *IEEE Commun. Surveys and Tutorials*, vol. 9, no. 3, 2007, pp. 58–74.
- [5] S. Fahmy and M. Kwon, "Characterizing Overlay Multicast Networks," *Proc. 11th IEEE Int'l. Conf. Network Protocols*, 2003, pp. 61–70.
- [6] J. C.-I. Chuang and M. A. Sirbu, "Pricing Multicast Communication: A Cost-Based Approach," *Telecommun. Systems*, 1998, pp. 281–97.
- [7] C. Bettini et al., "A Survey of Context Modelling and Reasoning Techniques," *Pervasive and Mobile Computing*, vol. 6, no. 2, 2010, pp. 161–80.
- [8] C. Jacob et al., "Bio-Inspired Context Gathering in Loosely Coupled Computing Environments," *1st Bio-Inspired Models of Network, Information and Computing Systems*, 2006, pp. 1–6.
- [9] B. Beamon, "Evaluation of First Order Bayesian Networks for Context Modeling and Reasoning," *2010 8th IEEE Int'l. Conf. Pervasive Computing and Communications Wksp.*, vol. 29, 2010.
- [10] N. Coutinho et al., "Context-Aware Selection in Multicast Environments," *2010 IEEE Symp. Computers and Commun.*, 2010, pp. 646–52.
- [11] J. Simoes et al., "Context-Aware Control for Personalized Multiparty Sessions in Mobile Multihomed Systems," *MobiMedia '09*, 2009, pp. 1–1.
- [12] N. Coutinho and S. Sargento, "Abstract Multiparty Transport: An

---

Analytical Study," May 2011, [www.av.it.pt/ssargento/internal\\_reports/report\\_AMT\\_analytical.pdf](http://www.av.it.pt/ssargento/internal_reports/report_AMT_analytical.pdf).

- [13] T. Zhang, K. Yang, and H.-H. Chen, "Topology Control for Service-Oriented WMNS," *IEEE Wireless Commun.*, vol. 16, no. 4, Aug. 2009.

### Biographies

NUNO COUTINHO ([nunocoutinho@ua.pt](mailto:nunocoutinho@ua.pt)) concluded in 2008 his five-year Integrated M.Sc. in electronics and telecommunications engineering at the University of Aveiro. His Master's dissertation, *Intelligence in Mobility Decisions*, was about the selection of the best access network according to context information about user and network. Since October 2008 he has been a Ph.D. student at the University of Aveiro and joined the Celfinet Innovation Department from October 2008 to February 2009. Currently, he is a researcher associated with the Institute of Telecommunications, involved in several national projects (MuMoMgt, GEN-CAN, UbiquiMesh, GTI-CANE) and European projects (C-CAST). His main research interests are related to future network management architectures, multicast, context awareness, quality of experience, and network coding.

SUSANA SARGENTO ([susana@ua.pt](mailto:susana@ua.pt)) received her Ph.D. in 2003 in electrical engineering. She joined the Department of Computer Science of the University of Porto in September 2002, and has been at the University of Aveiro and the Institute of Telecommunications since February 2004. She is also guest faculty of the Department of Electrical and Computer Engineering at Carnegie Mellon University since August 2008, where she performed a faculty exchange during her sabbatical (2010-2011). She has been involved in several national and international projects, taking leadership of several activities in the projects, such as the QoS and ad hoc networks integration activity in the FP6 IST-Daidalos Project. She has recently been involved in several FP7 projects (4WARD, Euro-NF, C-Cast, WIP, Daidalos, C-Mobile), national projects, and CMU|Portugal projects (DRIVE-IN with Carnegie Mellon University). She also has strong involvement with national and international companies, such as Portugal Telecom Inovao, Nokia Siemens Networks, CISCO, Alcatel Lucent, France Telecom, Deutsche Telekom, and Huawei, both in projects and in exchange of students. Her main research interests are in the areas of next generation and future networks, more specifically QoS, mobility, and self- and cognitive networks.

---

# Vertical and Horizontal Circuit/Packet Integration Techniques for the Future Optical Internet

**Chaitanya S. K. Vadrevu, University of California, Davis**

**Massimo Tornatore, Politecnico di Milano**

**Chin P. Guok and Inder Monga, Energy Sciences Network**

**Biswanath Mukherjee, University of California, Davis**

---

## Abstract

Hybrid circuit/packet networks where circuit and packet networks coexist are becoming attractive to support future Internet applications. They support both packet/IP services and circuit/wavelength services. Packet services include traditional data services such as VPN, VoIP, and email, while dynamic circuit services include end-to-end bandwidth-intensive applications such as terascale science experiments. In present hybrid networks, such as ESnet, the bandwidth boundary between the circuit and packet sections of the network is fixed. However, a flexible boundary between the circuit and packet sections will enable cost-efficient bandwidth management in the network. Our study investigates two methods to dynamically migrate capacity between the circuit and packet sections, called vertical stacking and horizontal partitioning, and serves as a tutorial. In vertical stacking, the backup capacity of wavelength circuits can be dynamically exchanged between packet and wavelength services while ensuring survivability. The backup capacity can be used to protect wavelength services in the event of a failure and route packet traffic otherwise. In horizontal partitioning, the excess capacity on links in the packet section can be loaned to circuit services. We have conducted experiments using a snapshot of real traffic on ESnet with horizontal partitioning. Control mechanisms for our approaches that can be operational in ESnet are presented.

---

**T**he use of optical networks for supporting bandwidth-hungry services is an attractive proposition to ensure wide-area reach and huge amount of inexpensive bandwidth. Two services which will be prevalent in future telecom backbone networks are packet/IP services and circuit/wavelength services [1]. Packet services, including traditional data services such as VPN, teleconference, and data backups, are well established in carrier networks. Wavelength services including bandwidth-intensive applications, such as terascale scientific experiments, characterized by strict QoS requirements, require circuit-switched technologies to deliver guaranteed bandwidth, and are managed directly at the optical layer. Wavelength services currently contribute to a small fraction of carrier traffic today, but are expected to grow as applications such as e-science and cloud services emerge [2–4]. Thus, important and timely research is needed to enable telecom service providers to design cost-effective hybrid circuit/packet networks by jointly supporting packet and wavelength services on the same network infrastructure.

Hybrid circuit/packet networks consist of circuit networks

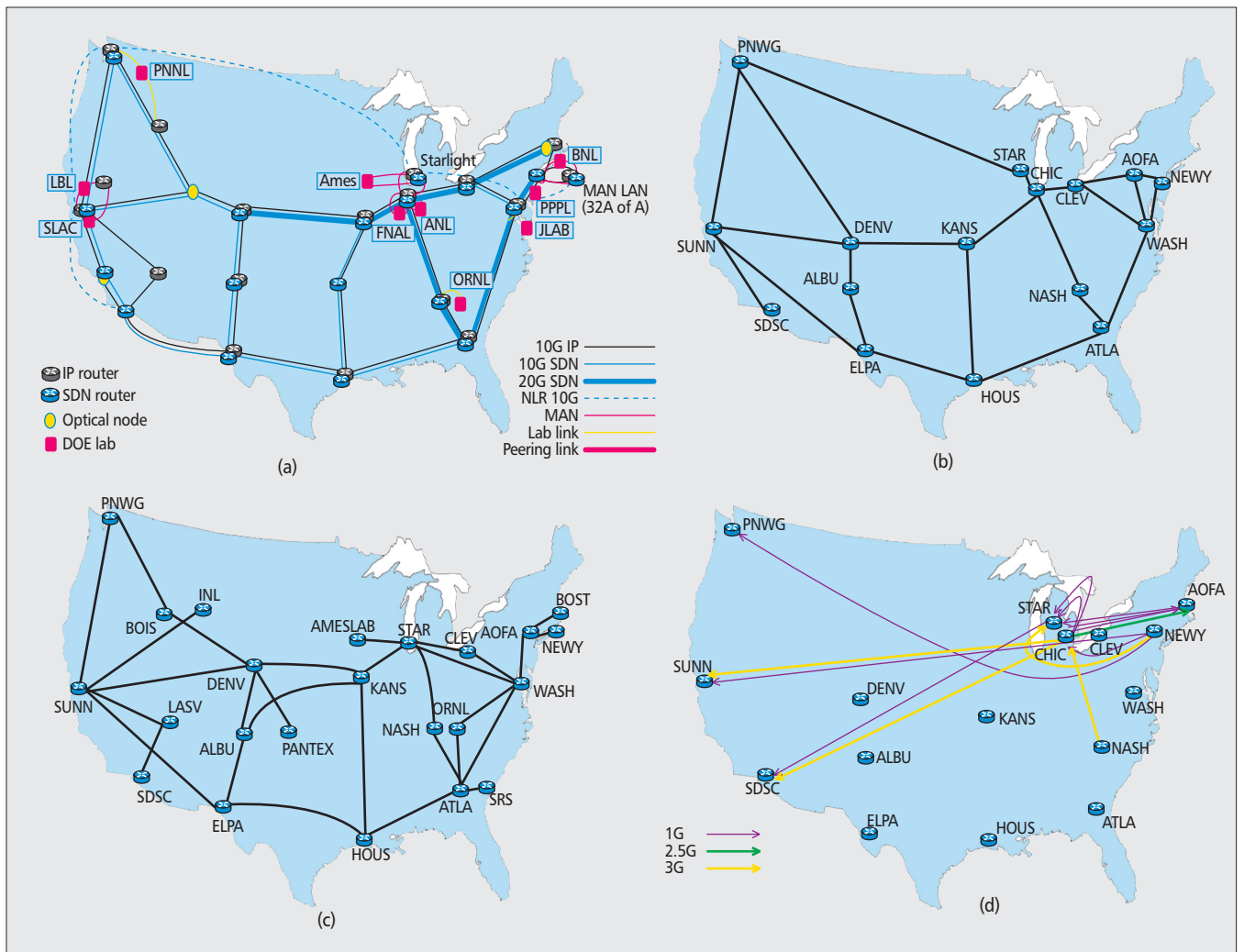
coexisting with packet networks; usually the packet network is overlaid on top of the circuit network, such as an IP-over-wavelength-division multiplexing (WDM) network. However, in some cases such as the U.S. Department of Energy's (DoE's) Energy Sciences Network (ESnet) [5], the circuit and packet networks are deployed side by side (e.g., they have common end-node sites and equipment), but they are logically separate and may have physically disjoint links. In both network partitions, the packet and wavelength services are supported by leasing capacity over wavelengths.

In ESnet, the bandwidth partitioning between the two networks is relatively static. However, it would be very desirable to have dynamic partitioning of the bandwidth between the circuit and packet sections so that the bandwidth can easily be migrated from the packet section to the circuit section, and vice versa [6, 7]. For example, traffic variations over a day can be exploited by accommodating large data transfers for terascale science applications during the night over the circuit section, leaving its exploitable capacity for packet traffic during peak hours of the day.

Also, protection of the carried traffic can benefit vastly by using a dynamic network resource partitioning scheme among the circuit and packet services; for example, wavelength circuits are often protected by dedicated backup circuits [8],

---

*Massimo Tornatore was with the Computer Science Department, University of California, Davis.*



**Figure 1.** Energy Sciences Network (ESnet) topologies: a) ESnet; b) ESnet science data network (SDN) topology; c) ESnet packet network (IP) topology.

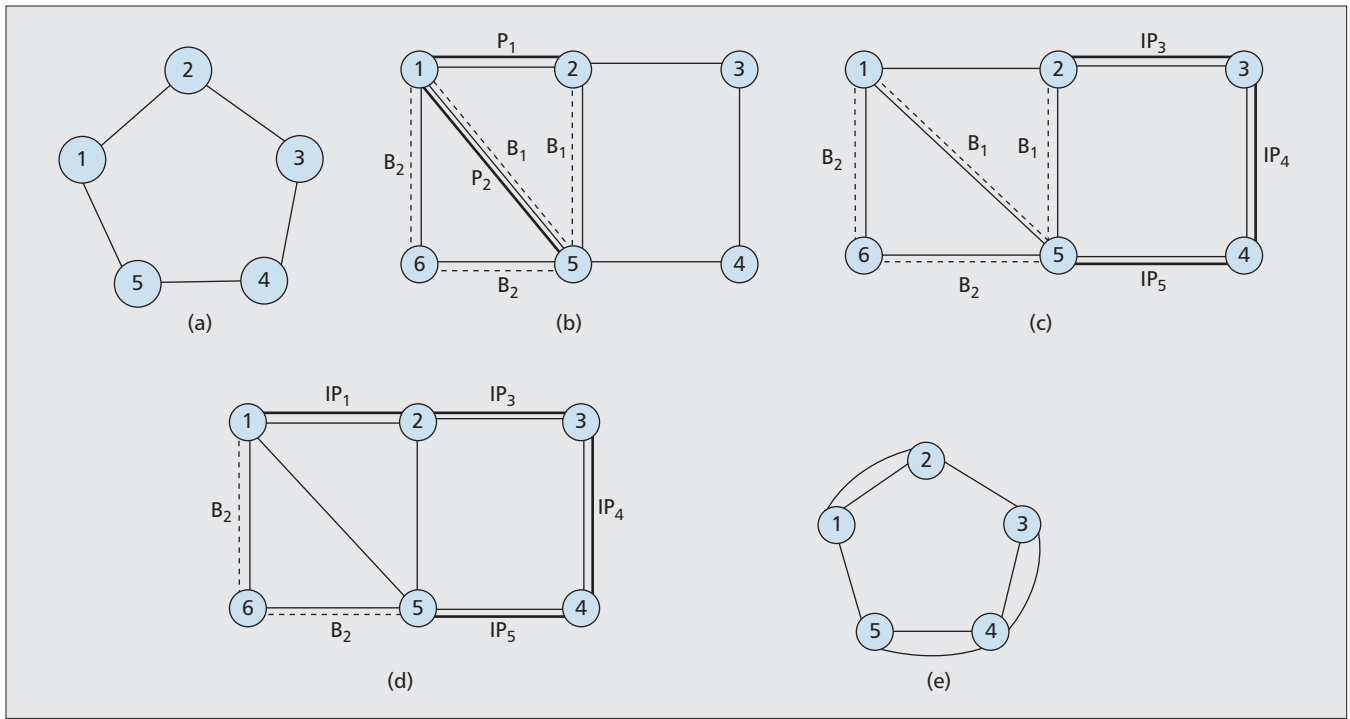
which are generally idle and unutilized unless there is a failure in the network. Especially with the upcoming deployment of 100G transmission systems, huge amounts of backup resources will be underutilized. The network capacity could be better utilized if idle backup circuits of wavelength traffic could be loaned to packet services. Similarly, the packet network capacity could be loaned to circuit traffic under extreme circumstances (multiple and/or concurrent failures) to ensure rerouting for a high-priority subset of the circuit traffic. We need novel design approaches that allow the idle wavelength backup capacity to be loaned to packet services, and vice versa, without sacrificing the survivability of both services [8].

In this study, we describe two complementary approaches to enable dynamic partitioning of capacity between a packet network and a circuit network (or, equivalently, between the circuit and packet sections of a network, as in the ESnet), which we classify as *vertical stacking* and *horizontal partitioning* and present as a tutorial article. We describe the characteristics of ESnet, which provides a relevant case study of a network that jointly carries circuit and packet services. We focus on the packet-circuit capacity integration using vertical stacking, and mechanisms for capacity migration from wavelength services to packet services in packet-over-circuit networks are disclosed. We introduce horizontal partitioning, and describe how to migrate capacity from the packet section to the circuit section. We outline state-of-the-art control mechanisms that allow some degree of capacity partitioning. We then conclude the study.

### Case Study of a Hybrid Network: ESnet

With increasing demands for packet and wavelength services, hybrid networks supporting both services are becoming important. We present the U.S. DoE ESnet as an example hybrid network supporting mixed packet and wavelength services. ESnet is a high-speed hybrid circuit/packet network serving thousands of U.S. DoE scientists at over 40 institutions as well as connecting to more than 100 other networks. ESnet provides high-bandwidth reliable connections that link researchers at national laboratories, universities, and other research institutions, enabling them to collaborate on some of the world's most important scientific research challenges, including energy, climate science, and the origin of the universe.

In ESnet [5], the circuit network and packet network are deployed side by side (e.g., they have common end-node sites and equipment), but they are logically separate and may have physically disjoint links. Figure 1a visualizes the long-distance topology of ESnet, showing the partition between the IP (packet) section of the network and the circuit network in ESnet, known as the science data network (SDN). While SDN provides dynamic and scheduled circuit services, the ESnet IP core network (Fig. 1c) transports IP packets. ESnet's SDN is based on traffic-engineered circuits that can be likened to a comparable dynamic wavelength circuit network in terms of flow handling end to end.



**Figure 2.** Illustrative examples of backup capacity sharing: a) sample IP topology; b) sample physical topology; c) capacity sharing (CS): non-survivable routing of IP demands using WL backup capacity; d) capacity sharing with survivability (CSS): survivable routing of IP demands using WL backup capacity; e) capacity sharing with survivability and capacity assurance (CSSCA) for IP topology: augmented IP topology.

The SDN topology shown in Fig. 1b comprises SDN nodes and the logical links between those SDN nodes. The SDN nodes comprise optical cross-connects (OXCs) and routers. Note that these logical SDN links are mapped over a physical network of optical fibers by leasing capacity from the Level 3 network [10]. Bandwidth partitioning between the circuit and packet sections is relatively fixed in ESnet.

The rest of the study describes two possible approaches for dynamic capacity partitioning in ESnet (and more generally in any packet/circuit core network), including some state-of-the-art solutions on ESnet and some promising research on dynamic capacity partitioning. We hope our article serves as a tutorial on dynamic capacity migration techniques with ESnet as an example and inspires readers to pursue these techniques further.

### Vertical Stacking: Packet-over-Circuit Network

In a packet-over-circuit network, a packet network is overlaid on top of a circuit network (e.g., an IP-over-WDM network), where an IP topology consists of IP routers interconnected by optical WDM circuits (lightpaths), while the physical topology consists of OXCs or reconfigurable optical add-drop multiplexers connected by physical fiber links. The lightpaths that support IP traffic (IP lightpaths) have to be mapped over the physical topology. Also, circuits for wavelength (WL) services (WL lightpaths) may need to be provisioned at the physical layer, and if some degree of protection is required, these circuits are generally protected by a dedicated backup circuit. Evidently, network capacity can be better utilized if idle backup circuits of wavelength traffic can be loaned to IP services [6]. We refer to this approach for dynamic capacity partitioning between circuit and packet services by reusing the WL backup capacity as dynamic partitioning by *vertical stacking*.

### Sharing Backup Capacity between IP and Wavelength Services

The idle WL backup circuits can be used to support IP demands, and the backup capacity can be dynamically partitioned between IP and WL services depending on the demand requests. In case of a failure (e.g., a fiber cut), IP demands over the backup circuit can be preempted and (possibly) rerouted, while the backup capacity can be restored for serving the primary WL traffic.

Note that in order to successfully reroute the IP traffic over alternate paths, the IP topology must remain connected even after a failure on the physical topology has affected one or more of the IP lightpaths [9]. How to design a survivable IP-over-lightpath topology is a classical multilayer design problem. Multiple IP lightpaths can be routed over a single fiber in the physical topology. Hence, failure of a single physical link can result in multiple failures in the IP layer and disconnect the IP topology [9]. To ensure connectivity of the IP topology under all single-link-failure cases, we ensure that, for all proper cuts of the IP topology, the number of lightpaths in the cut-set on any physical link is less than the size of the cut-set [9]. Additional IP links may need to augment the IP topology to also ensure guaranteed capacity for full restoration of IP requests on top of ensuring connectivity of IP topology.

For dynamic partitioning of backup circuit capacity, we add another item of design complexity. The backup circuits are also among the set of candidate lightpaths for IP demands. In fact, only those backup circuits that do not disconnect the IP topology in case of a failure can be loaned by the WL services to IP services. Hereafter, we assume that WL circuits are already assigned, and IP lightpaths are provisioned on top of them. Integrated approaches for jointly provisioning both IP and WL services will lead to greater cost savings.



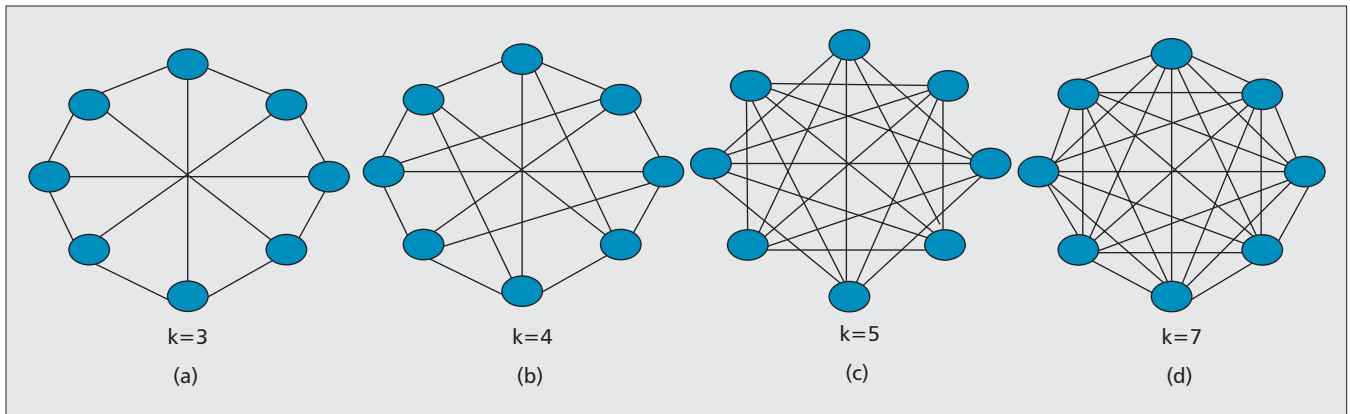


Figure 3. Eight-node wheel networks of increasing connectivity.

### Approaches Enabling Backup Capacity Sharing

We explain a few approaches using some illustrations [6]. Figure 2a shows a sample IP topology to be supported over the physical topology. In Fig. 2b, a sample six-node physical topology with three WLs per link is shown. The primary WL circuits are denoted by P (solid lines) and the backup circuits by B (dashed lines). Note that, in Fig. 2b, two primary WL circuits, P<sub>1</sub> and P<sub>2</sub>, have already been provisioned. The backup capacities for these WL paths are provisioned via routes B<sub>1</sub> and B<sub>2</sub>, respectively. We observe that we can either use existing backup WL paths in the physical topology or establish new IP lightpaths while provisioning the IP demands, (1–5) and (1–2), in Fig. 2a.

We describe four different solutions: no capacity sharing (NCS), capacity sharing (CS, Fig. 2c), capacity sharing with survivability (CSS, Fig. 2d), and capacity sharing with survivability and capacity assurance (CSSCA, Fig. 2e). In the NCS approach, we do not allow sharing of backup capacity between IP and WL services, and establish new lightpaths for each and every IP layer demand. In the second and third approaches, there is sharing of wavelength backup capacity between IP and WL services, but, while in the CSS approach the survivability of the IP topology is strictly enforced, the survivability is sacrificed in the CS approach in order to maximize the sharing of backup capacity. As an example, in the CS approach (Fig. 2c), both backup WL circuits, B<sub>1</sub> and B<sub>2</sub>, are used for supporting IP requests. If the physical link (1–5) fails, both the backup circuit B<sub>1</sub> and primary circuit P<sub>2</sub> fail. When the primary circuit P<sub>2</sub> fails, its backup circuit B<sub>2</sub> becomes unavailable, and the IP traffic over this backup circuit B<sub>2</sub> gets preempted. Thus, both the backup circuits B<sub>1</sub> and B<sub>2</sub> become unavailable for supporting IP traffic during failure of the physical link (1–5), and hence the IP topology gets disconnected. Thus, the solution in Fig. 2c does not provide a survivable mapping of IP connections, while the solution of the CSS approach (Fig. 2d), which uses only one backup WL circuit B<sub>2</sub>, is survivable. We need to consider failures of backup circuits not only due to failures of the links over which they are passing but also due to failures of their primaries. For CSSCA in Fig. 2e, survivability of IP topology is enhanced compared to CSS by also ensuring sufficient capacity for rerouting preempted IP requests during physical link failures on top of IP topology connectivity as in CSS while maximizing sharing of WL backup circuits. The IP topology is augmented with additional IP links to generate sufficient capacity. CSSCA can be input with selected IP requests for capacity assurance. However, connectivity for the entire IP topology is ensured. Consider that IP requests (1–2), (3–4), and (4–5) in Fig. 2a need capacity assurance, so the IP topology is augmented with additional IP links, as shown in Fig. 2e. The routing of the augmented IP

topology over physical topology should be carefully chosen. A careful choice of backup circuits is crucial to ensure the survivability of IP topology. In [6], design techniques for this problem have been investigated.

### Illustrative Numerical Examples

Simulation experiments using the 14-node NSFnet physical topology equipped with 30 WL channels per fiber link to illustrate cost savings in backup capacity sharing between IP and WL services are presented. As IP topologies<sup>1</sup> to be mapped over the NSF physical topology, we consider the 8-node wheel networks shown in Fig. 3 characterized by an increasing connectivity index. Each edge in the IP topology is a demand request of unit WL capacity, and the IP topologies in Fig. 3 denote the traffic matrix in our experiments. Results in [6] are preliminarily discussed and significantly expanded here. The background WL service requests are assumed to be 50 percent of the IP requests.<sup>2</sup>

Table 1a shows the cost benefits in supporting IP traffic due to WL backup circuit sharing. Large savings in wavelength channels are achieved by the CS and CSS approaches compared to the NCS approach. New lightpaths are established in NCS for every IP request, and it is most expensive as there is no reuse of backup WL capacity. CS and CSS reuse backup capacity and are capacity efficient. CSS incurs only small additional expenditure compared to CS with guarantee on survivability of the IP topology under all single link failures. The percentage of cost savings using CSS over NCS in our experiments is in the range of 37–45 percent. Table 1b shows that CSS has 30 percent cost savings over NCS on the total WL channels due to WL backup sharing between IP and WL services. CSSCA has 17–35 percent less expenditure than NCSSCA, and the cost savings are again due to WL backup capacity sharing. CSSCA has 40 percent additional expenditure compared to CS, which is due to additional lightpaths in the augmented IP topology. SP-NCSSCA in Table 1b denotes shared path protection for WL services and capacity assurance for IP topology with no backup capacity sharing. We notice that CSSCA has 12–18 percent less expenditure compared to SP-NCSSCA. Thus, cost benefits due to WL backup capacity sharing are greater than shared protection. As the proportion

<sup>1</sup> There are four different IP topologies with increasing connectivity from 3 to 7 ( $k$ -connectivity means that each node in the IP topology is connected to  $k$  other nodes).

<sup>2</sup> 0 percent WL service requests imply that no WL circuits have been pre-provisioned, and there is no chance of sharing capacity between IP and WL services, while 50 percent WL service requests imply that 50 percent of the source-destination couples requiring IP traffic also require WL traffic.

(a): Total number of wavelengths needed for supporting only IP traffic vs. connectivity of the IP topology.				
	3	4	5	7
NCS	24	31	45	58
CS	12	16	20	27
CSS	15	19	25	32
(b): Total number of wavelengths needed for supporting both IP and WL traffic vs. connectivity of the IP topology.				
	3	4	5	7
CSS	41	54	79	104
NCS	46	66	100	137
CSSCA	68	78	125	154
NCSSCA	80	106	155	201
SP-NCSSCA	78	95	154	187

**Table 1.** Performance comparison of CSS, NCS, CSSCA, NCSSCA, and SP-NCSSCA approaches.

of WL services increases, the savings due to WL backup sharing are expected to be greater.

In Fig. 4a, we explore degradation in IP services (i.e., a hit in bandwidth) during failures. Those requests that have strict QoS requirements can be assured full restorable capacity, but the rest can be exposed to reduced bandwidth under failures to minimize cost. CSSCA — 0.3 indicates that 30 percent of IP requests are ensured full bandwidth for restoration, but only connectivity of IP topology is ensured for the rest. Note that by accepting degradation, a huge amount of cost savings can be achieved. CSSCA — 0.6 can ensure guaranteed restoration for 60 percent of IP requests with only 40–60 percent of capacity resources. As connectivity of IP topology improves, the savings are greater.

Figure 4b presents the savings in operational expenditure (OPEX) in our scheme considering energy consumption of transponders and regeneration, which are assumed to be 34.5 W and 50 W per WL, respectively. Note that CSS achieves up to 20 percent energy savings over NCS, which is significant. Thus, WL backup sharing not only gives significant capital expenditure (CAPEX) savings but also considerable OPEX savings. CSSCA has larger energy consumption over NCS due to the additional augmented IP links.

### Horizontal Partitioning: Packet Section Side by Side with Circuit Section

Here we discuss the opportunity of when the capacity in the circuit section is exhausted, the excess capacity from the packet section can be borrowed to support WL services [7] and vice versa. We need to ensure the survivability and QoS requirements for both IP and WL services, while migrating the capacity. Unlike the previous case of vertical stacking, now the IP and WL sections are deployed side by side (not one on top of the other), so dynamic capacity partitioning applies horizontally instead of vertically. In ESnet, the packet and circuit sections use separate and distinct physical resources (WLS

on fiber). Horizontal partitioning enables this fixed capacity barrier between circuit and packet sections of the network to be overcome through dynamic capacity migration.

### A Case Study for Horizontal Partitioning in ESnet

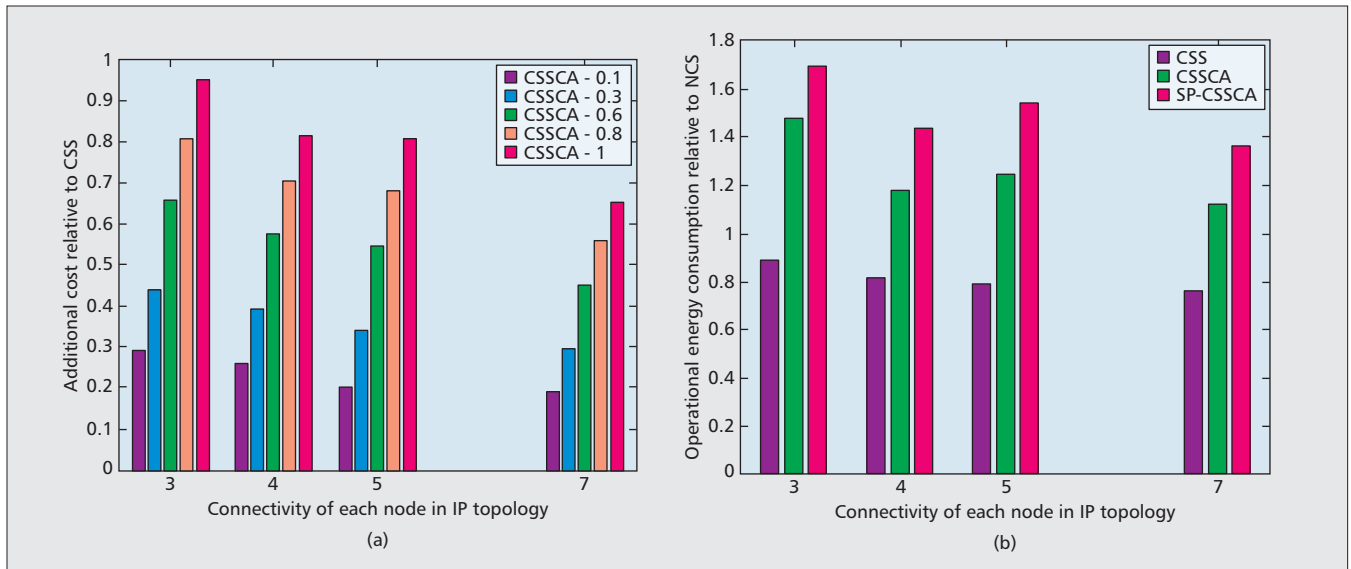
Below, we describe a specific case of horizontal bandwidth migration from the packet section to the circuit section. The scenario arises when we try to provision dedicated protection to all SDN reservations in ESnet. In this case study, we use a snapshot of the ESnet topology and traffic from April 2010, focusing mainly on the circuit section (SDN). More up-to-date traffic distributions and logical topology connectivity of ESnet are available at [weathermap.es.net](http://weathermap.es.net). Based on the traffic profiles in April 2010, we noticed 15 active circuit services in SDN. The SDN topology comprised 16 SDN nodes in April 2010 with the logical links as shown in Fig. 1b. The SDN links are bidirectional and of 10 Gb/s capacity leased over WLS from the Level 3 network [10]. There are 35 such bidirectional links in the SDN network, and multiple links exist between certain pairs of nodes. The 15 active reservation requests are shown in Fig. 1d and are of 1 Gb/s, 2.5 Gb/s, and 3 Gb/s capacity. There are 10 1G reservation requests, one 2.5G reservation request, and four 3G reservation requests. The total capacity lit up in SDN from the Level 3 fiber network is 700 Gb/s. ESnet [5] has established primary circuits for routing these 15 reservations, and the capacity used is only 85 Gb/s out of 700 Gb/s total. The rest of the capacity (615 Gb/s) in the SDN network is idle.

In our solution, we wanted to protect these 15 SDN reservations by providing dedicated protection (since the network has a lot of excess capacity: 615 Gb/s); design techniques are discussed in [7]. We can protect 13 out of the 15 active SDN reservations with dedicated backup circuits, using only 218.5 Gb/s of the 700 Gb/s SDN capacity. Note that there is only one logical SDN link connecting the SDSC node to the rest of the SDN network (Fig. 1b), and two out of the 15 active reservations are to SDSC. Thus, we cannot provision backup circuits for the reservation requests to SDSC with the existing SDN capacity. However, by borrowing capacity from just two IP links, LASV-SDSC and SUNN-LASV, from the packet network (Fig. 1c), we can ensure 2-connectivity at SDSC and offer protection for all 15 active SDN reservation requests. The available capacity on the IP links LASV-SDSC and SUNN-LASV was 622.5 Gb/s as of April 2010, and the demand requests to SDSC are of total 4 Gb/s capacity. Thus, we were able to partially protect the two reservation requests to SDSC using our approach. However, capacity on these IP links can be upgraded to offer full protection.

### Enabling Technologies and Control Mechanisms

In order to exploit the advantages of dynamic partitioning of capacity between IP and WL services, new control mechanisms are needed to support the migration of IP traffic over WL circuits and of WL traffic over the IP networks.

In its present mode of operation, ESnet provides some capability to partition network capacity between the two classes of services (IP and WL) [5]. In ESnet, the IP and SDN networks are managed logically as a single network (i.e., virtually partitioned as two different use networks) using routing protocol metrics and different classes of service. In ESnet, both the IP and SDN networks are defined to be in the same Open Shortest Path First with Traffic Engineering (OSPF-TE) area, with Resource Reservation Protocol with TE (RSVP-TE) and multiprotocol label switching with TE (MPLS-TE) enabled on



**Figure 4.** Illustration of benefits of vertical stacking schemes: a) performance of CSSCA with increased proportion of degraded IP services; b) energy consumption of CSS, CSSCA, and SP-CSSCA vs. NCS.

all interfaces. This enables packet (IP) and circuit (SDN) traffic to utilize either the IP or SDN networks in a fixed manner.

To ensure that the IP network is preferred over the SDN network for IP traffic, the OSPF metrics for each link are generally configured using the following rule set:

- 10 Gb/s IP circuits are set to  $100 \times$  round-trip time (RTT).
- 10 Gb/s SDN circuits are set to  $10 \times$  the IP link metric (if both IP and SDN links are attached to the same adjacent nodes) or  $1000 \times$  RTT if there is no parallel IP link.
- 2 Gb/s IP circuits (e.g., aggregated Ethernets) are set to  $2500 \times$  RTT.
- 1 Gb/s IP circuits are set to  $5000 \times$  RTT.
- Commodity exchange point circuits are set to  $30,000 \times$  RTT.
- Low-speed (i.e.,  $< 1$  Gb/s) links are set to  $60,000 \times$  RTT, and low-speed backup links are set to  $65,000 \times$  RTT.

Using these rules, IP traffic will always prefer the IP network, unless it is bipartitioned, in which case IP packets can traverse the SDN network where necessary. To manage circuit traffic on SDN, MPLS label switched paths (LSPs) with explicit route objects (EROs) are used to define the strict hop-by-hop SDN links of the circuit path. With IP packet and circuit traffic capable of utilizing either IP or SDN links, different classes of service (i.e., QoS) are used to isolate and ensure some level of bandwidth guarantees in a failure scenario. To this end, ESnet supports four different classes of service:

- Network control: Used specifically for network management and not available to the user
- Expedited forwarding: Used exclusively for in-profile circuit traffic
- Best effort: Used for IP packet traffic
- Scavenger: Used for low-priority scavenger tagged IP packet traffic and out-of-profile (i.e., over-subscribed) circuit traffic

The different classes of service are enforced by different queue sizes in the network, which are configured identically to the transmit rates as shown in Table 2 [5]. Based on the QoS setting, a failure in the IP section will displace up to 20 percent of traffic on an SDN link where necessary. Conversely, up to 20 percent of circuit traffic can be provisioned and guaranteed on any IP link, if necessary. A much more flexible partitioning of the capacity among the two network sections is desirable to enable load balancing and more efficient protection, and promptly respond to sudden changes in traffic

	Network control	Expedited forwarding	Best effort	Scavenger
IP c	5%	20%	74%	1%
SDN	5%	74%	20%	1%

**Table 2.** Quality of service queue percentages.

demands. Research as well as experimentation is needed to understand these problems.

## Conclusion

In this study, we discuss two approaches to dynamically partition capacity between the circuit and packet sections of a hybrid network, known as vertical stacking and horizontal partitioning. Vertical stacking — as in a packet-over-circuit network — enables backup circuits (e.g., wavelengths) to support packet (such as IP) traffic while ensuring survivability of both packet and circuit services. In horizontal partitioning, we have shown that in hybrid networks of the ESnet type, capacity over links in the packet section can be borrowed to support or protect circuit traffic. In this manner, bandwidth can be flexibly migrated between the packet and circuit sections to achieve significant cost savings. Finally, we discuss some state-of-the-art control mechanisms (e.g., in ESnet) that enable a limited degree of capacity partitioning between packet (IP) and circuit (wavelength) services. However, control mechanisms that enable a high degree of capacity partitioning as discussed need to be developed.

## References

- [1] J. Simmons, *Optical Network Design and Planning*, Springer, 2008.
- [2] A. Chiu et al., "Network Design and Architectures for Highly Dynamic Next-Generation IP-Over-Optical long Distance Networks," *IEEE/OSA J. Lightwave Tech.*, vol. 27, 2009, pp. 1878-90.
- [3] J. Baliga et al., "Green Cloud Computing: Balancing Energy in Processing, Storage, and Transport," *Proc. IEEE*, vol. 99, 2011, pp. 149-67.
- [4] J. Berthold et al., "Optical Networking: Past, Present, and Future," *IEEE/OSA J. Lightwave Tech.*, vol. 26, 2008, pp. 1104-18.
- [5] C. Guok et al., "A User Driven Dynamic Circuit Network Implementation," *Proc. Distributed Autonomous Network Management Systems*

- Wksp., New Orleans, LA, Dec. 2008.
- [6] C. S. K. Vadrevu *et al.*, "Integrated Design for Backup Capacity Sharing Between IP and Wavelength Services in IP-Over-WDM Networks," *IEEE/OSA J. Opt. Commun. and Networking*, vol. 4, no. 1, 2012.
- [7] C. S. K. Vadrevu *et al.*, "Hybrid Circuit/Packet Networks with Dynamic Capacity Partitioning," *Proc. 3rd ITU-T Kaleidoscope Conf.*, Pune, India, Dec. 2010.
- [8] D. Zhou and S. Subramaniam, "Survivability in Optical Networks," *IEEE Network*, vol. 14, 2000, pp. 16–23.
- [9] E. Modiano and A. Narula-Tam, "Survivable Lightpath routing: A New Approach to the Design of WDM-based Networks," *IEEE JSAC*, vol. 20, May 2002, pp. 800–09.
- [10] [http://www.level3.com/downloads/Level\\_3\\_Network\\_map.pdf](http://www.level3.com/downloads/Level_3_Network_map.pdf).

## Biographies

CHAITANYA S. K. VADREU ([skchaitanya.vadrevu@gmail.com](mailto:skchaitanya.vadrevu@gmail.com)) is currently working as a software engineer and network researcher at Silver Spring Networks, a leading smart-grid networking company located in Silicon Valley, California. He received Ph.D. and M.S. degrees in computer science from the University of California, Davis (UC Davis) in 2012 and 2009, respectively. His research interests include smart grid networks, software-defined networking, network survivability and optimization, backbone optical WDM networks, wireless networks, and network security. He was a co-recipient of best paper award at IEEE ANTS 2011, a semi-finalist for the Corning best student paper award at OFC/NFOEC 2011, and co-recipient of the best poster paper award at IEEE ANTS 2009. He has held visiting research positions at Nanyang Technological University, Singapore, BBN Technologies, and Lawrence Berkeley National Laboratory.

MASSIMO TORNATORE ([tornator@elet.polimi.it](mailto:tornator@elet.polimi.it)) is currently working as an assistant professor in the Department of Electronics and Information of Politecnico di Milano, where he received a Ph.D. degree in information engineering in 2006 and a Laurea (M.Sc. equivalent) degree in October 2001. He also holds an appointment as visiting assistant professor in the Department of Computer Science at the University of California, Davis, where he served as a postdoc researcher in 2008 and 2009. He started his career in 2001 as an intern at CoreCom Optical Network Laboratory where he worked in collaboration with PSTS (Pirelli Submarine Telecom Systems) and TILAB (Telecom Italia Labs). During his Ph.D., he visited the Networks Laboratory at UC Davis and the Optical Networking group of CTTC laboratories. He is author of more than 130 conference and journal papers, and his research interests include design, protection, and energy efficiency in optical transport and access networks and group communication security. He has been a co-recipient of five Best Paper Awards from IEEE conferences.

CHIN GUOK ([chin@es.net](mailto:chin@es.net)) joined ESnet in 1997 as a network engineer, focusing primarily on network statistics. He was a core engineer in the test-

ing and production deployment of MPLS and QoS (Scavenger Service) within ESnet. He was the principal investigator of the ESnet On-Demand Secure Circuits and Advanced Reservation System (OSCARS) project, which enables end users to provision guaranteed bandwidth virtual circuits within ESnet, and continues to serve as the technical lead for the service. He was a core contributor to the DICE (DANTE, Internet2, CANARIE, ESnet) Inter-Domain Controller protocol, and is an active member of the Open Grid Forum Network Services Interface Working Group. He also serves as a co-chair to the Open Grid Forum On-Demand Infrastructure Service Provisioning Research Group. He has an M.S. in computer science from the University of Arizona and a B.S. in computer science from the University of the Pacific.

INDER MONGA ([inder@es.net](mailto:inder@es.net)) serves as chief technologist and area lead of network engineering, tools, and research at ESnet. He plays a key role in developing and deploying advanced networking services for collaborative and distributed "big-data" science. He has helped contribute to multiple standards in the distributed systems community with currently active roles as co-chair of the Network Services Interface working group in the Open Grid Forum. He also drives a number of initiatives in the global research and education community and is co-chair of the Next-Generation Architecture and Distributed Topology Exchange working group at the Global Lambda Integrated Facility (GLIF) consortium. His research interests include network virtualization, software-defined networking, energy efficiency, and distributed computing. He currently holds 17 patents and has over 15 years of industry and research experience in telecommunications and data networking at Wellfleet Communications, Bay Networks, and Nortel. He earned his undergraduate degree in electrical/electronics engineering from the Indian Institute of Technology, Kanpur, before graduate studies at Boston University's Electrical and Electronics Computer Science Department.

BISWANATH MUKHERJEE [F] ([bmukherjee@ucdavis.edu](mailto:bmukherjee@ucdavis.edu)) holds the Distinguished Professorship at UC Davis, where he has been since 1987, and served as chairman of the Department of Computer Science during 1997 to 2000. He was Technical Program Co-Chair of OFC 2009. He served as Technical Program Chair of IEEE INFOCOM '96. He is the Editor of Springer's Optical Networks Book Series. He serves or has served on the editorial boards of seven journals, most notably *IEEE/ACM Transactions on Networking* and *IEEE Network*. He was Steering Committee Chair and General Co-Chair of the IEEE Advanced Networks and Telecom Systems (ANTS) Conference. He is co-winner of the Optical Networking Symposium Best Paper Awards at IEEE GLOBECOM 2007 and IEEE GLOBECOM 2008. He is the author of the textbook *Optical WDM Networks* (Springer, 2006). He served a five-year term as a Founding Member of the Board of Directors of IPLocks, Inc., a Silicon Valley startup company. He has served on the Technical Advisory Board of a number of startup companies in networking, most recently Teknovus, Intelligent Fiber Optic Systems, and LookAhead Decisions, Inc. (LDI).

---

# A Tale of the Tails: Power-Laws in Internet Measurements

**Aniket Mahanti, University of Auckland**  
**Niklas Carlsson, Linköping University**  
**Anirban Mahanti, NICTA**  
**Martin Arlitt, HP Labs and University of Calgary**  
**Carey Williamson, University of Calgary**

---

## Abstract

Power-laws are ubiquitous in the Internet and its applications. This tutorial presents a review of power-laws with emphasis on observations from Internet measurements. First, we introduce power-laws and describe two commonly observed power-law distributions, the Pareto and Zipf distributions. Two frequently occurring terms associated with these distributions, specifically heavy tails and long tails, are also discussed. Second, the preferential attachment model, which is a widely used model for generating power-law graph structures, is reviewed. Subsequently, we present several examples of Internet workload properties that exhibit power-law behavior. Finally, we explore several implications of power-laws in computer networks. Using examples from past and present, we review how researchers have studied and exploited power-law properties. We observe that despite the challenges posed, power-laws have been effectively leveraged by researchers to improve the design and performance of Internet-based systems.

---

**P**ower-laws are observed in many naturally occurring phenomena (e.g., earthquakes, precipitation, topography), as well as in many human-related behaviors (e.g., citations, urban population, wealth). Power-laws have been observed in many aspects of information systems, including software systems and computer networks. Early examples include memory referencing behavior in virtual memory systems, database queries, and file usage patterns in file systems. More recently, several characteristics of the Internet and the web have also been claimed to exhibit power-law characteristics such as the number of visitors to a web site [1], the number of hyperlinks to a web page [1], the sizes of web objects [2], the number of links to routers on the Internet [3], and the number of friends of users on online social networks [4].

Power-law properties typically appear in high variance distributions wherein observations span many orders of magnitude, particularly if there is a pronounced skew of the distribution. Compared to exponential distribution, which has been widely used in mathematically modeling telecommunication systems, power-law distributions decay more slowly. The presence of power-laws indicate that arbitrarily large values can occur with a non-negligible probability, and therefore, rather than ignoring these extreme values as “outliers,” it is useful to study their statistical prevalence if sufficiently many such samples are present in a large dataset.

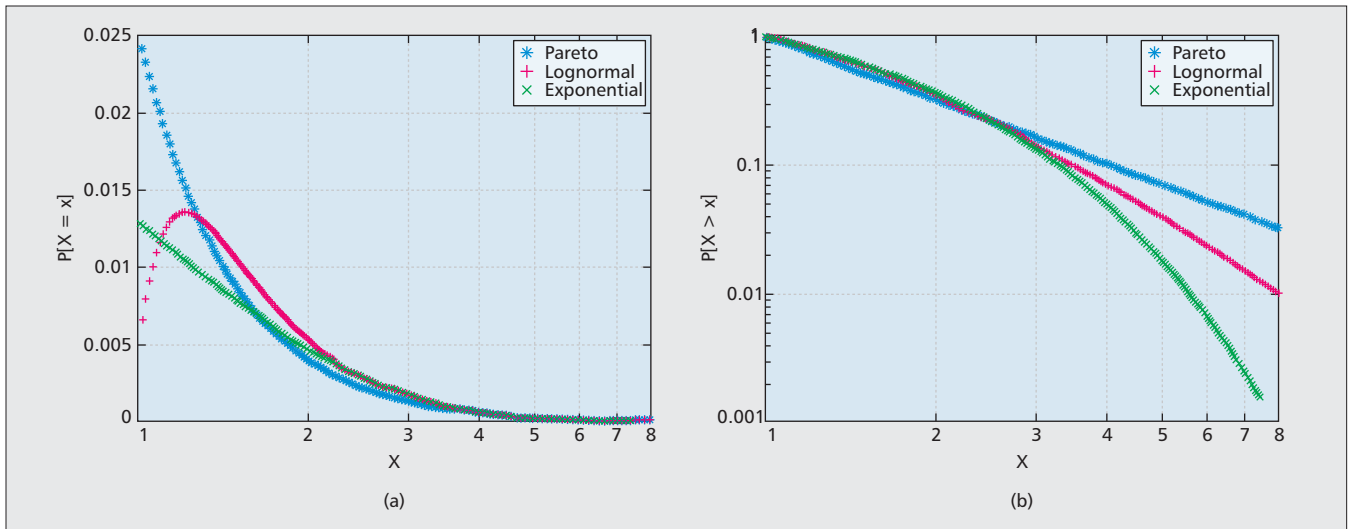
The apparent abundance of power-law distributions in computing (and other) literature has drawn significant interest in understanding the origin and implications of these power-law properties. For example, it has led to improved web caching

policies, better traffic routing and load balancing techniques, smarter search schemes, and sophisticated network topology generators. The ubiquity of power-laws has also sparked interest in developing models that generate power-law distribution, often with the goal of gaining insights on the processes behind the occurrence of power-laws. The presence of power-laws and the accuracy of these models have been debated [5]. This debate has been fueled by the discovery of measurement artifacts and the difficulty of deploying proper sampling techniques in large-scale systems. Due to the presence of many other highly skewed distributions, another active discussion topic is how to best identify the presence of power-laws from measurement data [2].

In this article, we review power-law relationships reported in the Internet measurement literature. We define power-law relationships in general, discuss approaches to identifying the presence of power-laws, and discuss two commonly used power-laws, the Pareto and Zipf distributions. We provide examples of Internet measurements that suggest power-law behavior and discuss several examples from the literature highlighting how researchers have leveraged power-laws in an effective way to improve the design and performance of Internet-based systems and applications.

## Power-Law Relationships

A power function is a scale-invariant function,  $f(x)$ , of the form  $f(x) = \alpha x^{-\eta}$ , where  $\alpha$  and  $\eta$  are positive constants, and  $\eta$  is called the scaling exponent. Taking logarithms on both sides of the power function produces  $\log(f(x)) = -\eta \log(x) + \log(\alpha)$ .



**Figure 1.** Comparison of the Pareto vs. lognormal and exponential distributions. In this example, the Pareto, exponential, and lognormal distributions have shape parameters of 1.67, 0.96, and 0.98, respectively. Note the scale of the axes in the plot; Figure 1a uses log-scale on the  $x$  axis, while Fig. 1b uses log-scales on both the  $x$  and  $y$  axes.

This expression exhibits a linear relationship with slope  $-\eta$  and  $y$ -intercept  $\log(\alpha)$ . When plotted on a log-log scale, the function appears as a straight line. This observance is often considered the distinctive feature of a power-law relationship.

In the computing literature a dataset is often said to follow a power-law if for large values the distribution follows the power function. More formally, a distribution is considered power-law if [2]

$$f(x) \sim x^{-\eta} \quad (1)$$

where  $\sim$  is used to indicate asymptotic proportionality; that is,

$$\frac{f(x)}{x^{-\eta}} \rightarrow c,$$

for some constant  $c > 0$ , when  $x \rightarrow \infty$ . In other words, the power-law distribution exhibits the power function for large values of  $x$ , typically referred to as the *tail of the distribution*.

### Pareto Distribution

One commonly observed power-law distribution in Internet traffic measurements is the Pareto distribution. A random variable  $X$  is said to follow a Pareto distribution if the complementary cumulative distribution function (CCDF) indicating the probability of occurrence of an event being greater than  $x$  is inversely proportional to a power of  $x$ ; that is,  $P[X > x] \propto x^{-\kappa}$ , where  $\kappa$  is called the shape parameter. A property of power functions is that the integral of a power function is also a power function. Due to this property, it is easy to show that the Pareto distribution (which itself has a power-law shape) and the power-law distribution are related by  $\kappa = \eta - 1$ .

Figure 1 illustrates the Pareto, exponential, and lognormal distributions. Figure 1a shows the probability distribution function. Figure 1b shows the CCDF on doubly logarithmic scales. We used the following shape parameters: 1.67 for Pareto, 0.96 for exponential, and 0.98 for lognormal. We note that the tail of the Pareto distribution gradually tapers off when compared to the exponential distribution. Note that in Fig. 1b, the lognormal distribution appears to exhibit a linear relationship. In fact, there has been some debate on how to best determine whether a dataset follows lognormal, power-law, or other related distributions [1, 2]. In many cases, it is indeed difficult to ascertain whether or not a distribution is power-law unless we observe a straight line across several

orders of magnitude on a log-log scale. These debates have also resulted in development of more sophisticated methods for identification of power-laws [1].

### Zipf Distribution

Another classical example of a power-law is the Zipf distribution, which was first used for modeling word frequencies in written texts, but has since been used to model the skewed popularities for library books, movies rentals, and web objects. The Zipf distribution is a discrete distribution, defined in the rank-frequency domain by Zipf's law, which states that when items are ranked ( $\mathbb{R}$ ) in descending order of their popularities, the frequency ( $\mathbb{F}$ ) of the item is inversely proportional to the rank of the item:

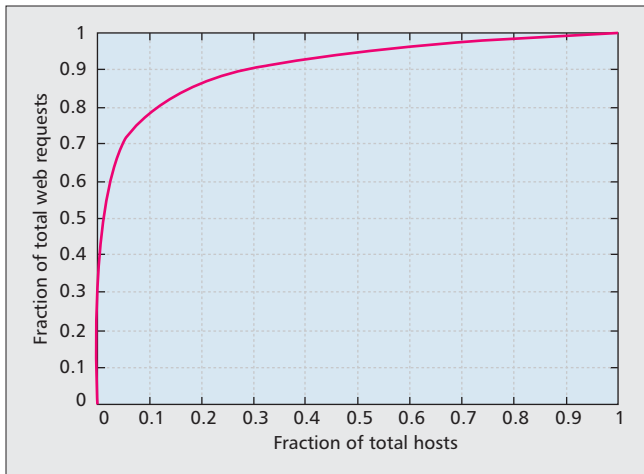
$$\mathbb{F} \propto \mathbb{R}^{-\theta} \quad (2)$$

The Zipf distribution exhibits a straight line with slope  $-\theta$  on a log-log rank-frequency plot. The value of  $\theta \approx 1$  for a pure Zipf distribution, but other values are possible while still exhibiting a straight-line behavior. Degenerate forms of the Zipf distribution, in which the behavior is piecewise linear, or only linear for a portion of the plot, are also often seen in Internet measurements. For example, the popularity of files in a peer-to-peer file sharing system have been found to be Zipf-like with deviation from the expected straight line for the most popular files arguably because of users' "fetch-at-most-once" approach to file sharing [6].

The Zipf distribution may be considered to be a discrete interpretation of the Pareto distribution. It can be represented by transforming the axes of the Pareto distribution. Thus, the Zipf distribution can be written as a Pareto distribution as follows:  $\mathbb{R} \propto \mathbb{F}^{(-1/\theta)}$ . To summarize, the Pareto, power-law, and Zipf parameters are related as

$$\kappa = \eta - 1 = \frac{1}{\theta} \quad (3)$$

The Zipf distribution has a strong skew of references to a small but highly popular set of items. For example, it is not uncommon for a small subset of the items (e.g., 10–20 percent) to account for most of the referencing activity (e.g., 80–90 percent). The exact trade-off depends on the shape parameter. In general, in many empirical studies similar skews have been observed, as shown for example in Fig. 2 for hosts



**Figure 2.** The Pareto principle: The illustration shows the distribution of requests to the WWW 2007 conference Web server over a one-year period. We observe that the top 10 percent of the hosts account for 80 percent of the total web requests. It exemplifies the Pareto principle, where most of the web requests are made by a few hosts.

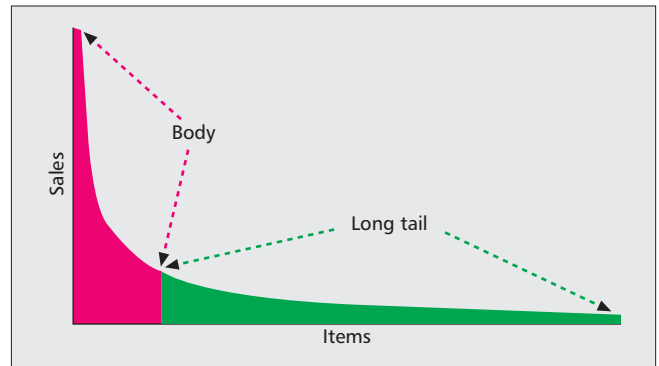
making requests to a web server. This phenomenon is often referred to as *Pareto's Law*, the *Pareto Principle*, or the 80/20 rule in the literature.

### Heavy-Tailed Distributions

Typical empirical distributions from Internet measurements can be divided into two parts: the *body* (small to medium-sized values that are responsible for much of the distribution) and the *tail* (large-sized values that are responsible for the rest). A probability distribution is said to have a heavy tail if the tail is not exponentially bounded. As illustrated in Fig. 1b, the Pareto distribution is a heavy-tailed probability distribution. A tail can be Pareto distributed (and heavy-tailed) even if the body of a distribution does not follow the power-law distribution. Such distributions are analyzed by looking at the tail of the distribution. Heavy tails are a subject of interest because they highlight the presence of large-sized values. Change in occurrence of these (less frequent) large-sized values affects the distribution more than change in the (abundant) small-sized values. This has an impact on modeling of the empirical distribution.

### The Long Tail

The *long tail* is a manifestation of power-law relationships. A long tail exemplifies the statistical property that there are many more low-frequency events compared to a Gaussian (normal) distribution. For example, it has been argued that keywords used for searches often have a long tail. This means that if we order the keywords, from most popular to least popular, based on how many times a keyword was used, we would find that there are only a few keywords that are often used and that there is a very long list of infrequently used keywords. Since the list of keywords is very large, these infrequently occurring keywords could together account for a large fraction of keyword searches seen by a search engine. This term came into popular parlance from an article written by Chris Anderson in *Wired* magazine (October 2004) where he argued that online businesses such as Amazon, eBay, and Netflix have successfully leveraged the long tail. Anderson argued that these online businesses carry a wide variety of products, each of which may appeal to only a few customers. This is in contrast to standard retailers, which mostly offer popular items because they are restricted by the size of their store inventory. Figure 3



**Figure 3.** The long tail of item sales: This illustration is based on the proposition by Chris Anderson that sales of niche items (area in green) collectively can earn more revenue than sales of the few popular items (area in red). As more sales are derived from the tail section, the body of the distribution becomes smaller. This illustration has been adapted from the public domain picture by Hay Kranen available at Wikipedia.org.

shows an illustration of the long tail phenomenon; the niche products sold would be in the green region or the long tail of the sales popularity distribution. As more sales are derived from the tail region, the body of the distribution becomes smaller. We note that the large number of items and their low individual popularity pose some technical challenges as discussed later in this article.

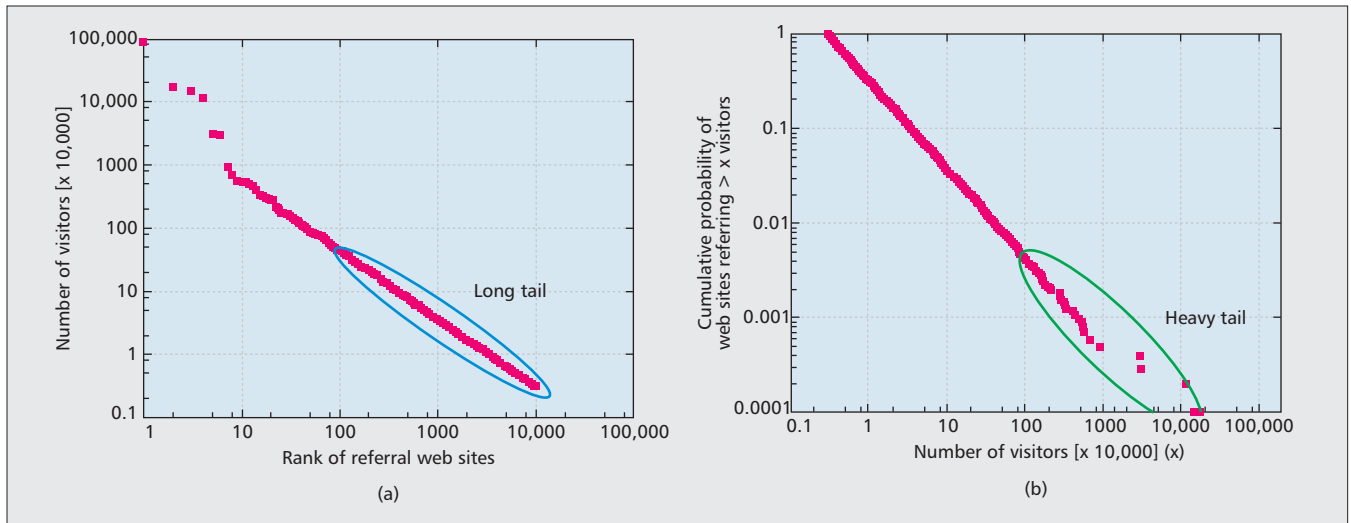
### Keeping Track of the Tails

The observant reader may have noticed that the long and heavy tails associated with the power-law distribution, as described above, refer to two different ends of the distribution. This is often forgotten and simply discussed as *the tail*. While both the Pareto and Zipf distributions are power-law, we note that the way these distributions are plotted (and defined) often focuses on two different parts of the distributions. In particular, the tail of the Pareto probability distribution refers to the rare (but probable) occurrences of events with high values, whereas the tail of the Zipf distribution refers to the many occurrences of events with small values. This subtle but important difference is illustrated in Fig. 4. Here, we note that the body of one distribution makes up for the tail of the other distribution, and vice versa. What further confuses this discussion is that a subclass of heavy-tailed probability distributions, defined in the probability domain, is sometimes referred to as long-tailed. This is in sharp contrast to the long tails referred to in the popular literature, which typically refers to the rank-frequency domain.

### Power-Law Examples

We review some power-law examples from Internet measurements. One widely reported example of power-laws is web objects access frequencies [7]. Another widely reported example is the web-object size distribution, which has been shown to follow the Pareto distribution [2]. There has also been some debate as to whether web-object sizes are power-law or follow other related distributions [2].

More recently, studies of YouTube video popularity (e.g., [8]) have found that video popularity, both at an edge network and as observed by the YouTube servers, appears to follow Zipf-like distributions. There are, however, noticeable differences at the tail of the distribution. Power-law characteristics have also been identified for other user-generated video sharing services, with the number of short-term video views exhibiting power-law characteristics, while long-term video



**Figure 4.** Tracking the tails: The illustration shows the distribution of visitors arriving at YouTube from referring web sites. Figure 4a shows that the number of visitors to YouTube is directly proportional to the ordered rank of the referring web site. Figure 4b shows that there are a few sites that provide a bulk of the referrals to YouTube. The data was collected from Compete.com, which omitted values for referrals that resulted in less than 3000 visitors arriving at YouTube.

views are better modeled using a power-law distribution with an exponential cutoff.

Other works have analyzed Internet TV workloads and found that the popularity of TV channels can be approximated using a Zipf-like distribution [9]. This observation highlights user proclivity toward watching the same channel. The channel popularity distribution followed the Pareto principle, with the top 10 percent of channels accounting for nearly 80 percent of viewers. While the audience demographics changed, the Pareto principle was found to be consistently true through different times of the day.

Various Internet-related power-law structures have also been identified. One such example is the number of (online) friends per user in online social networks, which has been shown to follow power-law distributions [4]. Another interesting observation for these networks is that there typically is a highly connected (core) group of users. This significantly reduces the number of friend-of-friends needed to connect to arbitrary people. Similar observations have been made for real-world networks (e.g., the web and offline social networks). For these networks, the highly connected core allows for efficient dissemination of information and data.

### A Generative Model for Power-Laws

Despite being widely observed, the origin of power-laws is an open problem and an active discussion topic. Generative models have been developed in order to understand the underlying processes that cause the observed power-laws to occur. The *preferential attachment* (or “rich-get-richer”) model [10] is one particularly popular generative model, although other models have also been developed [2]. The preferential attachment model has received much attention in the context of graph structures in which the vertices have a power-law degree distribution.

As an example, consider web pages and the hyperlinks among them. The web pages may be viewed as vertices of a graph and the hyperlinks as directed edges between the vertices. A simple preferential attachment model is as follows. Suppose that we begin with a single page with a hyperlink to itself. At each time step, a new page is created, and this page is assumed to create a new hyperlink. The new link is formed to one of the existing pages, chosen uniformly at random with probability  $p < 1$  and chosen proportionally to the number of

incoming links with probability  $1 - p$ . Iterating this process for many vertex additions generates a power-law graph.

Figure 5 shows a comparative illustration of a power-law and a random graph, each consisting of 150 vertices. The power-law graph was created using the preferential attachment model, where each vertex creates one outbound edge at a time. The random network is based on the Erdos-Renyi model where the probability that any two vertices are connected have an equal probability  $p$ . The figure illustrates how the high-degree vertices in the power-law graph can be critical for good connectivity and may make the network sensitive to attacks (e.g., targeted node elimination). Similar to the power-law structures discussed here, and as observed in various social and physical networks, preferential attachment and rich-get-richer behavior have been considered as a potential explanation for content popularity and web server workloads, among many other things.

### Some Implications of Power-Laws

#### Web Caching

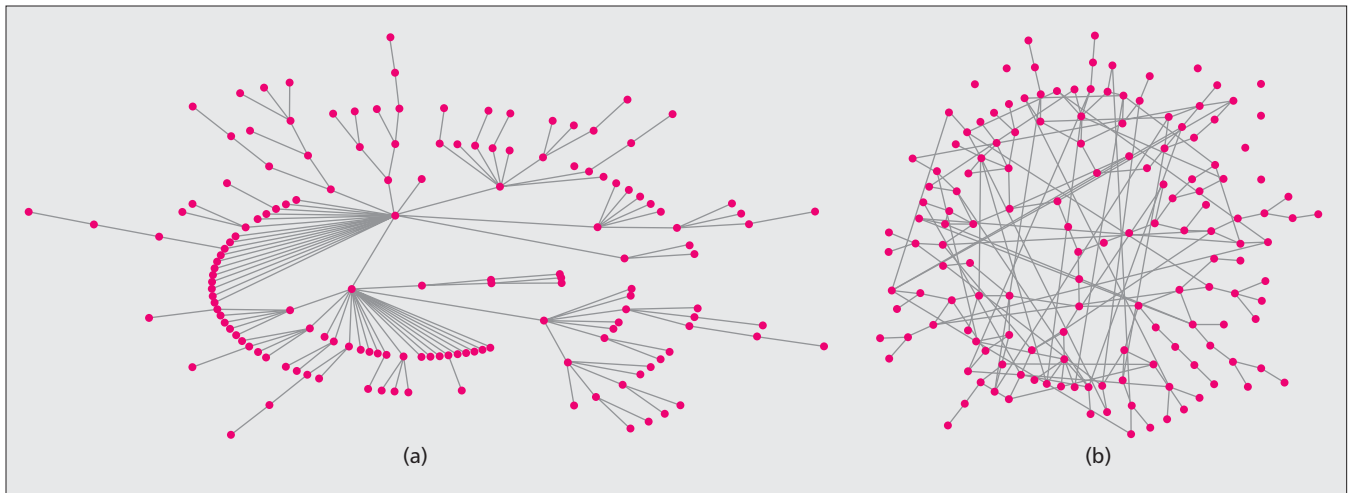
Effective web caching relies on the presence of a non-uniform popularity distribution of web objects and their sizes. Web accesses have been shown to follow Zipf’s law [7]. This property has proved important in the design of web cache architectures, since it allows designers to calculate approximate cache sizes to achieve desired hit rates. The appropriate cache size along with the appropriate replacement policy could achieve high cache hit rates.

Zipf’s law can be useful for predicting the probability of access of an object. Researchers have found that deploying caching hierarchies may be undesirable as they suffer from diminishing returns on hit rates. This is because the objects most worth caching are cached multiple times at levels closest to the users. Furthermore, deeper caching hierarchies may increase document access latencies. Content delivery networks (CDNs) can take advantage of more proactive delivery schemes.

#### Search Schemes

Power-law node connectivity distribution has helped improve search in the web. As described in the previous section, the web can be considered a directed graph of web pages and hyperlinks. Measurements of the structure of the web graph





**Figure 5.** Comparison of power-law and random graphs: Each graph consists of 150 vertices. A vertex is represented by a red dot and the edge is shown using a solid grey line. The graphs were simulated using the NetworkX package in Python, and visualized using Graphviz.

have identified the presence of key nodes. On any given topic, some web pages may have a high out-degree and others may have high in-degree [11]. Web pages that work as information aggregators typically have high out-degree and are referred to as hubs, whereas web pages with high in-degree are typically referred to as authorities. The idea of hubs and authorities was used for ranking web pages in one of the early search mechanisms [11]. Similar approaches have also been exploited for searching in online social networks and peer-to-peer (P2P) systems.

#### *The Long Tail and Business Practices*

Online businesses have taken advantage of mildly popular items to increase sale volumes. Both Amazon and Netflix have developed smarter recommendation schemes that expose users to items of personal interest based on purchasing history of the user and other users with similar interest. This allows them to potentially recommend niche content to a customer. This is in contrast to search engines that use popularity measures to rank web pages, and users formulate their decisions based on the top-ranked pages.

#### *The Long Tail and System Design*

The long tails observed in power-law can impact system efficiency. For example, a new object storage system has been designed to optimize Facebook's Photos application with the aim of serving the long tail (requests for less popular photos) [12]. These optimizations are important, as requests from the long tail accounted for a significant amount of their traffic, and the low individual request rates and high miss rates caused most of these requests to be served from the main photo storage server rather than by Facebook's content delivery network (CDN).

The long tail also poses challenges in the context of P2P file sharing systems such as BitTorrent. In particular, mildly popular files may not have sufficient popularity to have an active torrent. One approach to improve the resulting file availability problem is to group multiple files into bundles such that the bundles become sustainable torrents [13]. There are both static and dynamic bundling approaches, including adaptive bundling policies, that take the current file popularities into consideration.

#### *Measurement Issues*

Large-scale graphs such as the Internet topology or online social networks present many measurement challenges. Some

of these challenges have helped fuel the debate about the authenticity of the power-law nature of Internet graphs. For example, the scale of these networks often limits the fraction of the network that can be captured. It has been suggested that the bias in the partial crawling of online social graphs, which results in only a subset of the graph being captured, can underestimate the power-law scaling exponent [4]. To alleviate this problem, recently a multidimensional random walk algorithm [14] was proposed, which captured dynamic real world networks better and reduced scaling exponent estimation errors.

Other researchers have suggested that incomplete measurement data (e.g., using traceroute) result in missing large numbers of Internet topology connections, causing it to exhibit power-law behavior. The debate regarding whether the power-law degree distribution is an integral property of the Internet topology or an artifact of biased sampling has attracted significant attention [3, 5, 15]. Such deliberations point toward the challenges in measurement and accentuates the need for appropriate sampling techniques.

The hazards of improper sampling have also been investigated and discussed in the contexts of access patterns and other workloads that have been argued to possess power-law characteristics. Ideally, content popularity measurements should be based on probability sampling methods with known biases, such that the underlying distribution can be recreated based on the samples. Unfortunately, probability sampling is often difficult to apply to large-scale dynamic systems. The impact of sampling methods is embodied by the differences in the content popularity distribution observed when applying different definitions of popularity, sampling methods, or the length of the measurement interval.

#### *Conclusions*

Power-laws are apparent in several aspects of Internet measurements. Power-laws pose some challenges; however, they have been efficaciously leveraged by researchers in design and optimization of Internet-based systems. We describe a simple generative power-law model, although several other models exist in the literature that can lead to power-law behavior. We touch upon the ongoing debate regarding the authenticity of the power-law nature of some Internet attributes. Nevertheless, these deliberations point to the significance of power-laws in computer networks, which cannot be ignored.

---

## References

- [1] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-Law Distributions in Empirical Data," *SIAM Review*, vol. 51, no. 4, 2009, pp. 661–703.
- [2] M. Mitzenmacher, "A Brief History of Generative Models for Power Law and Lognormal Distributions," *Internet Mathematics*, vol. 1, no. 2, 2003, pp. 226–51.
- [3] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On Power-law Relationships of the Internet Topology," *Proc. ACM SIGCOMM Conf.*, Cambridge, MA, 1999.
- [4] A. Mislove et al., "Measurement and Analysis of Online Social Networks," *Proc. ACM SIGCOMM Int'l. Measurement Conf.*, San Diego, CA, 2007.
- [5] Q. Chen et al., "The Origin of Power Laws in Internet Topologies Revisited," *Proc. IEEE INFOCOM*, New York, NY, 2002.
- [6] K. P. Gummadi et al., "Measurement, Modeling, and Analysis of a Peer-to-Peer File-Sharing Workload," *Proc. ACM Symp. Operating Systems Principles*, Bolton Landing, NY, 2003.
- [7] L. Breslau et al., "Web Caching and Zipf-Like Distributions: Evidence and Implications," *Proc. IEEE INFOCOM*, New York, NY, 1999.
- [8] P. Gill et al., "Youtube Traffic Characterization: A View from the Edge," *Proc. ACM SIGCOMM Int'l. Measurement Conf.*, San Diego, CA, 2007.
- [9] M. Cha et al., "Watching Television Over an IP Network," *Proc. ACM SIGCOMM Int'l. Measurement Conf.*, Vouliagmeni, Greece, 2008.
- [10] A.-L. Barabasi and R. Albert, "Emergence of Scaling in Random Networks," *Science*, vol. 286, no. 5439, 1999, pp. 509–12.
- [11] J. Kleinberg, "Authoritative Sources in a Hyperlinked Environment," *Proc. ACM-SIAM Symp. Discrete Algorithms*, Philadelphia, PA, 1998.
- [12] D. Beaver et al., "Finding a Needle in Haystack: Facebook's Photo Storage," *Proc. USENIX Symp. Operating System Design and Implementation*, Vancouver, Canada, 2010.
- [13] D. S. Menasche et al., "Content Availability and Bundling in Swarming Systems," *Proc. ACM Conf. Emerging Networking Experiments and Technologies*, Rome, Italy, 2009.
- [14] B. Ribeiro and D. Towsley, "Estimating and Sampling Graphs with Multidimensional Random Walks," *Proc. ACM SIGCOMM Int'l. Measurement Conf.*, Melbourne, Australia, 2010.
- [15] W. Willinger, D. Alderson, and J. Doyle, "Mathematics and the Internet: A Source of Enormous Confusion and Great Potential," *Notices of the AMS*, vol. 56, no. 5, 2009, pp. 586–99.

## Biographies

ANIKET MAHANTI (aniket.mahanti@gmail.com) is a lecturer in the Department of Computer Science at the University of Auckland, New Zealand. He holds a B.Sc. (Honors) in computer science from the University of New Brunswick, Canada, and his M.Sc. and Ph.D. in computer science from the University of Calgary, Canada. His research interests include Internet traffic measurement and performance evaluation.

NIKLAS CARLSSON is an assistant professor at Linköping University, Sweden. He received his M.Sc. degree in engineering physics from Umeå University, Sweden, and his Ph.D. in computer science from the University of Saskatchewan, Canada. He has previously worked as a postdoctoral fellow at the University of Saskatchewan and as a research associate at the University of Calgary. His research interests are in the areas of design, modeling, characterization, and performance evaluation of distributed systems and networks.

ANIRBAN MAHANTI is a principal researcher at NICTA, Australia. He received his B.E. degree in computer science and engineering from the Birla Institute of Technology (at Mesra), India, and his M.Sc. and Ph.D. degrees in computer science from the University of Saskatchewan. His research interests are in the areas of network measurements, network protocols, performance evaluation, and distributed systems.

MARTIN ARLITT is a senior research scientist at Hewlett-Packard Laboratories (HP Labs) in Palo Alto, California, where he has been working since 1997. His general research interests are workload characterization and performance evaluation of distributed computer systems. He is the creator of the ACM SIGMETRICS GreenMetrics workshop, and Chair of the IEEE Computer Society's Committee of Special Technical Communities. Since 2001 he has been living in Calgary, Alberta, where he is currently an adjunct assistant professor in the Department of Computer Science at the University of Calgary.

CAREY WILLIAMSON is a professor in the Department of Computer Science at the University of Calgary. He holds a B.Sc. (Honors) in computer science from the University of Saskatchewan, and a Ph.D. in computer science from Stanford University, California. His research interests include Internet protocols, wireless networks, network traffic measurement, network simulation, and web performance.

Register for  
Your Full Conference  
Package by  
**19 February**  
and **Save Up to \$115!**

# GO BEYOND.

Beyond Copper. Beyond 100G. Beyond Next Gen.

750+ Technical Presentations/550+ Exhibits: Cloud and Data Center Networking • Space Division Multiplexing • 100G/400G Network Design and Optimization • 1Tb and Beyond Optical Networking • Wavelength Agile Access Networks • Flexible Grid Networks • Virtualization and Software Defined Networks (SDN) • High-Speed Photonic Integration for Coherent Detection • Convergence of Optical and Wireless Networks • And More!

[www.ofcnfoec.org](http://www.ofcnfoec.org)

**Advancing optical solutions in telecom, datacom, computing and more!**

Sponsored by:



Technical Conference 17-21 March  
Exposition 19-21 March  
Anaheim Convention Center  
Anaheim, CA USA

SAMSUNG



The Next Big Thing Is Here

Samsung GALAXY S III

 /SamsungMobileUSA

Copyright © 2012 Samsung Telecommunications America, LLC. Samsung and Galaxy S are registered trademarks of Samsung Electronics Co., Ltd. Other company names, product names and marks mentioned herein are the property of their respective owners and may be trademarks or registered trademarks. Appearance of phone may vary. Phone screen images simulated.