

一种最小化编码节点的网络编码优化算法

郝琨^{①②} 金志刚^{*③}

^①(天津大学计算机科学与技术学院 天津 300072)

^②(天津城市建设学院电子与信息工程系 天津 300384)

^③(天津大学电子信息工程学院 天津 300072)

摘要: 网络编码能有效地提升多播网络的传输性能,但编码的引入增加了节点的计算开销。为了克服网络编码带来的额外开销,该文提出了在代数网络编码框架下的网络编码优化模型,并在此模型基础上给出了基于改进遗传算法的最小化编码节点算法(MCN, Minimizing Coding Nodes)。MCN 在简单遗传算法的基础上增加了一些新的策略,避免了局部性问题和降低了算法寻优时间。模拟实验结果表明,MCN 是有效的而且运行的更快,输出的网络编码方案所需要的编码节点也更少。同时将 MCN 应用到具有实际意义的网络中,同传统的网络编码相比,吞吐率仍可达到 25% 以上,而网络的平均延迟和网络开销却大大减少。

关键词: 网络编码; 遗传算法; 多播速率

中图分类号: TP393

文献标识码: A

文章编号: 1009-5896(2011)02-0260-06

DOI: 10.3724/SP.J.1146.2010.00438

An Optimization Algorithm of Network Coding for Minimizing Coding Nodes

Hao Kun^{①②} Jin Zhi-gang^③

^①(School of Computer Science and Technology, Tianjin University, Tianjin 300072, China)

^②(Electronic & Information Engineering Department, Tianjin Institute of Urban Construction, Tianjin 300384, China)

^③(School of Electronics and Information Engineering, Tianjin University, Tianjin 300072, China)

Abstract: Although network coding is an effective technology to improve the performance of multicast communication, encoding of node brings the additional overhead. In order to overcome this limitation, this paper proposes a network coding optimization model under the framework of algebraic network coding. And an algorithm called the MCN (Minimizing Coding Nodes) is proposed, which is based on the improved genetic algorithm. In MCN some new methods are introduced into the simple generic algorithm in order to avoid locality problem and to reduce optimization time. The experimental results show that MCN is effective and it runs faster, and that the output network coding scheme requires less coding nodes. Moreover, when it is applied to the actual meaningful network, it can guarantee the same network throughput, and much lower average delay and network overhead as the traditional network.

Key words: Network Coding (NC); Genetic Algorithm (GA); Multicast rate

1 引言

网络编码^[1,2]是信息论领域中的重要突破,其基本思想是允许网络节点对来自不同链路的信息进行编码组合,使网络的传输效率达到网络的最大流限,并被广泛的应用在P2P^[3]网络、无线Mesh网络^[4]等领域。然而,传统的网络编码是对网络中所有可能的

中间节点进行编码,编解码的引入增加了节点额外的计算和存储开销,同时也给网络带来了延迟。因此,在利用网络编码带来的好处同时,如何尽可能地降低各种开销就成为一个非常有意义的问题,即网络编码优化问题。

网络编码的优化问题^[5]是指在给定的网络拓扑结构上,基于某个优化目标,在保证达到理论多播速率的前提下,尽可能地降低网络的各种开销。目前优化研究方法主要有3种:基于信息流分解^[6]和简单网络^[7]的最优化方法,线性规划方法^[8]及基因法^[9-11]等。文献[6]按照网络信息流的共性,采用最

2010-05-05 收到, 2010-09-14 改回

国家自然科学基金重大研究计划(90604013), 国家 863 计划项目(2008AA01A320)和天津市高等学校科技发展基金(20090802)资助课题

*通信作者: 金志刚 zgjin@tju.edu.cn

小子树分解技术，得出对于双源， M 个接收节点的网络，编码节点的上界为 $M-1$ ，但对信源个数大于2的情况下没有给出明确的结论。文献[7]采用简单网络的优化方法，将网络转换为节点入度出度之和不超过3的简单网络，但简单网络优化方法将原网络扩大化，得出的最小代价与实际网络的不等。文献[8]运用线性规划理论对网络编码的代价进行了分析，提出了最小代价网络编码的理论模型，但该模型很难在实际的网络中得到应用。文献[9,10]证明了网络代价优化问题是NP难问题，并提出了应用简单遗传算法进行网络编码最小编码边的优化，方法是半分布式的，随着问题规模的增大，算法单次循环需要的时间剧增，得到解的质量下降。文献[11]在一般的遗传算法中加入了一些新的处理，实验证明优化后的网络编码方案所消耗的资源进一步减少，但没有考虑优化后的方案对实际网络性能的改善。

基于以上内容，本文提出了一种在代数框架下的最小化编码点的网络编码优化算法(MCN, Minimizing Coding Nodes)，该算法是基于改进的遗传算法，目的在于保证达到最大多播速率的前提下，尽量减少参与网络编码的节点数量，从而减少网络编码的开销。

2 网络编码优化模型

一个给定的组播网络使用网络编码时，并不需要每一个中间节点进行编码，真正需要编码的节点数目是有限的。在分析了组播网络中所需编码节点的上界和下界后，利用代数网络编码的方法解决多播网络中编码优化的问题并建立了相应的网络模型。

2.1 组播网络中编码节点数目分析

一个有向无循环网络 $G = (V, E)$ ，其中 V 是节点的集合， E 表示边集，在网络 G 上的网络编码问题可以用 $N(G, S, T, h)$ 表示^[12]，其中 S 为源节点， T 为目的节点集， h 为表示从信源节点到任意一个目的节点 $t \in T$ 的数据包数。将需要进行编码的节点称为编码节点，不需要进行编码的点称为转发节点。

定理 1^[12] 对任意的 $\epsilon > 0$ ，假设在一个具有 $|v|$ 个节点的网络 $G = (V, E)$ 上其组播编码问题 $N(G, S, T, h)$ 所需要的最小编码节点数目为 $\text{Min}(N)$ ，则在计算量小于 $|v|^{1-\epsilon}$ 内能找到或逼近 $\text{Min}(N)$ 的算法是个 NP 难问题。

定义 1^[12] 如果某条边 e 上对应传输的数据，依赖于两个或多个数据包的组合，则称 e 为一个编码边；如果某条边 e 上传输的数据，仅依赖于一个数据包，则 e 称为一个转发边。如果某个非信源节点

的某条输出边为编码边，则该节点为编码节点；如果某非信源节点所有的输出边都是转发边，则称该节点为转发节点。

定理 2 $N(G, S, T, h)$ 存在可行解的充分必要条件即从源节点 s 到每个目的节点的最小割至少等于 $h^{|T|}$ 。一个可行的网络编码需要的编码节点数目最多为 $h^3 k^2$ ，其中 $k=|T|$ 表示目的节点的数目，所需要最小的编码节点数目为 $\Omega(h^2 k)$ ，即 $\Omega(h^2 k) \leq \text{Min}(N) \leq h^3 k^2$ 。

2.2 网络模型

代数型编码方法是由文献[13]提出的，其核心思想就是将网络中节点的输入信息与输出信息之间的关系利用矩阵 $M = A(I - F)^{-1} B^T$ 的形式表示出来。其中 A 表示源节点的输入矩阵， B 表示目的节点的接收矩阵， F 表示邻接矩阵， I 表示单位矩阵。当考察网络编码的多播形式时，只要保证各个目的节点对应的网络转移矩阵是满秩的，目的节点即可以准确地恢复原始信息。基于代数型网络编码方法，本文提出了网络编码优化模型。

设一个有向无环网络 G 中任意一条边的容量为单位容量，组播速率为 R ，如果某条链路的容量超过了单位容量，则用平行边来表示。点集 V 中引入一个虚拟源节点 S' ，边集 E 中引入 R 条平行边 e_1, e_2, \dots, e_R 连接 S' 和 S 。为了使两个节点仍能够唯一地表示一条链路，可以在所有拆分的平行边上增加一个辅助节点，这样网络 G 被转化为网络 $G' = (V', E')$ 。

网络 G' 中所有节点的输出链路为该节点输入链路分配的二进制编码向量。用来表示输出链路是如何处理输入链路发送来的数据。如图 1 所示：节点 v 入度为 3，所以输出链路使用 3 位的二进制向量表示。链路 y_1 的编码向量 $a_1=011$ 表示将输入链路 x_2, x_3 上的数据进行编码， y_2 的编码向量 $a_2=000$ 表示不发送任何数据。

当整个网络 G' 中所有的输出链路上的二进制编码向量确定后，多播网络的信息流也就确定了。在每个接收节点处验证网络的多播速率。根据代数型网络编码方法，为网络图 G' 中的所有不为 0 的向

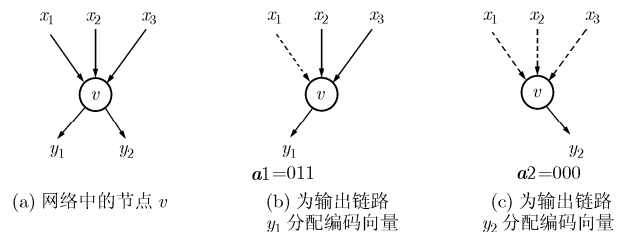


图 1 网络模型中的一个节点

量分配一个链路系数 ξ_i , 每 d 个连接建立传输矩阵。每传输矩阵为 $R \times R$ 维表示源节点与目的节点之间的关系, $p(\xi)$ 为这些传输矩阵行列式的积。在所有的接收者处检查 $p(\xi)$, 当 $p(\xi)$ 不等于 0 时, 即可以确定整个网络的多播速率^[13]。

实际上在这一网络模型中已经给出了编码方案, 针对一个给定的节点 v , 只需要统计每个输出向量中 1 的个数, 当输出向量中 1 的个数最多为 1 个, 即该边上传输的数据最多依赖一个数据包, 表示该输出边是转发边。如果节点 v 所有的输出边均不是编码边而是转发边, 则该点不是编码点从而不进行编码。如果节点 v 的某个输出边的二进制向量中 1 的个数大于 1, 即表示该边上传输的数据依赖于两个或多个数据包的组合, 则该边为编码边该节点为编码节点。编码节点统计过程如图 2 所示。

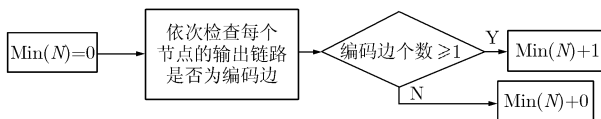


图2 编码节点统计过程

3 MCN 算法

计算需要最少编码点问题是一个 NP 问题, 遗传算法在解决特别复杂问题有明显的优势。MCN 算法是基于简单的遗传算法 (Simple Genetic Algorithm, SGA) 之上的。给定的网络拓扑, 拥有多条输入链路的节点均在网络编码优化的范围内, 单个输入链路的节点对接收到的信息直接进行转发。优化方案由拓扑转换、数据处理和多播速率的验证等部分组成。数据处理部分是根据转换后的网络拓扑生成网络编码方案的数据结构, 最后由遗传算法输出编码节点数量最小的编码方案。优化算法的处理过程如图 3 所示。

遗传算法在实际应用中的主要问题是局部性问题和早熟收敛。为了避免局部性问题本文加入了一些新的解决方案: (1) 在初始种群中加入所有中间节点都进行网络编码的初始成员, 即加入了所有输出链路上的二进制向量的所有元素均为 1 的编码方案, 以保证在最坏的情况下也可以输出可用的编码方案。(2) 在每次循环的开始补充一部分新成员参与



图3 优化算法处理过程

之后的遗传和变异环节, 提高种群的多样性, 实验结果显示这样基本上解决了局部性问题。新的解决方案的加入, 增加了算法的复杂度, 但是现在的计算机计算能力很强, 对算法的执行效率的影响实际上是可以忽略不计的。

每个染色体的长度为^[9]: $m = \sum_{v \in V} d_{in}(v) d_{out}(v)$, $d_{in}(v)$ 为节点 v 的入度, $d_{out}(v)$ 为节点出度。

每个个体即染色体的适应度按下式计算:

$$F(y) = \begin{cases} \text{编码节点个数}, & y \text{ 可行} \\ \infty, & y \text{ 不可行} \end{cases} \quad (1)$$

其中 $c_i \in C$, C 为编码节点的集合, 染色体 y 的可行性即输出的编码方案能否满足最大多播速率, 通过计算 $p(\xi)$ 来确定 y 的可行性, 每一个传输矩阵 $M_i = A(I - F)^{-1} B^T$ ($1 \leq i \leq d$), 当且仅当 $p(\xi)$ 的值不为 0 时, 生成的方案可以满足所有的接收者都达到预期的多播速率, 则该个体或编码方案 y 是可行的, 可以继续参与执行后续的遗传操作。当 $p(\xi)$ 的值为 0 时, 表示染色体 y 对应的编码方案不可行, 则舍弃。

在基因的遗传阶段, 交叉操作采用简单交叉。因为应用二元变异可有效地保持同一基因位上等位基因的多样性, 可以有效地预防了早熟收敛问题。因此对于二进制字符串组成的染色体, 采用二元变异方法即应用同或/异或逻辑运算方法针对两条染色体进行运算。有效地克服早熟收敛, 提高遗传算法的运算速度。

算法 1 基于遗传算法的最小化编码节点 MCN 算法

- (1) 算法初始化;
- (2) 随机生成初始种群 (加入所有中间结点都进行编码的初始成员);
- (3) While 进化终止条件为假;
- (4) 计算每个个体的适应度;
- (5) 从父代群体中选择个体;
- (6) 进行交叉、二元变异操作生成子代群体;
- (7) 在子代群体中增加一些新成员替换父代群体;
- (8) End While;
- (9) 输出最优个体。

4 仿真与算法性能评价

仿真从算法有效性、收敛性和运行时间方面对优化算法自身的性能进行了分析, 同时从提高网络吞吐量、降低网络延迟和网络开销方面评价了 MCN 算法对实际网络性能的改善。本文将文献[9]中的提出优化算法 SGA 与 MCN 算法进行比较。使用

Matlab 对不同节点数目的网络结构进行仿真实验。其中一部分拓扑是用人工的方法将 3, 7 个蝶形结构^[1]进行叠加, 如图 4 所示 3 个蝶形结构进行叠加, 来验证 MCN 算法的有效性。另外一部分拓扑是由 BRITE 软件随机产生具有实际网络特性的网络拓扑, 用来测试算法的收敛速度及算法的性能。随机生成的拓扑结构为(20, 79, 4, 4)(40, 132, 6, 5)(节点, 链路数, 目的节点个数, 速率)。种群规模为 150, 交叉概率为 0.8, 变异率为 0.01, 进化代数为 500。同时, 为了验证改进的优化方法能很好地应用在实际的网络中, 应用 NS 网络仿真软件对 MCN 算法优化后的编码方案进行了仿真, 评价指标包括吞吐量、平均延迟和网络开销。仿真场景为 1000 m×1000 m 的矩形, 节点随机分布, 上层业务类型为 CBR, 数据包长度为 8 个 byte。该场景节点密度程度适中, 连通性较好, MAC 协议为 802.11。

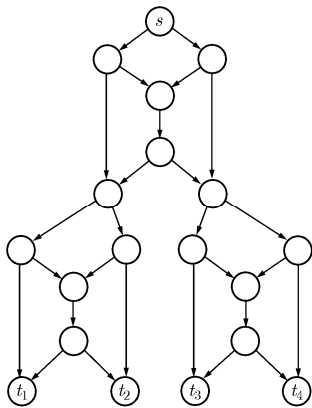


图 4 3 个蝶形结构的叠加

4.1 算法的有效性验证

表 1 给出了 MCN 与 SGA 在 4 组拓扑条件下多次实验的平均结果, 平均结果列表示编码节点数量的平均值, 最佳结果列是所有实验中输出的网络编码方案的编码节点数的最小值。从实验结果可以看出对 3,7 个蝶形结构进行测试时, 能够准确地得出最小编码节点数, 在输出编码方案上是有效的。而且数据显示在对随机拓扑进行优化时, 能够得到较小的编码节点数。

4.2 收敛性比较

为了体现改进的遗传算法的优点, 对随机拓扑 (20, 79, 4, 4) 上进一步完成了一些实验, 对算法的收敛性进行了测试。网络拓扑如图 5 所示, 其中 s 为源节点, t_1, t_2, t_3, t_4 为目的节点。从图 6 中可以看出, MCN 在进化的开始几代种群的平均适应值快速下降, 在较少的代数内接近最优值。而 SGA 收敛的

表 1 MCN, SGA 算法有效性验证结果比较

拓 扑	MCN		SGA	
	最佳结果	平均结果	最佳结果	平均结果
3 个蝶形结构叠加	3	3.8	3	4.7
7 个蝶形结构叠加	7	7.5	8	9.7
(20,79,4,4)	3.5	4.3	5.1	7.2
(40,132,6,5)	8.6	10.3	10.1	11.6

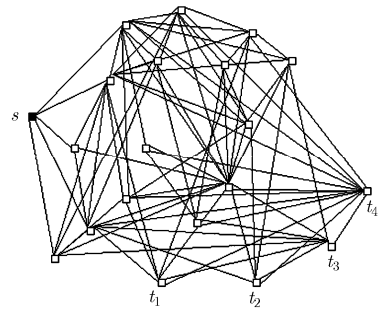


图 5 20 个节点的实际网络拓扑

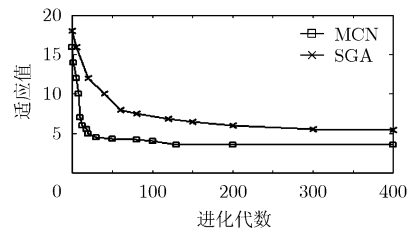


图 6 收敛性能比较

速度缓慢, 最终陷入了局部性。在每次循环开始时随机增加一些新成员, 增加了种群的多样性, 避免了局部性问题。同时二元变异也提高了算法的收敛速度。

4.3 运行时间比较

表 2 给出了 MCN 与 SGA 在不同的拓扑结构上每代的平均运行时间, 不难发现 MCN 中每代一般只需要较少的时间就可以得到最优解。而且随着节点数目的增加, SGA 每代所需要的时间急剧的增加, 当网络规模达到 40 个节点的时候每次循环需要 72.6 s, 这样为 40 个节点的网络产生一个优化结果需要几个小时。这主要是由于 MCN 只是针对网络中有多个输入的节点进行筛选, 降低了问题的复杂度。同时, 二元变异算子增加了 MCN 的寻优能力, 而且在每次循环的开始都增加新的成员增加了群体的多样性, 早熟情况得到了明显的改善。这为算法的实际应用提供了极大的方便。

表 2 每代运行的时间

网络中节点数量	每代运行的时间(s)	
	MCN	SGA
15	2	2.7
20	3.1	4.6
25	4.8	7.4
30	6.5	14.5
35	11.2	32.5
40	15.7	72.6

4.4 吞吐量

图 7 表示网络吞吐量与网络规模的关系。在随机拓扑条件下, 分别对节点个数为 20, 40, ..., 120 的网络进行的吞吐量的比较。如图所示, 随着网络规模的扩大网络编码在提高网络吞吐量方面有很大的优势, 同传统的路由方式相比吞吐率提高 25% 以上, 而且经 MCN 优化后的编码方案能够得到与传统的网络编码 NC(所有入度大于等于 2 的节点均为编码点) 同样的吞吐量, 是因为 MCN 是以实现多播理论容量为前提的。而在达到相同吞吐率的情况下进行编码的节点数比传统的编码方案中节点数目要少得多, 大大降低了编码消耗。

4.5 平均延迟

图 8 为 MCN 算法优化的编码方案和传统的网络编码方案 NC 在延迟性能上的比较。在网络节点

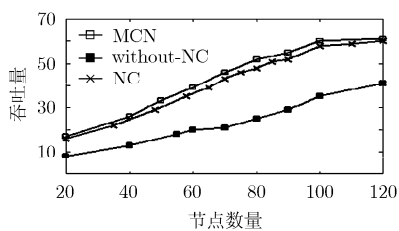


图 7 网络吞吐量的比较

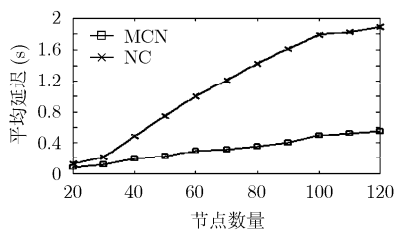


图 8 平均延迟评价

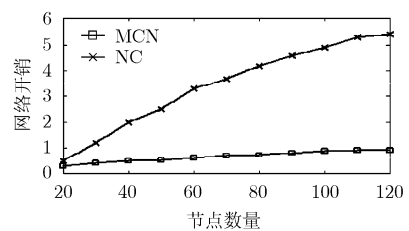


图 9 网络开销评价

数目少时, 节点之间的连接比较稀疏, 两种方案参与编码的节点数目差别不是太大, 延迟差别不是很明显。但是随着网络规模的增大, 传统的网络编码方案中, 参与编码的节点数目迅速增加, 延迟逐渐的增大。而 MCN 算法优化后的编码方案参与编码的节点数目较少, 数据包的等待延迟、编解码延迟较少, 可见 MCN 算法能够充分的发挥它的优越性。

4.6 网络开销评价

图 9 为 MCN 算法优化后的编码方案和传统的网络编码方案 NC 在网络开销比较。MCN 算法优化后的网络编码方案, 参与编码的节点数目少, 相应的网络开销比较小, 而且上升的幅度很小, 表现出良好的性能。随着网络规模的增加, 传统的网络编码方案中参与编码的节点数目增加, 必然会增加网络的整体开销, 降低网络的性能。

5 结束语

本文在代数网络编码的框架下提出了网络编码优化模型, 在此模型基础上给出了最小化编码节点的网络编码优化算法。算法的实现基于一般的遗传算法, 针对遗传算法的不足在一般遗传算法基础上加入了一些新的解决方案。实验结果表明, 在保证多播理论容量的前提下, 改进的遗传算法能够显著的减少编码节点的数量, 从而有效的降低了网络编码的代价。同时, 应用到实际的网络中, 可以有效的提高网络的吞吐率、降低延迟和网络的开销。

参考文献

- [1] Yeung R W, Li S-Y R, and Cai N, *et al.* Network Coding Theory [M]. Hanover: NOW Publishers Inc, 2006: 12-18.
- [2] Ahlswede R, Cai N, Li S-Y R, and Yeung R W. Network information flow[J]. *IEEE Transactions on Information Theory*, 2000, 46(4): 1204-1216.
- [3] Kim M, Lima L, Zhao F, Barros J, and Médard M, *et al.* On counteracting Byzantine attacks in network coded peer-to-peer networks[J]. *IEEE Journal on Selected Areas in Communications: Special Issue on Mission-Critical Infrastructure*, 2010, 28(5): 692-702.
- [4] 覃团发, 廖素芸, 罗会平等. 支持网络编码的无线 Mesh 网络路由协议[J]. *北京邮电大学学报*, 2009, 32(1): 15-19.
- [5] 黄政, 王新. 网络编码中的优化问题研究[J]. *软件学报*, 2009, 20(5): 1349-1361.
- [6] Huang Zheng and Wang Xin. Research on the optimization problems in network coding[J]. *Journal of Software*, 2009, 20(5): 1349-1361.
- [6] Fragouli C and Soljanin E. Information flow decomposition

- for network coding [J]. *IEEE Transactions on Information Theory*, 2006, 52(3): 829-848.
- [7] Langberg M, Sprintson A, and Bruck J. The encoding complexity of network coding[J]. *IEEE Transactions on Information Theory*, 2006, 52(6): 2386-2397.
- [8] Bhattad K, Ratnakar N, and Koetter R. Minimal network coding for multicast[C]. IEEE International Symposium on Information Theory. Melbourne Australia: IEEE Communication Society, 2005: 1730-1734.
- [9] Kim M, Médard M, and O'Reilly V. Evolutionary approaches to minimizing network coding resources[C]. Proc. of the IEEE INFOCOM 2007. Anchorage, 2007: 1991-1999.
- [10] M Kim, Médard M, and O'Reilly U-M. An evolutionary approach to inter-session network coding[C]. Proc. of the IEEE INFOCOM 2009. Rio de Janeiro, 2009: 450-458.
- [11] 邓亮, 赵进, 王新. 基于遗传算法的网络编码优化[J]. 软件学报, 2009, 20(8): 2269-2279.
- Deng Liang, Zhao Jin, and Wang Xin. Genetic algorithm solution of network coding optimization[J]. *Journal of Software*, 2009, 20(8): 2269-2279.
- [12] 樊平毅. 网络信息论[M]. 第一版, 北京: 清华大学出版社, 2009, 2: 121-123.
- Fan Ping-yi. *Network Information Theory*[M]. Tsinghua University Press, 2009, 2: 122-123.
- [13] Koetter R and Médard M. An algebraic approach to network coding[J]. *IEEE/ACM Transactions on Networking*, 2003, 11(5): 782-795.
- 郝 琨: 女, 1979年生, 讲师, 博士, 研究方向为网络编码、网络性能优化.
- 金志刚: 男, 1972年生, 教授, 博士生导师, 研究方向为网络管理与安全、网络性能评价.