Layering As Optimization Decomposition: Current Status and Open Issues

Mung Chiang, Steven H. Low, A. Robert Calderbank, John C. Doyle

Abstract—Network protocols in layered architectures have historically been obtained on an ad-hoc basis, and much of the recent cross-layer designs are conducted through piecemeal approaches. Network protocols may instead be holistically analyzed and systematically designed as distributed solutions to some global optimization problems in the form of generalized Network Utility Maximization (NUM), providing insight on what they optimize and structures of the network protocol stack. This paper presents a short survey of the recent efforts towards a systematic understanding of "layering" as "optimization decomposition", where the overall communication network is modeled by a generalized NUM problem, each layer corresponds to a decomposed subproblem, and the interfaces among layers are quantified as functions of the optimization variables coordinating the subproblems. Furthermore, there are many alternative decompositions, each leading to a different layering architecture. Industry adoption of this unifying framework has also started. Here we summarize the current status of horizontal decomposition into distributed computation and vertical decomposition into functional modules such as congestion control, routing, scheduling, random access, power control, and coding. Key messages and methodologies arising out of many recent work are listed. Then we present a list of challenging open issues in this area and the initial progress made on some of them.

Keywords: Adaptive coding, Cross-layer design, Congestion control, Distributed algorithm, Lagrange duality, MAC, Network utility maximization, Optimization, Power control, Reverse engineering, Routing, TCP/IP, Scheduling, Stochastic control, Wireless ad hoc networks.

I. OVERVIEW

Layered architectures form one of the most fundamental and influential structures of network design. It adopts a modularized and often distributed solution approach to network coordination and resource allocation. Each layer controls a subset of the decision variables, and observes a subset of constant parameters and the variables from other layers. Intuitively, layered architectures enable a scalable, evolvable, and implementable network design while introducing potential risks to manageability of the network. There are clearly more than one way to "divide and conquer" the network design problem. From a data-plane performance point of view, some layering schemes may be more efficient or more fair than others. Focusing on such resource allocation functionalities and using only performance metrics, this paper examines the question of "how to" and "how

M. Chiang is with Electrical Engineering Department, Princeton University, chiangm@princeton.edu. S. H. Low is with Computer Science and Electrical Engineering Departments, Caltech, slow@caltech.edu. A. R. Calderbank is with Electrical Engineering and Mathematics Departments, Princeton University, calderbk@princeton.edu. J. C. Doyle is with Control and Dynamic Systems Department, Caltech, doyle@cds.caltech.edu.

not to" layer. The limitations of our focus, in terms of semantics functionalities and "network X-ities" metrics, will be discussed at the end of the paper.

Each layer in the protocol stack hides the complexity of the layer below and provides a service to the layer above. While the general principle of layering is widely recognized as one of the key reasons for the enormous success of data networks, there is little quantitative understanding to guide a systematic, rather than an ad hoc, process of designing layered protocol stack for wired and wireless networks. One possible perspective to rigorously and holistically understand layering is to integrate the various protocol layers into a single coherent theory, by regarding them as carrying out an asynchronous distributed computation over the network to implicitly solve a global optimization problem. Different layers iterate on different subsets of the decision variables using local information to achieve individual optimality. Taken together, these local algorithms attempt to achieve a global objective. Such a framework of "layering as optimization decomposition" exposes the interconnection between protocol layers and can be used to study rigorously the performance tradeoff in protocol layering, as different ways to modularize and distribute a centralized computation. Even though the design of a complex system will always be broken down into simpler modules, this theory will allow us to systematically carry out this layering process and explicitly trade off design objectives.

The key idea in "layering as optimization decomposition" is as follows. Different *decompositions* of an optimization problem, in the form of a generalized Network Utility Maximization (NUM), are mapped to different *layering* schemes in a communication network, with each decomposed subproblem in a given decomposition scheme corresponds to a layer, and functions of primal or Lagrange dual *variables* coordinating the subproblems correspond to the *interfaces* among the layers. Since different decompositions lead to alternative layering architectures, we can also tackle the question "how to and how not to layer" by investigating the pros and cons of decomposition techniques. Furthermore, by comparing the objective function values under various forms of optimal decompositions and suboptimal decompositions, we can seek "separation theorems" among layers: conditions under which layering incurs no loss of optimality. Robustness of these separation theorems can be further characterized by sensitivity analysis in optimization theory: how much will the differences in the objective value (between different layering schemes) fluctuate as constant parameters in the generalized NUM formulation are perturbed.

The mentality of "network as an optimizer" and "protocol as a distributed solution" to some global optimization problem

(in the form of the basic NUM) has been successfully tested in trials for Transmission Control Protocol (TCP) [20]. The key innovation from this line of work [22], [27], [28], [36], [37], [39], [46] is to view TCP/IP network as an optimization solver and each variant of congestion control protocol as a distributed algorithm solving a specified basic NUM, where the objective is the sum of source utilities as functions of rates, the constraints are linear flow constraints, and optimization variables are source rates. Other recent results also show how to reverse engineer Border Gateway Protocols (BGP) as solving the Stable Path Problem [16], and contention-based Medium Access Control (MAC) protocols as game-theoretic selfish utility maximization [31], [50]. Starting from a given protocol originally designed based on engineering heuristics, reverse engineering discovers the underlying mathematical problems being solved by the protocols and demonstrates the application of derived insights through forward engineering improvements of the protocols.

These reverse engineering successes provide one of the justifications to employ generalized versions of NUM for systematic cross-layer design. Furthermore, utility of allocated resources to end users and elasticity of application traffic can both be modeled through general utility functions. As optimization's objective, utility functions provide a metric to define optimality of resource allocation efficiency, while different shapes of utility functions lead to optimal resource allocations that satisfy some definition of fairness (*e.g.*, α-fair utilities parameterized by $\alpha > 0$: $U(x) = (1 - \alpha)^{-1} x^{1-\alpha}$ [39] leads to α-fair resource allocation). In general, utility functions can be coupled across the users. They may depend on not just rates, but also other metrics such as reliability, latency, jitter, and energy.

While the application needs give rise to the objective function, *i.e.*, network utility to be maximized, restrictions in the communication infrastructure are translated into many constraints of a generalized NUM problem. The resulting problem may be a very difficult nonconvex optimization with integer constraints. These generalized NUM problems put the end user utilities at the "driver's seat" for network design. For example, benefits of innovations in physical layers through better modulation and coding schemes are now characterized by the enhancement to applications rather than just the drop in bit error rates, which the users do not directly observe. An optimal solution to a generalized NUM formulations automatically establishes the benchmark for all layering schemes. Indeed, layering is a human engineering effort, the problem itself does not have any pre-determined layering architecture.

How to attain an optimal solution to a generalized NUM in a modularized and distributed way then becomes an overarching question. Vertical decompositions across modules and horizontal decompositions across disparate network elements can be conducted systematically through the theory of decomposition for nonlinear optimization. Implicit or explicit message passing quantifies the amount of information sharing and decision coupling required for a particular decomposition. There are many different ways to decompose a given problem, each of which corresponds to a different layering architecture. Even a different representation of the same NUM problem can lead to different decomposability structures even though the optimal

solution remains the same. These decompositions, *i.e.*, layering schemes, have different characteristics in efficiency, robustness, asymmetry of information and control, and tradeoff between computation and communication. Some are "better" than others depending on the criteria set by the network users and managers. A systematic exploration in the space of alternative decompositions is possible, where each particular decomposition represents a holistically designed protocol stack.

Given the layers, crossing layers is tempting. For example, layers can be crossed for wired or wireless networks in at least the following ways:

- Information may be passed from one layer to another. For example, a TCP proxy at the base station of a wireless cellular network may be informed by the physical layer that a packet loss is due to channel fading and not congestion.
- Information from one layer may be used in another layer to either adapt its existing algorithm or create new diversity. For example, if the medium access layer informs the routing layer about its performance, multipath routing may be used to provide spatial diversity.
- Tasks may be jointly accomplished across the layers. For example, joint routing in the network layer and data compression in application layer may leverage the spatial redundancy in the sensor network measurements to reduce the network traffic load.
- Tasks may be re-divided among the layers. For example, error correction is performed in different forms in each of the application, transport, network, link, and physical layers. The task of ensuring the accuracy of the received bits may be re-allocated across the layers and some error checking functions may be removed from certain layers.

As evidenced by the large and ever growing number of papers on cross layer design over the last few years, we expect that there will be no shortage of cross layer ideas based on piecemeal approaches. The growth of the "knowledge tree" on cross layer design has been exponential. However, any piecemeal design jointly over multiple layers does not bring more structured thinking process than the ad hoc design of just one layer. What seems to be lacking is a level ground for fair comparison among the variety of cross layer designs, a unified view on how to and how not to layer, basic principles rigorously quantified, and fundamental limits on the impacts of layer-crossing on network performance and robustness metrics.

"Layering as optimization decomposition" provides a candidate for such a unified framework. It attempts at shrinking the "knowledge tree" on cross layer design rather than growing it. It is important to note that "layering as optimization decomposition" is *not* the same as the generic phrase of "cross-layer optimization". What is unique about this framework is that it views the network as the optimizer itself, puts the end user application needs as the optimization objective, provides the globally optimal performance benchmark, and leads to a systematic design of decomposed solution to attain the benchmark. Carrying the intellectual thread from "forward engineering" (solve a given problem) to "reverse engineering" (find the problem being solved by a given protocol) one step further to "design for optimizability", it may be that the difficulty of solving a particular set of subproblems also illustrates that the given decompo-

sition was conducted in a wrong way and suggests that better alternatives exist.

The power of "layering as optimization decomposition" has been illustrated through many case studies carried out by various research groups in the last couple of years, generating considerable general insights in addition to the specific cross-layer designs. The summary lists of key messages and methodologies in the next section illustrate the conceptual simplicity in this rigorous and unifying framework, which is more important than any specific cross layer design derived from the framework, such as some examples summarized in the Appendix.

Industry adoption of "layering as optimization decomposition" has already started. For example, insights from reverse-engineering TCP has lead to an improved version of TCP implemented over the last several years: FAST (Fast AQM Scalable TCP) [13], [20]. Putting end-user application utilities as the objective function has lead to a new way to leverage innovations in the physical and link layers beyond the standard metrics such as bit error rate, *e.g.*, in "FAST Copper" Project (here FAST stands for Frequency, Amplitude, Space, Time) for an order-of-magnitude boost to rates in fiber/DSL broadband access systems [14].

II. Current Status

A. Network Utility Maximization

The basic NUM problem is the following formulation [22], known as monotropic programming and studied since 1960s. TCP variants have recently been reverse engineered to show that they are implicitly solving this problem, where source rate vector $\mathbf{x}$ is the only optimization variables, and routing matrix $\mathbf{R}$ and link capacity vector $\mathbf{c}$ are both constants:

$$\begin{array}{ll} \text{maximize} & \sum_s U_s(x_s) \\ \text{subject to} & \mathbf{R}\mathbf{x} \preceq \mathbf{c}. \end{array} \tag{1}$$

Utility functions U_s are often assumed to be smooth, increasing, concave, and depends on local rate only, although recent investigations have removed some of these assumptions for applications there they are invalid. Utility functions can be picked based on any of the following five grounds: reverse-engineering (a given protocol description implicitly dictates the underlying utility function), user perception behavior models, application traffic elasticity, efficiency of resource allocation, and fairness among competing users.

Many of the papers on "layering as optimization decomposition" are special cases of the following generic problem [5], one of the possible formulations of a generalized NUM for the entire protocol stack:

$$\begin{array}{ll} \text{maximize} & \sum_s U_s(x_s, P_{e,s}) + \sum_j V_j(w_j) \\ \text{subject to} & \mathbf{R}\mathbf{x} \le \mathbf{c}(\mathbf{w}, \mathbf{P}_e), \\ & \mathbf{x} \in \mathcal{C}_1(\mathbf{P}_e), \ \ \mathbf{x} \in \mathcal{C}_2(\mathbf{F}), \\ & \mathbf{R} \in \mathcal{R}, \ \ \mathbf{F} \in \mathcal{F}, \ \ \mathbf{w} \in \mathcal{W}. \end{array} \tag{2}$$

Here, x_s denotes the rate for source s and w_j denotes the physical layer resource at network element j. The utility functions U_s and V_j may be any nonlinear, monotonic functions. $\mathbf{R}$ is the routing matrix, and $\mathbf{c}$ are the logical link capacities

as functions of both physical layer resources $\mathbf{w}$ and the desired decoding error probabilities $\mathbf{P}_e$. The issue of signal interference and power control can be captured in this functional dependency. The rates must also be constrained by the interplay between channel decoding reliability and other error control mechanisms like ARQ. This constraint set is denoted as $\mathcal{C}_1(\mathbf{P}_e)$. The issue of rate-reliability tradeoff and coding is captured in this constraint. The rates are further constrained by the medium access success probability, represented by the constraint set $\mathcal{C}_2(\mathbf{F})$ where $\mathbf{F}$ is the contention matrix. The issue of packet collision and medium access control is captured in this constraint. The sets of possible physical layer resource allocation schemes, of possible scheduling or contention based medium access schemes, and of single-path or multi-path routing schemes are represented by $\mathcal{W}, \mathcal{F}, \mathcal{R}$, respectively. The optimization variables are $\mathbf{x}, \mathbf{w}, \mathbf{P}_e, \mathbf{R}, \mathbf{F}$. Holding some of the variables as constants and specifying some of these functional dependencies and constraint sets will then lead to a special class of this generalized NUM formulation.

A deterministic fluid model is used in the above formulations. Stochastic network utility maximization is an active research area, as discussed in Section III, where stochastic models are imposed at session, packet, channel, and topology levels, raising new questions such as stochastic stability, average optimality, and outage performance.

Whether it is the basic, general, or stochastic NUM, there are three steps in the process: first formulate a specific NUM problem, then devise a modularized and distributed solution following a particular decomposition, and finally explore the space of alternative decompositions that provide a choice of layered protocol stack and coupling across the layers.

In general, there are two types of objective functions: sum of utility functions by end users, which can be functions of rate, reliability, delay, jitter, or power level, and a network-wide cost function by network operators, which can be functions of congestion level, energy efficiency, network lifetime, or collective estimation error. Some of these utility functions may not have an additive structure. Maximizing a weighted sum of all utility functions, which is the focus of this paper, is only one of the possible formulations. An alternative is multi-objective optimization to characterize the Pareto-optimal tradeoff between the user objective and operator objective. Another formulation is game-theoretic between users and operators, or among users or operators themselves.

B. Reverse Engineering of Individual Layers

In the terminology of the standard seven layer reference model, it is well-known that physical layer algorithms try to solve the data transmission problem formulated by Shannon: maximizing data rate subject to vanishing error probability constraints. Recent progress have put protocols in layers 2-4 of the standard reference model on a mathematical foundation as well:

- The congestion control functionality of TCP has been reverse engineered to be implicitly solving the basic NUM problem (1). While *heterogeneous* congestion control protocols do not solve an underlying NUM problem, its

equilibrium and dynamic properties can still be analyzed through a vector field representation and Poincare-Hopf index theorem [51], which show that bounded heterogeneity implies global uniqueness and local stability of network equilibrium.

- IGP of IP routing is known to be variants of shortest path routing solvers, and the policy-based routing protocol in BGP has recently been modeled as the solution to the Stable Path Problem [16].
- Scheduling based MAC protocols are known to be solving variants of maximum weight matching, and random access (contention based MAC) protocols have recently been reverse engineered as a non-cooperative selfish utility maximization game [31], [50].

C. Forward Engineering: Cases of Systematic Cross-layer Design

Following is a *non-exhaustive* list of some of the recent publications using "layering as optimization decomposition" [1], with four examples to be explained in some detail in the Appendix. In all these cases, a NUM problem that is more complicated than the basic NUM represents a more general networking problem encompassing more than congestion control, and some functions of the Lagrange dual variables act as the "layering variables".

- Jointly optimal congestion control and adaptive coding or power control [5], [30]
- Jointly optimal congestion and contention control [4], [21], [32], [52], [58], [59]
- Jointly optimal congestion control and scheduling [12]
- Jointly optimal routing and scheduling [26]
- Jointly optimal routing and power control [42], [55]
- Jointly optimal congestion control, routing, and scheduling [3], [34]
- Jointly optimal routing, scheduling, and power control [9], [54]
- Jointly optimal routing, resource allocation, and source coding [57]
- TCP/IP interactions [53], [18] and jointly optimal congestion control and routing [23]
- Network lifetime maximization [40]

D. Alternative Decompositions

The basic idea of decomposition is to divide the original large problem into smaller subproblems, which are then coordinated by a master problem by means of some kind of signalling. Most of the existing decomposition techniques can be classified into *primal decomposition* and *dual decomposition* methods [2]. The former is based on decomposing the original primal problem, whereas the latter based on decomposing the Lagrange dual of the problem. Primal decomposition methods have the interpretation that the master problem directly gives each subproblem an amount of resources that it can use; the role of the master

[1] We apologize in advance for any references we may have missed and would appreciate any information about other citations.

[2] This is not to be confused with primal-dual interior-point algorithm, or primal driven network control, or primal penalty function approach.

problem is then to properly allocate the existing resources. In dual decomposition methods, the master problem sets the price for the resources to each subproblem which has to decide the amount of resources to be used depending on the price; the role of the master problem is then to obtain the best pricing strategy. Primal decomposition and dual decomposition can in fact be inter-changed by introducing auxiliary variables [43].

Almost all the papers in the vast, recent literature on NUM use a standard dual-based distributed algorithm. Contrary to the apparent impression that such a decomposition is the only possibility, there are in fact many alternatives to solve a given network utility problem in different but all distributed manners [43], including multi-level and partial decompositions. Each of the alternatives provides a possibly different tradeoff among three important considerations: convergence speed, amount and asymmetry of message passing's communication overhead, and architecture of distributed computation. There is no universally "best" way to distribute the solution process across a network: which alternative is the most desirable depends on the specific problem formulation and application.

Coupling for generalized NUM can happen not only in constraints, but also in the objective function, where the utility of source s, $U_s(x_s, \{x_i\}_{i \in I(s)})$, depends on both its local rate x_s and the rates of a set of other sources with indices in set $I(s)$. If U_s is an increasing function of $\{x_i\}_{i \in I(s)}$, this coupling models cooperation in a clustered system, otherwise it models competition such as power control in wireless network or spectrum management in DSL. Such coupling in the objective function can be decoupled [49] by first introducing auxiliary optimization variables and consistency equality constraints, thus shifting coupling in objective to coupling in constraints, then introducing "consistency prices" to decouple the consistency constraints. These consistency prices are iteratively updated through local message passing.

E. Key Messages

More than just an ensemble of specific cross-layer designs for existing protocol stacks, "layering as optimization decomposition" is a *mentality* that views networks as optimizers, a *common language* that allows researchers to quantitatively compare alternative network architectures, and a suite of *methodologies* that facilitates a systematic design approach for modularized resource allocation. Some of the key messages and methodologies that have been obtained from many case studies are outlined in the following lists.

- Protocols in layers 2,3,4 can be *reverse engineered*. Reverse engineering in turn leads to better design in a rigorous manner.
- There is a *unifying approach* to cross-layer design, as summarized in Section I of this paper.
- Loose coupling through "layering price" can be optimal, and congestion price (or queuing delay, or buffer occupancy) is often the right "layering price" for stability and optimality, with important *exceptions* as well.
- There are many *alternatives* in decompositions, leading to different divisions of tasks across layers and even different time-scales of interactions.

- *Convexity* of the generalized NUM is the key to devising a *globally optimal* solution.
- *Decomposability* of the generalized NUM is the key to devising a *distributed* solution.

F. Key Methodologies

- Dual decomposition for linear coupling constraints.
- Consistency pricing for coupled objective functions.
- Descent lemma for proof of convergence of dual-based distributed subgradient algorithm.
- Stability proof through Lyapunov function construction, singular perturbation theory, and passivity argument.
- Log change of variables to turn multiplicative coupling into linear coupling, and to turn nonconvex constraints to convex ones.
- Sufficient conditions on curvature of utility functions for it to remain concave after a log change of variables.
- Construction of conflict graph, contention matrix, and transmission modes in contention based MAC design.
- Maximum differential congestion pricing for node-based back-pressure scheduling (part of the connections between distributed convex optimization and stochastic control).

III. OPEN ISSUES

Despite the variety of progress made along this research direction, there remain a number of important open issues as outlined in this section under six groups.

A. Modeling Challenges

BGP for inter-AS routing is still difficult to be fully incorporated in the generalized NUM framework. Similarly, an optimization-based, unifying view on wireless ad hoc network routing is lacking. Much further work remains to be done to model utility functions in specific applications, especially inelastic, real-time applications such as VoIP and streaming media [19]. In a more refined physical/link layer model, the option of forwarding rather than re-encoding at intermediate nodes must be considered, as well as retransmission through ARQ.

B. Transient Behavior Characterization

For certain applications, if the resource allocation (*e.g.*, window size, signal-to-interference-ratio) for a user drops below a threshold during the transient, the user may be disconnected. In such cases, the whole idea of equilibrium becomes meaningless. Bounding transient behavior, as well as providing tight estimates of the rate of convergence for popular iterative, distributed algorithms, remain under-explored topics.

C. Alternative Decomposition

Even a different representation of the same primal problem may change the duality and decomposability structures even though it does not change the optimal solution. It remains an open issue on how to systematically explore the space of alternative vertical and horizontal decomposition, thus the space of alternative network architectures, for a given set of requirements on rate of convergence, symmetry of computational load distribution, and amount of explicit message passing.

D. Stochastic NUM

When sessions (*i.e.*, flows, connections, end-users) arrive and depart, packets come in bursts, channels vary over time, and topology subject to change, new formulations of stochastic NUM become necessary, presenting new challenges on stability and performance characterization. Most of the known results concern stochastic stability and validity of the deterministic fluid model, with little characterization on the distribution of utility induced by the distributions of stochastic models at various levels.

Session level. For Poisson arrivals of sessions with exponentially distributed file size, [1], [10], [38] showed that, for certain classes of utility functions under the time-scale separation assumption [3], the stability region of the basic NUM is the largest possible, which is the capacity region formed by the fixed link capacities. Then [33], [47] extended this stochastic stability result to the case without the time-scale separation assumption. Extensions have recently been carried out to other models [48], [56], with fluid limits and diffusion approximations proposed as well [24], [25]. In [35], stochastic stability and optimality (in the expected sense) for general, constrained convex optimization is proved for any concave utilities and without time-scale separation.

Packet level. There have been two major approaches that appeared over the last two years: translating on-off HTTP session utility into transport layer TCP utility (mapping from microscopic to macroscopic model) [2], and showing many-flow asymptotical validation of fluid model (justifying the transition from microscopic to macroscopic model) [11], [45].

Channel level. Channel variations offer both the challenge to prove stability/optimality for existing algorithms and the ability to do opportunistic transmission and scheduling. In [3], *stability* and *optimality* are established for dual algorithms under channel-level stochastic for any convex optimization where the constraint set has the following structure: a subset of the variables lie in a polytope and other variables lie in a convex set that vary according to an irreducible, finite-state Markov chain. "Layering as optimization decomposition" type of algorithms that only require instantaneous knowledge of the current channel state (*e.g.*, queue-lengths) remain stable and optimal (in the expected sense).

Topology level. Very little has been explored on this topic, which is important for battery based or highly mobile wireless ad hoc networks.

E. Nonconvex NUM

Non-zero duality gaps may arise due to a variety of reasons: integer constraints (*e.g.*, in single path routing, admission control, scheduling, algebraic coding, constellation size), nonconcave utilities (*e.g.*, power efficiency or some empirically verified utility curves), and constraints describing nonconvex sets. A nonzero duality gap means that the standard dual-based distributed subgradient algorithm, and in general dual decomposition approaches, may lead to suboptimal and even infeasible primal solutions and instability in cross layer interactions.

---

[3] Here time-scale separation means that the resource allocation algorithm converges before the number of sessions changes.

This very difficult problem can be tackled through a combination of well-established and more recent optimization techniques (*e.g.*, sum-of-squares programming [44] and geometricsignomial programming [6]). For example, there have been three recent approaches to solve nonconcave utility maximization over linear constraints:

1) [29] proposes a distributed, suboptimal heuristics (for sigmoidal utilities) called "self-regulating" heuristics, which is shown to avoid link congestion caused by sigmoidal utilities. It attains the optimal rate allocation $\mathbf{x}^*$ in the asymptotic case when the proportion of sources with nonconcave utilities vanishes.

2) [17] determines optimality conditions for the dual-based distributed algorithm to converge globally (for all nonlinear utilities). The engineering implication is that appropriate overprovisioning of link capacities will ensure global convergence of the dual-based distributed algorithm even when user utility functions are nonconcave.

3) [15] develops an efficient but centralized method to compute the global optimum (for a wide class of utilities that can be transformed into polynomial utilities), using the sum-of-squares method.

F. Network X-ities

Protocol design and layering architecture are not just for maximizing the efficiency of performance metrics, such as throughput, latency, distortion, but also robustness metrics, such as evolvability, scalability, and manageability. Interactions among layers introduce the risks of losing robustness against unforseen demands arising over time or significant growth over space. Despite the importance in practical network operations, these network X-ities remain as important yet fuzzy notions, and a quantified foundation for them is long overdue [8]. Intuitively, "design by decomposition" enhances scalability and evolvability, but may present risks to manageability such as diagnosability and optimizability. Quantifying network X-ities and trading-off network X-ities with performance metrics in layered protocol stack design is a long-term, challenging direction.

IV. CONCLUSION

"Layering as optimization decomposition" is a unifying framework for understanding and designing distributed control and cross-layer resource allocation in wired and wireless networks. It has been developed by various research groups over the last several years, and is now emerging to provide a mathematically rigorous and practically relevant approach to quantify the risks and opportunities of modifying existing layered network architecture. It shows that network protocols in layers 2, 3, and 4 can be reverse-engineered as implicitly solving some optimization-theoretic or game-theoretic problems. By distributively solving generalized NUM formulations through decomposed subproblems, we can systematically generate layered protocol stacks. There are many alternatives for both horizontal decomposition into disparate network elements and vertical decomposition into functional modules (*i.e.*, layers). While queuing delay or buffer occupancy is often used as the "layering

price", it may sometimes lead to unstable interactions. A variety of techniques to tackle coupling and nonconvexity issues have become available. A more detailed survey of the recent efforts to establish 'layering as optimization decomposition' as a common "language" for systematic network design can be found in [7].

REFERENCES

[1] T. Bonald and L. Massoulie, "Impact of fairness on Internet performance", *Proc. ACM Sigmetrics*, pp. 82-91, 2001.

[2] C. S. Chang and Z. Liu, "A bandwidth sharing theory for a large number of HTTP-like connections," *IEEE/ACM Trans. on Networking*, vol. 12, no. 5, Oct. 2004.

[3] L. Chen, S. H. Low, M. Chiang, and J. C. Doyle, "Optimal joint congestion control, routing, and scheduling design for wireless ad hoc networks," *Proc. IEEE INFOCOM,* April 2006.

[4] L. Chen, S. H. Low, and J. C. Doyle, "Joint congestion control and medium access control for wireless ad hoc networks," *Proc. IEEE INFOCOM,* March 2005.

[5] M. Chiang, "Balancing transport and physical layers in wireless multihop networks: Jointly optimal congestion and power control" *IEEE J. Sel. Area Comm.,* vol. 23, no. 1, pp.104-116, Jan. 2005.

[6] M. Chiang, "Geometric programming for communication systems", *Trends and Foundations in Communications and Information Theory*, vol. 2, no. 1, pp. 1-156, Aug. 2005.

[7] M. Chiang, S. H. Low, R. A. Calderbank, and J. C. Doyle, "Layering as optimization decomposition", To appear in *Proceedings of IEEE*, 2006.

[8] M. Chiang and M. Yang, "Towars X-ities from a topological point of view: Evolvability and scalability", *Proc. Allerton Conf.* Oct. 2004.

[9] R. L. Cruz and A. Santhanam: "Optimal routing, link Scheduling, and power control in multihop wireless networks," *Proc. IEEE INFOCOM*, April 2003.

[10] G. de Veciana, T. J. Lee, and T. Konstantopoulos, "Stability and performance analysis of network supporting elastic services", *IEEE/ACM Trans. on Networking*, vol. 9, no. 1, pp. 2-14, Feb. 2001.

[11] S. Deb, S. Shakkottai, and R. Srikant, "Asymptotic behavior of Internet congestion controllers in a many-flow regime", *Math. Operations Research*, 2005.

[12] A. Eryilmaz and R. Srikant, "Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control," *Proc. IEEE INFOCOM*, March 2005.

[13] TCP FAST Project netlab.caltech.edu.

[14] FAST Copper Project www.princeton.edu/fastcopper.

[15] M. Fazel and M. Chiang, "Nonconcave network utility maximization through sum of squares method", *Proc. IEEE CDC*, Dec. 2005.

[16] T. G. Griffin, F. B. Shepherd, and G. Wilfong, "The stable path problem and interdomain routing," *IEEE/ACM Trans. on Networking,* vol. 10, no. 2, pp. 232-243, April 2002.

[17] P. Hande, S. Zhang, and M. Chiang, "Distributed rate allocation for inelastic flows", Submitted to *IEEE/ACM Trans. on Networking*, preliminary version in *Proc. IEEE INFOCOM*, March 2005.

[18] J. He, M. Chiang, and J. Rexford, "TCP/IP interaction based on congestion prices: Stability and optimality", *Proc. IEEE ICC*, June 2006.

[19] J. Huang, Z. Li, M. Chiang, and A. K. Katsaggelos, "Pricing-based rate control and joint packet scheduling for multi-user wireless upllink video streaming", *Proc. IEEE Packet Video Workshop*, April, 2006.

[20] C. Jin, D. X. Wei, and S. H. Low. TCP FAST: motivation, architecture, algorithms, performance. *Proc. of IEEE INFOCOM*, March 2004.

[21] K. Kar, S. Sarkar, and L. Tassiulas, "Achieving proportional fairness using local information in Aloha networks," *IEEE Trans. on Auto. Control*, vol. 49, no. 10, pp. 1858–1862, October 2004.

[22] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *J. Op. Res. Soc.*, vol. 49, no. 3, pp.237–252, March 1998.

[23] F. P. Kelly and T. Voice, "Stability of end-to-end algorithms for joint routing and rate control", *Computer Communications Review*, vol, 35, no. 2, pp. 5-12, 2005.

[24] F. P. Kelly and R. J. Williams, "Fluid model for a network operating under a fair bandwidth-sharing policy", *Annals of Applied Probability*, vol. 14, no. pp. 1055-1083, 2004.

[25] P. Key and L. Massoulie, "Fluid models of integrated traffic and multipath routing", To appear in *Queuing Systems*, 2006.

[26] M. Kodialam and T. Nandagopal, "Charaterizing achievable rates in multi-hop wireless networks: The joint routing and scheduling problem", *Proc. ACM Mobicom*, Sept. 2003.

[27] S. Kunniyur and R. Srikant. End-to-end congestion control: utility functions, random losses and ECN marks. *IEEE/ACM Trans. on Networking*, 2003.

[28] R. J. La and V. Anantharam, "Utility-based rate control in the Internet for elastic traffic," *IEEE/ACM Trans. on Networking*, vol. 10, no. 2, pp. 272-286, April 2002.

[29] J. W. Lee, R. R. Mazumdar, and N. Shroff, "Non-convex optimization and rate control for multi-class services in the Internet," *Proc. IEEE INFO-COM*, Hong Kong, China, March 2004.

[30] J. W. Lee, M. Chiang, and R. A. Calderbank, "Price-based distributed algorithm for optimal rate-reliability tradeoff in network utility maximization", To appear *IEEE Journal of Selected Areas in Communications*, 2006.

[31] J. W. Lee, M. Chiang, and R. A. Calderbank, "Utility-optimal medium access control: reverse and forward engineering," *Proc. IEEE INFOCOM*, April 2006.

[32] J. W. Lee, M. Chiang, and R. A. Calderbank, "Jointly optimal congestion and contention control in wireless ad hoc networks", To appear in *IEEE Communication Letters*, 2006.

[33] X. Lin and N. B. Shroff, "On the stability region of congestion control",, *Proc. Allerton Conference*, 2004.

[34] X. Lin and N. Shroff, "Impact of imperfect scheduling in cross layer design," *Proc. IEEE INFOCOM*, March 2005.

[35] J. Liu, M. Chiang, and H. V. Poor, "Stability of distributed dual algorithm for convex optimization", Submitted to *IEEE Conf. Control and Decision*, 2006.

[36] S. H. Low, "A duality model of TCP and queue management algorithms," *IEEE/ACM Trans. on Networking*, vol. 11, no. 4, pp. 525–536, Aug. 2003.

[37] S. H. Low and D. E. Lapsley, "Optimization flow control, I: basic algorithm and convergence," *IEEE/ACM Trans. on Networking*, vol. 7, no. 6, pp. 861–874, December 1999.

[38] L. Massoulie and J. W. Roberts, "Bandwidth sharing and admission control for elastic traffic", *Telecommunication Systems*, vol. 15, pp. 185-201, March 2000.

[39] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Trans. on Networking*, vol. 8, no. 5, pp. 556–567, Oct. 2000.

[40] H. Nama, M. Chiang, and N. Mandayam, "Utility lifetime tradeoff in self regulating wireless sensor networks: A cross-layer design approach", *Proc. IEEE ICC*, June 2006.

[41] T. Nandagopal, T. Kim, X. Gao, and V. Bharghavan, "Achieving MAC layer fairness in wireless packet networks," *Proc. ACM Mobicom*, Boston, USA, Aug. 2000.

[42] M. J. Neely, E. Modiano, and C. E. Rohrs, "Dynamic power allocation and routing time varying wireless networks", *IEEE J. Sel. Area Comm.*, vol. 23, no. 1, pp. 89-103, Jan. 2005.

[43] D. Palomar and M. Chiang, "Alternative decompositions for distributed maximization of network utility: Framework and applications", *Proc. IEEE INFOCOM*, April 2006.

[44] P. A. Parrilo, "Semidefinite programming relaxations for semialgebraic problems," *Mathematical Programming Series B*, vol. 96, no. 2, pp. 293-320, 2003.

[45] S. Shakkottai and R. Srikant, "Mean FDE models for Internet congestion control under a many-flows regime", *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 1050-1072, June 2004.

[46] R. Srikant, *The Mathematics of Internet Congestion Control*, Birkhauser 2004.

[47] R. Srikant, "On the positive recurrence of a Markov chain describing file arrivals and departures in a congestion-controlled network", *IEEE Computer Communications Workshop*, 2004.

[48] A. L. Stolyar, "Maximizing queueing network utility subject to statbility: greedy primal-dual algorithm", *Queueing Systems*, vol. 50, no. 4, pp. 401-457, 2005.

[49] C. W. Tan, D. Palomar, and M. Chiang, "Decoupling coupled network utility maximization", *Proc. IEEE ICASSP*, May 2006.

[50] A. Tang, J. W. Lee, J. Huang, M. Chiang, and R. A. Calderbank, "Reverse engineering MAC", *Proc. IEEE WiOpt*, April 2006.

[51] A. Tang, J. Wang, S. H. Low, and M. Chiang, "Equilibrium of heterogeneous congestion control protocols", *Proc. IEEE INFOCOM*, March 2005.

[52] X. Wang and K. Kar, "Cross-layer rate control for end-to-end proportional fairness in wireless networks with random access," *Proc. ACM Mobihoc*, May, 2005.

[53] J. Wang, L. Li, S. H. Low, and J. C. Doyle. "Cross-layer optimization in TCP/IP networks," *IEEE/ACM Trans. on Networking*, vol. 13, no. 3, pp. 582-595, August 2005.

[54] Y. Xi and E. Yeh, "Optimal distributed power control, routing, and congestion control in wireless networks", Submitted to *IEEE Trans. Inform. Theory*, 2005.

[55] L. Xiao, M. Johansson, and S. Boyd, "Joint routing and resource allocation via dual decomposition," *IEEE Trans. Comm.*, vol. 52, no. 7, pp. 1136-1144, July 2004.

[56] H. Ye, "Stability of data networks under optimization-based bandwidth allocation", *IEEE Trans. Auto. Control*, vo.. 48, no. 7, pp. 1238-1242, July 2003.

[57] W. Yu and J. Yuan, "Joint source coding, routing, and resource allocation for wireless sensor networks", *Proc. IEEE ICC*, May 2005.

[58] C. Yuen and P. Marbach, "Price-based rate control in random access networks," *IEEE/ACM Trans. on Networking*, vol. 13, no. 5, pp. 1027–1040, December 2005.

[59] D. Zheng and J. Zhang, "A stochastic primal-dual algorithm for joint flow control and MAC design in multihop wireless networks", Preprint, 2006.

V. APPENDIX: CASE STUDIES

Brief summaries of some case studies from our recent publications are provided to illustrate the ideas in Section II.C.

A. Jointly optimal congestion control and adaptive coding

Adaptive error correction channel coding in physical layer can change the "pipe sizes" of communication channels, adding another dimension in the "supply-demand" balance in NUM. Indeed, the concept of signal quality is absent from the basic NUM (1). The link "capacities" $\mathbf{c}$ in (1) implicitly assume fixed decoding error probabilities $\mathbf{P}_e = \{P_{e,l}\}$. In wireless networks, adaptive channel coding (including adaptive control between diversity-gain and multiplexing-gain in space-time coding) can change the rate-reliability *tradeoff*. A link can have a larger c_l and accommodate more flows, by increasing its decoding error probability $P_{e,l}$, or vice versa. A source may transmit at a higher rate if the end-to-end signal quality is allowed to degrade, *i.e.*, the end-to-end decoding error probability $P_{e,s} \approx \sum_{l \in L(s)} P_{e,l}$ is larger. Of course, each source's utility depends on both rate x_s and reliability $P_{e,s}$. In [30], we develop distributed algorithms to obtain the globally optimal rate-reliability tradeoff in the following NUM, with nonconvex constraints over variables $\{x_s, P_{e,l}\}$:

$$
\begin{array}{ll}
\text{maximize} & \sum_s U_s(x_s, P_{e,s}) \\
\text{subject to} & \sum_{s:l \in L(s)} x_s \leq c_l(P_{e,l}), \quad \forall l
\end{array}
\tag{3}
$$

where each function $c_l(P_{e,l})$ represents a nonlinear, generally nonconcave, and rather complicated dependency of a link's attainable throughput c_l on the desired decoding error probability $P_{e,l}$. We show that, when each link provides the same decoding error probability for all flows through the link (an "integrated policy" for dynamic reliability optimization), as in the formulation (3), a new distributed algorithm that uses pricing on both

rate and reliability can be proved to converge to global optimality, as long as the channel codes are strong enough. When each link provides different decoding error probabilities for different flows (a "differentiated policy"), which expands the model in (3), the problem becomes a nonconvex optimization with coupling among the terms, even for strong channel codes. Using various decoupling techniques and log change of variables, convergence to a global optimum can be proved for utility functions whose curvature is sufficiently negative $U''(x) \leq -U'(x)/x$, *i.e.*, if the traffic is elastic enough.

B. Jointly optimal congestion and contention control

In [31], we investigate joint end-to-end congestion control and per-link random access control in ad-hoc wireless networks. Using a generalized NUM formulation, we can accommodate multi-class services as well as exploit the tradeoff between efficiency and fairness of resource allocation by adjusting the types of utility functions. Define $L_{out}(n)$ as a set of outgoing links from node n, and $N_{to}^I(l)$ as the set of nodes whose transmission cause interference to the receiver of link l, excluding the transmitter node of link l. Each node decides to contend the medium with a persistence probability P^n, with each of its outgoing link's contention probability denoted as p_l. The effective link capacity becomes a product form as in the following generalized NUM with both source rates $\{x_s\}$ and persistence probabilities $\{p_l, P^n\}$ as optimization variables:

$$
\begin{aligned}
&\text{maximize} && \sum_s U_s(x_s) \\
&\text{subject to} && \sum_{s:l \in L(s)} x_s = c_l p_l \prod_{k \in N_{to}^I(l)} (1 - P^k), \quad \forall l \\
& && \sum_{l \in L_{out}(n)} p_l = P^n, \quad \forall n \\
& && 0 \leq P^n \leq 1, \quad \forall n
\end{aligned}
$$
(4)

Despite the inherent difficulties of nonconvexity and non-separability of the optimization problem, we show that, again under curvature negativity conditions on utility functions, we can develop a distributed algorithm, with limited message passing, that converges to the globally and jointly optimal rate allocation and persistence probabilities. These results can accommodate general concave utility function (the special case of logarithmic utility function is readily separable as shown earlier in [52]). Different from the other three sample case studies in the appendix, it is better to use a primal penalty function approach rather dual decomposition in this case, the engineering implication of which is that global congestion control and local contention control can operate on the same timescale.

C. Jointly optimal congestion control, routing, and scheduling

In multihop ad hoc wireless networks, end-to-end congestion control, routing, and scheduling among contending links are coupled. Route choices not only affect congestion control in the transport layer, but also determine schedulable regions at the physical layer. In [3], we model contention relations between wireless links as a conflict graph (first proposed in [41]), which indicates the set of links that mutually interfere and cannot be active simultaneously. This determines a feasible rate region Π. Consider an ad hoc wireless network with a set N of nodes and a set L of logical links. We assume fixed physical layer resource

allocations so that each logical link l has a fixed capacity c_l when it is active. The feasible rate region at the link layer is the convex hull of the corresponding rate vectors of independent sets of the conflict graph. Let x_i^k be the flow rate generated at node i for destination k. Let f_{ij}^k be the amount of capacity of link (i, j) allocated to the flows on that link for final destination k. Consider the following generalized NUM in variables $\{x_s\}$ (where x_s is a shorthand for x_i^k) and $\{f_{ij}^k\}$:

$$
\begin{aligned}
&\text{maximize} && \sum_s U_s(x_s) \\
&\text{subject to} && x_i^k \leq \sum_{j:(i,j) \in L} f_{ij}^k - \sum_{j:(j,i) \in L} f_{ji}^k, \quad \forall i, j, k \\
& && f \in \Pi.
\end{aligned}
$$
(5)

Dual decomposition of the above joint congestion control, routing, and scheduling NUM then leads to a jointly optimal cross-layer design where a source adjusts its sending rate based on the congestion prices generated locally at the node, the backpressure from the differential price of neighboring nodes is used to perform optimal scheduling, and routing is automatically generated by the scheduling decision. The implication of this particular decomposition alternative to network architecture is that routing is essentially absorbed into congestion control and scheduling. We also prove that it converges arbitrarily close to the system optimum, and remain stable and optimal (on average) when the schedulability constraint set is modulated by a Markov chain representing time-varying channels.

D. TCP/IP interactions

TCP reverse engineering assumes fixed routing **R**. The routing is computed by variants of IP and updated on a different timescale, based on traffic condition in the network. There is hence a feedback loop where IP routing decision at time t determines flow rates $\mathbf{x}(t)$ and congestion prices $\boldsymbol{\lambda}(t)$ through the capacity constraints in (1), and the congestion prices $\boldsymbol{\lambda}(t)$ affects routing in the next time instance. First consider the case where TCP converges faster than each IP update. It is shown in [53] that an equilibrium of TCP/IP, if exists, indeed solves NUM over *both* source rates $\mathbf{x}$ and routes $\mathbf{R}$, provided congestion prices are used as link costs in the shortest-path computation. Since the routing matrix is discrete, the NUM problem is no longer a convex optimization problem. An equilibrium exists if and only if this NUM and its Lagrange dual have zero duality gap. When there is a non-zero duality gap, the gap can be interpreted as the penalty (in utility) in not splitting the traffic.

In [18], we further consider three alternative timescale separations for the joint congestion control and shortest-path routing dynamics based on congestion price. Analytic characterizations and simulation experiments demonstrate how the step size of the congestion-control algorithm affects the stability of the system models, and how the timescale of each control loop and homogeneity of link capacities affect system stability and optimality. In particular, the stringent conditions on capacity configuration for TCP/IP interaction to remain stable suggests that congestion price, on its own, would be a poor "layering price" for TCP and (dynamic routing based) IP in practice. Alternative traffic engineering methods should be considered.