Big Data: Transforming the Design Philosophy of Future Internet

Hao Yin, Yong Jiang, Chuang Lin, Yan Luo, and Yunjie Liu

Abstract

Big data opens the era of the fourth paradigm for science discovery through data-driven computing. This new paradigm applies to the design of the future Internet, which currently faces issues in supporting new applications, efficient resource utilization, and continuous evolvement. We observe several technological transformations in network architecture, services, and applications, and point out the grand opportunities for designing future Internet architecture, communication models, and resource management mechanisms enabled by the availability of massive network data. In particular, we envision in the future Internet: 1) computational complexity replaces state complexity in the control plane; 2) data intelligence enables user choices and rewards innovations; and 3) correlations from data analytics help solve inherently hard optimization problems. Finally, we identify the key challenges in data-driven Internet design and outline future research directions.

The advent of the Internet has been a transformative event, changing our lives economically, technologically, and socially. While its overall architecture is an undeniable success, the current Internet is too expensive, too complicated to manage, too prone to vendor lock-in, and too rigid to evolve [1]. The latest trends in network expansion, resource provisioning, and usage scenarios have introduced new challenges in network science and engineering, summarized as follows.

Availability: The availability of both network infrastructure and services is essential as the scale and variety of network applications outweigh network speed upgrades. For example, video content streaming will represent 86 percent of global Internet traffic by 2016, but the Internet was historically designed for data file transfers, thus having difficulties in providing highly available video services. Furthermore, mobile Internet and heterogeneous devices impose many challenges on achieving availability. The future Internet should support extensible network infrastructure, versatile connected network devices, and uninterrupted web and mobile services. Security threats must be contained and eliminated promptly to minimize their adverse effects. The future Internet architecture should be resilient and able to recover from faults, maintaining its availability.

Efficiency: The future Internet is to deliver customized information in an increasingly effective manner. Social network applications help personalize content and services, make the content consumption highly selective, and require the network to be agile in content delivery. Facebook's 1.3 billion users install about 20 million applications each day [2]. Internet data center (IDC) investment is increasing, and networks have become software programmable thanks to decoupled control and data plane protocols (e.g., OpenFlow). Such software defined networks (SDNs) enable large, real-time, and automatic network resource scheduling and control [3]. To achieve resource efficiency and economic viability, the Internet architecture will maximize its utilization by allowing choices of services and matching application requirements to infrastructure through network programmability.

Evolvability: The future Internet is an ever evolving composite of devices, services, and usage models. The shaping of the Internet architecture is an open-ended process as infrastructures and applications change steadily. Since we have no access to an accurate prediction of future network services, any proposed future Internet architecture must be able to evolve over time [4]. Its adaptability and flexibility are crucial to deal with changing requirements, incorporate new functions, and accommodate continuous innovations.

Emerging trends in network technologies, computing technologies, and data intensive network science enable fundamental redesign of the future Internet that is highly available, efficient to operate, and agile to innovate. To cope with new challenges and take advantage of technological advances, we make the following key observations on the directions and opportunities for the future Internet.

Computational intelligence: Use computational complexity to compensate for or replace network state complexity. Rather than maintaining complex network states in every device for control and management, the future Internet computes the network status and applies policies at runtime programmatically on network devices with simplified data plane. Along this direction, SDN embeds computational intelligence through controllers and controller applications, which manipulate the data plane in SDN devices. The operation of a network has become a centralized application

Hao Yin, Yong Jiang, and Chuang Lin are with Tsinghua University.

Yan Luo is with the University of Massachusetts Lowell.

Yunjie Liu is with the Jiangsu Future Networks Innovation Institute.

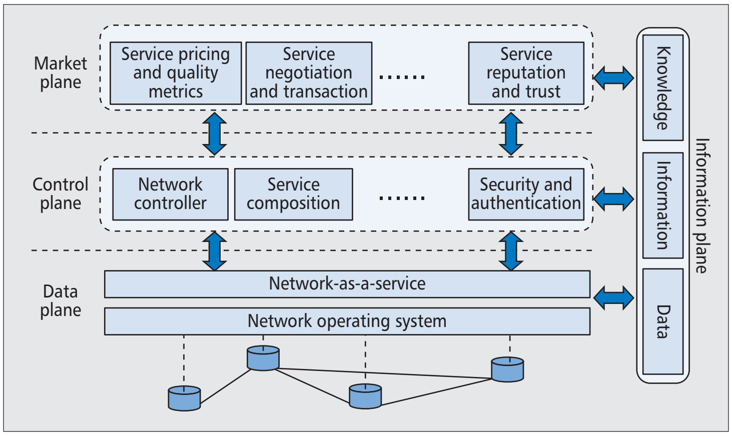

Figure 1. Data driven Internet architecture with four planes: data, control, information, and market.

running on top of network operating systems and abstractions of network devices (including switches, routers, middleboxes, etc.).

Data intelligence: Data obtained at scale bring both entropy and intelligence. Network data analytics follows the data-driven scientific discovery paradigm: comprehensive data outweigh sampled data, correlation rather than causality is sought from data analysis, and the efficiency of the network correlation discovery is emphasized over accuracy. For example, users' quality of experience (QoE) is a direct feedback on network performance and more valuable than low-level network metrics. Leveraging such QoE to infer network properties and optimize web applications is more beneficial than traditional network metrics [5]. New network designs should expose more effective quality metrics to objectively compare alternative network innovations [6].

Effectiveness intelligence: Big data in network science has enabled the methodology shift from mathematical approaches to effectively solvable, data-driven approaches for addressing issues in network resource management. This methodological change dramatically simplifies the problem formulation and speeds up the critical decision-making process, reducing cost and improving network performance. We call this direction *effectiveness intelligence* as it signifies the new trend in network design and optimization: transforming an NP hard problem to a practically solvable problem using correlation inferred from data rather than causality.

Big Data Transforming the Design Philosophy of Future Internet

The Changes of Internet Architecture

Computational Complexity Replacing State Complexity — The control plane is overwhelmed with network state information. For example, multiple routing processes on a router depend on a huge number of control knobs: route metrics, access control lists (ACLs), policies, and so on. The network state information such as dynamic states in forwarding information bases (FIBs), port settings, policies, packet filters, and timers, as well as their mutual dependencies, may cause detrimental effects: faults, instability, inconsistency, and so on. Some nontechnical restrictions such as ownership impede innovations.

Disruptive changes in the control plane determine how the future Internet can be simplified in network management and operation. The ongoing transformation is evident in data plane abstraction, control plane abstraction, as well as computational network control.

SDN architecture decouples the control and data planes by abstracting network devices with flows and actions applied on flows. Such data plane abstraction makes it possible to construct uniform control protocols (e.g., OpenFlow) and network operating systems (e.g., NOX) on top of which the control plane can be operated programmatically as opposed to via manual configurations. As a result, network intelligence and state are logically centralized in a software-based SDN controller, which maintains a global view of the network and can programmatically fulfill complex control functions.

Increasing computational capability at the network control plane drives the fundamental changes in how networks are controlled. Current distributed control planes make it challenging to realize new sophisticated control functions because distributed algorithms and protocols make it difficult to sustain state dynamics, resulting in poor extensibility. Incremental additions of ad hoc control components prevent consensus on the right protocol. On the contrary, a centralized optimization algorithm is more robust, and configuration-free protocols can provide better reachability and allow fine-grained flow control. As a result, designing a centralized control plane becomes more desirable for implementing sophisticated control logics. This control plane is logically central and can be implemented with multiple instances on multiple hardware platforms to ensure its reliability.

Data-Driven Network Design — Network design shares much synergy with social and economical functioning in how the constituents interact and influence each other. From the perspective of economics, in a perfect competition market, consumers are willing to buy cheaper and better products, while providers are encouraged to improve productivity and supply better products. It is the transparency of economical data such as pricing and quality of products that enable vendor selection and resource optimization.

We envision four essential planes of future Internet that enable the choice of network design alternatives and eventually lead to cost-effective, reliable, and agile network architecture. As illustrated in Fig. 1, the four planes are control, data, information, and economics. The data plane forwards packets and schedules flows based on rules, and the network operating systems and abstractions of network devices serve as the interface to the control plane. The control plane runs centralized network control services to apply forwarding, scheduling, and security policies onto the data plane. The information plane collects network measurement data and provides them to the control plane, which utilizes real-time network snapshots for responding to significant events and/or statistical data for periodical network tuning. The market plane hosts a marketplace where users, services, and infrastructures interact, allowing users to compare cost and quality, negotiate transactions, and pursue the most suitable matches to maximize the tangible metrics. The market plane requires information completeness, symmetry, and transparency from the information plane. The knowledge learned from the information plane drives economical functions in the market plane to shape Internet architecture advancement.

The design of the future Internet will inevitably benefit from user choices based on cost and user experiences, because choices can drive the competition and innovation necessary for future networks. The market discipline is imposed on the network operators and service providers, requiring their investment in facilities and innovation in services. Users of the network can select from a range of alternative services that

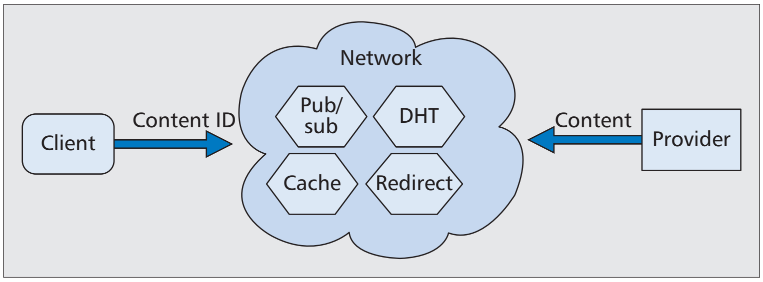

Figure 2. New communication models.

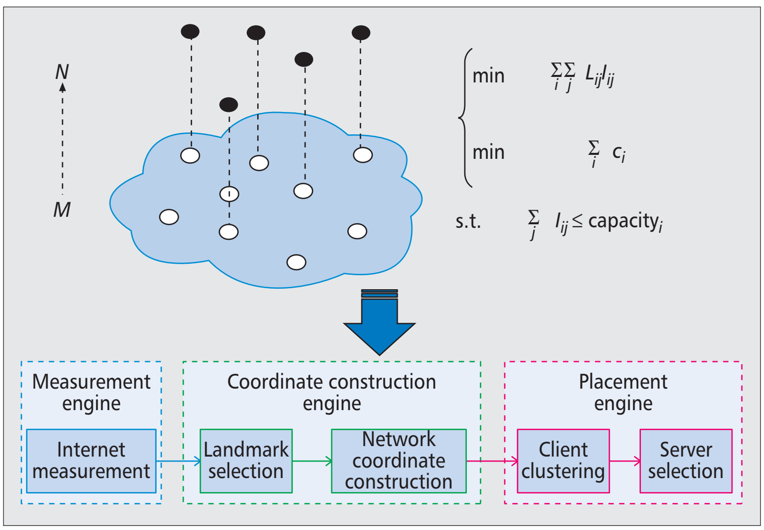

Figure 3. Change in scientific paradigm.

may differ in functionality, performance, and cost. Meanwhile, the choice mechanism also promotes the robustness and security of a network by encouraging diverse solutions at all levels of the architecture. ChoiceNet [7] is a recent example of how network architecture can provide choice as a core principle.

New Communication Models

Increasing computational power, outpacing network I/O speed [8], makes it possible to transform conventional communication models. Network operators progressively integrate computing power to alleviate the bottleneck of bandwidth resource in a network system. For example, in content distribution, application acceleration and load balancing can generate more revenue and optimize resource allocation of scarce network bandwidth. Using computing resources to compensate for network resources makes a communication model more scalable and efficient on existing infrastructure.

The Internet communication model has begun shifting from routing-centric to content/service/X-centric. Conventional routing centric communication models assume a "dumb data pipe" due to the scarcity of CPU and memory resources in network devices. With the increasing ubiquitous computing power present in the Internet, network controllers and middleboxes are exploiting such computational capabilities. As shown in Fig. 2, information centric networking (ICN) has changed how clients communicate with content providers [9]. In ICN, content identifiers, rather than server IP addresses, become the handles of requests and replies. In such a way, ICN naturally supports various functionalities including content distribution, multicast, and mobility, all of which leverage computing power. Similarly, DMap manages dynamic identifier-to-locator mapping for supporting mobility and content delivery [10].

Correlation Replaces Causality for Resource Management

Internet resource management problems are typically described with mathematical models, many of which are NP hard. The network topology of the Internet is too complicated to be measured and described. As a result, network management, task scheduling, and resource optimization are difficult to model with an ideal framework without assumptions far from reality. Even with such models, one often faces state explosion when attempting to solve them due to the distinct targets from multiple dimensions such as user requirements, network bandwidth, and geolocation. The stakeholders have objectives that may be competitive and adverse to each other, and each of these parties vies for its particular interest. Thus, it is difficult to discover the internal causality among all the competing factors in the Internet. It is desirable to consolidate the objectives and obtain an effective solution as opposed to seeking the optimal solution in an unmanageable time span, because benefits from new resource management strategies are immediate and bring new subsequent changes in user behavior, which may invalidate the original problem formulation.

We describe the implications of such transformation using a server placement example as follows. An effective media server placement strategy is a classical facility location problem with the aim of choosing M replicas or hosting services among N potential sites, as shown in the upper part of Fig. 3. It is typically modeled as a multi-objective optimization problem to solve, with the objectives of minimizing total costs and delay. Such models suffer from three major limitations:

- The assumption of a fixed candidate pool is invalidated by the rapid growth of cloud computing platforms and data centers, which introduces choices of pervasive hosting resources.
- Most of the existing solutions do not scale well with the number of server candidates and the number of clients: the K-mean problem or a facility location problem is known to be NP-hard.
- Traditional solutions fail to discover inherent rules and potential trends, and thus cannot give insight on the choice of an ideal number of servers.

Enabled by the correlation discovered through big data analytics, we reconsider the work-server placement example and propose a novel scheme called NetClust [11], shown in the lower part in Fig. 3. NetClust takes advantage of the latest network coordinate techniques to obtain global network information for server placement, and leverages a clustering algorithm to determine the correlation between deployment cost and service performance.

Challenges Faced by Data-Driven Design Philosophy

Data-Driven Internet Architecture Design

Data-driven Internet architecture remains in its early stage and faces many open questions. As essential elements of the future Internet architecture, the four planes (4P) in the envisioned data-driven Internet (Fig. 1) are dynamic in definitions, immature in theoretical foundations, and versatile in implementations.

The separation of the control and data planes has gained much traction. Most of the prior work focused on the performance of the data plane and enhanced features of the control plane until the emergence of SDN, which introduces horizontal partition of the packet processing procedures. The success of SDN brings an interesting set of new problems such as network control verification and debugging, therefore calling for broader employment of computational intelligence in network architecture thanks to the advent of massive network measurement and monitoring data. It is challenging to derive the right abstraction of control plane and data plane, and leverage them to control network operations and states.

At the information plane, the increasing volume and variety of network data and the knowledge derived from them (including network traffic variation, user access patterns, data center capacity and utilization, content generation and customization, etc.) reflect the natural interactions among user, content, and network. Many studies [7, 11] have shown that correlations learned from data can be utilized to effectively improve network resource allocation, reduce cost, and maximize revenue without resorting to formulation of NP-hard problems. Data-driven knowledge discovery of network operation and optimization guides the design of future Internet. It remains an open question as to how to expose information in a transparent manner while preserving privacy. Data heterogeneity in time scale and units requires comprehensive data curation. Accuracy-tolerant data processing also needs effective approximation and learning.

The market plane serves as the arena where users reward services and technologies that are deemed innovative and helpful. From advertisement to content delivery, the Internet has been closely engaged with economical behaviors. However, economics-driven Internet design is still in its infancy. The current Internet architecture lacks transparent pricing, as well as transaction and selection systems for users and providers to match their interest in experiences, quality, and costs. In particular, the network core is far more closed than the edge with respect to service selection. The obstacles include the unwillingness of Internet service providers to expose information, and users' inaccessibility to a large selections of services with well defined marketable metrics. Once an open, transparent, fair marketplace for network technologies and services is established, the Internet design will evolve in accordance with basic economics principles and advance through economical corrections.

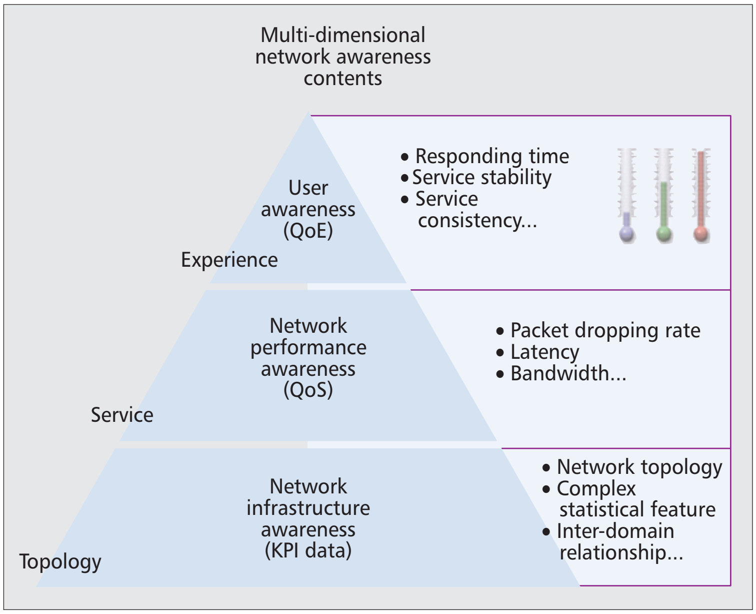

Figure 4. SITEWARE platform for multi-dimensional network measurement.

Data Acquisition

Data acquisition for information plane is required to address the volume, velocity, variety, and validation of big data. The objectives are to obtain sufficient data for understanding network infrastructure topology, network service performance, and user experiences, which are three crucial components of Internet design. Progress has been made in data models and tools for at-scale network measurements. A network coordinate system [12] models the Internet as a geometric space and characterizes the position of any node in the Internet by a coordinate in this space. IPlane [13] utilizes the Internet's routing topology to build an atlas of the Internet, and estimates the properties of the paths between arbitrary end hosts on the Internet based on the path composition technique, which derives these estimates by composing the inferred properties of either path segments or links in the atlas. However, their data acquisition schemes are not ready for comprehensive, constant, or diverse measurements.

We have established a big-data-based Internet measurement platform named SITEWARE shown in Fig. 4. By means of distributed network measuring, this platform orchestrates a variety of measurement tools, methods, and theories to evaluate network performance. Specifically, we develop our awareness of network infrastructure with topology information (AS relationship, statistical features, etc.), network performance with quality of service (QoS) data (packet loss, latency, bandwidth, etc.), and user experience with QoE information (e.g., response time and service stability). From bottom up, our corresponding goals are to optimize the Internet architecture, protocol, and resource management, and task scheduling for improved user experiences. By using this layered measuring platform, we can leverage a cross-layer method to systematically identify and address a fault. For example, if QoE is below expectation, there must be a problem in the corresponding QoS metrics. If a QoS metric is lower than a threshold, it implies that there must be an infrastructure-related problem, for example, changes in AS relationship or a bottleneck link between the source and destination.

Internet measurement is a perpetually challenging task for technical, economic, political, and policy reasons. Technically, it is challenging to obtain complete, multi-dimensional, and accurate data about the Internet due to its heterogeneity, scale, and dynamics in operations. Network malfunctions, failures, state changes, and mobility result in instability of Internet data. Accurate and large-scale Internet data collection remains a major undertaking because of the dynamics of Internet operations. Addressing issues such as data validation, sampling rate, representativeness, and security is indispensable to Internet-scale data acquisition. Economically, many network operators use autonomous systems(ASs) to shield their private information from the public. With respect to the policy reasons, Internet measurement has to be compliant with national interests and ethical requirements.

Data Utilization

As data obtained from Internet measurements feature multi-source, heterogeneity, inconsistency of entity, incompleteness, and inaccuracy, data curation needs theories and tools to support data representation, storage, integration, fusion, retrieval, and extraction. Data volume keeps expanding, and faults and

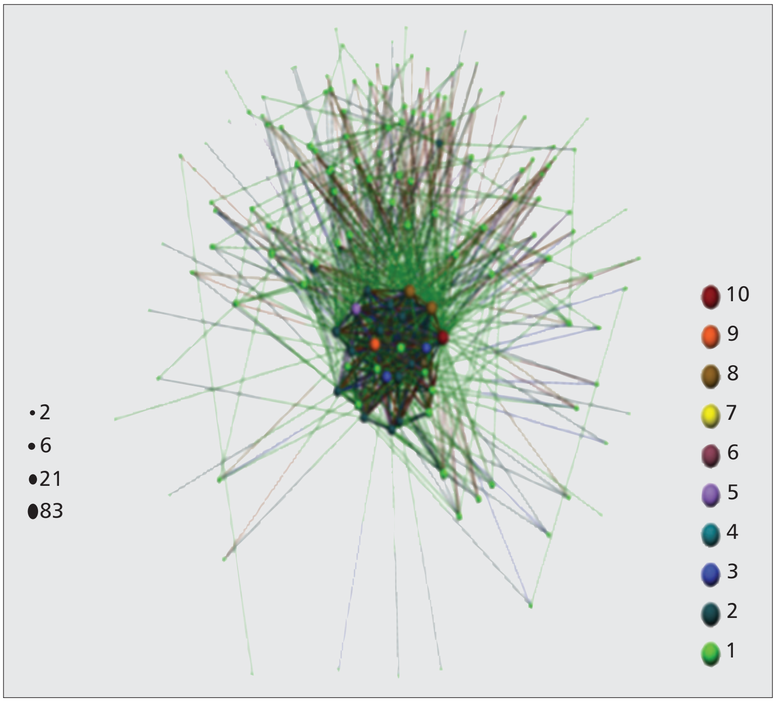

Figure 5. AS topology.

malfunctions occur frequently, complicated by the lack of a unified network information collection mechanism. The state information obtained by different operators or entities needs to be shared and assembled for a global view of the Internet; however, their proprietary interfaces prevent the community from sharing data. Timely or even real-time responses to interactive inquiries about urgent Internet events are critical. A common ontology is desirable for mapping diverse data sources to measured objects [14] and describing measurement data with accompanying metadata the structure of which can be defined with web ontology languages. Structured, semi-structured, and unstructured data about the Internet call for novel solutions to database indexing problems.

Big data reveals details we never could when we were limited to smaller quantities, and provide a clear view of the granularities, categories, and hierarchies that samples cannot disclose. Approaches such as data visualization, predicative analysis, and actionable instructions are instrumental to discovering the potential value from the data.

Visualization: Data might be obscure and hide their potential value until they are made amenable to human perception. Data visualization exhibits the correlations and implications of raw data with images and dimensions so that our eyes can see and our brains can understand the trends and connections for us to act on. For example, Fig. 5 illustrates 97 percent of the observable ASs in the Chinese Internet and their statistical features [15]. Nodes representing ASs in China are laid out across a series of two-dimensional concentric circles with their diameters inversely proportional to the corresponding coreness values, which imply the robustness of an AS. The sizes of nodes are exponentially proportional to the corresponding node degrees, and their colors help differentiate nodes of diverse coreness value.

Predictive analytics: A data-intensive computing paradigm provides an effective way to identify the mingled factors in the Internet by searching for correlations. Utilizing various machine learning and data mining algorithms, one can derive knowledge from the data to support decisions on resource management, task scheduling, production recommendation, anomaly detection, and so on.

Actionable instructions: The value of data depends highly on its usefulness in producing actionable items. These actions are directly tied to the benefits to users, ISPs, content providers, and all of society. They are intended to improve Internet performance metrics, management and control mechanisms, resource allocation policies, and so on. The complex and sometimes counter-intuitive relationships discovered from predicative analytics should furnish tangible mechanisms for Internet operators to work with, just like the interest rate and currency supply in macroeconomics.

Conclusion and Future Work

Big data brings unprecedented opportunities for reshaping the future Internet. This position article is a significant first step in highlighting this paradigm shift to address future Internet design challenges. A data-driven approach takes advantage of the massive data and ubiquitous computing capabilities to transform network architectures, communication models, and resource management. Centralized network control such as in SDN replaces distributed and autonomous subsystems. Emerging communication models such as content-centric networks compensate for inadequate network bandwidth with computational intelligence. Discovery of the correlations among the Internet components allows for scalable resource optimization, an inherently hard problem.

Data-driven network design calls for the effort to address data acquisition, curation, and utilization, and, most important, the Internet economics where transparent data empower choices of network technologies and services. There are abundant open issues in theories, mechanisms, and tools for harnessing data and fostering knowledge in this big data era, the ultimate goal of which is to derive actionable items for network design in the coming decades.

Acknowledgment

This work has been partially supported by the National Basic Research Program of China (No.2011CB302601 and No.2012CB315801), the National Natural Science Foundation of China (No.61222213 and No.61170290), and a grant from Intel Corporation.

References

[1] M. Casado *et al.*, "Fabric: A Retrospective on Evolving SDN," *HotSDN '12*, Aug. 2012, pp. 85–90.
[2] http://www.statisticbrain.com/facebook-statistics.
[3] ONF, "Software-Defined Networking: The New Norm for Networks," Apr. 2012.
[4] D. S. Han *et al.*, "XIA: Efficient Support for Evolvable Internetworking," *NSDI '12*, Apr. 2012.
[5] X. Liu *et al.*, "A Case for A Coordinated Internet Video Control Plane," *ACM SIGCOMM '12*, Aug. 2012, pp. 359-70.
[6] A. Balachandran *et al.*, "A Quest for an Internet Video Quality-of-Experience Metric," *HotNet '12*, Oct. 2012, pp. 97–102.
[7] T. Wolf *et al.*, "Choice as a Principle in Network Architecture," *ACM SIGCOMM '12*, Aug. 2012, pp. 105–06.
[8] D. S. Han, *Supporting Long Term Evolution in an Internet Architecture*, Carnegie Mellon Univ., 2012.
[9] B. Ahlgren *et al.*, "A Survey of Information-Centric Networking," *IEEE Commun. Mag.*, vol. 50, no. 7, July 2012, pp. 26–36.
[10] T. Vu *et al.*, "DMap: A Shared Hosting Scheme for Dynamic Identifier to Locator Mappings in the Global Internet," IEEE ICDCS'12, June 2012, pp. 698–707.
[11] H. Yin *et al.*, "NetClust: A Framework for Scalable and Pareto-Optimal Media Server Placement," *IEEE Trans. Multimedia*, vol. 15, no. 8, Dec. 2013, pp. 2114–24.

[12] T. S. E. Ng and H. Zhuang, "Global Network Positioning: A New Approach to Network Distance Prediction," *IEEE INFOCOM '02*, June 2002, pp. 170–79.
[13] H. V. Madhyastha *et al.*, "IPlane Nano: Path Prediction for Peer-to-Peer Applications," NSDI '09, vol. 9, Apr. 2009, pp. 137–52.
[14] http://ercim-news.ercim.eu/en77/special/unified-access-to-internet-measurement-data.
[15] H. Yin *et al.*, "Discovering a Large Scale Internet Topology: Complementary and Contrast View," *TrustCom '12*, June 2012, pp. 1491–98.

Biographies

HAO YIN (h-yin@tsinghua.edu.cn) is a professor in the Research Institute of Information Technology (RIIT) at Tsinghua University. His research interests span broad aspects of multimedia networks, future networks, and big-data-driven network science and engineering. Some of his research results have been widely used in industry and adopted by industry standards. He has published over 100 papers in refereed journals and conferences.

YONG JIANG (scottjiang@mail.tsinghua.edu.cn) is a postdoctoral fellow in the RIIT at Tsinghua University. He received his Ph.D. degree in computer science from Beijing University of Posts and Telecommunications in 2012. His research interests focus on future Internet architecture, network economics, and network resource optimization.

CHUANG LIN is a professor in the Department of Computer Science and Technology at Tsinghua University. His research interests include computer networks, performance evaluation, network security analysis, and Petri net theory and its applications.

YAN LUO is an associate professor in the Department of Electrical and Computer Engineering at the University of Massachusetts Lowell. His research focuses on software defined networking, heterogeneous computing architecture, and embedded systems.