神经网络在安全领域的应用

——以密钥交换协议为例

网络空间安全　赵正一

摘要

随着技术的进步和信息管理系统的日益强大，加强信息安全性的问题也变得越来越关键。在过去几年中，出于各种目的,大量使用通信网络给我们的生活带来了新的严重安全威胁，并增加了违法行为可能造成的潜在损害。随着各机关组织对计算机网络环境的依赖日益增加，他们越来越容易受到安全漏洞的攻击。如今，私营和公共部门比以往任何时候都更依赖于它们拥有的信息。违反信息安全性可能会危害整个系统的工作，并造成严重的损害。神经网络的进步为该问题提供了有效的解决方案。在这里，安全性问题被认为是使网络上的通信保持私密性的问题。换句话说，安全网络仅允许预期的收件人截取并阅读发给她/他的消息。所以可以预见的是，建立一个有效的通信信道是至关重要的。近些年来，一个叫做神经密码学的新技术逐渐崭露头角，它通过相互学习机制构建会话密钥，以此代替密钥交换协议。本文对神经网络在安全领域的应用进行了概述，并针对在密钥交换协议上的应用进行了研究分析。

关键字：密钥交换协议；神经网络；安全信道

Abstract

With the advancement of technology and the increasing strength of information management systems, the issue of strengthening information security has become more and more critical. In the past few years, the extensive use of communication networks for various purposes has brought new and serious security threats to our lives and increased the potential damage that may be caused by illegal activities. Through the increasing reliance of various agencies and organizations on

the computer network environment, they are becoming more and more vulnerable to security breaches. The private and public sectors are now more dependent on the information they have than ever before. Violation of information security may endanger the work of the entire system and cause serious damage. The advancement of neural networks provides an effective solution to this problem. Here, the security issue is considered to be the issue of keeping the communication on the network private. In other words, the secure network only allows the intended recipient to intercept and read the message sent to her/him. So it is foreseeable that the establishment of an effective communication channel is solid. In recent years, a new technology called neural cryptography has gradually emerged. It uses a mutual learning mechanism to convert session keys to replace exchange exchange protocols. This article gives an overview of the application of neural networks in the security field, and conducts research and analysis on the application of key exchange protocols.

Keywords：Key Exchange Protocol; Neural Network; Security Channel

目录

1 背景介绍

随着现代互联网的不断发展，人们通常认为，可以通过加密通信来保护通信双方的通信安全，但是通过实践说明，这样的观点是行不通的。因此，对通信安全的研究不仅仅包括加密，还包括流量安全。它的本质在于隐藏信息。这样的隐藏分为两种，一种是狭义上的隐藏，即把重要的信息藏匿于某种介质中，通过在公开信道传输介质来保证重要信息的传播 [1]。这一概念属于信息隐藏的范畴，这里按下不表。除此之外，广义上的隐藏还包含了通信信道加密，可以看作是把通信信道进行了某种隐藏。

怎么搭建一种安全的信道是值得我们研究的问题。通常，我们会采用密钥交换协议来沟通会话密钥，通过某种对称密钥算法，来构建一个安全的通信信道。密钥交换协议在确保通过不安全的通信信道进行交流的网络通信中起着至关重要的作用。查阅文献 [1][2] 我们可以发现，现在已经发展出了大量的密钥交换协议方案。其中大多数的安全性是基于传统的 Diffie-Hellman（DH）协议来开展的。但是在量子计算机的存在下，这种基于数论的密钥交换协议变得不再安全。因此，我们需要一种能够高效快速自动生成对称密钥的算法，来代替传统的基于 DH 的密钥交换协议。

本文研究了一种基于神经密码学的密钥交换协议。神经密码学是通过通信双方构建相同的神经网络，并将这种固定的神经网络拓印至某种特殊的电路中。通过相互学习机制，连续进行沟通数据，通过多轮学习，最终达到双方的神经网络权重相等的结果。由于这种学习方式的黑盒特征，攻击者很难找到一种攻击方式来对神经密码学进行有效的攻击。通信双方可以将权重矩阵作为沟通的会话密钥，用于后续的信息交流。

这是神经网络在信息安全领域的一个重要应用，这一方法的提出更深层次的打破了人工智能领域和信息安全领域的壁垒 [3]。对以后继续探索神经网络与信息安全相融合探索了新的途径。

2 相关工作

神经密码学的研究已经开展了很多，在这里本文简要回顾了关于神经密码学的相关工作。

标准树奇偶校验机 对于树奇偶校验机的研究工作已经有很多了。Mislovaty 证明了两个树奇偶校验机的同步可以通过相互学习的规则来实现。在这个基础上，已经同步了的树奇偶校验机可以用来做加密密钥交换协议。从这些研究成果之后，Ruttor[4] 等人分析了整个树奇偶校验机结构中的同步过程，并发现有两个主要的原因导致它同步的发生。相互吸引的移动和相互排斥的移动。基于这些移动，Ruttor 等人证明了两个树奇偶校验

机的同步时间取决于突触深度。在部分论文中，作者认为树奇偶校验机之间单一权重值的同步可以类似于赌徒问题。这也使得树奇偶校验机的同步可以通过贪心算法的扩展定理证明。

改进的协议 为了使用基于树奇偶校验机的神经密码学，前人提出了各种通过扩展树奇偶校机的基本概念从而得出的模型。Santhanalakshmi[5] 等人提出了通过扩展基本神经密码学来交换组密钥的新协议。作为密钥交换的扩展构建模块，Volkmer 提出了一种基于树奇偶校验机的经过身份验证的密钥交换协议。此外，前人通过实验表明，可以通过秘密边界来实现带有身份验证的密钥交换。如上所述，由于在资源受限的环境中难以使用传统的神经密码学算法。有人提出了迷你树奇偶校验机来实现嵌入式系统中的密钥交换。对于基于树奇偶校验机的密钥交换协议的实际使用，Volkmer 和 Wallner 提出了一种重密钥算法，用于通过重用先前交换过的密钥来生成新的密钥。

某种攻击下的安全性 为了分析树奇偶校验机模型的安全性，已经提出了各种攻击模型，如简单攻击、图形攻击、多数攻击和模拟攻击。在各种攻击情形下的安全性分析中，参与者可以通过增加突触深度来防止攻击。但是，如果增加突触深度，树奇偶校验机的效率会降低。也就是说，树奇偶校验机的同步时间会随着突触深度的增加而增加。为了研究满足合理安全性的树奇偶校验机参数，Salguero 等人通过改变内部参数对集合攻击的安全性进行了实验分析。但是，他们仅仅考虑了集合攻击，因此还不能得出他们的最佳参数是否满意足针对多数攻击或者针对模拟攻击的合理安全性。

树奇偶校验机的某种变体 为了提高效率和安全性，已经提出了各种密钥交换协议。这些协议根据方法可以大致分为三类：调整输入值、调整输出值和重建模型体系结构。在调整输入值的情况下，隐藏单元用于生成输入值，一时的输入值发生混乱。由于攻击者无法知道两个参与者的隐藏单元，因此公共的输入值会改变部分为私密的输入值。因此由于输入值的机密性，攻击者比以前更难以攻击树奇偶校验机。另一方面，隐藏单元可以间接影戏那个输入值的生成。为了加速双向的同步学习，参与者生成与其隐藏单元有关的输入值，而不是随机值。在干扰输出值的情况下，提出了一种名为 DTMP[6] 的机制。DTMP 允许两个参与者将预先定义好的噪声添加到计算出的输出值中，从而使试图恶意使用公共输出值的攻击者无从下手。但是这一机制还需要另外一个条件，即两个参与者必须实现进行协商才能产生相同的噪声。

我们通过总结以上工作，发现所有的神经密码学方案都建立在树奇偶校验机上，所以后文将对这一核心概念树奇偶校验机进行详细介绍。

3 模型介绍

3.1 密钥交换协议

建立安全通道的能力是现代通信研究中最具挑战性的领域之一。密码学的基本任务之一就是生成密钥交换协议 [7]。双方都以私钥开始，并使用公共协议传输其加密的私钥，经过一些转换之后，这些私钥会生成一个公共密钥。Diffie-Hellman 密钥交换协议 [8] 适用于生成公共密钥的典型协议。

所有已知的安全密钥交换协议都是用单向功能，这一功能通常是基于数论的，尤其是基于难以分解大质数的乘积。通常，N 位长度的密钥在两个通信方之间传输，并通过某种函数转换位公共密钥。密码学理论的基本问题之一就是：是否有可能建立一个不依赖于数论的安全密码系统；是否可以传输少于 N 比特长度的密钥；是否可以生成很长的密钥，可以直接用于一次性的流密码。

3.2 人工神经网络

人工神经网络的诞生源自于真实的神经网络 [9][10]。人们通过研究神经元的特性发现，学习可以被看作是在神经元之间建立新的连接或者是对已经有的类似于连接的物质进行修改的过程。所以对于人工神经网络来说，就要建立起权重与输入相乘最终得到某个结果的等式。其中 W 是权重矩阵，X 是输入矩阵，o 是对应的输出。

神经网络是有多层神经元群组成的，由泰勒展开公式 [11] 可知，一个特定的函数可以分解为多个非线性函数的加和，也可以说多个非线性函数加和的时候，可以模拟出某一个特定的函数。

$$f(x) = \frac{f(x_0)}{0!} + \frac{f^{'}(x_0)}{1!}(x - x_0) + ... + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + R_n(x) \tag{1}$$

基于这个想法，我们将输入的 x，匹配某个权重 w，再增添某个偏置 b，就可以得到对于输入 x 的线性函数。再将结果输入某个激活函数中，得到最终结果。

$$f(x) = sign(wx + b) \tag{2}$$

我们将由输入空间到输出空间的到的函数称为感知机。多个神经元以全连接形式层次相连，形成的网络称为前馈神经网络，也称为多层感知机（MLP），由泰勒展开可以得知，MLP 理论上可以模拟所有函数 [12]。其中模型为 $y = F(x)$，训练数据为 $D = x_i, y_i{}_{i=1}^{n}$，预测数据为 $\hat{y}_i = F(x_i)$，训练目标为 $\min(|\hat{y}_i - y_i|)$

3.3 反向传播

通过多层感知机，我们可以得到最终的函数如下 [13][14]

$$y = F(x) = f_3(W_3, f_2(W_2, f_1(W_1, x))) \tag{3}$$

并且通过梯度下降法优化目标

$$E = \frac{1}{2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4}$$

梯度是误差对于权重的偏导数，误差通过下面的公式更新参数

$$W^{t+1} = W^t - \eta_t \frac{\partial E}{\partial W} \tag{5}$$

由于偏导数存在链式法则，我们可以通过从后向前反向传播的方式计算梯度

$$\frac{\partial E}{\partial W_1} = \frac{\partial E}{\partial f_3} \cdot \frac{\partial f_3}{\partial f_2} \cdot \frac{\partial f_2}{\partial f_1} \cdot \frac{\partial f_1}{\partial W_1} \tag{6}$$

但是梯度下降法存在以下问题 [15]：目标函数通常不是标准的凸函数，所以非常容易陷入局部最优解而不是全局最优解；网络层数增多后，容易出现梯度消失或者梯度爆炸问题

3.4 树奇偶校验机

树奇偶校验机（Tree Parity Machine, TPM）[16] 由 K 个隐藏层单元组成，其输入向量是 X，初始权重是 W，输出单元是 σ_k。

$$X \subset x_{ij} \in \{-L, \cdots, 0, \cdots, +L\}$$

$$W \subset w_{ij} \in \{-L, \cdots, 0, \cdots, +L\}$$

$$\sigma_k = \text{sign}(\sum_{j=1}^{N} w_{ij}x_{ij}) \tag{7}$$

其中，sign(·) 函数代表取自变量的数学符号，用来表示自变量的正负形，所以有：

$$\text{sign}(\cdot) \in \{-1, 0, +1\}$$

TPM 最终的输出为 τ：

$$\tau^{A/B} = \prod_{k=1}^{K} \sigma_k^{A/B} \tag{8}$$

TPM 的构建是一个动态的过程，通过多轮对比最终构建并同步好彼此的神经网络。其构建流程如下 [17][18]

Algorithm 1 TPM 双方同步按照以下流程执行

1: 随机初始化权重矩阵 $w_{ij}^{A/B}$;
2: **while** $w_{ij}^{A} \neq w_{ij}^{B}$ **do**
3: 随机生成相同的输入矩阵 x_{ij};
4: 计算 $\sigma_i^{A/B}$ 和 $\tau^{A/B}$ 的结果;
5: **for all** τ^A 和 τ^B，比较他们的结果，并按照如下结果继续执行 **do**
6: $\tau^A = \tau^B$：更新彼此的权重;
7: $\tau^A \neq \tau^B$：回到第三步;
8: **end for**
9: **end while**

TPM 的学习规则如上述流程图所示 [19][20][21]，在进行到第六步的时候，需要按照特定的规则来更新神经网络的权重，通常我们会使用以下三种更新方法

1. Hebbian 算法

$$w_i^+ = g(w_i + \sigma_i x_i \theta(\sigma_i, \tau^A)\theta(\tau^A, \tau^B)) \tag{9}$$

2. Anti-Hebbian 算法

$$w_i^+ = g(w_i - \sigma_i x_i \theta(\sigma_i, \tau^A)\theta(\tau^A, \tau^B)) \tag{10}$$

3. Random walk 算法

$$w_i^+ = g(w_i + x_i \theta(\sigma_i, \tau^A)\theta(\tau^A, \tau^B)) \tag{11}$$

其中，$\theta(x, y)$ 代表了 x 和 y 的相等关系，g 函数表示某种运算法则使得运算后的结果仍然保持在原运算空间内 [22][23]：

$$\theta(x, y) = \begin{cases} 1, & x = y \\ 0, & x \neq y \end{cases} \tag{12}$$

4 实验验证

4.1 实验设计

本文所做的实验实在如下环境开展的

1. 操作系统：Ubuntu 20.04.2
2. 语言环境：Python 3.7.0

3. 库函数：bumpy：1.19.5

4.2 实验结果分析

最终结果如下图所示。Alice 和 Bob 在更新了 718 轮权重后完成了权重对齐，并且将相等的权重矩阵作为协商好的会话密钥使用。

```
$ python TreeParityMachine.python

  Creating machines : k = 100, n = 10, l = 10
  Using hebbian update rule.
  Synchronization = 100 % Updates = 718
  Machines have been synchronized.
  Time taken = 12.529 seconds.
```

5 总结和展望

本文对神经密码学近些年的研究工作进行了概述，并着重分析了树奇偶校验机的工作流程。通过实验验证，证明了树奇偶校验机的工作能力。神经网络目前在安全领域的应用仍然很少，但是神经网络的学习潜力将随着运算能力的普遍提升而越来越强。将神经网络与机器学习的相关技术应用到安全领域是未来发展的方向。

参考文献

[1] Habibzadeh F, Habibzadeh P, Yadollahie M, et al. On the information hidden in a classifier distribution[J]. Scientific reports, 2021, 11(1): 1-11.

[2] Young S, Schatzmann J, Weilhammer K, et al. The hidden information state approach to dialog management[C]//2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07. IEEE, 2007, 4: IV-149-IV-152.

[3] Wang Z, Ji Q. Classifier learning with hidden information[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 4969-4977.

[4] Katzenbeisser S, Petitcolas F A P. Digital watermarking[J]. Artech House, London, 2000, 2.

[5] Cox I, Miller M, Bloom J, et al. Digital watermarking and steganography[M]. Morgan kaufmann, 2007.

[6] Kundur D, Hatzinakos D. Digital watermarking using multiresolution wavelet decomposition[C]//Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181). IEEE, 1998, 5: 2969-2972.

[7] Günther C G. An identity-based key-exchange protocol[C]//Workshop on the Theory and Application of of Cryptographic Techniques. Springer, Berlin, Heidelberg, 1989: 29-37.

[8] Kaufman C, Hoffman P, Nir Y, et al. Internet key exchange protocol version 2 (IKEv2)[R]. RFC 5996, September, 2010.

[9] Liu F, Huo W, Han Y, et al. Study on network security based on PCA and BP neural network under green communication[J]. IEEE Access, 2020, 8: 53733-53749.

[10] Kang M J, Kang J W. A novel intrusion detection method using deep neural network for in-vehicle network security[C]//2016 IEEE 83rd Vehicular Technology Conference (VTC Spring). IEEE, 2016: 1-5.

[11] Shihab K. A backpropagation neural network for computer network security[J]. Journal of Computer Science, 2006, 2(9): 710-715.

[12] Kang M J, Kang J W. Intrusion detection system using deep neural network for in-vehicle network security[J]. PloS one, 2016, 11(6): e0155781.

[13] Hecht-Nielsen R. Theory of the backpropagation neural network[M]//Neural networks for perception. Academic Press, 1992: 65-93.

[14] LeCun Y, Touresky D, Hinton G, et al. A theoretical framework for back-propagation[C]//Proceedings of the 1988 connectionist models summer school. 1988, 1: 21-28.

[15] LeCun Y, Boser B E, Denker J S, et al. Handwritten digit recognition with a back-propagation network[C]//Advances in neural information processing systems. 1990: 396-404.

[16] Zhang Y, Xiang T, Hospedales T M, et al. Deep mutual learning[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 4320-4328.

[17] Scholz R W. Mutual learning as a basic principle of transdisciplinarity[C]//Transdisciplinarity: Joint Problem-solving among Science, Technology and Society. Proceedings of the International Transdisciplinarity 2000 Conference. Workbook II: Mutual Learning Sessions, Haffman, Zürich. 2000: 13-7.

[18] Béguin P. Design as a mutual learning process between users and designers[J]. Interacting with computers, 2003, 15(5): 709-730.

[19] Volkmer M, Wallner S. Tree parity machine rekeying architectures[J]. IEEE Transactions on Computers, 2005, 54(4): 421-427.

[20] Rosen-Zvi M, Klein E, Kanter I, et al. Mutual learning in a tree parity machine and its application to cryptography[J]. Physical Review E, 2002, 66(6): 066135.

[21] Volkmer M, Schaumburg A. Authenticated tree parity machine key exchange[J]. arXiv preprint cs/0408046, 2004.

[22] Jeong S, Park C, Hong D, et al. Neural cryptography based on generalized tree parity machine for real-life systems[J]. Security and Communication Networks, 2021, 2021.

[23] Chen T, Huang S H. Tree parity machine-based one-time password authentication schemes[C]//2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). IEEE, 2008: 257-261.