Towards the Limits of Network Performance

Coding and Information Theoretic Approaches for Efficiency and Robustness

by Tracey Ho

Many features in modern networks have their roots in technologies developed for the Internet. The recent concept of network coding is an exception; and goes beyond traditional forwarding of immutable packets, viewing network operation instead as flows of information that can be operated on mathematically within the network. In particular, network nodes may perform mathematical operations across packets received on different incoming links so as to form new packets containing coded combinations of the original packets' information, all of which will be subsequently decoded at other nodes in the network. Traditional store-and-forward operation, where network nodes simply route packets unchanged from incoming to outgoing links, is a special case of this more general network coding framework.

> "...fundamentally improve network performance, particularly in uncertain, adversarial, or resource-constrained scenarios. "

Early work on network coding by myself and others had demonstrated promising aspects of network coding under a small collection of network scenarios. Recent work from my group at Caltech, supported by the Lee Center, has broadened the range of network coding techniques as well as the range of applicable network types and scenarios. Our research investigates the use of network coding and other information theoretic approaches so as to fundamentally improve network performance, particularly in uncertain, adversarial, or resource-constrained scenarios.

One application of network coding is in network error correction, where redundant information is coded across multiple network links. This enables reliable communication over a network even in cases where individual links cannot be made reliable; for instance, in case of adversarial attacks or arbitrary faults, for which it is not sufficient to do error correction on a link by link basis. Prior work had considered network error correction for the case of single-source networks that are homogenous in the sense that each link (or packet) has equal capacity and is equally vulnerable to errors. We broadened the scope to include multiple-source networks and heterogeneous links, and thereby showed that in-network error correction coding allows redundant capacity to be shared among multiple sources[1]. Notably, such coding makes it possible to achieve the same transmission rates as if each source were given exclusive use of all the redundant capacity. This is illustrated in Figure 1, where if we allocate part of the network to carry information from source 1 only, and the remainder of the network carries information from source 2 only, then either one of the sources, but not both, is able to communicate reliably at rate 1 under a single link error. Coding at the intermediate nodes allows the two sources to use shared network capacity to send redundant information, so that both sources can simultaneously communicate reliably at rate 1 under a single link error. We also showed that non-homogenous networks require a wider range of different error correction coding strategies compared to the homogenous case. In a homogenous network, the error correction capacity can be achieved by linear network coding, and for a single source and sink, it is sufficient to do coding at the source and simple forwarding at intermediate nodes. However, in the non-homogenous case, we showed that linear coding is suboptimal in general, and that forwarding at intermediate nodes is not sufficient even for a single source and sink—instead, depending on the topology, coding, partial error detection or partial error correction at intermediate nodes could be required.

Besides security against errors, coding in networks is also useful for information security against eaves-

dropping, where coding is used to ensure that an eavesdropper observing a subset of network links receives no information about the message being communicated. For this problem, we also showed[2] that non-homogenous networks require a wider range of different coding strategies compared to homogenous networks, and that it is more complex to find the capacity limits and optimal strategy in the general case.

One limitation of purely information theoretic approaches is that they do not take advantage of computational limitations of adversaries. Cryptographic techniques exploit the assumption of computationally bounded adversaries, but at the same time place a higher computational burden on the communicating nodes, which becomes an important limiting factor on performance in networks with computationally limited nodes, such as wireless and sensor networks. Our recent research is merging information theoretic coding and cryptographic techniques for robustness against a wide range of natural impairments and adversarial attacks. We have developed hybrid strategies, which take advantage of the assumption of computationally bounded adversaries, to provide secure and efficient communication when some or all of the communicating nodes are computationally limited. In such strategies, different packets may undergo different operations at a node, and different nodes may employ different techniques based on their topological location and capabilities. For instance, we developed an adaptive network error correction strategy which, when used in conjunction with probabilistic cryptographic verification of subsets of packets processed by network nodes, can achieve a higher rate of reliable information compared to purely information theoretic or purely cryptographic approaches[3]. Our eventual goal is to provide practical guidance on how to combine various information theoretic coding techniques and cryptographic techniques optimally under different network scenarios and constraints such as limitations in power and processing speed of different network nodes; and to develop efficient algorithms that adapt dynamically in the presence of uncertain and changing network conditions.

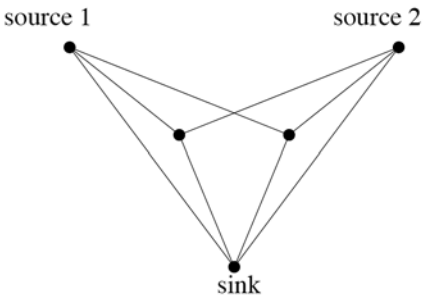

Figure 1: Example showing that in-network error correction coding is needed to share redundant capacity across multiple sources.

Another application of network coding is in distributed network storage. By distributing important information across different storage nodes with sufficient redundancy, information retrieval can be made robust to node failures, loss of connectivity, data corruption and so on. There is a fundamental trade-off between the dissemination/storage costs and the reliability of information retrieval. Similar to the robust transmission problem, the cost-reliability trade-off in the robust storage problem can be improved by use of coding. While exact optimization of the cost-reliability trade-off, given an arbitrary network topology and failure/mobility model, is very complex, we have obtained exact solutions for some special cases (e.g. the case where storage costs dominate and the storage budget is large enough to support a high probability of successful information retrieval from a random subset of storage nodes of given size[4]) and approximate approaches for other cases. We are also investigating regeneration of stored information following failures or errors among the storage nodes. We have found new constructions of practical systematic codes that can be regenerated with minimum data downloaded from other storage nodes, and our ongoing work seeks to obtain a characterization of the set of all possible systematic codes for a given number of storage nodes, code rate and repair bandwidth, as well as an efficient method to systematically enumerate these codes.

“merging information theoretic coding and cryptographic techniques”

My group is also studying the design of new practical wireless networking techniques using network coding and information theoretic approaches. For example, two-way relaying scenarios where two or more terminal nodes exchange information via a common relay node provide a canonical instance of the usefulness of network coding, and can be used as building blocks for practical code construction in general networks. We showed how to generalize any given network layer cod-

ing strategy to a family of joint physical-network layer strategies where the relay effectively propagates physical-layer soft information to improve decoding at the terminals, and also showed how the optimal strategy is affected by the channel parameters and relative power constraints of the relay and terminals[5]. Another result we obtained was a new class of random medium access control protocol, which allows each user to transmit at multiple data rates and uses successive interference cancellation so that multiple packets can be received simultaneously. We found that the proposed protocol can achieve a significant throughput gain over existing random medium access protocols under a Gaussian wireless model. We plan to consider the case of fading wireless channels as the next step towards realizing the benefits of this new approach in practice. LEE

Tracey Ho is Assistant Professor of Electrical Engineering and Computer Science.

Read more at: http://www.its.caltech.edu/~tho

References

[1] S. Vyetrenko, T. Ho, M. Effros, J. Kliewer, and E. Erez, "Rate regions for coherent and noncoherent multisource network error correction," in *IEEE ISIT*, June 2009.

[2] T. Cui, T. Ho, and J. Kliewer, "On secure network coding over networks with unequal link capacities," *IEEE Transactions on Information Theory*, 2009, submitted.

[3] S. Vyetrenko, A. Khosla, and T. Ho, "On combining information theoretic and cryptographic approaches to network coding security against the pollution attack," in Asilomar Conference on Signals, Systems, and Computers, invited paper, November 2009.

[4] D. Leong, A. Dimakis, and T. Ho, "Distributed storage allocation for high reliability," in *IEEE International Conference on Communications*, 2009, to appear.

[5] T. Cui, T. Ho, and J. Kliewer, "Memoryless relay strategies for two-way relay channels," *IEEE Transactions on communications*, **57** (10). pp. 3132-3143, October 2009.