УДК 004.056.5
DOI: 10.21122/2309-4923-2025-2-55-59

ГОЛИКОВ В.Ф., РАДЮКЕВИЧ М.Л., ШУЛЯК Д.В.

ДВУХСТОРОННЯЯ АТАКА НА СИНХРОНИЗИРУЕМЫЕ ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ, ФОРМИРУЮЩИЕ ОБЩИЙ СЕКРЕТ

Научно-производственное республиканское унитарное предприятие
«Научно-исследовательский институт технической защиты информации»
г. Минск, Республика Беларусь

Аннотация. *Рассматривается уязвимость технологии формирования общего криптографического ключа с помощью синхронизируемых искусственных нейронных сетей (ИНС) относительно нового типа атаки, названной двухсторонней. Суть атаки заключается в том, что криптоаналитик, прослушивая открытый канал связи, создает для этого две идентичные искусственные нейронные сети, одну из которых он синхронизирует с сетью одного из легальных абонентов, а вторую – с сетью другого легального абонента. Сравнивая вектора весовых коэффициентов атакующих сетей, криптоаналитик имеет возможность определить момент наступления полной синхронизации сетей легальных абонентов и значение сформированного секретного числа. Далее исследуются возможности двухсторонней атаки при различных моделях формирования секретного числа.*

Ключевые слова: *синхронизируемые искусственные нейронные сети, атака, общий криптографический ключ, отложенный перебор*

Введение

В работах [1-4] излагается технология формирования общего секретного ключа шифрования у абонентов, имеющих аутентифицированный открытый канал связи. Технология основывается на использовании синхронизируемых искусственных нейронных сетей TPM структуры (рисунок 1).

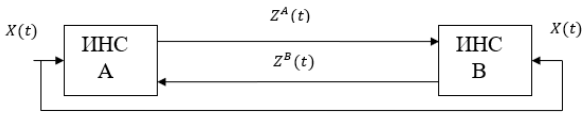

Рисунок 1. Схема синхронизации ИНС

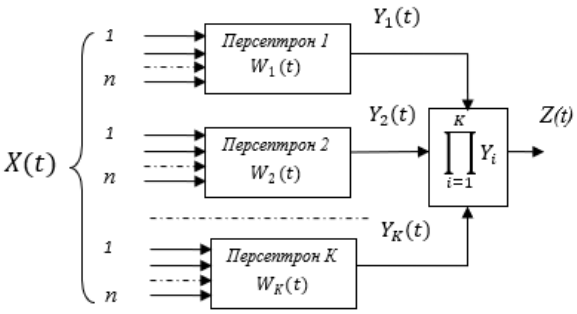

Рисунок 2. Структура ИНС

На этом рисунке: ИНС А, ИНС В – искусственные нейронные сети абонентов *A* и *B*, формирующие общий криптографический ключ; $X(t)$ – вектор синхронизирующих случайных воздействий, t – номер такта синхронизации, $Z^A(t)$, $Z^B(t)$ – выходные величины сетей *A* и *B* соответственно. Архитектура и параметры всех сетей идентичны и представлены на рисунке 2.

Каждый персептрон имеет n входов, на каждый из которых поступает случайное число $x_{ij}(t)\in[-1,1]$ (одна из компонент $X(t)$, где $j = 1,2,\ldots,n$; $i = 1,2,\ldots,K$). Каждый персептрон описывается вектором весовых коэффициентов $W_i(t)$ с компонентами $w_{ij}(t)\in[-L,L]$, где L – целое положительное число. Выходные величины персептронов $Y_i(t)\in[-1,1]$ перемножаются и образуют выходные значения сетей $Z(t)\in[-1,1]$.

Начальные значения весовых коэффициентов персептронов сетей А и В: $W^A(0), W^B(0)$, где $W(0) = W_1(0)\|W_2(0)\|\ldots\|W_K(0)\|$ выбираются абонентами случайно, независимо друг от друга и сохраняются в секрете. Подавая синхронно на входы своих сетей вектор $X(t)$, абоненты *A* и *B* вычисляют выходные величины сетей $Z^A(t)$ и $Z^B(t)$, обмениваются ими и корректируют значения весовых коэффициентов персептронов своих сетей таким образом, что через некоторое число тактов t_{AB} наступает равенство векторов весовых коэффициентов $W^A(t_{AB}) = W^B(t_{AB})$. Значение полученного вектора и является общим секретным числом для *A* и *B*. Поскольку величина t_{AB} является случайной и неизвестной, то абоненты *A* и *B* заранее договариваются о длительности синхронизации d, так чтобы с большой вероятностью выполнялось условие $t_{AB} \le d$.

Односторонняя атака

Если третий абонент E имеет возможность прослушивать канал связи и пытается вычислить сформированный секрет, то ему необходимо подключить к каналу связи свою сеть с параметрами, идентичными параметрам сети A или B, за исключением начальных значений весовых коэффициентов персептронов, которые для E неизвестны, и попытаться синхронизировать свою сеть с сетью A или B. Такую атаку будем называть односторонней. Схема данной атаки представлена на рисунке 3.

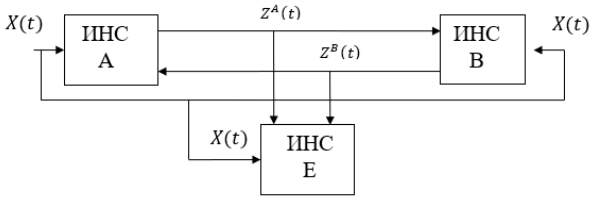

Рисунок 3. Односторонняя атака

Атакующая сеть E, используя значения $X(t)$, вычисляет $Z^E(t)$ и сравнивает его с перехваченными $Z^A(t)$ и $Z^B(t)$, корректирует значения весовых коэффициентов персептронов своей сети по определенному алгоритму и через некоторое число тактов t_{EA} добивается равенства

$$W^E(t_{EA}) = W^A(t_{EA}). \quad (1)$$

В зависимости от выбранного абонентом E алгоритма коррекции различают несколько видов атак. Наиболее эффективной считается «геометрическая атака» [3]. Исследования показали, что независимо от вида атаки обеспечивается

$$P(t_{AB} \le d) > P(t_{EA} \le d). \quad (2)$$

Выражение (2), однако, совсем не означает, что в процессе атаки обязательно произойдет событие $(t_{AB} \le d, t_{EA} \le d)$, т.е. будет иметь место успешная атака, что приведет к выполнению (1). Тем не менее в статье [5] показано, что любой выбранный тип атаки, реализованный в отложенном варианте, будет успешным при сравнительно небольших затратах вычислительных мощностей. Этот тип атак назван «отложенный перебор». Суть метода заключается в том, что абонент E, прослушивая канал связи между A и B, запоминает значения $X(t)$, $Z^A(t)$, $Z^B(t)$, d и организует многократное повторение синхронизаций своей сети с сетью A или B. Причем легальная сеть в каждой синхронизации повторяется в соответствии с записанными параметрами, а злоумышленная сеть меняет начальные значения вектора весовых коэффициентов $W^E(0)$. В работе [5] показано, что необходимый объем отложенного моделирования (количество моделируемых синхронизаций) равен

$$m = \frac{\ln(1-\gamma)}{\ln(1-P^1_{EA})},$$

где $P^1_{EA} = P(t_{EA} \le d))$ – вероятность успешной синхронизации сети E в одной попытке, γ – вероятность того, что из m проведенных синхронизаций хотя бы одна закончилась успешно (вероятность γ задается близкой к 1). Как показывают практические расчеты, объем отложенного перебора относительно невелик. Например, при $P^1_{EA} = 10^{-4}$ и $\gamma = 0{,}98$ имеем $m \approx 3{,}9 \cdot 10^4$.

Для обеспечения более высокой секретности формируемого общего ключа против этой атаки в статье [6] предлагается способ усиления секретности, основанный на интеграции результатов многократных синхронизаций (ИМС). Суть метода [6] заключается в том, что A и B проводят r синхронизаций, не проверяя успех каждой синхронизации. В качестве итоговой секретной бинарной последовательности используют свертку полученных r последовательностей. В качестве свертки можно использовать побитовое сложение соответствующих битов этих последовательностей. Абонент E, не имея хэш-значение результата отдельно взятой синхронизации, не может произвести проверку правильного результата отложенного перебора в отношении каждой синхронизации, а только в отношении свертки r синхронизаций (интегрированного результата r синхронизаций), хэш-значение которой оглашается абонентами A и B. Поэтому E вынужден, осуществляя перебор в отношении совокупности из r синхронизаций, проверять идентичность полученного результата с объявленным хэш-значением. Вероятность успеха при этом в каждой i-той синхронизации равна P^1_{EAi}, а в отношении свертки $P_{EA,r} = \prod_{i=1}^{r} P^1_{EAi} = (P^1_{EA})^r$, а объём моделирования при этом равен

$$m = \frac{\ln(1-\gamma)}{\ln(1-(P^1_{EA})^r)}. \quad (3)$$

Из (3) следует, что объём отложенного моделирования экспоненциально возрастает с ростом r. В рассмотренном ранее численном примере $P^1_{EA} = 10^{-4}$ и $\gamma = 0{,}98$, $r = 10$ имеем $m \approx 10^{40}$.

Этот способ демонстрирует надежную безопасность формируемого ключа по отношению к атаке «отложенный перебор» и стимулирует поиск новых методов взлома.

Двухсторонняя атака

В данной статье рассматривается новый тип атак, названный двухсторонней атакой. Схема данной атаки представлена на рисунке 4.

Для реализации этой атаки абонент E строит две ИНС: ИНС E_1 и ИНС E_2. Параметры этих сетей должны быть идентичны параметрам сетей ИНС A и ИНС B за исключением векторов весовых коэффициентов, которые выбраны случайно и независимо друг от друга и равны: $W^A(0)$, $W^B(0)$, $W^{E1}(0)$, $W^{E2}(0)$. A и B синхронизируют свои сети за время d, проверяют равенство весовых коэффициентов

$W^A(d) = W^B(d)$ сравнением хэш-значений этих чисел и успешно заканчивают процесс, при этом $W^A(d) = W^B(d) = W^{AB}(d)$, где $W^{AB}(d)$ – общий секретный ключ. Абонент E организует атаку «отложенный перебор» двумя сетями E_1 и E_2. При этом он синхронизирует сеть E_1 с сетью А, а сеть E_2 с сетью В. В процессе синхронизации абонент E сравнивает значения весовых коэффициентов своих сетей между собой, т.е. проверяет равенство $W^{E1}(t) = W^{E2}(t)$. Если на момент d окажется, что: $W^{E1}(d) = W^{E2}(d)$, а это видно прямым сравнением векторов $W^{E1}(d)$,$W^{E2}(d)$ на стороне E, то это означает, что $W^{E1}(d) = W^A(d)$, а $W^{E2}(d) = W^B(d)$. Следовательно, $W^A(d) = W^B(d)$, $W^{E1}(d) = W^{E2}(d) = W^{AB}(d)$. Т.е. сформированный секретный ключ известен абоненту E. Графически рассматриваемый процесс представлен на рисунке 5. На этом рисунке показаны условные траектории изменения векторов весовых коэффициентов сетей A, B, E_1, E_2 в процессе синхронизации. Изображена ситуация, при которой произошла синхронизация сетей E_1 с A, E_2 с B и A с B, т.е. $t_{E1A} \le d$, $t_{E2B} \le d$, $t_{AB} \le d$.

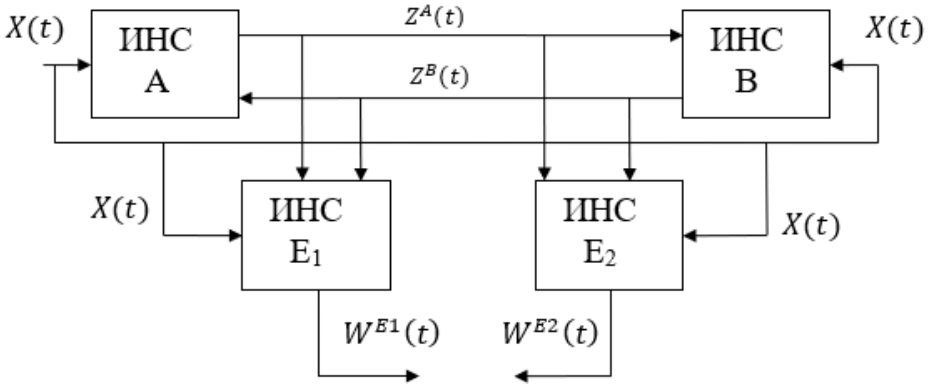

Рисунок 4. Двухсторонняя атака

Если синхронизм E_1 с A и E_2 с B за d тактов не достигнут, то абонент E меняет начальные значения весовых коэффициентов своих сетей: $W^{E1}(0)$ и $W^{E2}(0)$ и повторяет синхронизацию, как предусматривает атака «отложенный перебор».

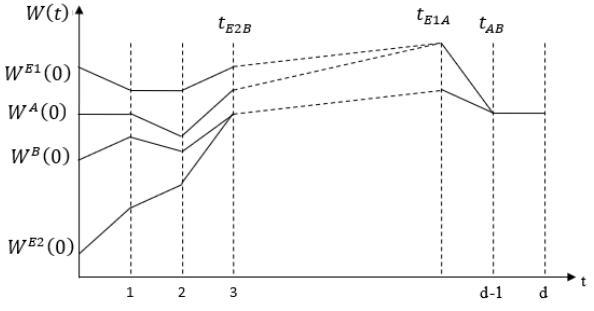

Рисунок 5. Траектории изменения векторов весовых коэффициентов сетей в процессе синхронизации

При атаке одной сетью E_1 вероятность успешной синхронизации абонентов E_1 с А в одной попытке равна

$$P^1_{E1A} = P(t_{E1A} < d).$$

При атаке двумя сетями вероятность успешной синхронизации абонентов E_1 с A и E_2 с B и как итог обнаружение синхронизации E_1 с E_2 в одной попытке равна

$$P^1_{E1E2} = P(t_{E1A} < d, t_{E2B} < d).$$

Так как $P^1_{E1E2} = P(t_{E1A} < d)P(t_{E2B} < d / t_{E1A} < d)$, где второй сомножитель есть условная вероятность, то следует, что вероятность успешной атаки двух сетей меньше, чем одиночной атаки. Эта вероятность тем меньше, чем меньше корреляция $W^{E1}(t)$ с $W^{E2}(t)$. При слабой корреляции можно считать, что $P(t_{E2B} < d / t_{E1A} < d) \approx P(t_{E2B} < d)$, а так как $P(t_{E1A} < d) \approx P(t_{E2B} < d)$, то $P^1_{E1E2} = P^2(t_{E1A} < d) = P^2(t_{E2B} < d)$.

Это делает применение двухсторонней атаки не актуальным для E, тем более, что одиночная атака позволяет E добиваться дискредитации формируемого секретного ключа после относительно небольшого объема моделирования, как было показано выше.

Рассмотрим далее эффективность двухсторонней атаки против метода ИМС.

При двухсторонней атаке согласно её алгоритму у абонента E появляется возможность осуществлять атаку отложенного перебора относительно каждой синхронизации А с В в отдельности до полного успеха. Для этого E имеет записанные параметры атакуемой синхронизации А с В: $X(t)$, $Z^A(t)$, $Z^B(t)$, d и критерий остановки моделирования в нужный момент, при котором $W^A(d) = W^B(d)$. Как указывалось выше, абонент E, сравнивая значения весовых коэффициентов своих сетей между собой, т.е. проверяя равенство $W^{E1}(t) = W^{E2}(t)$, обнаруживает наступление $W^{E1}(d) = W^A(d)$, $W^{E2}(d) = W^B(d)$, а значит и $W^{E1}(d) = W^{E2}(d) = W^{AB}(d)$.

Такой критерий отсутствует при односторонней атаке. Таким образом, абонент E успешно атакует каждую i-тую синхронизацию A с B. Объём моделирования при этом равен:

$m_i = \frac{\ln(1-\gamma)}{\ln(1-P^1_{E1E2})}$, где $P^1_{E1E2} = P(t_{E1A} < d)P(t_{E2B} < d) \approx (P^1_{EA})^2$.

При атаке на r синхронизаций суммарный объём моделирования равен:

$$M = \sum_{i=1}^{r} m_i = \sum_{i}^{r} \frac{\ln(1-\gamma)}{\ln(1-P^1_{E1E2})}.$$

Считая $m_1 = m_2 = m_3 = \ldots = m_r = m$, получим

$$M = r\,\frac{\ln(1-\gamma)}{\ln(1-P^1_{E1E2})}.$$

Т.е. объём атакующего моделирования при двухсторонней атаке на метод ИМС зависит от r линейно, что указывает на более высокую эффективность этой атаки по сравнению с односторонней атакой. В рассматриваемом численном примере $P^1_{EA} = 10^{-4}$ и $\gamma = 0{,}98$, $r = 10$ имеем $P^1_{E1E2} \approx (P^1_{EA})^2$, $M \approx 3{,}9 \cdot 10^9$. Т.е. объём атакующего моделирования при двухсторонней атаке меньше, чем необходимый объём атакующего моделирования при односторонней атаке примерно в 10^{30} раз.

Возможно, что методы формирования общего секретного числа, использующего технологию синхронизируемых искусственных нейронных сетей, изложенные в [7, 8], окажутся более безопасными к двухсторонней атаке, однако это требует дополнительных исследований.

Заключение

Таким образом, двухсторонняя атака на процесс формирования общего секретного числа, использующего технологию ИМС, может существенно снизить её защищенность и требует дополнитель-ного усиления безопасности применяемых методов.

ЛИТЕРАТУРА

1. **Kanter, I.** The Theory of Neural Networks and Cryptography / Ido Kanter, Wolfgang Kinzel // The Phisics of Communication. – 2003. – P. 631–642. – DOI: 10.1142/9789812704634_0044

2. **Kinzel, W.** Neural Cryptography / W. Kinzel, I. Kanter // ICONIP'02 : proc. of the 9th Intern. Conf. on Neural Information Processing, Nov. 18–22, 2002, Orchid Country Club, Singapore. – Singapore, 2002. – Vol. 3. – P. 1351–1354. – DOI: 10.1109/ICONIP.2002.1202841.

3. **Ruttor, A.** Dynamics of neural cryptography / A. Ruttor, I. Kanter, W. Kinzel // Physical Review E – Statistical, Nonlinear, and Soft Matter Physics. – 2007. – Vol. 75, iss. 5. – P. 056104.

4. **Урбанович, П.П.** Использование гиперкомплексных чисел в протоколе согласования криптографических ключей на основе нейронных сетей / П.П. Урбанович, Н.П. Шутько // Журнал Белорусского государственного университета. Математика. Информатика. – 2024. – № 2. – С. 81–92 .

5. **Голиков, В.Ф.** Атака на синхронизируемые искусственные нейронные сети, формирующие общий секрет, методом отложенного перебора / В.Ф. Голиков, А.Ю. Ксеневич // Доклады БГУИР. – 2017. – № 8. – С. 48–53.

6. **Радюкевич, М.Л.** Усиление секретности криптографического ключа, сформированного с помощью синхронизируемых искусственных нейронных сетей / М.Л. Радюкевич, В.Ф. Голиков // Информатика. – 2020. − Т. 17, № 1. – С. 102–108. – DOI: 10.37661/1816-0301-2020-17-1-102-108

7. **Радюкевич, М.Л.** Комбинированный метод формирования криптографического ключа с помощью синхронизируемых искусственных нейронных сетей / М.Л. Радюкевич, В.Ф. Голиков // Доклады БГУИР. – 2021. – Т. 19, № 1. – С. 79–87. – DOI: 10.35596/1729-7648-2021-19-1-79-87

8. **Радюкевич, М.Л.** Комбинированный метод формирования криптографического ключа с секретной модификацией результатов синхронизации искусственных нейронных сетей / М.Л. Радюкевич // Системный анализ и прикладная информатика. – 2021. – № 3. – С. 51–58. – DOI: 10.21122/2309-4923-2021-3-51-58

REFERENCES

1. **Kanter, I.** The Theory of Neural Networks and Cryptography / Ido Kanter, Wolfgang Kinzel // The Phisics of Communication. – 2003. – P. 631–642. – DOI: 10.1142/9789812704634_0044

2. **Kinzel, W.** Neural Cryptography / W. Kinzel, I. Kanter // ICONIP'02 : proc. of the 9th Intern. Conf. on Neural Information Processing, Nov. 18–22, 2002, Orchid Country Club, Singapore. – Singapore, 2002. – Vol. 3. – P. 1351–1354. – DOI: 10.1109/ICONIP.2002.1202841

3. **Ruttor, A.** Dynamics of neural cryptography / A. Ruttor, I. Kanter, W. Kinzel // Physical Review E – Statistical, Nonlinear, and Soft Matter Physics. – 2007. – Vol. 75, iss. 5. – P. 056104.

4. **Urbanovich, P.P.** Usage of hypercomplex numbers in a cryptographic key agreement protocol based on neural networks / P.P. Urbanovich, N.P. Shutko // Journal of the Belarusian State University. Mathematics and Informatics. – 2024. – Vol. 2. – P. 81–92.

5. **Golikov, V.F.** Attack on synchronized artificial neural networks, forming a common secret by deferred search / V.F. Golikov, A.Y. Ksenevich // Doklady BGUIR. – 2017. – Vol. 8. – P. 48-53. (In Russ.)

6. **Radziukevich, M.L.** Enhancing the secrecy of a cryptographic key generated using synchronized artificial neural networks / M.L. Radziukevich, V.F. Golikov // Informatics. – 2020. – Vol. 17, № 1. – P. 102-108. – DOI: 10.37661/1816-0301-2020-17-1-102-108

7. **Radziukevich, M.L.** Combined formation of a cryptographic key using synchronized artificial neural networks / M.L. Radziukevich, V.F. Golikov // Doklady BGUIR. – 2021. – Vol. 19, № 1. –P. 79-87. – DOI: 10.35596/1729-7648-2021-19-1-79-87

8. **Radziukevich, M.L.** A combined method of formation of a cryptographic key with secret modification of the results of synchronization of artificial neural networks. System analysis and applied information science. – 2021. – Vol. 3. – P.51-58. (In Russ.). – DOI: 10.21122/2309-4923-2021-3-51-58

HOLIKAU U.F., RADZIUKEVICH M.L., SHULIAK D.V.

TWO-WAY ATTACK ON SYNCHRONIZED ARTIFICIAL NEURAL NETWORKS FORMING A COMMON SECRET

Scientific Production Republican Unitary Enterprise "Research Institute for Technical Protection of Information"
Minsk, Republic of Belarus

Abstract. *The vulnerability of the technology for generating a common cryptographic key using synchronized artificial neural networks (ANN) is considered in a relatively new type of attack called two-way. The essence of the attack is that a cryptanalyst, listening to an open communication channel, creates two identical artificial neural networks, one of which he synchronizes with the network of one of the legal subscribers, and the second with the network of another legal subscriber. By comparing the vectors of weight coefficients of attacking networks, the cryptanalyst is able to determine the moment of full synchronization of the networks of legal subscribers and the value of the generated secret number. Next, we examine the possibilities of a two-way attack using various models of secret number generation.*

Keywords: *synchronized artificial neural networks, attack, shared cryptographic key, delayed brute force*

Голиков Владимир Федорович, Доктор технических наук, профессор. Сфера научных интересов: защита информации, криптография.

Uladzimir F. Holikau, Doctor of Technical Sciences, Professor. Area of scientific interests: information security, cryptography.

E-mail: gvf147@mail.ru

Радюкевич Марина Львовна, Научно-производственное республиканское унитарное предприятие «Научно-исследовательский институт технической защиты информации», г. Минск, Республика Беларусь. Кандидат технических наук, начальник испытательной лаборатории по требованиям безопасности информации. Сфера научных интересов: защита информации.

Maryna L. Radziukevich, Scientific Production Republican Unitary Enterprise "Research Institute for Technical Protection of Information", Minsk, Republic of Belarus. Ph.D. of Engineering Sciences, head of the testing laboratory for information security requirements. Area of scientific interests: information security.

E-mail: 1218a@list.ru

Шуляк Дмитрий Викторович, Научно-производственное республиканское унитарное предприятие «Научно-исследовательский институт технической защиты информации», г. Минск, Республика Беларусь. Магистр технических наук, ведущий инженер испытательной лаборатории по требованиям безопасности информации. Сфера научных интересов: защита информации.

Dmitry V. Shuliak, Scientific Production Republican Unitary Enterprise "Research Institute for Technical Protection of Information", Minsk, Republic of Belarus. Master of Engineering Science, leading engineer of the testing laboratory for information security requirements . Area of scientific interests: information security.

E-mail: shuliak.dv@gmail.com