

有损网络中多代混合网络编码性能分析研究^①王伟平^② 费柳凤 宋 虹

(中南大学信息科学与工程学院 长沙 410083)

摘 要 对单代网络编码(SG-NC)和多代混合网络编码(MGM-NC)两种编码方案的解码性能进行了理论分析,得到了两种方案的各代数据解码概率的计算公式。理论分析结果表明解码率与代的大小、冗余数据包多少、冗余数据包的发送方式、链路丢包率、混合集大小有关;相同的代大小、链路丢包率、链路长度下,用同样的发送冗余,MGM-NC 取得的平均解码率高于 SG-NC;在 MGM-NC 中,同一编码混合集中低代数据的解码率高于高代数据的解码率;且冗余数据包采取单代冗余发送时,低代与高代数据的解码率差异更加明显。模拟测试结果验证了理论分析的正确性。

关键词 网络编码,多代混合,混合集,解码率

0 引 言

网络编码^[1]是一种融合了路由与编码的信息交换技术,其核心思想是中间节点对各条信道上收到的信息进行线性或者非线性的处理,再转发给下游节点,使得网络节点不仅能进行简单的存储转发,还能够对数据进行编码处理,从而节省网络带宽,增加网络的吞吐量^[2,4]。网络编码一提出就得到了国际学术界的广泛关注,尤其是在无线传输领域中。Deb^[5]、Ramamoorthy^[6]等人首先对有广播特性的无线网络的编码性能进行了理论研究,将网络编码应用在无线网络中,以提高网络的吞吐量和抗丢包的能力,减少重传次数^[7,8]。随着网络编码理论的日益成熟,网络编码在实际网络环境中的应用得到了广泛关注。Chou 等人^[9]首先提出了一个实际网络编码方案——随机线性网络编码方案,Ho 等人^[10]证明了随机线性网络编码方案可以达到最大多播容量,接着 Katti^[11]等人提出了另一种基于异或操作的网络编码方案。这些网络编码方案先把需要发送的数据进行分组,并将每一个分组称为一个代(generation)。源端按代进行数据传输,将整个代的分组进行随机线性编码后发送。每一代中包含的数据包数 K 是固定的,只对同一代内的数据包进行网络编码

操作。这种编码方案被称为单代网络编码(single-generation network coding,SG-NC)方案。

在一些具有高丢包率的无线网络环境中,由于链路的丢包接收节点不能接收到足够多的编码包,导致该代数据无法被解码。用 SG-NC 方案时,如果当代的数据包传完后该代不能被解码,那么这一代的编码包将被丢弃。这样,一个编码包的丢失可能会导致整个代都无法解码,这种代价是很大的。为了提高每代数据的解码率,Halloush 等人^[12,13]提出了一种新的编码方案——多代混合网络编码(multi-generation mix network coding,MGM-NC)方案。不同于 SG-NC 方案,MGM-NC 方案允许不同代间的数据包进行编码操作。MGM-NC 方案的主要目的在于通过提高每代数据的成功传输率来提高代的解码率。同时,Halloush 等人^[14]通过实验验证了使用 MGM-NC 方案可以使得同一混合集中低代的解码率高于高代的解码率,该方案具有优先传输功能。但是,上述研究只是用实验来验证了这种特性,缺乏严格的理论分析和证明。本文从理论上分析了 MGM-NC 方案各代数据的解码率与混合集大小、代大小、冗余方式、链路丢失率等因素间的定量关系,结果表明,在相同的冗余下,(1)采用单代冗余 MGM-NC 方案时,低代数据的解码率高于高代数据;(2)采用 m 代冗余 MGM-NC 方案时,各代的解

① 国家自然科学基金(61173169,61103203)和教育部新世纪优秀人才计划(NCET-10-0798)资助项目。

② 女,1969年生,博士;研究方向:网络编码,网络信息安全;联系人,E-mail:wpwang@mail.csu.edu.cn
(收稿日期:2013-01-23)

码率相同,并且其总解码率要高于单代冗余 MGM-NC 和 SG-NC 方案;(3)采用 MGM-NC 方案,各代的解码率随着混合集的大小的增加而增大;(4)增大代大小可以增加代的解码率。在此基础上,还得到了在一定网络环境下,要保证传输质量所需要的冗余大小。模拟测试结果验证了上述理论分析结果的正确性。

1 SG-NC 方案和 MGM-NC 方案

为了便于分析,首先剖析一下单代网络编码(SG-NC)方案和多代混合网络编码(MGM-NC)方案的基本思想。

1.1 SG-NC

SG-NC 方案见图 1。该方案通常将原数据进行分组,每一组的数据包数都是固定的,每一组数据称

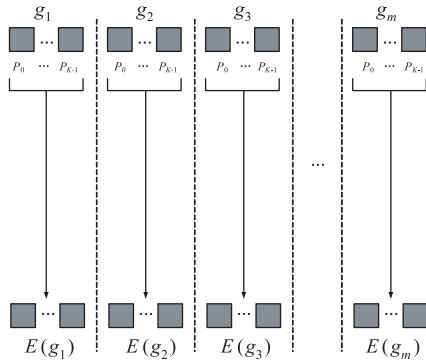


图 1 单代网络编码方案

为一个代,属于同一代的数据包都被标记相同的代号。节点只对同一代内的数据包进行随机线性编码,不同代内的数据包的编码解码是相互独立的,因此,对于某一大小为 K 的独立代,只有当接收节点接收到 K 个相互线性独立的编码包时,该代才能被成功解码。假设原数据被分为 $g_1, \dots, g_m$ 共 m 代,每一代包含 K 个数据包。对于任一独立代 g_i ,通过从足够大小的有限域内 ($F_q = 2^8$) 随机选择 K 个的编码系数,来生成 K 个线性独立的编码包 E_x ($0 \leq x < K$)。假设用 $(P_0, P_1, \dots, P_{K-1})$ 表示该代的 K 个原始包,用 C_x 表示跟编码包 E_x 相关的编码系数向量,其中 $C_x = \{c_{x,0}, \dots, c_{x,K-1}\}$ 。因此,SG-NC 方案的编码操作可用式

$$E_x = \sum_{j=0}^{K-1} c_{x,j} \cdot P_j \quad (1)$$

表示。生成的 K 个线性独立的编码包可用下面的形式表示:

$$\begin{bmatrix} E_0 \\ \vdots \\ E_{K-1} \end{bmatrix} = \begin{bmatrix} c_{0,0} & \cdots & c_{0,K-1} \\ \vdots & \ddots & \vdots \\ c_{K-1,0} & \cdots & c_{K-1,K-1} \end{bmatrix} \begin{bmatrix} P_0 \\ \vdots \\ P_{K-1} \end{bmatrix} \quad (2)$$

其编码操作如图 1 所示,每一代都是独立编码的,只有当接收节点能接收到 K 个线性独立的编码包时,该代数据才能被解码。

1.2 MGM-NC

MGM-NC 方案同样将原数据分为很多代,不同于 SG-NC 方案,这种编码方式会将若干个代组成一个混合集。编码时,属于同一混合集的不同代可以编码在一起。为了清楚说明编码方式,设一个混合集共有 m 个代,每一代包含 K 个数据包,同一混合集中的不同代用不同的代号 l 表示, $1 \leq l \leq m$ 。根据代号的递增顺序对各代进行排序,把代号小的代称为低代,代号大的代称为高代。当节点需要发送某混合集中代号为 h 的代时,它将该混合集中 l 小于等于 h 的代的数据包一起进行线性编码发送,把编码后的数据包称为属于 h 代的编码包。这样,同一混合集的编码包中,高代的编码包中就包含了所有低代数据包的信息。图 2 给出了属于同一混合集的 m 代的数据编码示意。

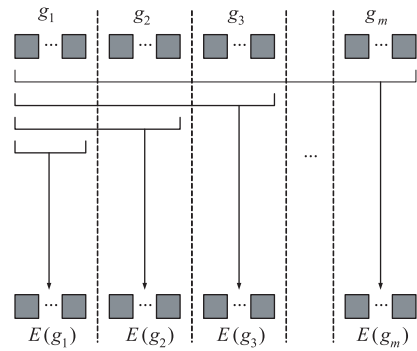


图 2 多代混合网络编码方案

用 MGM-NC 方案时,编码向量的大小由编码在一起的数据包的数量决定,即由需要发送的代在混合集中的位置决定。如果节点即将编码发送代号为 l 的代,那么编码向量的大小就是 lK 。用 P_j 表示属于同一编码集合的数据包,即 $P_{(l-1)K} \sim P_{lK-1}$ 是属于代号为 l 的数据包。同样,对于同一编码混合集中第 l 代,用 C_x^l 表示属于第 l 代的编码包 $E_{l,x}$ 的系数向量,其中 $C_x^l = \{c_{x,0}^l, c_{x,1}^l, \dots, c_{x,lK-1}^l\}$ 。因此,MGM-NC 方案的编码操作可表示如下:

$$E_{l,x} = \sum_{j=0}^{l \cdot K - 1} c_{x,j}^l \cdot P_j \quad (3)$$

其中 x 为编码包的包号, $E_{l,x}$ 表示属于第 l 代的第 x 个编码包。对应于同一编码混合集中第 l 代的编码包生成的 K 个线性独立的编码包可用下面的形式表示:

$$\begin{bmatrix} E_{l,0} \\ \vdots \\ E_{l,K-1} \end{bmatrix} = \begin{bmatrix} c_{0,0}^l & \cdots & c_{0,l \cdot K - 1}^l \\ \vdots & \ddots & \vdots \\ c_{K-1,0}^l & \cdots & c_{K-1,l \cdot K - 1}^l \end{bmatrix} \begin{bmatrix} P_0 \\ \vdots \\ P_{l \cdot K - 1} \end{bmatrix} \quad (4)$$

从以上两种编码方案中我们不难看出, SG-NC 是 MGM-NC 混合集大小为 1 时的特例。

2 MGM-NC 性能分析

直觉上, 在 MGM-NC 中, 由于高代的编码包会包含低代的信息, 高代数据对低代的解码有一定的帮助。因此, 代的解码就有两种方式: 一种是每个代在当前代的编码包传送完后就得到解码, 称为立即解码。例如在传完第一代 g_1 的编码包 $E_{1,x}$ 后, 接收节点已经收到了足够多的线性独立的数据包 (即收到了多于 K 个线性独立的数据包), 这时, 所有属于 g_1 代的原始数据包都能够解码。另一种解码方式是当前代的编码包传完后接收节点没有收到足够多的属于 g_1 代的编码包, 无法解码出该代数据, 但可以通过高代的编码包的帮助而解码, 称为滞后解码。例如在传输完所有属于 g_1 代的编码包后, 接收节点接收到的编码包数少于 K 个, 那么这一代数据是无法解码出来的, 但是当传输属于 g_2 的编码包时, 由于编码包仍然包含了 g_1 的信息, 所以只要属于 g_2 的编码包发送完后, 接收节点接收到的两个代的总编码包数大于 $2K$, 这两代数据就都能解码。下面我们将从理论上对这两种编码方案的解码率进行分析。

2.1 MGM-NC 传输模型

本文将通过一个传输模型来分析在非可靠网络中 MGM-NC 的传输性能。假设从源节点 s 到目的节点 r 为端到端传输, 端到端丢包率均为 ε 。

源节点采用 MGM-NC 方式, 对原始数据进行编码传输, 一个混合集包含 m 个代, 每代由 K 个原始数据包组成, 当节点收到一个属于代 l 的编码包时, 它将缓存中当前所有最大代号不大于 l 的编码包同该编码包再编码在一起, 发送给下一跳节点。假设源节点 s 共有 n 代数据要发送给目的节点 r , 显然, 对于每一个要发送的代, 源节点都将生成至少 K 个

相互独立的编码包。考虑到链路的丢包问题, 为了提高目的节点的接收率, 源节点采用发送一定数量冗余编码包的方式。我们将每一代所发送的冗余包的数量与每代数据包数量之间的比值称为冗余度。

冗余包的发送分为单代冗余和 m 代冗余两种方式:

(1) 单代冗余是指源节点针对任意代发送 $K + R$ 个编码包, 即对每代数据发送了 R 个该代的冗余编码包。

(2) m 代冗余是指源节点针对同一个混合集来发送冗余包, 对同一个混合集中前 $m - 1$ 个代依然每个代发送 K 个编码包, 在第 m 代发送 $K + mR$ 个编码包。这里 mR 个冗余编码包是对同一混合集中所有代的原始数据包的线性编码。

为了方便对单代冗余和 m 代冗余的理解, 图 3 给出了两种冗余方式的示意图。

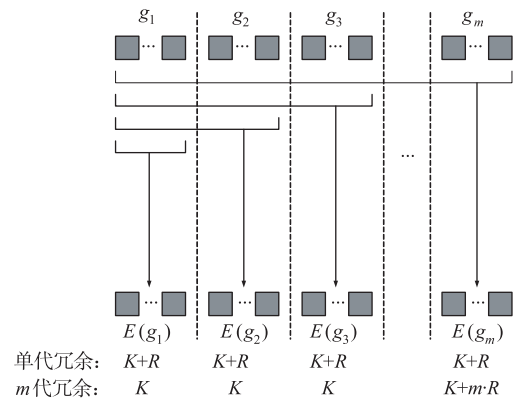


图3 单代冗余和 m 代冗余两种发送方式

2.2 解码性能分析

定理1 若采用多代混合编码方案, 属于同一混合集的 m 代数据的低代数据总是先于高代数据解码, 即若第 i 代数据不能解码, 则第 j 代数据必定也无法解码 ($i < j \leq m$)。

证明: 在多代混合编码传输中, 无论是采用单代冗余还是 m 代冗余方案, 在同一混合集中, 属于第 j 代的编码包包含了第 i 代的数据包 ($j > i$), 所以, 显然当第 i 代的数据包无法解码时, 第 j 代数据肯定也无法解码。

定理2 在同一编码混合集中, 设 $i < j \leq m$, 在接收第 j 代编码包前, 第 i 代数据未能解码, 当接收完第 j 代编码包后, 若第 j 代数据包不能解码, 则这时第 i 代数据肯定也不能解码。

证明: 由于属于第 j 代的编码包是由第 j 代以及

低于 j 代的原始数据包线性编码得到,若第 j 代数据包不能解码,意味着这些编码包中存在未知数据包,显然,这时这些编码包不能帮助低于 j 代的原始数据的解码,因此第 i 代数据依然无法解码。

定理 1 和定理 2 意味着在多代混合网络编码传输中,若某个时刻接收节点可以解码数据包,表示这之前收到的所有数据包都能解码;若某个时刻接收节点存在不可解码的某个代的数据包,此时该代编码包对前面代数据包的解码还没有帮助。

为了便于分析多代网络编码中各代数据的解码概率,首先在此给出以下符号定义:

$f(M, N)$: 表示接收节点从源节点发送的 M 个编码包中成功接收到任意 N 个的概率。

p_i : 表示接收节点在接收完第 i 代编码包后,能立即解码出前 i 代的原始数据包的概率,即立即解码概率。

p_i' : 表示接收节点在接收完第 i 代编码包后,不能解码第 i 代的原始数据包,必须在高代编码包的帮助下才能解码的概率,即滞后解码概率。

$P(i)$: 表示第 i 代数据解码率。显然有

$$P(i) = \begin{cases} p_i + p_i' & (i < m) \\ p_i & (i = m) \end{cases} \quad (5)$$

采用不同的冗余方式,对每一代的解码率影响不一样。在本文中,由于编码系数的选取的有限域为 $F_q = 2^8$,这个大小足够保证不同的编码包之间的线性独立性,因此我们假设只要接收节点能接收足够数量的有用编码包,就一定能够解码。下面将分析采用以上两种冗余方式时,每一代的解码率。

(1) 单代冗余

在单代冗余 MGM-NC 中,冗余包是在每一代发送的,且冗余包等效于该代的编码包,因此,任何一个代都有可能用自己的冗余包帮助低代数据解码。假设链路丢包率为 ε ,那么有

$$p_i = \sum_{j=1}^i p_{i-j} \cdot q_{i,j} \quad (p_0 = 1) \quad (6)$$

其中

$$q_{i,j} = \begin{cases} \sum_{t=K}^{K+R} f(K+R, t) & (j=1) \\ \prod_{t=1}^{j-1} \sum_{i_t=\max(t \cdot K - (j-t) \cdot R - \sum_{m=0}^{t-1} i_m, 0)}^{t \cdot K - 1 - \sum_{m=0}^{t-1} i_m} f(K+R, i_t) & (j > 1, i_0 = 0) \end{cases}$$

$$f(M, N) = \binom{M}{N} \cdot (1 - \varepsilon)^N \cdot \varepsilon^{M-N}$$

$$p_i' = \sum_{r=0}^{i-1} \sum_{j=i-r+1}^{m-r} p_r \cdot q_{j+r,j} \quad (7)$$

定义 p_0 恒为 1。

式(6)中 $q_{i,j}$ 表示接收到第 i 代的编码包后,共解码了 j 代数据的概率,包括第 i 代和前面未解码的 $j-1$ 代。根据定理 1 和定理 2,这 j 代数据就是第 $i-j+1$ 代至第 i 代。

$q_{i,1}$ 意味着收到第 i 代编码包后,要解码第 i 代数据,只要确保在 $K+R$ 个当代数据包中收到数据包超过 K 个。

当 $j > 1$ 时,意味着收到第 i 代编码包后,需确保解码第 $i-j+1$ 至 i 的共 j 代数据,其中, i_v 代表第 $i-j+v$ 代成功接收到的编码包数目 ($1 \leq v \leq j$),即 i_1 至 i_j 分别表示第 $i-j+1$ 至第 i 代接收到的编码包数目。对于第一个未解码的代 g_{i-j+1} ,接收节点接收到的编码包数量必须满足 $K - (j-1)R \leq i_1 \leq K-1$,最小值 $K - (j-1)R$ 保证了后面的 $j-1$ 代的冗余包能帮助代 g_{i-j+1} 成功解码, $K-1$ 是代 g_{i-j+1} 没有解码时接收到的该代编码包的最大值,因为若 $i_1 > K-1$,该代就是可解的了。 i_2 的上界是 $2K - i_1 - 1$,这是为了保证 $i_1 + i_2 \leq 2K - 1$,不然这两代在收完 i_2 代的编码包后就能解码。同样,对于每一个未解码的代接收到的编码数据包数 i_v 都有上下界限制。

式(7)表示的是第 i 代的滞后解码概率。由于第 i 代不能立即解码也分多种情况,公式中 p_r 表示在接收到前 r 代编码包后这 r 代数据都能够解码的概率, $q_{j+r,j}$ 为接收到第 $j+r$ 代编码包后,共解码了 j 代的概率,即表示第 r 代后的 $j-1$ 代均不能解码,直到收到第 $j+r$ 代编码包后才完全解码的概率。

根据式(5)-(7),计算得到了单代冗余 MGM-NC 传输策略的各代数据立即解码、滞后解码和总的解码概率。

图 4 显示了在单代冗余 MGM-NC 方案中,当

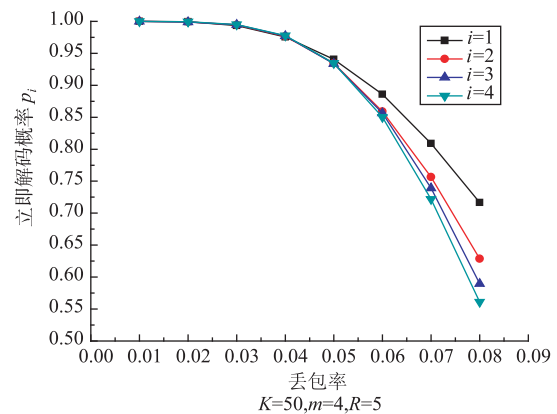


图 4 单代冗余 MGM-NC 中,各代立即解码概率

$K=50, m=4, R=5$ (即 10% 的冗余度) 时, 随着链路丢失率的增加各代能立即解码的概率都呈现出下降的趋势, 高代立即解码率下降更快, 原因是比较直观的, 第 1 代的编码包数只要大于 K 个就能保证第 1 代解码, 而第 i 代成功接收编码包数必须保证整体接收编码包数大于 iK 。

图 5 和图 6 分别表示各代的滞后解码概率和总的解码概率。可以看出对单代冗余 MGM-NC 方案而言, 属于同一混合集中的低代数据的解码率高于高代数据的解码率。这个结论直接可以由定理 1 和定理 2 来解释, 第 i 代数据解码必然保证了同一编码集合中低于 i 代的数据解码, 但低代数据解码时并不能保证高代数据解码。

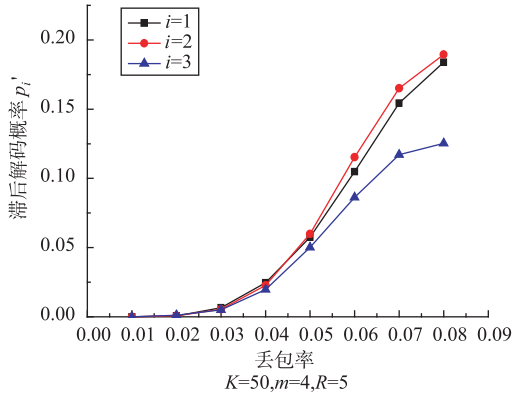


图 5 单代冗余 MGM-NC 中, 滞后解码概率

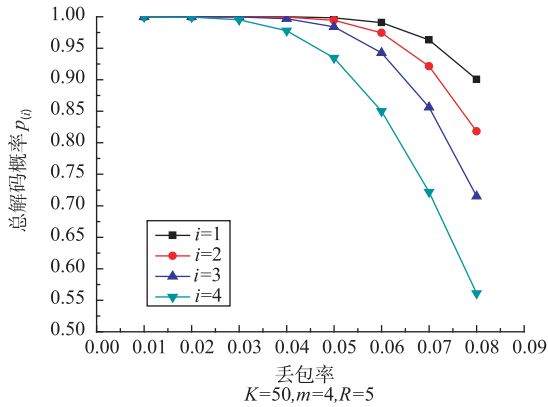


图 6 单代冗余 MGM-NC 中, 各代的解码率

从图 6 还可以看出, 当链路丢失率增大, 冗余不够充分时, 这种解码率的差异更加明显。图中当链路丢失率达到 8%, 10% 的冗余显得很不充分, 第 1、2、3、4 代的解码率分别为 90.5%、81.1%、72.1%、57.4%。

(2) m 代冗余

采用 m 代冗余 MGM-NC 方式时, 同样分析了立即解码概率 p_i 和滞后解码概率 p_i' 。

不同于单代冗余发送, 在 m 代冗余中除了第 m 代, 每代只发送了 K 个编码包, 因此第 i 代立即解码意味着包含第 i 代以及之前的所有编码包都要能成功接收到。

当 $i = m$ 时, 则只要确保总的接收编码包数目大于 $m \cdot K$ 个, 所以有下列公式:

$$p_i = \begin{cases} [(1 - \varepsilon)^K]^i & (i < m) \\ \sum_{i=m \cdot K}^{m \cdot K + m \cdot R} f(m \cdot K + m \cdot R, i) & (i = m) \end{cases} \quad (8)$$

$$p_i' = \sum_{r=0}^{i-1} p_r \left\{ \sum_{j_1=\max(K-m \cdot R, 0)}^{K-1} f(K, j_1) \cdot \sum_{j_2=(m-r) \cdot K-j_1}^{(m-r-1) \cdot K+m \cdot R} f((m-r-1) \cdot K + m \cdot R, j_2) \right\} \quad (9)$$

另外 p_i' 的计算与单代冗余不同, 当出现某一代无法立即解码情况时, 意味着该代和之后的所有代的数据都会滞后到接收完第 m 代编码包才能解码。

如式(9)所示, p_r 是前面 r 代数据立即解码的概率, 从第 $r+1$ 代开始不能立即解码, j_1 是第 $r+1$ 代接收到的编码包的数量, j_2 是剩下的 $m-(r+1)$ 代接收到的总编码包的数量, 它的值一定要能保证前面的代都能解码。

根据式(8)和(9), 图 7、图 8 分别记录的是在 m 代冗余 MGM-NC 方案下, 当 $K=50, m=4, R=5$ 时, 随着链路丢失率的变化, 各代立即解码概率和滞后解码概率的变化。可以看出相比于单代冗余, m 代冗余方案每一代都没有冗余编码包, 所以其立即解码概率比单代冗余低很多。

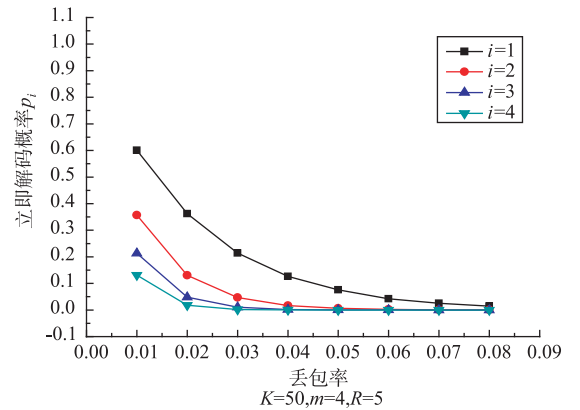


图 7 m 代冗余 MGM-NC 中, 各代立即解码概率

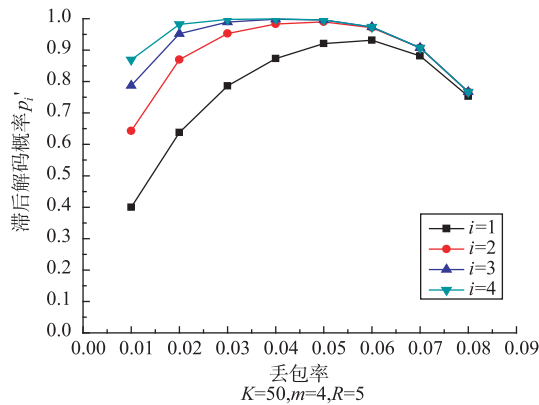


图 8 m 代冗余 MGM-NC 中, 各代滞后解码概率

从图 9 可以看出在运用 m 代冗余 MGM-NC 方案时, 各代的解码率曲线是几乎重叠在一起了, 说明每一代的解码率都很接近。

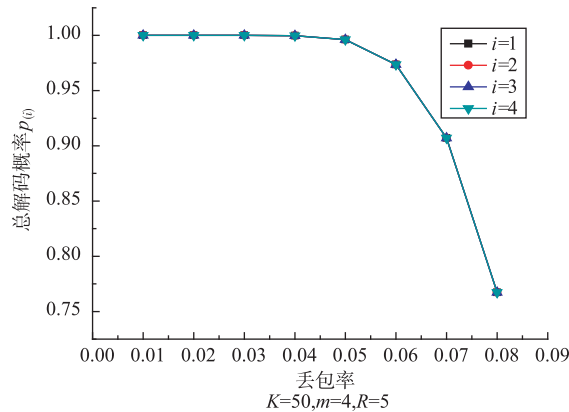


图 9 m 代冗余 MGM-NC 中, 各代的解码率

我们观察了各代解码率的差异, 当冗余包的总数超过了 m 代的总的丢包数, 即只要 $m \cdot K + m \cdot R$ 个编码包中收到大于等于 $m \cdot K$ 个, 各代都能解码, 这种情况的概率为 $P_c = \sum_{i=m \cdot K}^{m \cdot K + m \cdot R} f(m \cdot K + m \cdot R, i)$ 。而各代之间的差异主要是在不能满足上述条件下,

前面的代收到足够的数据包能解码造成的。第 i 代的解码概率与 P_c 之间的差异为 $[(1 - \varepsilon)^K]^i \cdot \sum_{j=0}^{(m-i) \cdot K - 1} f((m-i) \cdot K + m \cdot R, j)$ 。当 $K=50, m=4, R=5, \varepsilon=0.04$ 时, 第 1 代到第 4 代与 P_c 之间的差异值分别为 $6.7 \times 10^{-5}, 2.3 \times 10^{-7}, 5.5 \times 10^{-10}, 0$, 这些差异值是非常小的。改变各参数的取值大小时, 该差异值仍然可以忽略不计。

前面已给出了单代冗余和 m 代冗余 MGM-NC 两种方案的解码率公式。实际上, 在 MGM-NC 中, 当 m 取值为 1, 即混合集大小为 1 时, 该编码方案就是 SG-NC 方案。因此, SG-NC 方案也可以看作是 MGM-NC 的特例。

3 模拟验证

本文模拟验证了两种编码方案的性能分析结果。模拟的网络环境为一个源节点、一个目的节点的端到端传输, 模拟测试中, 取 $K=50, \varepsilon=0.08, R=0.1K$, 源节点向目的节点发送了 10000 个数据包。

实验测试了 SG-NC 和 MGM-NC 两种方案下各代的解码率。表 1 和表 2 分别记录了单代冗余 MGM-NC 和 m 代冗余 MGM-NC 方案, 在不同混合集大小下, 各代解码率的计算值和模拟测试值。可以看出模拟测试结果与理论计算结果是一致的。

在单代冗余 MGM-NC 中, 同一混合集中低代的解码率总是高于高代解码率; 而在 m 代冗余 MGM-NC 中, 同一混合集中各代的解码率相差很小。两者随着混合集大小 m 的增大, 各代的解码率都增加。

在同样冗余的情况下, 可以看到 m 代冗余方案的平均解码率要大于单代冗余, 而单代冗余方案中低代的解码率要高于 m 代冗余方案, 高代的解码率要低于 m 代冗余方案。这就意味着单代冗余方案低代的数据具有优先解码的特性。

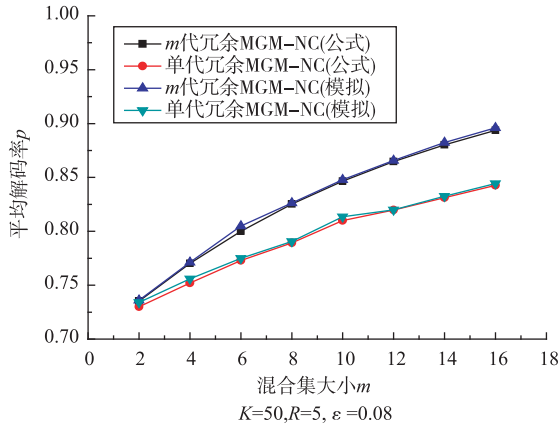
表 1 单代冗余 MGM-NC 的解码率 ($K=50, R=5, \varepsilon=0.08$)

解码率	$m=1$		$m=2$		$m=3$		$m=4$	
	公式	模拟	公式	模拟	公式	模拟	公式	模拟
g_1	0.724	0.7249	0.830	0.830	0.880	0.877	0.905	0.901
g_2			0.630	0.635	0.757	0.756	0.811	0.818
g_3					0.583	0.582	0.721	0.715
g_4							0.574	0.561
平均解码率	0.724	0.7249	0.730	0.732	0.738	0.738	0.752	0.749

表 2 m 代冗余 MGM-NC 的解码率 ($K=50, R=5, \varepsilon=0.08$)

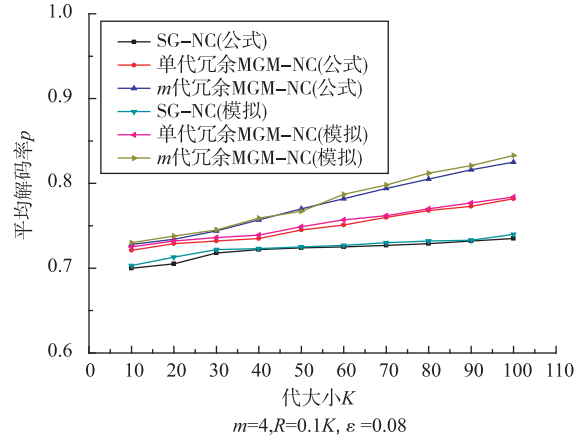
解码率	$m=1$		$m=2$		$m=3$		$m=4$	
	公式	模拟	公式	模拟	公式	模拟	公式	模拟
g_1	0.724	0.725	0.7354	0.7404	0.7532	0.7524	0.7705	0.7671
g_2			0.735	0.7402	0.7531	0.7521	0.7702	0.7669
g_3					0.7528	0.7519	0.7701	0.7667
g_4							0.7701	0.7667
平均解码率	0.724	0.725	0.7352	0.7403	0.753	0.7521	0.7702	0.7668

图 10 显示了混合集大小 m 取值变化时,单代冗余和 m 代冗余两种方案的平均解码率变化情况。仿真结果与理论计算值基本是一致的。结果表明,随着 m 的增加,单代冗余和 m 代冗余两种方案的平均解码率都增加。但是随着 m 的增加,该代数据的编解码复杂度也会增加,缓存需求和解码延迟也会增加。若代中数据包数为 K ,则解码矩阵大小为 $K \cdot m \times K \cdot m$,解码最大延迟为接收到所有 $K \cdot m$ 数据包,方能整体解码。因此在实际应用中, m 值不宜取太大,应根据应用背景进行选取。

图 10 混合集大小 m 的变化对平均解码率的影响

在前面的测试中, K 固定为 50,实际上, K 的大小对解码率也有很大的影响。图 11 给出了在冗余度一定情况下(即 $R=0.1 \cdot K$), K 的变化对三种编码方案的平均解码率的影响。不论采用哪种编码方案, K 越大平均解码率越高。这是因为随着 K 的增大,随机选择的 K 个编码向量线性独立的可能也会增加,因此每一代的解码率也会增加。

容易推出, K 个编码向量线性独立的概率 ρ 与有限域大小 F_q 之间的关系为 $\rho = 1 - \frac{K \cdot (K-1)}{2} \left(\frac{1}{2^q}\right)^{K-1}$ 。在相同 q 值下,随着 K 大小

图 11 K 的变化对平均解码率的影响

的增加, K 个编码向量线性独立的概率也增加。当 $q=8, K=2$ 时,编码向量线性独立的概率为 0.996;当 $q=8, K=50$ 时, K 个编码向量线性相关的概率为 1.1×10^{-115} , ρ 值则无限趋近于 1。因此,在实际应用中,取有限域大小 $F_q=2^8$ 就足够了。

图 12 显示的是,在 K, R, m 固定的情况下,随着链路丢失率的变化,三种编码方案的平均解码率的比较。从图中可以看出,在相同网络环境下,平均解

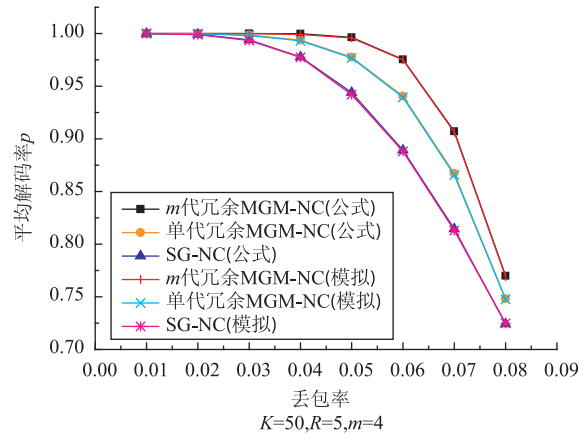


图 12 三种方案的平均解码率比较

码率由高到低分别为 m 代冗余 MGM-NC、单代冗余 MGM-NC、SG-NC 方案。图中模拟验证的结果与公式计算的结果基本一致。

我们还可以很直观地看出,不论采用哪种编码方案,随着链路丢失率的增加,接收节点的平均解码率都会下降。MGM-NC 方案的平均解码率总是要高于 SG-NC 方案。在 10% 的冗余度情况下, m 代冗余 MGM-NC 的平均解码率最大增益约为 10%,而单代冗余 MGM-NC 的平均解码率增益约为 5%。

图 13 显示的是,在 K, m 固定的情况下,随着链路丢失率的变化,为了保证每代的平均解码率至少在 90%,每代所需增加的冗余比例(即 R/K 的值)。从图中可以看出,在相同链路丢失率下,要保证 90% 的平均解码率, m 代冗余 MGM-NC 所需的冗余比例最小,SG-NC 所需的冗余比例最大。尤其是当链路丢失率较大时,采用多代混合编码方案的优势更明显。当 $\varepsilon = 0.4$ 时,采用 m 代冗余,单代冗余和 SG-NC 三种方案所需的冗余编码包的比例分别为 73%, 76%, 81%。使用多代混合网络编码可以减少冗余包的发送数量,提高网络的吞吐量。

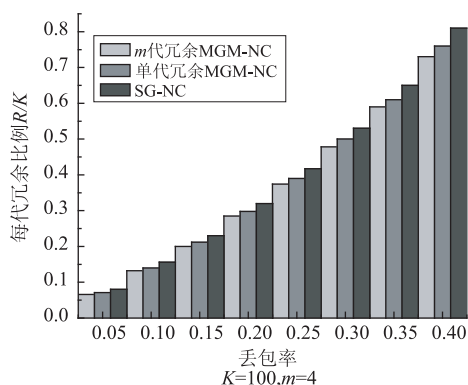


图 13 为了保证平均解码率至少在 90%, 三种方案所需的冗余比例

4 结论

本文的主要贡献是从理论上分析并推导了 MGM-NC 和 SG-NC 两种方案各代的解码率公式,分析了无线有损网络中多代混合网络编码 MGM-NC 的解码性能,并分析了代大小、冗余方式、链路丢失率等因素对每代解码率的影响。模拟结果验证了推导公式的正确性。

分析结果表明,在有差错的网络中,相同冗余度

情况下,MGM-NC 方案可以获得高于 SG-NC 方案的平均解码率; m 代冗余 MGM-NC 的平均解码率要高于单代冗余 MGM-NC;单代冗余 MGM-NC 方案的低代数据的解码率高于高代解码率。

本文分析得到的公式与结果,可以为有明确服务质量要求的传输,计算在一定丢包传输环境下所需的冗余传输数据量,从而确定满足服务质量的编码传输方式。同时,本研究的结果也为不同优先级别的数据流采用单代冗余 MGM-NC 提供了理论依据。由于单代冗余 MGM-NC 中,低代数据的解码率总是要高于高代的解码率,可以将该编码方案应用于流媒体编码传输中,例如在可分级视频编码 (scalable video coding, SVC) 中,视频信息被分成优先级别不同的数据层,在编码传输时,应用单代冗余 MGM-NC 可以将高优先级数据层放在低代中,从而使其获得较高的解码率。另一方面,尽管多代混合编码在同样冗余传输中获得了较好的解码率,在实际应用中是否采用多代混合编码方案,需要另外考虑多代混合编码在存储和计算量上的增加。

参考文献

- [1] Allswede R, Cai N, Li S R, et al. Network information flow. *IEEE Transactions on Information theory*, 2000, 46 (4): 1204-1216
- [2] Li S R, Yeung R, Cai N. Linear network coding. *IEEE Transactions on Information theory*, 2003, 49 (2): 371-381
- [3] Wu Y, Chou P A, Jain K. A comparison of network coding and tree packing. In: *Proceedings of International Symposium on Information Theory*, Chicago, USA, 2004. 145
- [4] Robins G, Zelikovsky A. Improved steiner tree approximation in graphs. In: *Proceedings of the 11th Annual ACM-SIAM Symposium on Discrete Algorithms*, San Francisco, USA, 2000. 770-779
- [5] Deb S, Effros M, Ho T, et al. Network coding for wireless applications: a brief tutorial. In: *Proceedings of International Workshop on Wireless Ad-hoc Networks*, London, UK, 2005.
- [6] Ramamoorthy A, Shi J, Wesel R. On the capacity of network coding for wireless networks. In: *Proceedings of the 41st Annual Allerton Conference on Communication Control and Computing*, Monticello, USA, 2003
- [7] Fragouli C, Boudec J, Windmer J. Network coding: an instant primer. *ACM SIGCOMM Computer Communication Review*, 2006, 1 (36): 63-68
- [8] Koetter R, Kschischang F. Coding for errors and erasures

- in random network coding. *IEEE Transactions on Information Theory*, 2008, 54(8):3579-3591
- [9] Chou P, Wu Y, Jain K. Practical network coding. In: Proceedings of the 41st Annual Allerton Conference on Communication, Control and Computing, Monticello, USA, 2003. 1001-1003
- [10] Ho T, Medard M, Shi J, et al. A random linear network coding approach to multicast. *IEEE Transactions on Information Theory*, 2006, 52(10):4413-4430
- [11] Katti S, Rahul H, Hu W, et al. XORs in the air: practical wireless network coding. *IEEE/ACM Transactions on Networking*, 2008, 16(3):497-510
- [12] Halloush M, Radha H. Network coding with multigeneration mixing: a generalized framework for practical network coding. *IEEE Transactions on Wireless Communications*, 2011, 10(2):466-473
- [13] Halloush M, Radha H. Network coding with multigeneration mixing. In: Proceedings of 42nd Annual Conference on Information Science and Systems, Princeton, USA, 2008. 515-520
- [14] Halloush M, Radha H. Performance evaluation: priority transmission using networking coding with multigeneration mixing. In: Proceedings of 43rd Annual Conference on Information Science and Systems, Baltimore, USA, 2009. 424-429

Analysis and study of the performance of multi-generation mix network coding in lossy networks

Wang Weiping, Fei Liufeng, Song Hong

(School of Information Science and Engineering, Central South University, Changsha 410083)

Abstract

The theoretical analysis of the decoding performance of single-generating network coding (SG-NC) and multi-generation mix network coding (MGM-NC) was carried out, and the two coding schemes' decoding probability formulas of original packets in different generations were obtained. The analysis results show that the decoding probability is related to the generation size, the redundant packets amount, the redundant transmission policy, the link lose probability and the size of mix set. With the same generation size and the same redundant transmission amount, the average decoding probability of the MGM-NC scheme is greater than that of the SG-NC scheme. And in the same mix set, the generation with a lower number has a greater decoding rate than the generation with a higher number. Especially when the redundant packets are send generation by generation, this difference is significant. The simulation results also validate the results of the theoretical analysis.

Key Words: network coding, multi-generation mix, mix set, decoding probability