Трофимов Павел Владимирович

Ордена трудового красного знамени

ФГБОУ ВО «Московский технический университет связи и информатики»

Студент

**Криптографические системы, основанные на нейронных сетях**

Модель искусственных нейронных сетей подходит для построения хэш-функций на её основе, искусственный нейрон выступает в качестве базового блока при построении искусственной нейронной сети. Используя модель искусственного нейрона, у которого **n** входов, **n** весовых характеристик по одной на каждый вход. Смещение, подаваемое на вход функции активации и единственное выходное значение (рис. 1).



*Рисунок 1 - Модель искусственного нейрона*

Имея известный входной вектор вычислить выходное значение легко, однако обратная задача получения входного вектора нейрона по известному выходному значению является трудной, маловероятной для решения. Используя эти данные можно сделать вывод о том, что искусственные нейронные сети обладают свойством однонаправленной функции.

Особую роль в криптографии играют однонаправленные функции, которые в общем случае не являются биективными.

Однонаправленной называется такая функция ***f***, для которой легко определить значение функции ***y=f(x)***, но практически невозможно отыскать для заданного ***y*** такое ***x***, что ***y=f(x).***

Для построения криптографических систем защиты информации чаще используются однонаправленные функции, для которых обратное преобразование существует и однозначно, но вычислительно нереализуемо. Они называются вычислительно необратимыми функциями.

В качестве примера однонаправленной функции ***y=f(x)*** рассмотрим широко известную функцию дискретного возведения в степень: ***y=αx(mod p)***, где ***x*** – целое число от 1 до ***p-1*** включительно, а вычисление производится по модулю ***p***, где ***p*** – очень большое простое число; ***α*** – целое число ***(1<α<p).***

Помним, что простым числом называется целое число, которое не делится ни на какие числа, кроме себя самого и единицы.

Пример:

Для примера возьмем небольшое простое число ***p=7***; тогда для осуществления преобразований можно выбрать примитивный элемент ***α=3***, так как ***α1(mod 7)=3***, ***α2(mod 7)=32(mod 7)=9(mod 7)=2***, ***α3(mod 7)=6***, ***α4(mod 7)=4***, ***α5(mod 7)=5***, ***α6(mod 7)=1***.

Функция ***y=αx(mod p)*** вычисляется сравнительно просто, а обратная к ней функция ***x=logyp*** является вычислительно сложной практически для всех ***(1<y<p)*** при условии, что не только ***p*** велико, но и (***p-1***) имеет большой простой множитель (лучше всего, если это будет другое простое число, умноженное на 2). В связи с этим такую задачу называют задачей нахождения дискретного логарифма или задачей дискретного логарифмирования.

Задача дискретного логарифмирования состоит в том, что для известных целых ***α***, ***p***, ***y*** необходимо найти целое число ***x***. Однако алгоритм вычисления дискретного логарифма за приемлемое время пока не найден. Поэтому модульная экспонента считается однонаправленной функцией.

По современным оценкам теории чисел при целых числах ***α≈2664*** и ***p≈2664*** решение задачи дискретного логарифмирования потребует около **1026** операций, что имеет в **103** раз большую вычислительную сложность, чем задача разложения на множители. При увеличении длины чисел разница в оценках сложности задач возрастает.

Следует отметить, что пока не удалось доказать, что не существует эффективного алгоритма вычисления дискретного логарифма за приемлемое время. Исходя из этого, модульная экспонента отнесена к однонаправленным функциям условно, что, однако, не мешает с успехом применять ее на практике.

Одним из первых применений однонаправленных функций было решение задачи обеспечения безопасности и использования пароля, по которому осуществляется доступ пользователя к ресурсам и услугам в автоматизированных системах.

Открытое значение ***y*** вместе с именем пользователя может быть помещено в список паролей доступа, хранящихся в, базе данных. Законный пользователь для получения доступа в автоматизированную систему предъявляет свое число ***x***. Вычисляется по этому числу значение однонаправленной функции ***y=αx(mod p)*** и сравнивает с хранящимся значением ***y***. При совпадении этих значений пользователь становится идентифицированным и получает требуемый доступ.

Кроме однонаправленных функций, не имеющих вычислительно простого обратного отображения даже для законных пользователей, знающих секретную ключевую информацию, в криптографии широко используются однонаправленные функции, для которых знание секретного ключа дает возможность законному пользователю вычислительно просто находить обратное отображение. Они получили название однонаправленных функций с потайным ходом, иногда их называют однонаправленными функциями с лазейкой.

Существует ряд научных работ, посвященных созданию алгоритмов хэширования, использующих искусственную нейронную сеть. Используемая в них искусственная нейронная сеть изображена на рисунке 2.



*Рисунок 2 - Модель искусственной нейронной сети, используемой для построение хеш-функции*

Используется трехслойная нейронная сеть прямого распространения и хаотическое отображение. Обычно в качестве хаотического отображения применяется некоторое количество итераций кусочно-линейной функции с контролирующим параметром. Хаотичность отображения при этом определяется значением параметра. Исследования показывают, что при правильном выборе количества итераций хаотического отображения выходное значение функции будет сильно отличаться при незначительных изменениях во входных параметрах. В приведенных работах в скрытом слое нейронов хаотическое отображение не итерируется, для того, чтобы сократить количество операций, требуемых для вычисления функции хэширования, а на входном и выходном слоях нейронной сети производится T (T ≥ 50) итераций. На основе описанной выше искусственной нейронной сети и генератора ключей строится алгоритм хэширования. Генератор ключей преобразует ключ пользователя в набор весовых характеристик, смещений и управляющих параметров, для каждого слоя. Алгоритм переводит данные произвольной длины в 128-битное значение хэш-функции. Для этого данные сначала добиваются до кратности блоку длиной 1024 бита, который и подается на вход нейронной сети, по следующему алгоритму: в последний некратный блок дописываем одну единицу и остальные нули. После этого каждый блок подаётся на вход нейронной сети. Результат вычисляется по формуле (1):

|  |  |
| --- | --- |
| $$H\_{M}=K\_{M\_{n-2}}⨁H\_{M\_{n-1}}=\left(K\_{M\_{n-3}}⨁H\_{M\_{n-2}}\right)⨁H\_{M\_{n-1}}=\cdots =\left(K⨁H\_{M0}\right)⨁H\_{M1}⨁\cdots ⨁H\_{M\_{n-1}}$$ | ((1) |

Модель искусственной нейронной сети подходит для задач шифрования.

Задача шифрования будет рассмотрена на примере алгоритма AES. AES (Advanced Encryption Standard) является итеративным блочным шифром, который был выбран NIST в качестве международного стандарта и замены DES. В настоящее время является наиболее широко распространенным блочным шифром. Rijndael является итеративным блочным шифром, шифрование или расшифрование блока данных осуществляется c помощью раундовой функции. Три стандартные версии AES называются AES-128, AES-192 и AES-256 и отличаются друг от друга длиной ключа (128, 192, и 256 бит) и количеством раундов (10, 12, и 14) Их безопасность была тщательно исследована АНБ. Тем не менее, в последнее время появились некоторые виды атак, которые, как было доказано, достаточно эффективно применяются к AES. Чтобы противостоять этим атакам, можно сделать некоторые изменения, внедрив нелинейную нейронную сеть в AES.

Нейронные сети используются для классификации и аппроксимации функций/выделения задач. Они устойчивы к некоторым неточностям, для которых имеется много доступных данных для обучения, но к которым не могут быть применены жесткие правила. Попытаемся реализовать Rijndael-криптосистему с помощью ИНС. Эта криптосистема имеет менее сложное строение, чем AES и не линейна в эксплуатации. Нелинейной должна быть нейронная сеть с обратной связью, что позволило бы выполнить шифрование/расшифрование открытого текста/зашифрованного текста с высокой производительностью и очень низким уровнем ошибок. Идея состоит в том, чтобы разработать такую нелинейную искусственную нейронную сеть. Нелинейность необходима для уменьшения вероятности взлома алгоритма. Уменьшение вероятности взлома достигается с помощью нелинейной функции активации, также свойство нелинейной аппроксимации сети является полезным для практического применения.

Многослойный персептрон (MLP) является наиболее популярным видом ИНС прямого распространения, его свойством является способность эмулировать любое соотношение входов/выходов. MLP отвечает здесь за частоту обновления, так как он обновляет веса после того, как весь блок обучающих данных представлен полностью. Блочная адаптация является более надежной с того момента, когда шаг обучения усреднен по всем обучающим шаблонам.

MLP имеет следующие параметры. Для процесса шифрования, могут быть использованы следующие настройки:

входной вектор, являющийся открытым текстом;

выходной — зашифрованный текст из алгоритма AES;

первоначальные веса будут ключом шифрования;

нелинейная функция активации каждого нейрона будет (log-сигмойда), что дает выходные значения между (0 — 1).

Для процесса расшифровки, могут быть использованы следующие параметры:

входным вектором является шифртекст;

выходные значения — открытый текст;

Первоначальные веса будут ключом для расшифровки;

нелинейная функция каждого нейрона будет (сигмовидной), что дает выход между (0 — 1).

Вход и выход можно масштабировать, чтобы иметь возможность совмещаться с функцией активации. В фазе обучения, Rijndael-алгоритм производит зашифрованный текст из определенного входного текста. ИНС принимает входной текст в качестве входного и выходного текста в качестве цели, и самообучается для достижения такого же выходного текста. Модель структуры искусственной нейронной сети называется «последовательно-параллельная модель», где алгоритм Rijndael влияет на динамическое поведение нейронной модели. Фаза операций будет многослойной ИНС прямого распространения с конечными весами, используемыми для получения выходного сигнала шифрования/расшифрования. На этом этапе будет продолжаться производство открытого текста/зашифрованного текста до тех пор, пока новый ключ не начнет использоваться в системе, тогда ИНС необходимо будет переобучиться на новый ключ. В шифровании и расшифровании, в качестве нелинейной функции активации для каждого нейрона был использован гиперболический тангенс. Свойство сигмоиды производит выходные значения в диапазоне (от -1 до +1). Данные, используемые в AES, находятся в диапазоне (0 — 255). Таким образом, чтобы искусственная нейронная сеть была совместима с данными AES (текстом, зашифрованным текстом, ключом шифрования) на этапе функции активации, есть необходимость преобразовать данные в диапазон (0 — 1). Это достигается путем использования коэффициента масштабирования (1/256) таким образом, чтобы стать совместимым с выходом функции активации. Коэффициент может быть использован на выходе каждого нейрона повторно для преобразования выходного диапазона обратно в (0 — 255). Для простоты воспользуемся только версией AES, имеющей 128-битный ключ. AES использует блок данных длиной 128 бит для шифрования/расшифрования. Таким образом, искусственная нейронная сеть будет использовать данные в форме байт, так что и вход, и выход будет размером 16 байт. Искусственная нейронная сеть должен иметь по крайней мере один скрытый слой в 16 нейронов для достижения размера входного ключа длины 16 байт. Для обучения искусственной нейронной сети имеем топологию (16 - 16 - 16 - 1). Она принимает вектор из 16 байт в качестве входных данных и производит вектор из 16 байт как обученный текст. Веса каждого слоя являются ключом (16 байт) вектор. В этой искусственной нейронной сети шифрование не будет производить нужный текст, потому что минимальный градиент будет достигнут без достижения целевой производительности. Рисунок отражает обучение процессов шифрования и дешифрования на основе 16-ти байт входного текста.

Тесты показывают, что выходные тексты на нейро-AES в шифровании и расшифровании не идентичны выходным текстам шифрования и расшифрования AES. Одним из основных способов решений этой проблемы является сокращение входных и целевых векторов искусственной нейронной сети. Это решение будет иметь тенденцию к сокращению времени подготовки и достижению целевой производительности. Идея сокращения заключается в создании более чем одной параллельной операции искусственной нейронной сети. Каждая операция искусственной нейронной сети будет работать с частью входных и выходных векторов.

На данный момент опробованы несколько топологий искусственной нейронной сети и обнаружено, что искусственная нейронная сеть с топологией (4 - 16 - 16 - 1) (рис. 5) является лучшим выбором. Нейро-криптосистема состоит из 4 слоев, работающих параллельно. Каждый из слоев искусственной нейронной сети принимает входной вектор длиной 4 байта. Топология желаемой искусственной нейронной сети в процессе расшифрования (рис. 4) / шифрования (рис. 3) выглядит следующим образом.



Рисунок 3 – Обучение шифрованию из 16 байт входного текста



Рисунок 4 – Обучение расшифрованию из 16 байт входного текста

Слой 1 с 4 нейронами, где каждый нейрон получает один байт из входного вектора (открытый текст/зашифрованный текст).

Слой 2 с 16 нейронами, где каждый нейрон суммирует вес всех 8 нейронов в 1-м слое.

Слой 3 с 16 нейронами, где каждый нейрон суммирует вес всех 8 нейронов во 2-м слое.

Слой 4 (выходной слой) с 1 нейроном, который суммирует вес всех 8 нейронов в третьем слое.

Начальные веса искусственной нейронной сети являются ключами для процесса расшифрования/шифрования. Таким образом, ключ длиной 16 байт из искусственной нейронной сети был взят следующим образом:

- 4x1 матрица в качестве начального веса от входа к слою 1;

- 16x4 матрица (представляет 16 байт из ключа и повторяется 4 раза) в качестве начальных весов от слоя 1 к слою 2;

- 16x16 матрица (представляет те же 16 байт в слое 1, но повторяется 16 раз) в качестве начального веса от слоя 2 к слою 3;

- 1x16 матрица (представляет то же самое в слое 1 от входа, но в транспонированном виде), в качестве первоначального веса от слоя 3 до выходного слоя.



Рисунок 5 – ИНС [4 — 16 — 16 — 1]

Уравнение выхода этой искусственной нейронной сети (2):

|  |  |
| --- | --- |
| $output=\tanh((\sum\_{k=1}^{16}W\_{1k} \*tanh⁡(\sum\_{k=1}^{16}\sum\_{j=1}^{16}W\_{kj} \*tanh⁡(\sum\_{j=1}^{16}\sum\_{i=1}^{4}W\_{ji} \*tanh⁡(\sum\_{i=1}^{4}W\_{i1} \*input))) ))$, | (2) |

где **output** — открытый текст/зашифрованный текст (в соответствии со способом), **w** — симметричный ключ. Тангенциальные сигмоиды передаточных функций в скрытых слоях сжимают бесконечный входной диапазон в конечный выходной диапазон.

Древовидные Машины Четности (ДМЧ) или Tree Parity Machines (TPM) - это особый вид многоуровневой нейронной сети прямого распространения, которая имеет следующий вид (рисунок 6).



*Рисунок 6 - Древовидная Машина Четности (ДМЧ) или Tree Parity Machines (TPM)*

Она состоит из одного выходного нейрона, **K** скрытых нейронов и **K**×**N** входных нейронов. Входные нейроны принимают двоичные значения (3)

|  |  |
| --- | --- |
| $x\_{ij}\in \left\{-1,+1\right\}$. | (3) |

Веса между входными и скрытыми нейронами равны (4)

|  |  |
| --- | --- |
| $w\_{ij}\in \left\{-L, …, 0, …, +L\right\}$. | (4) |

Значение каждого скрытого нейрона есть сумма произведений входного значения и весового коэффициента (5), (6)

|  |  |  |  |
| --- | --- | --- | --- |
| $$σ\_{i}=sgn(\sum\_{j=1}^{N}w\_{ij}x\_{ij})$$ | (5), | $$sgn\left(x\right)= \left\{\begin{array}{c}-1 if x\leq 0,\\1 if x>0.\end{array}\right.$$ | (6). |

Значение выходного нейрона есть произведение всех скрытых нейронов: $τ= \prod\_{i=1}^{K}σ\_{i}$, выходное значение также двоичное.

Для обмена ключами между двумя абонентами наиболее часто используется алгоритм Диффи-Хеллмана. Его более безопасная замена основана на синхронизации двух древовидных машин четности. Синхронизация этих машин похожа на синхронизацию двух хаотических осцилляторов в теории хаотических связей (chaos communications).

Динамика двух сетей и их весовых коэффициентов нашла применение в явлении, где сети синхронизируют состояния с идентичными весовыми коэффициентами, зависящими от времени. Эта концепция быстрой синхронизации по взаимному обучению может быть применена к протоколу обмена секретным ключом через публичный канал. А сгенерированный ключ может быть использован для шифрования и расшифрования передаваемого сообщения.

Алгоритм не оперирует большими числами и методами из теории чисел, следовательно, приводит к быстрой синхронизации открытого ключа. Безопасность нейрокриптографии постоянно подвергается сомнению, ведь из-за того, что метод основан на стохастическом процессе, есть небольшой шанс, что злоумышленник синхронизируется с ключом.

Также, было установлено, что защищенность обычных криптографических систем можно улучшить, увеличив длину ключа. В нейрокриптографии вместо ключа увеличивается синаптическая длина **L**. Это увеличивает сложность атаки экспоненциально, в то время как затраты абонентов на дешифрацию растут полиномиально.

Таким образом, взлом подобной системы является NP-сложной задачей. Есть утверждения, что исходный алгоритм нейросинхронизации может быть сломан, по крайней мере, тремя видами атак: геометрической, вероятностным анализом и генетическими алгоритмами.

Хотя данная реализация небезопасна, идеи случайной синхронизации могут привести к абсолютно безопасной схеме. Например, можно значительно улучшить защищенность обычных криптографических систем, увеличив длину ключа. Протокол обмена ключами, основанный на синхронизации двух древовидных машин четности, заключается в следующем.

У каждого абонента (А или В) есть своя ДМЧ. Их синхронизация происходит следующим образом:

1. Задаём случайные значения весовых коэффициентов

2. Выполняем следующие шаги, пока не наступит синхронизация

3. Генерируем случайный входной вектор X

4. Вычисляем значения скрытых нейронов

5. Вычисляем значение выходного нейрона

6. Сравниваем выходы двух ДМЧ:

7. Выходы разные: переход к п. 3

8. Выходы одинаковые: применяем выбранное правило к весовым коэффициентам

9. После полной синхронизации (веса $w\_{ij }$обоих TPM одинаковые), А и В могут использовать веса в качестве ключа.

Этот метод известен как двунаправленное обучение. Для обновления весовых коэффициентов могут использоваться следующие правила.

|  |  |
| --- | --- |
| - Правило положительного подкрепления:$w\_{i}^{+}=w\_{i}+σ\_{i}x\_{i}Θ\left(σ\_{i}τ\right)Θ\left(τ^{A}τ^{B}\right)$; | (7) |
| - Правило отрицательного подкрепления:$w\_{i}^{+}=w\_{i}-σ\_{i}x\_{i}Θ\left(σ\_{i}τ\right)Θ\left(τ^{A}τ^{B}\right)$; | (8) |
| - Случайное блуждание:$w\_{i}^{+}=w\_{i}+x\_{i}Θ\left(σ\_{i}τ\right)Θ\left(τ^{A}τ^{B}\right)$. | (9) |

Таким образом можно сказать что модель искусственных нейронных сетей подходит для построения хэш-функций Имея известный входной вектор вычислить выходное значение легко, а обратная задача является трудной, маловероятной для решения. Это позволяет сделать вывод, что искусственные нейронные сети обладают свойством однонаправленной функции. Однонаправленные функции. не являются биективными, однонаправленной называется функция, для которой легко определить её значение, но практически невозможно отыскать обратное значение. Для построения криптографических систем защиты используются однонаправленные функции, так называемые, необратимые функции. Пока не удалось доказать, что эффективного алгоритма вычисления дискретного логарифма за приемлемое время не существует. Одним из первых применений однонаправленных функций было решение задачи обеспечения безопасности и использования пароля, по которому осуществляется доступ пользователя к ресурсам и услугам в автоматизированных системах. Кроме однонаправленных функций в криптографии широко используются однонаправленные функции с лазейкой, для которых знание секретного ключа дает возможность вычислительно просто находить обратное отображение.

Следовательно, модель искусственной нейронной сети подходит для задач шифрования. В настоящее время AES является наиболее широко распространенным блочным шифром шифрование или расшифрование блока данных осуществляется c помощью раундовой функции. Три стандартные версии AES называются AES-128, AES-192 и AES-256 и отличаются друг от друга длиной ключа, 128, 192, и 256 бит соответственно и конечно количеством раундов, 10, 12, и 14. Безопасность AES была тщательно исследована и подтверждена, однако в последнее время появились атаки, которые оказались достаточно эффективными по отношению к AES. Нейронные сети устойчивы к некоторым неточностям, для которых имеется много доступных данных для обучения, но к которым не могут быть применены жесткие правила. Rijndael-криптосистема на основе ИНС имеет менее сложное строение, чем AES и не линейна. Нелинейной должна быть нейронная сеть с обратной связью, что позволяет выполнить шифрование открытого текста и расшифрование зашифрованного текста с высокой производительностью и очень низким уровнем ошибок. Уменьшение вероятности взлома достигается с помощью нелинейной функции активации.

Многослойный персептрон является наиболее популярным видом ИНС, он способен эмулировать любое соотношение входов/выходов, отвечает здесь за частоту обновления. Блочная адаптация является более надежной с того момента, когда шаг обучения усреднен по всем обучающим шаблонам. Для процесса шифрования, могут быть использованы входной вектор, выходной вектор, первоначальные веса и нелинейная функция активации каждого нейрона. Для процесса расшифровки используются выходной вектор, входные значения, первоначальные веса и нелинейная функция. Модель структуры искусственной нейронной сети это последовательно-параллельная модель. Фаза операций многослойна, нейронная сеть прямого распространения с конечными весами, нужна для получения выходного сигнала шифрования/расшифрования.

Итак, ДМЧ - это особый вид многоуровневой нейронной сети прямого распространения, состоящий из одного выходного нейрона, скрытых нейронов и входных нейронов. Входные нейроны принимают двоичные значения, веса между входными и скрытыми нейронами принимают значения от –**L** до +**L**.Значение выходного нейрона это произведение всех скрытых нейронов. Для обмена ключами используется алгоритм Диффи-Хеллмана, его более безопасная замена это синхронизация двух древовидных машин четности. Динамика двух сетей и их весовых коэффициентов может быть применена к протоколу обмена секретным ключом через публичный канал, а сгенерированный ключ может быть использован для шифрования и дешифрования передаваемого сообщения. Защищенность криптографических систем можно улучшить, увеличив длину ключа, а в нейрокриптографии увеличением синаптической длины. Различают, по крайней мере, три вида атак, геометрический, вероятностный анализ и генетические алгоритмы. Для обновления весовых коэффициентов используют три правила, правило положительного подкрепления, правило отрицательного подкрепления и случайное блуждание.