Hash Function Generation by Neural Network Summary

作者：李靖东

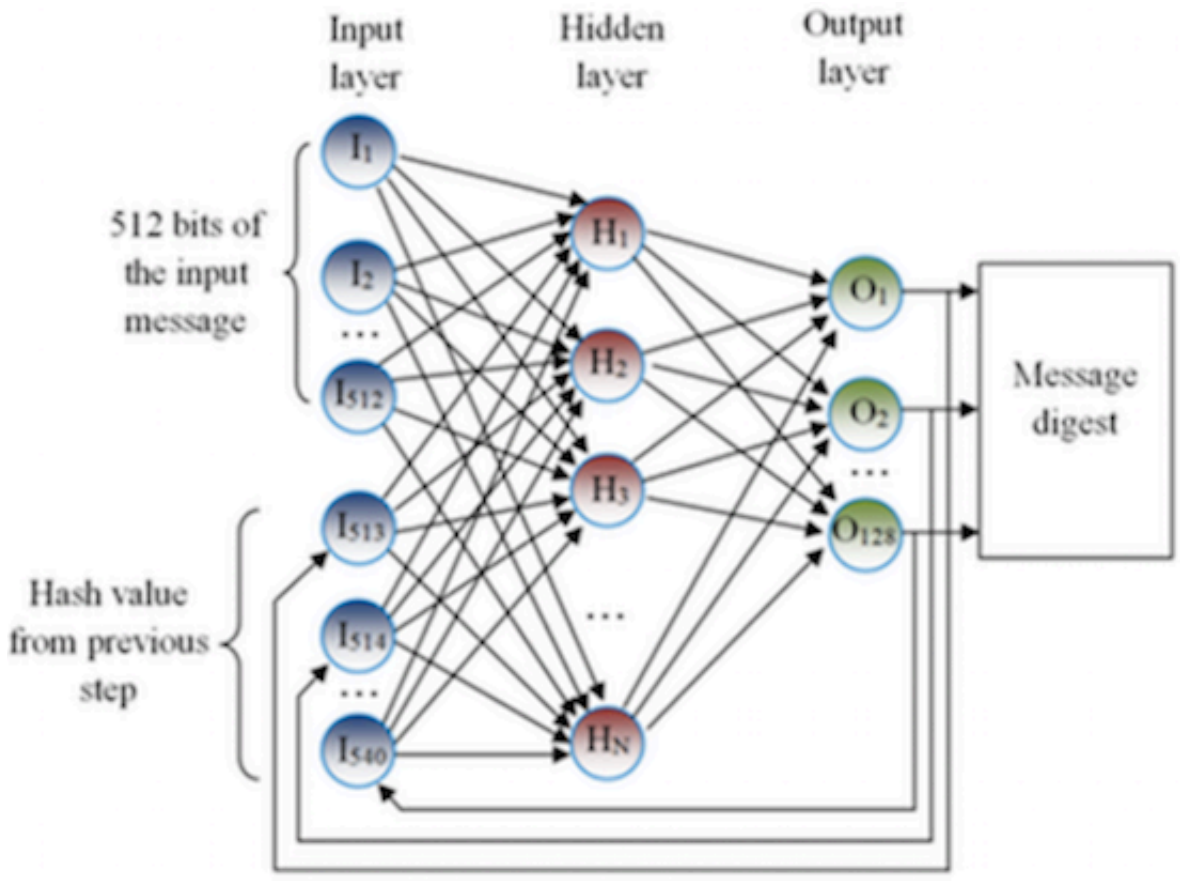
1.动机

随着互联网的不断发展，越来越多的文本和个人信息在网络上不断传递，随之而来的便是信息安全问题。Hash算法(例如：MD5，SHA-1)能够将输入的文本字符串信息转化为固定长度的数字字符串，能够起到压缩数据和保护信息的作用。作者尝试使用**ANN(artificial neural network)** 来拟合出一个符合标准的hash函数，为了评价拟合出的hash函数的效果，引入密码学中的雪崩效应作为评价指标：

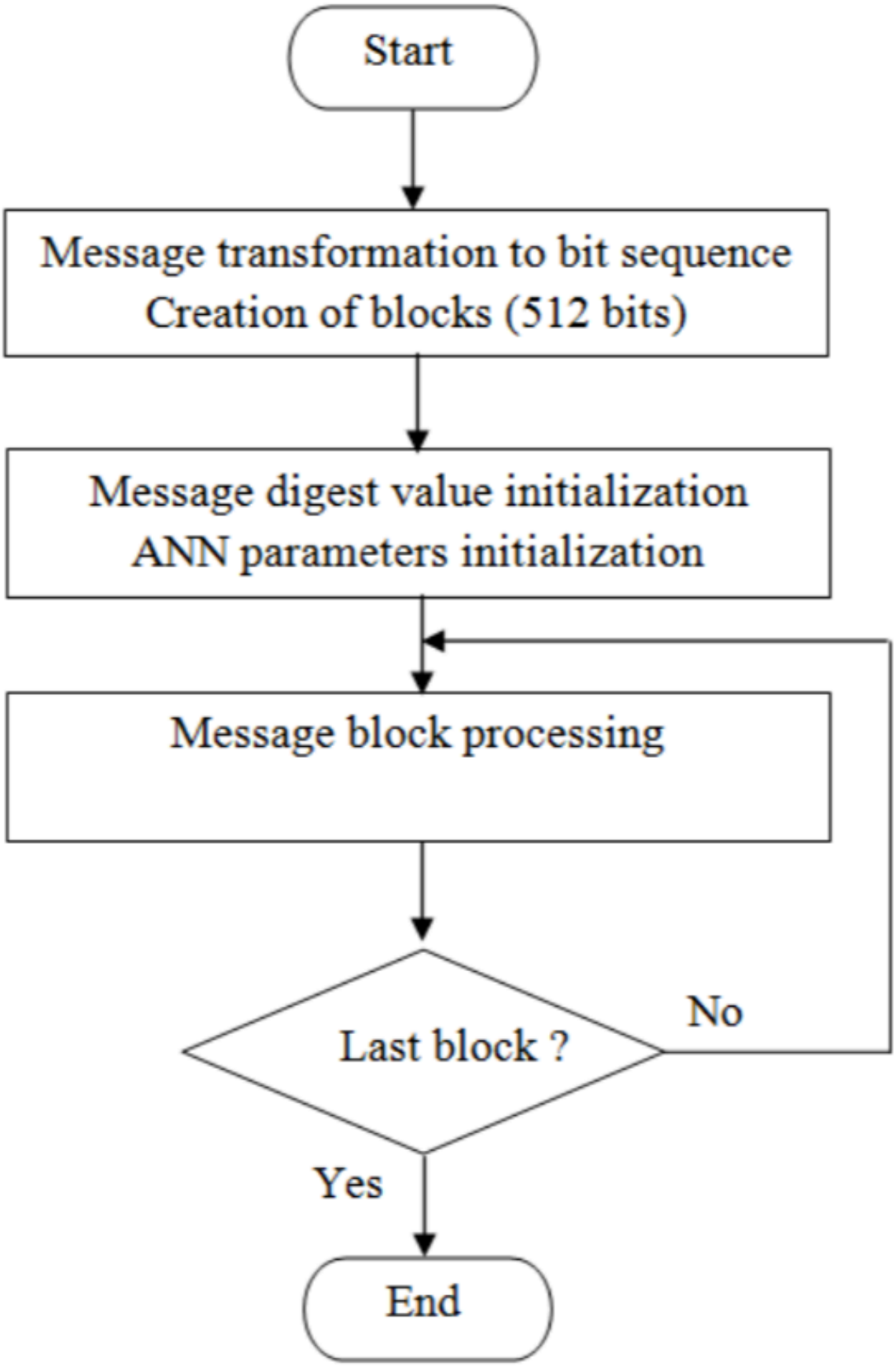
> 在密码学中，**雪崩效应（avalanche effect）**指加密算法（尤其是块密码和加密散列函数）的一种理想属性。雪崩效应是指当输入发生最微小的改变（例如，反转一个二进制位）时，也会导致输出的不可区分性改变（输出中每个二进制位有50%的概率发生反转）。合格块密码中，无论密钥或明文的任何细微变化都必须引起密文的不可区分性改变。(引自wiki)

2.模型

作者使用了一个包含三层神经元的神经网络，如下图所示：



具体流程如下图所示，第一次输入使用的Hash value需要单独初始化，随后用产生的结果替代



3.实验

实验目的：为了找到最优的隐藏层神经元个数

实验设计：在不改变ANN输入层和输入层神经元个数的情况下，将隐藏层神经元个数从10至130每次增加10个，更改输入文本中的某些字符，比较生成的key值位数的变化（以雪崩效应为评价指标）

实验输入：

Testing text - The number of characters			
English	**The number of characters**	**Slovak**	**The number of characters**
TextEn1	3587	TextSvk1	3912
TextEn2	11,658	TextSvk2	12,104
TextEn3	19,782	TextSvk3	20,025
TextEn4	32,451	TextSvk4	33,147
TextEn5	61,456	TextSvk5	61,259

实验结果：

TABLE II. COMPARISON of ANN's for 5 CHANGED CHARACTERS in ANALYZED ENGLIS TEXTS

Number of neurons	**Number of changed characters in the hash value**	
	5 changed characters in the text	**10 changed characters in the text**
10	6	7
20	9	9
30	11	10
40	11	11
50	10	9
60	12	11
70	9	8
80	11	10
90	12	13
100	11	12
110	10	9
120	12	12
130	11	12

TABLE III. COMPARISON of ANN's for 5 CHANGED CHARACTERS in ANALYZED SLOVAK TEXTS

Number of neurons	Number of changed characters in the hash value	
	5 changed characters in the text	10 changed characters in the text
10	7	7
20	8	7
30	9	10
40	8	9
50	9	11
60	11	11
70	10	11
80	9	11
90	12	12
100	11	12
110	12	12
120	11	12
130	13	12

4.结论

作者展示了如何利用ANN来拟合一个符合密码学约束的Hash函数，并且做了相关的优化实验。

5.评论

本文是一篇只有5页的demo，所以很多方面讲的不够具体，而且整个过程存在一些不足之处：

- 对于ANN模型中输入层为什么要添加额外的Hash Value没有做解释
- 实验中没有和现有的一些Hash函数的结果进行对比